

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Use of Networks, Graphs and Topology in Materials Modeling

Permalink

<https://escholarship.org/uc/item/2j48q66v>

Author

Zhang, Weiyi

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Use of Networks, Graphs and Topology in Materials Modeling

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

Doctor of Philosophy

in

Materials Science and Engineering

by

Weiyi Zhang

September 2022

Dissertation Committee:

Dr. Alex Greaney, Chairperson
Dr. Pingyun Feng
Dr. Juchen Guo

Copyright by
Weiyi Zhang
2022

The Dissertation of Weiyi Zhang is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor, without whose help, I would not have been here.

To all the support.

ABSTRACT OF THE DISSERTATION

Use of Networks, Graphs and Topology in Materials Modeling

by

Weiyi Zhang

Doctor of Philosophy, Graduate Program in Materials Science and Engineering

University of California, Riverside, September 2022

Dr. Alex Greaney, Chairperson

Atomistic simulations have become a prominent tool in chemistry, physics, and materials science for its capability interpreting experimental measurement, predicting material properties and designing new compounds from atomic-level perspective, for its capability helping visualize the topology of materials and their change during a process or reaction. It has been noted that atomic simulations are providing new data and exciting insights into various phenomenon that cannot be obtained readily in any other way—theory, or experiment[25]. Having a precise image of the structure or the dynamical process at this scale provides a brand-new approach understanding how the properties of a material is determined.

This thesis is comprised of four distinct computational projects that are united by the the analysis of networks, graphs or bonding typologies to gain fundamental insights into the phenomena studied. In the first of these I studied the topology of stacked and rippled graphene sheets to elucidate the enlarged d-spacing of hard carbon used for anodes in sodium ion batteries. In the second project, I examined the electrochemical stability of

highly concentrated aqueous LiCl electrolytes which have been found to have dramatically improved stability at low temperature. For this I studies the network of H-bonding in *ab initio* molecular dynamic (AIMD) simulations of the electrolyte revealing subtle changes in the electrolyte's solvation structure. In doing so I identify a new theoretical framework to guide design of new aqueous electrolytes. In the third and fourth projects, I go beyond studying existing material, and used atomic model to design new materials. In the third project I demonstrate the possibility of high proton conductivity in Prussian Blue analogues through use of graph theory to analyze the water hydrogen bonding network. The final project was concerned with the design of flexible organic ligands for Metal-Organic Frameworks (MOFs). It is a time-consuming work to calculate ligands' flexibility by either experimentally synthesizing and characterize or simulate deforming process, which is solved by forecasting the stiffness of molecules from its geometry with Neural Network.

Contents

List of Figures	x
List of Tables	xiii
1 Introduction	1
1.1 Effect of Curvature to Multi-Layer Graphite	2
1.2 Electrochemical Stabilization in LiCl Electrolytes	3
1.3 Organic Ligand in Metal-Organic Framework	4
1.4 Water-Network in Prussian Blue Analogues	5
1.5 Methods	5
1.5.1 Molecular Dynamic and Density Function Theory Simulation	5
1.5.2 Geometrical and Topological Representations	7
1.5.3 Algorithms	8
2 Effect of Curvature to Multi-Layer Graphite Inter-Layer Spacing for Ion Insertion	10
2.1 Background	10
2.2 Methods	12
2.3 Results and discussion	14
2.4 Conclusion	18
3 Solvation Energy Difference in LiCl Solution at Different Temperature	19
3.1 Background	20
3.2 Methods	21
3.2.1 Computational Simulation	21
3.2.2 Water Diversity	24
3.2.3 Coordination Number	25
3.2.4 Solvation Energy	26
3.3 Results and Discussion	27
4 Percolation in Vacancy Network in Prussian Blue Analogs	38
4.1 Background	38

4.2	Methods	39
4.2.1	Monte Carlo Simulation	39
4.2.2	Vacancy Connectivity Analysis	41
4.3	Results and Discussion	42
5	Efficiently Forecasting Organic Ligands Mechanical Properties From Atomic Structure via Deep Learning	47
5.1	Background	48
5.2	Methods	51
5.3	Results and Discussion	61
5.4	Conclusion	64
5.5	Appendix	65
5.5.1	Molecule Orientation	65
5.5.2	Moment Variance	66
5.5.3	Performance Comparison of Different Order of Moments	67
5.5.4	Voxelization Parameters	68
5.5.5	Identifying Backbone	70
6	Conclusions	72

List of Figures

2.1	Simulation process of of $30 \times 10 \times 4$ scale graphene with $\lambda = 0.60$. As seen, while rippled, the curvature of each layer is different, in which case, the multi-layer structure could be seen as stacking of u shaped carbon sheets.	13
2.2	Measurement of interlayer spacing from atomic model. 2.2(a) The line crossing the nearest two atoms in two layers is not perpendicular to the graphene surface. 2.2(b) The projection to XoZ plane can dismiss the error from mismatch along y-axis direction. 2.2(c) For practical calculating the interlayer spacing, the projected 2-D atom position is fitted to curves functions and the normal distance of the two curves representing neighbor layers is calculated.	14
2.3	Interlayer spacing of deformed multi-layer graphite. 2.3(a) and 2.3(b) are interlayer spacing of double-layer and triple-layer graphene respectively. 2.3(c) is the outer and 2.3(d) is the inner spacing of 4-layer graphene. Orange, green and blue curves is representing the interlayer in the order of maximum, average and minimum.	15
2.4	2.4(a) shows the displacement of layers caused by shifting along curve. The tangent plane is horizontal at the vertex. 2.4(b) and 2.4(c) are result from triple layer graphene simulation. 2.4(b) shows that correlation energy of neighbor layers decreases while the level of rippling gets higher. 2.4(c) is non-Euclidean distance of two aimed atoms in two neighbored layers changing with λ , which could help understanding the shift in 2.4(a).	16
2.5	Showing the atomic structure of rippled graphene after insertion. 2.5(a)The brown atoms are carbons. And yellow atom in the figure representing sodium ion is at the most favorable sites for insertion of a 0.70 rippled $8 \times 3 \times 2$ bilayer graphene unit. 2.5(b) is fitted curve of structure, with largest inter-layer spacing locates at vertex as well.	17
3.1	Schematic of possible bonding configurations available to water.	25
3.2	Ion insertion operation and resulted optimized structure.	27
3.3	DDAA water molecules under two different temperature.	29
3.4	(g) and (h) The average numbers of water's O coordinating Li+ and Cl-coordinating Li+ as a function of distance revealed in AIMD simulations of 25 wt.% LiCl.	30

3.5	Number of acceptor bonds to H of water or Li+ as a function of distance from AIMD simulations. Solid lines are about the oxygen of water molecules that are not aqua ligands, and dashed lines denote the oxygen of water molecules that are aqua ligands of Li+.	31
3.6	The average number of O per Cl- (dashed lines) and H coordinating Cl- (solid lines) as a function of distance retrieved from AIMD simulations of 25 wt.% LiCl.	32
3.7	Key steps in the hydrogen evolution reaction (HER, left) and oxygen evolution reaction (OER, right).	33
4.1	Example structure of PBA and PBA with vacancy.	40
4.2	Two GCMC example results of two simulation box size: 4.2(a) and 4.2(c) for a 3x3x1 system and 4.2(b) for smaller case which is 3x1x1. Combined with the atomic structure, we can clearly find a path in the directed graph of water network topology from one vacancy to another if they are next to each other for most of the cases. 4.2(d) shows two transporting path found after proton insertion with RAPTOR simulation.	44
4.3	PBAs frame and pipe configuration.	45
4.4	Vacancy connectivity analysis. 4.4(a) and 4.4(b) showing percolation thresholds at different system scale and percolation probability change with vacancy fraction. 4.4(c) shows that accessible vacancies change much smoother than percolation probability	46
5.1	A example sequence of design steps for automated generation of a linker molecule: Starting from the carboxylate seed molecule editing rules are selected at random and applied to the molecule. Between each editing step the incomplete linker is structurally relaxed, and the list of editing rules that can be applied to the new partial molecule are determined. In the final editing step a second carboxylate group is added that acts as the candidate linker's second connection point in a MOF.	51
5.2	The simulation process. The whole experiment is separated to four parts: annealing, energy minimizing, equilibration, and deformation.	52
5.3	Example molecules, 5.3(a) molecule with mechano-isomerization which could be found by obvious drop or impulse of stiffness and 5.3(b) a typical molecule that geometry and as well stiffness changing smoothly during deformation, in the 3 rd depth with 3 rules applied. Three subplots from top are the force-, energy-, and stiffness/length-strain curves respectively. Stiffness curve is the derivative of force curve.	54
5.4	Voxelization example with $a = 0.5\text{\AA}$, $\sigma = 0.6$. 5.4(a) shows the molecule geometry; 5.4(b) is the voxelized space, and 5.4(c) is XOY surface section of voxelized space. 5.4(d) is a high resolution section image of XOY surface.	60
5.5	Histogram of 5.5(a) stiffness at 0 strain, 5.5(b) stiffness 1, and 5.5(c) stiffness 2. Red is molecules with high stiffness and black represents flexible molecules.	62

5.6	Orientation for moments calculation. 5.6(a) the two surface determining the XOY plane, and 5.6(b) the re-oriented system.	66
5.7	Comparison of carbon channel geometric moments variance of dataset and uncertainty during equilibrium. Up to 5 th order moments without the 1 st are shown in the figure.	67

List of Tables

2.1	Interlayer Spacing(\AA)	17
3.1	Energy change resulted from ion insertion. LT represents low/cryogenic temperature set up while RT means room temperature. the last column quantities represent a surrogate for the thermodynamic quantities. They indicate the trends in the thermodynamic variable for which they are surrogate with cooling, but not their accurate values.	36
4.1	Water and hydrogen bond number comparison of three vacancy topology with same number of vacancies.	46
5.1	Prediction AUC	63
5.2	Stiffness Prediction AUC	68
5.3	Comparison of voxelization performance with σ changing	69
5.4	Comparison of voxelization performance with different voxel size	70

Chapter 1

Introduction

Computational simulation of atomic structure offers new aspect to understand macroscopic phenomena or performance of materials for its capability showing how atoms react under certain condition in given local environment and the interaction between particles. The key to solve real problem with simulated results is aiming on correlating atoms movement and interaction to materials performance, which helps us find theoretical explanation of unique mechanisms and understand how to further improve properties or avoid certain situations.

Recent researches have proven the high potential of atomic simulations. However with computational resources limitation and the complexity of realm, the simulation can never reach the scale to reconstruct an actual process but a simplified representation. That makes the first and most obvious compromise has to be made is the restriction of simulated systems' size. Till now, a first principle simulation of hundreds of atoms for tens of picoseconds could be easily achieved. A trade-off of accuracy to the level of empirical interatomic

potentials would be necessary to model millions of atoms or a period for microseconds, which still falls far short in both length and especially time scale of experiments[4]. For some specific properties, for example those associated with long-range interaction – electrostatic or elastic, the convergence of simulation could be extremely slow.

Is there a shortcut to bypass the simulation of a mechanical or chemical process? This is one of the problem we are trying to solve here. If the performance or the properties of materials can be predicted from their geometric structure, the simulation could be simplified to an energy minimization. Moreover, if the relation is built, a convenient way designing materials can be discovered. The tunable structural factors can offer possibilities to first design the atomic structure and then attempt to synthesis the desired material. That above covered the two aspects of our research which are uncover the theoretical explanation of phenomena/mechanisms, and material design.

1.1 Effect of Curvature to Multi-Layer Graphite

Starting with a less complex system – hard carbon, a geometric analysis of its microscopic structure successfully uncovered the reason of relatively larger inter-layer spacing than traditional graphene electrolytes. Lithium-ion batteries has become the most commonly used electrochemical energy storage device with a long history. While the advancements being high capacity and long cycle life, the cost is one of its main shortcoming. Facing the limitation of lithium resources, a great replacement, sodium is found, which is one of the richest elements on earth, that at the meantime being more environment-friendly.

Both being alkali metal, sodium is very chemically similar to lithium, but with a larger atomic radius. As reported, the d-spacing of stacked graphene is too small for sodium ion insertion. Therefore, despite the advanced application in lithium batteries, it cannot be replicated onto sodium battery. However, another carbon-based material can be a great alternative, which is hard carbon. As insertion material for ions, hard carbon has maximum inter-layer spacing bigger than $0.37nm$ comparing to graphite's d-spacing $0.335nm$, which meets the requirement for sodium ion insertion as predicted by the theoretical calculations[51]. To search for more insertion media candidate, we wanted to find the root reason causing the decrease on inserting energy.

Dissecting hard carbon, the main difference from it to graphite is the existence of nanosized internal pores and the fact that these graphite platelets are curved instead of flat sheets. Despite the nanopores, we wanted to find if there are more available sites for sodium ion insertion coming from the enlarged d-spacing caused by the rippling.

1.2 Electrochemical Stabilization in LiCl Electrolytes

Although with water as an amazing solvent, sadly, aqueous electrolytes suffer from poor electrochemical stability. But luckily, our collaborator found that 25 wt.% LiCl (aq.) cooled to $-78\text{ }^{\circ}\text{C}$ is stable over 2.87 V, nearly double its stability window at room temperature. With the exciting experimental output, it inspires us that if we can find the key reason for the great performance, there is potential to discover more electrolyte material that with the advanced understanding of wider stability window. The discussion on atomic level is accomplished by AIMD simulation combined with structural analysis to compare

local environment of each component at two temperature setups. We were able to observe obvious structural difference between LiCl solution at room temperature and -78 Celsius, that the local environment of water molecules and ions affect solvation energy of extra H+ or OH- ions, which have further impact on HER and OER energy, so that the energy barrier become higher to make the more stable material. the chapter focus on the details of how the atomic structure changes and how the changes affects the energy barrier.

1.3 Organic Ligand in Metal-Organic Framework

Flexible metal-organic frameworks (MOFs) remain at the forefront of porous materials research due to their enormous potential for various technological applications[41]. The special breathing mechanism provides the effective way adsorb and release small molecules like gas or water. The capacity is mainly determined by pore size and surface area, while the breathing efficiency is highly relying on flexibility.

MOFs are a class of porous crystalline materials constructed from inorganic building units, which are interconnected by organic linkers. With the combination of various types of SBUs and way much higher complexity in organic ligands topology, there are unlimited possibility of different frameworks. Looking to achieve higher flexibility, both experimental and simulating approach are too time-consuming. A shortcut predicting flexibility directly from atomic structure can help us rapidly screening for good candidates for gas storage.

1.4 Water-Network in Prussian Blue Analogues

Proton as charge carrier can travel way much faster with hopping mechanism in water network than most of the metal ion and even than itself by diffusion, which offers a solution for high-performance battery electrode material. Prussian Blue and its analogues are a kind of the micro-pore materials benefit from the special proton transport mechanism. Recent research with premium charge and discharge efficiency has proven the existence of long connected water chain in PBAs to accelerate charge transportation and the conductivity has been found highly dependent on vacancy fraction and configuration. With a high variety of vacancy arrangement, there is no reason not trying to discuss the possibility further improve the connectivity of water in the frame.

PBAs are capable holding a great amount of water in two formation, therefore, the water network configuration can become too complex to work with. The first step is to simplify the analysis on water-net toward vacancy-net, following which, the connectivity of vacancies are accomplished with graph theory. We were able to identify the least number of vacancies a system should have promising a percolating path, and the increase trend on accessible vacancies are also discussed with graph algorithm.

1.5 Methods

1.5.1 Molecular Dynamic and Density Function Theory Simulation

The atomic structure used in this thesis are mostly from ab initio molecular dynamics(AIMD). So the method section starts with the introduction of traditional MD and

DFT to offer the insight on how reliable the model used for the following projects is. Molecular dynamics (MD) is a general term that describes any simulation method that evolves a system of particles in time according to Newton's second law. The "particles" in an MD simulation can be proteins, molecules, or atoms, depending on the desired resolution and computational limitations of the problem, while in the following projects, the particles are mostly atoms and rarely small inorganic molecules. A complete simulating process is mainly about finding low energy status under given condition and rules, for particles or the combination of particles are always tending to find and maintain at low energy states while there is no interference nor other external affect. The accuracy and consistency of the low energy state searching is dependent on how the energy is calculated from current configuration, that means if the definition went wrong, the correct structure can never been converged to, which leads to the choice of interatomic potential. Interatomic potential decides the forces on all of the atoms and the resulted energy of the whole system.

In traditional classical MD, the interatomic potential is parametrized either from experimental data or from more accurate simulations, which makes sure that the simulated result should be at the most cases consistent with experiment result. But with the simplified force simulating mathematical model, there would be no universal interatomic potential that satisfying all cases. While designing a computational experiment, as known as simulation, the first step is always finding proper potential parameters being used in similar scenario and calibrate them with corresponding experimental outcome.

Only involving atom interaction became too simple to model realistic experiments. More complex model is required with more parameters need to be taken into consideration

especially for the problems that are very sensitive to changes in the electronic charge distribution over the system, which are exactly most of the topic discussed in this thesis. That is where AIMD gets involved, with whom, the model evolves to the level of the electronic structure and the electronic charge distribution. The model is solved with classical mechanics for the nuclei and quantum mechanics for the electrons. In essence, within the framework of AIMD, we perform a DFT calculation at each timestep and move the atoms according to the forces generated by the converged electronic density for higher liability of simulated structure and reaction.

1.5.2 Geometrical and Topological Representations

Now that atomic structures have been obtained, more than visualizing the reaction or equilibrium box/space and attempting to understand what is happening with observation by eyes. A more subjective and automatic approach needs to be figured out to be trustworthy and efficient. To start with, the most straightforward solution is 3-d coordination of every single atom in the system with their species and charge, which is surely a general way to begin quantitative analysis with, containing most of the structural information of a given model. However, the shortages are pretty obvious too, with the feature space is with too high dimension. And there is no specific information regarding to any application or function.

To modify feature space for given projects, the information highly related is better to be represented directly. For example, when the water connectivity is the focus, the atomic level information is abundant. The smallest unit is enlarged to molecule level and other atoms in the system is not worth wasting resources. So the system was represented as

direct graph only keep the H-bond information and makes the following analysis way easier. Another example is when the local environment is being discussed. Atomic environmental vector or coordination number can be the best choices as representation in these scenarios, because all the information required is included and the needless part is gotten rid of. In this thesis, while discussing topics with various application, several kind of topological representation of simulated model is utilized, the detail and the reason for feature space design are described in each chapter.

1.5.3 Algorithms

At the meantime, with multiple projects about bonding relation, the representative graph can be discussed with graph theory, which offers a mature tool and platform for our research. The graph related algorithms we used are fully discussed in the past decades, i.e. Depth First Search, Breadth First Search or Shortest Path. Translating the model to graph can effectively accelerate calculating process than building an algorithm from ground.

Machine learning, or one advanced step forward – deep learning is a powerful tool digging into the hidden correlations. Given that the relationship from atomic structure to macroscopic property is always worth attention. And with current acknowledge, the theoretical understanding might be difficult to find. So, the appearance of deep learning provides another or even better option to understand observation and designing direction. The algorithm used in our project is fully connected neural network. The reason choosing FNN is keeping all information in the trained model to keep the accuracy and the functionality in different cases. A fully connected neural network consists of a series of fully connected layers. A fully connected layer is a function from one real number space to another. Each

output dimension depends on each input dimension. For a layer of neurons, the output y is obtained with as a matrix multiply from previous layer x :

$$y = \sigma(wx) \tag{1.1}$$

where w is trainable parameter matrix and σ the non-linearity applied component-wise.

Chapter 2

Effect of Curvature to Multi-Layer Graphite Inter-Layer Spacing for Ion Insertion

2.1 Background

There has been a long history of lithium-ion batteries(LIBs) as commonly used electro-chemical energy storage devices. Despite the high energy capacity and long cycle life[28], the limited lithium resources and rapid increase inquiry on larger-scale batteries lead to the necessity to find a replacement to fulfill the needs of portable equipment, mobile devices, electric vehicles, etc[31, 22, 54]. With the known resources thousands of times more than lithium, sodium has been considered a potential alternative given its similar electron configuration and intercalation chemistry[10, 8, 42]. However, the chemistry and

electrochemistry of electrode materials for Na-ion batteries are sufficiently different from that of their Li-ion counterparts, for practical batteries, more work needs to be done to find the suitable candidates.[1] The biggest challenge is the larger ionic radius[29], which makes graphite, the most common commercially insertion materials for Li^+ , not applicable for Na^+ under moderate condition.[24, 20]

Unlike graphite, disordered carbon, e.g., hard carbon, has been experimentally proven a good Li^+ insertion host with larger inter-layer spacing and lower insertion energy for higher ion capacity, which can reach 300 mAh g^{-1} [50] in hard carbon comparing to lithium ion capacity as 372 mAh g^{-1} with traditional graphite electrode[60]. To discuss the key factors promising larger inter-layer spacing in hard carbon, taking a detailed look into structure of the platelets, it was found that instead of large flat graphite, it is composed with parallel-stacked rippled graphene platelets[58], with a quite wide variation in micro-structure. High-resolution TEM images from Wang et al. showed rippled graphene platelets in hard carbon are with different size, stack layers, and curvature, respectively. Consequently, ion capacity and then battery performance vary with slight difference in micro-structure[46, 6], in which case, by understanding how structural factors affect d-spacing between layers could further improve SIBs performance and even is possible to find proper insertion materials for other alternatives of LIBs such as potassium-ion battery with K^+ radius even larger than Na^+ .

With molecular dynamic simulation, the atomic level structure of rippled graphite is modeled to see how d-spacing changed by the structural features considered as determining factor. Comparing the resulted structure by rippling differentiated sizes of graphite, we

were able to figure out what brings in non-monotonicity to the correlation between curvature and size of spacing, and find the best rippling wave length to explain high sodium ion capacity in disordered carbon materials.

2.2 Methods

From experimental analysis, from 1 to 4 layers graphite have been observed in hard carbon micro-structure, therefor simulations are done with 2- to 4-layer ABC-stacked graphite for further discussion. Curvature is introduced in by shrinking along the surface sinusoidally while maintaining surface area unchanged and non-Euclidean distance on graphene layer between carbon atoms the same as well, with amplitude A obtained from rippling wavelength λL from Equation 2.1 and updated coordinates determined by Equation 2.2 to 2.4. The resulted structure would be optimized with Universal Force Fields by LAMMPS(Large-scale Atomic/Molecular Massively Parallel Simulator)[57, 47].(Figure 2.1).

$$\int_0^{2\pi} \sqrt{1 + A^2 k^2 \sin^2 x} dx = kL; k = \frac{2\pi}{\lambda L} \quad (2.1)$$

$$\int_0^x \sqrt{1 + A^2 k^2 \sin^2 \theta} d\theta = kx_{\text{orig}} \quad (2.2)$$

$$y = y_{\text{orig}} \quad (2.3)$$

$$z = A \cos kx + z_{\text{orig}} \quad (2.4)$$

After reaching equilibrium, with atoms on each sheets locating on irregular curved surface, the challenge is to calculate distances between surfaces.

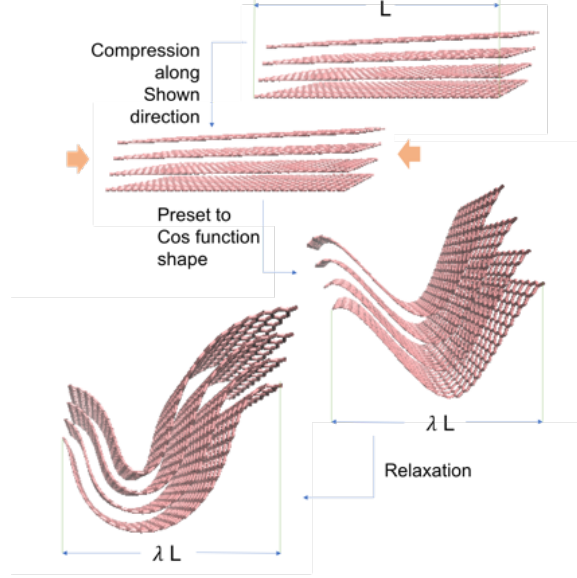


Figure 2.1: Simulation process of $30 \times 10 \times 4$ scale graphene with $\lambda = 0.60$. As seen, while rippled, the curvature of each layer is different, in which case, the multi-layer structure could be seen as stacking of u shaped carbon sheets.

Instead of by atom-atom distance directly, inter-layer spacing was found one of the key factor determining insertion energy, which is also more easily detectable experimentally (e.g. XRD). To accurately measure d-spacing, each layer was fit to up to 15 degree Fourier Function on XoZ plane. (Figure 2.2). The spacing in between fitted curving surface is the inter-layer spacing we are looking for.

$$\sum_{i=0}^{15} (A_i \cos(\frac{2i\pi}{\lambda L}(x - C)) + B_i \sin(\frac{2i\pi}{\lambda L}(x - C))) \quad (2.5)$$

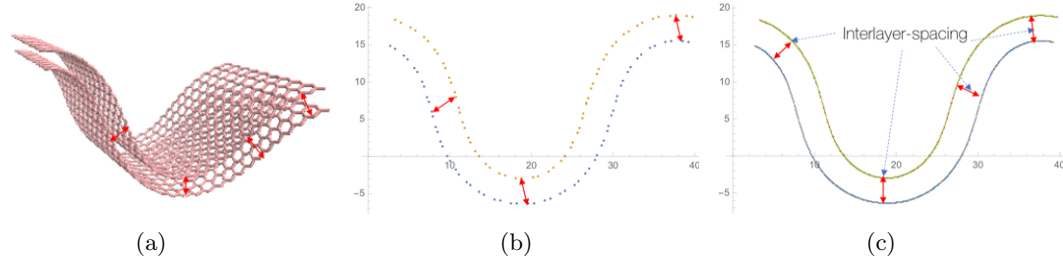


Figure 2.2: Measurement of interlayer spacing from atomic model. 2.2(a) The line crossing the nearest two atoms in two layers is not perpendicular to the graphene surface. 2.2(b) The projection to XoZ plane can dismiss the error from mismatch along y -axis direction. 2.2(c) For practical calculating the interlayer spacing, the projected 2-D atom position is fitted to curves functions and the normal distance of the two curves representing neighbor layers is calculated.

If the wavelength gets too short, after deforming the structure, it is impossible to avoid overlapping of layers, which also means that structure would be unrealistic, so we only run simulation with λ in range of 0.6 to 1.00.

2.3 Results and discussion

Starting with $30 \times 10 \times N$ model, maximum spacing is obviously larger after rippled in all the cases, and at the meantime, minimum spacing gets lower but not changing correspondingly. That results in the change of average varying in different situation. From Figure 2.3, if only 2 or 3 layers are stacked, after even slightly deformed, average spacing would increase, with the differential of 2-layer graphene a bit bigger than 3-layer. For 4-layer graphene, at some λ average is higher and other cases the opposite.

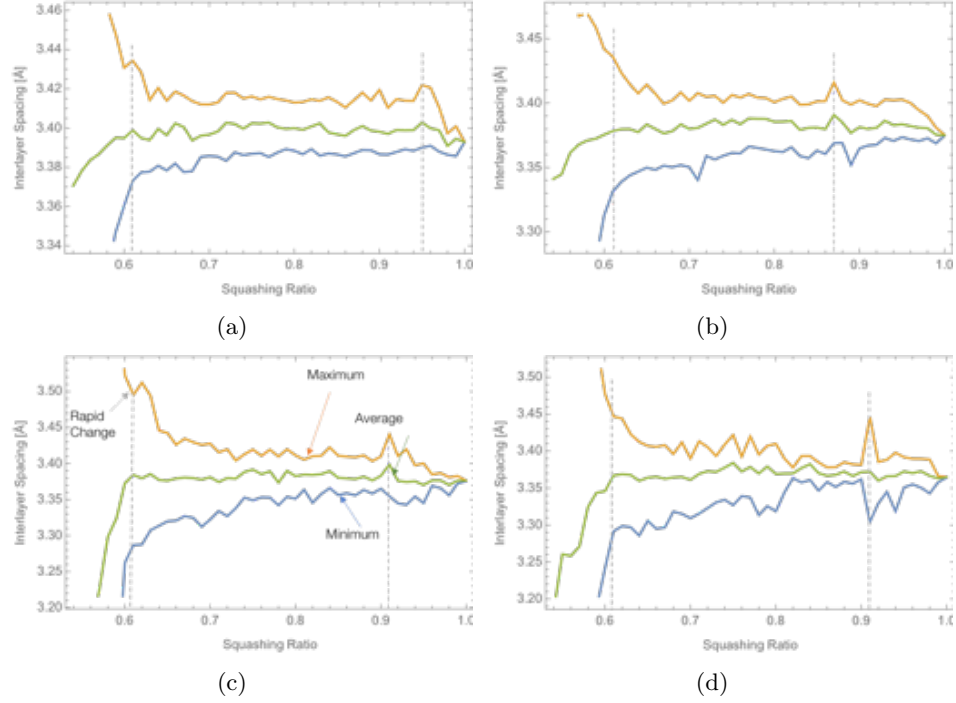


Figure 2.3: Interlayer spacing of deformed multi-layer graphite. 2.3(a) and 2.3(b) are interlayer spacing of double-layer and triple-layer graphene respectively. 2.3(c) is the outer and 2.3(d) is the inner spacing of 4-layer graphene. Orange, green and blue curves is representing the interlayer in the order of maximum, average and minimum.

The other noticeable feature of the data is the small peak around $\lambda = 0.9$, that means inter-layer spacing increases significantly at $\lambda = 0.95, 0.87$ and 0.91 for 2-, 3- and 4-layer respectively except for minimum of inner spacing in 4-layer structure, which decreases. This phenomenon is resulted by layer shifting along the curve (2.4(a)). While the two atoms in two layers getting closer in any direction, two layers would be forced to open up spacing between them relatively. Take 3-layer as example, two atoms at each of two neighbored layers around the vertex were chosen. Their non-Euclidean distance projected on the curved surface is used to show how layer shifts. Matching with the location of peak in spacing size, there is an obvious drop at $\lambda = 0.87$ exactly, which implies shifting causes

the distance between the two atoms around vertex reduces along curve if without interlayer spacing change. Therefore, these two atoms are tending to get away from each other, so the spacing especially at vertex would enlarge. Correlation energy between layers is another lateral proof, that while energy smoothly increases while λ decrease, there is a step around $\lambda = 0.87$.

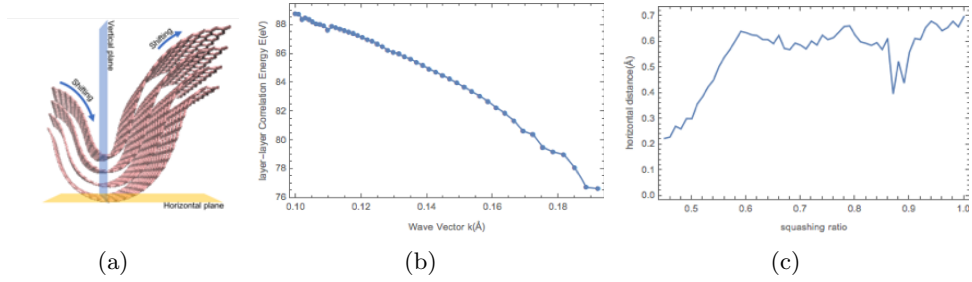


Figure 2.4: 2.4(a) shows the displacement of layers caused by shifting along curve. The tangent plane is horizontal at the vertex. 2.4(b) and 2.4(c) are result from triple layer graphene simulation. 2.4(b) shows that correlation energy of neighbor layers decreases while the level of rippling gets higher. 2.4(c) is non-Euclidean distance of two aimed atoms in two neighbored layers changing with λ , which could help understanding the shift in 2.4(a).

So, Generally speaking, the increase of the curvature could expand the maximum interlayer spacing and causes higher insertion possibility for all of quadruple, triple, double stacked layers respectively. With exact curvature, the effect could be more obvious. Besides, with less sheets paralleled stacked together, maximum spacing appears to be larger.

Considering the size of platelets varies in hard carbon as well, We had also run the same simulation on different original periodic box size. The number of unit cell is reduced to $20 \times 3 \times 2$, $15 \times 3 \times 2$, $12 \times 3 \times 2$, $10 \times 3 \times 2$ and $8 \times 3 \times 2$ comparing to the $30 \times 10 \times 2$ model. Given that data point number to fit in Fourier Function differs with size, the order of function changed from 15 for $30 \times 10 \times 2$ to 10, 8, 5, and 3 respectively. The summary of the statistical results in Table 2.1 is obtained with $\lambda = 0.6$. It is clear that spacing

especially the maximum increases comprehensively, in spite of minimum spacing decreases slightly.

	Max	Min	Ave
$30 \times 10 \times 2$	3.4384	3.3652	3.3961
$20 \times 10 \times 2$	3.4533	3.3465	3.3919
$15 \times 10 \times 2$	3.4792	3.2868	3.4007
$12 \times 10 \times 2$	3.6606	3.1890	3.4666
$10 \times 10 \times 2$	5.4321	3.0734	4.2690
$8 \times 10 \times 2$	5.7284	3.1171	4.2462

Table 2.1: Interlayer Spacing(\AA)

To prove larger interlayer spacing indeed able to increase ionic capacity. Sodium ion is deposited in different locations between two layers of a 0.7-ratio deformed $8 \times 3 \times 2$ structure. Then it was optimized with density functional theory simulation to get more accurate energy. The highest binding energy of Na ion in the rippled graphene model, which is 0.3603 eV, is found exactly at the vertex where the maximum inter-layer spacing occurs(Figure 2.5).

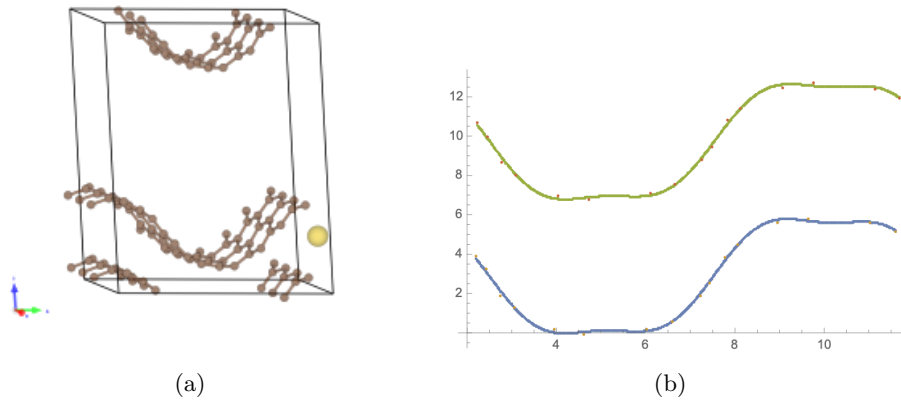


Figure 2.5: Showing the atomic structure of rippled graphene after insertion. 2.5(a)The brown atoms are carbons. And yellow atom in the figure representing sodium ion is at the most favorable sites for insertion of a 0.70 rippled $8 \times 3 \times 2$ bilayer graphene unit. 2.5(b) is fitted curve of structure, with largest inter-layer spacing locates at vertex as well.

By comparing the spacing before and after insertion, we found that the ion insertion will slightly further expand the spacing between layers, that implies the mechanism of sodium insertion in rippled graphene is similar to the lithium interaction found experimentally[39, 55]. The maximum, minimum and average spacing changed from 5.1266 Å, 3.1010 Å and 4.2462 Å to 5.2785 Å, 3.1645 Å and 3.9732 Å respectively.

2.4 Conclusion

Inter-layer spacing directly related to insertion energy of ion in graphene layers. Therefore, larger spacing makes hard carbon capable as insertion host for sodium[51], and gives SIBs chances for competitive or better performance than LIBs. One of the decisive factor is that platelets in disordered carbonaceous structure are bent unlike graphite. As discussed, if sizes, numbers of layer and curvature of small platelets in hard carbon could be altered, inter-layer spacing should be possible to control experimentally. The investigation on how the features affect interlayer spacing gives insight on further development of high performance electron materials for sodium-ion batteries.

Chapter 3

Solvation Energy Difference in LiCl Solution at Different Temperature

Aqueous electrolytes typically suffer from poor electrochemical stability; however, 25 wt.% LiCl (aq.) cooled to $-78\text{ }^{\circ}\text{C}$ is stable over 2.87 V, nearly double its stability window at room temperature. Integrated experimental and simulation results reveal that, upon cooling, Li^+ ions become less hydrated and pair up with Cl^- , ice-like water clusters form, and $\text{H}\cdots\text{Cl}^-$ bonding strengthens. Surprisingly, this low-temperature solvation structure does not strengthen water molecules' O–H bond, bucking conventional wisdom that increasing water's stability requires stiffening the O–H covalent bond. We propose a more general mechanism for water's low temperature inertness in the electrolyte: less favorable solvation of OH^- and H^+ , the byproducts of hydrogen and oxygen evolution reactions. The discussion on atomic level is accomplished by AIMD simulation combined with structural analysis to compare local environment of each component at two temperature setups. .

3.1 Background

Water is an amazing solvent; it is ubiquitous, readily dissolves a wide range of salts and remains liquid over a wide temperature span—properties that make water the essential biological solvent for life on earth[52]. As a solvent for battery electrolytes, water has the added advantages that it is inexpensive, non-flammable, and non-toxic. However, for batteries, water has one glaring disadvantage; the voltage for electrolysis of water in aqueous electrolytes is typically small, on the order of only 1.4 V. This fundamentally limits the operating voltage—and ultimately the energy density—of any batteries that use them[35, 61]. To put this into context, Li-ion batteries typically operate at around 4 V, and so to date these have predominantly relied on organic liquid electrolytes which are expensive and highly flammable. In this work we show that water’s poor electrochemical stability arises from water being too good a solvent, and this understanding provides a route to engineer new aqueous electrolytes with dramatically expanded electrochemical stability and has enabled us to demonstrate a low-temperature aqueous Li-ion battery[37, 53, 17].

We have investigated the electrolyte of 25 wt.% LiCl in water, a eutectic mixture that remains liquid down to -78 °C, allowing access to the anomalous behaviors of water at low temperature. In this highly concentrated electrolyte, there are roughly seven water molecules per Li⁺ ion; therefore, most water molecules are part of the first and second hydration shells of Li⁺. It is expected that water’s coordination to Li⁺ significantly weakens the water covalent O–H bonds. This is indeed the case at room temperature where the solution exhibits a narrow electrochemical stability window of 1.44 V. When the temperature is lowered to -78 °C, the structure of the solution changes to form more tetrahedral water

clusters, $\text{Li}^+\text{--Cl}^-$ pairs, and more directional $\text{O--H}\cdots\text{Cl}^-$ hydrogen bonding. Intriguingly, this is accompanied by a near doubling of the electrolyte’s electrochemical stability window. A widened stability window has been commonly attributed to the strengthening of the O--H covalent bond in water; however, our results show that the O--H bond is not strengthened. This contradiction indicates that the O--H bond strength is not the dominating factor in water’s stability at low temperatures. Instead, an often-overlooked factor in the electrolysis potential is dominant—the energy of the hydrogen and oxygen evolution reaction products, particularly the solvation energy of OH^- and H^+ . By chance, the local structural changes in 25 wt.% LiCl at low temperatures enables us to deconvolve the effects from O--H bonds and OH^- , H^+ solvation. The low-temperature structure silences the impact of O--H bonds but allows the solvation energies of OH^- and H^+ to play out, enabling us to attribute water’s expanded electrochemical stability window to the low-temperatures structure of the electrolyte being less hospitable for OH^- and H^+ solvation. Making water a less capable solvent increases its stability. This new understanding of how the local environment affects the electrochemical stability of aqueous electrolytes has profound implications across chemistry, physics, materials, and energy disciplines.

3.2 Methods

3.2.1 Computational Simulation

Ab initio molecular dynamics (AIMD) simulations were performed to reveal the structure of water and the solvation environment of the ions in both the 25 wt.% LiCl and 62 wt.% H_3PO_4 solutions at room and cryogenic temperatures. The 25 wt.% LiCl (62 wt.%

H3PO4) simulation was prepared by assembling 92 water molecules and 13 Li atoms and 13 Cl atoms (54 water and 18 H3PO4 molecules) in a square box with 12.5 Å sides using PACKMOL[33] (a 14.5 Å sized box for H3PO4) giving a density of 1.818 g/ cm³ matching that of the solution at room temperature.

In each case six different starting structures were created for each of these systems, and the results reported are the average of all of them. For the LiCl models, each of the assembled structures were simulated using classical molecular dynamics in the NVT (constant number N of particle, volume V and temperature T) ensemble for 3 nanoseconds (ns). From these trajectories, it was found that even at -78 °C the system had relaxed to equilibrium after 800 ps. The simulation was then switched to AIMD in the NVT ensemble with ab initio forces implemented to drive the dynamics. The system was simulated for 15 ps to equilibrate and a further 3 ps confirming convergence with the stabilized average potential energy. After this step, the system was simulated in the canonical ensemble for a further 5 ps period wherein the thermodynamic and structural properties were averaged.

The classical MD simulations were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[57] using the TIP4P41 potentials for water and with the Li⁺ and Cl⁻ ions modeled with a combination of the standard 12/6 Lennard-Jones potentials for van der Waals interactions and long range coulombic interactions. The Lennard-Jones potentials used energy and range of $\epsilon = 0.07647kcal/mol$, $\theta = 2.126\text{\AA}$, for Li⁺ and for Cl⁻, $\epsilon = 0.4928kcal/mol$, $\theta = 4.417\text{\AA}$. The long range tails of the coulomb interactions were modeled using Ewald summation (with tolerance, $ewald = 1.0E-5$).

The AIMD simulations were performed using the Vienna Ab initio Simulation Package (VASP) using the Perdew, Burke and Ernzerhof (PBE) generalized gradient approximation (GGA) for the exchange correlation functional, with projector augmented plane wave (PAW) pseudo potentials to model the atoms' core electrons. The AIMD simulations were performed using a time step of 0.5 fs. The H atoms in the system were modeled as deuterium (D). This is a well-established computational trick that moves the vibrational frequency of the light H atoms closer to those of the heavier atoms, thus enabling the use of a larger time step that speeds up calculations. At the very low temperatures that we model, the intramolecular vibrational modes would be frozen, and the water rigid. However, we wish to observe how the water molecules' O–H stretching frequency varies with their coordination environment to compare with the experimental FSRS data, so the H atoms were kept free meaning that the O–H bond vibrations were classically occupied. We used a frequency scaling correction when we computed the O–H stretching frequencies to account for the increased mass of the H atoms. The coordination environment of every water molecule was classified at each time step after equilibration with AIMD. The properties at each temperature reported in the main text were averaged from six different simulated systems with different initial random conditions.

Simulations for 62 wt.% H₃PO₄ were set up in a similar way but without the classical MD step. Instead, the generation of random arrangements using PACKMOL was followed by structural relaxation, and then 10 ps of AIMD simulation in the NVT ensemble. This time was sufficient to equilibrate the average potential energy, and the structural and thermodynamic properties were computed over a final 5 ps of simulation

time. The properties of the 62 wt.% H₃PO₄ systems at each temperature were averaged from six different simulated systems with different initial random conditions.

3.2.2 Water Diversity

Figure 3.1 shows a schematic of different bonding configurations available to a water molecule in LiCl aqueous solutions. Unlike dilute electrolytes or pure water, in these concentrated solutions the H atoms of each water molecule can act as a proton donor (D) in H-bonds to either an oxygen in another water molecule or a chloride ion. Similarly, a water's O atom can act as an acceptor (A) to either an H atom or a Li⁺ ion. However, we propound that if an O atom is serving as a double acceptor (AA) it is more likely to accept two protons rather than two Li⁺ ions or a Li⁺ and a proton. This is because when a water molecule accepts a Li⁺ ion, the bond angle to another accepted H or Li⁺ forces it to encroach on the Li⁺ ion's first coordination shell—a situation that is highly unfavorable due to strong electrostatic repulsion.

To identify the four types of water, H-bonding between water molecules are defined the same way in Chapter 4. And to distinguish the oxygen or hydrogen coordinated to other ions, they are simply classified by distances, with the threshold chosen based on structural factor – coordination number. Accordingly, all the hydrogen as doner and oxygen as acceptors are found to figure which one of DA, DAA, DDA or DDAA catagory one water molecule falls into.

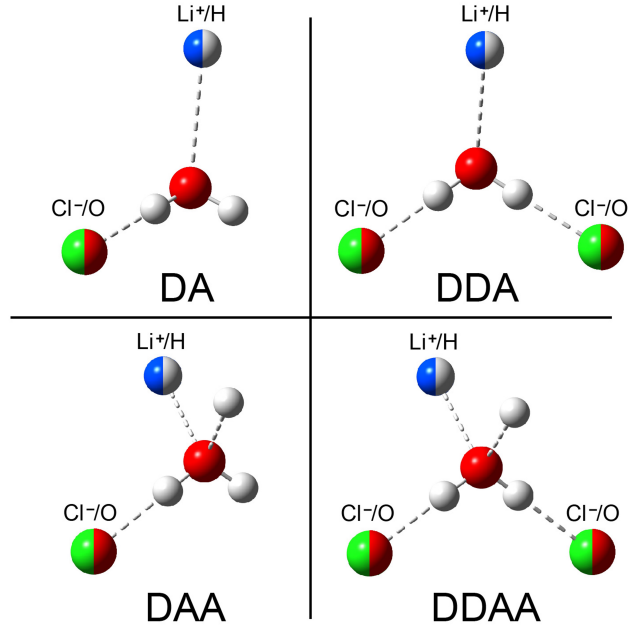


Figure 3.1: Schematic of possible bonding configurations available to water.

3.2.3 Coordination Number

Partial pair distribution function was calculated from the structural models using the relation with aiming pairs of atoms

$$G_c(r) = \frac{1}{r} \sum_i \sum_j \frac{b_i \times b_j}{\langle b \rangle^2} \delta(r - r_{ij}) \quad (3.1)$$

where i and j refer to two different types of atoms, b_i is the scattering power for atom i and $\langle b \rangle$ is the average scattering power. Similarly, the range-dependent coordination number, $n_{AB}(r)$, for the number of atoms of type A neighboring atoms of type B were computed

using a sum of error functions

$$n_{AB}(r) = \frac{1}{N_B} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{1}{2} (1 + \operatorname{erf}(\frac{r - r_{ij}}{\sigma})) \quad (3.2)$$

using a value of $\delta = 0.05\text{\AA}$, with N_X as the number of atoms of type X.

3.2.4 Solvation Energy

To observe the trend in the solvation energy of a proton in 25 wt.% LiCl solutions between room and cryogenic temperatures, a proton was added to the equilibrated structures of the simulated stoichiometric electrolyte and the AIMD simulation was continued to allow the system to equilibrate. For each simulated electrolyte structure, two simulations were performed with the proton inserted in different initial positions. The insertion sites were chosen manually to pick a chemically reasonable starting configuration. In all cases the a few steps of minimization were performed with the existing atoms frozen, to relax the position of the inserted proton, and then the full system was equilibrated with AIMD in the NVT ensemble using the same procedure as before. During the equilibration process we observed Grotthuss proton hopping as the excess proton found a low energy configuration coordinating between a water molecule and a chloride ion.

The trend in the solvation energy of OH^- was computed in a similar way by starting with the equilibrated stoichiometric electrolytes, but in this case by removing a proton from a randomly chosen water molecule that was a non-aqua ligand. To complement these calculations, we also computed the energy change for removing one of the water molecules that was not coordinating Li^+ . The energy of each system was found by averaging

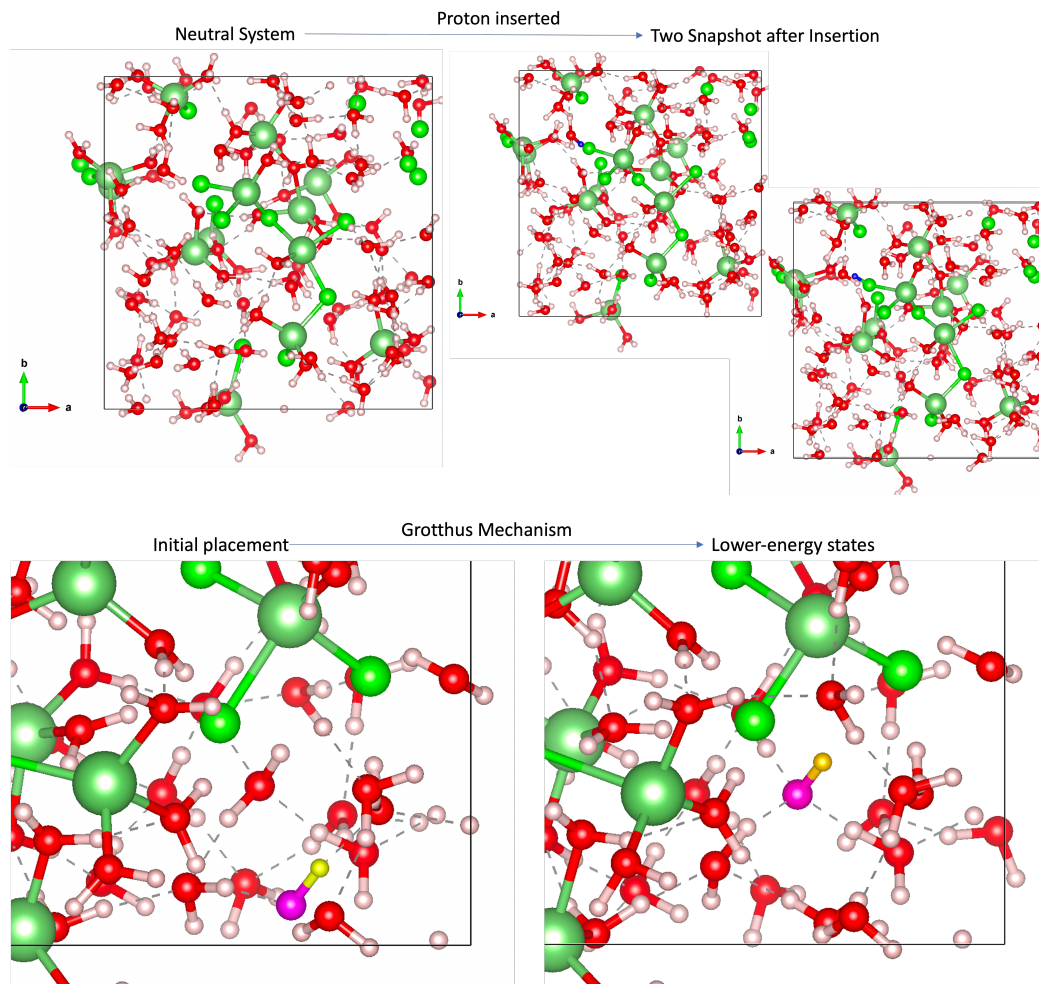


Figure 3.2: Ion insertion operation and resulted optimized structure.

the energy over 5 ps after the system had equilibrated.

3.3 Results and Discussion

To fully understand the coordination configurations of water molecules in 25 wt.% LiCl, we performed ab initio molecular dynamics (AIMD) simulations. Figure 3.5 shows the cumulative distribution of acceptor bonds that are made from O to either H or Li+,

and this has been plotted separately for the O in water molecules that are aqua ligands of Li⁺ and water molecules that are not solvating Li⁺. In room-temperature electrolytes, experimental results indicate that there are an average of 1.26 acceptor bonds per water. In corresponding AIMD simulations at room temperature, *sim*52% of water molecules are Li⁺ aqua ligands, and the average of 1.26 acceptor bonds per water occurs at a distance of $\sim 2\text{\AA}$; however, at this distance the number of acceptor bonds per aqua water molecule is only 1.02. Thus, the DAA and DDAA populations of water are predominantly free from Li⁺.

With observation that more DDAA or DAA water appear at high concentration of LiCl in water at room temperature, it is safe to assume the presence of Li⁺ is one of the key factor causing the decrease of ice-like structure in the solution. This is proven by AIMD simulation result that we can see clearly less DDAA (labeled blue in Figure 3.3) water in RT(room temperature) configuration while at the meantime more water molecules have chance coordinating to Li⁺ because of the reduction from 11 to 8 in Cl⁻ number bonded with Li⁺. As well, about twice number of Cl are found coordinating to two instead of one Li atom at room temperature, which leads to even less site for water molecules to form H-bonding with Li⁺.

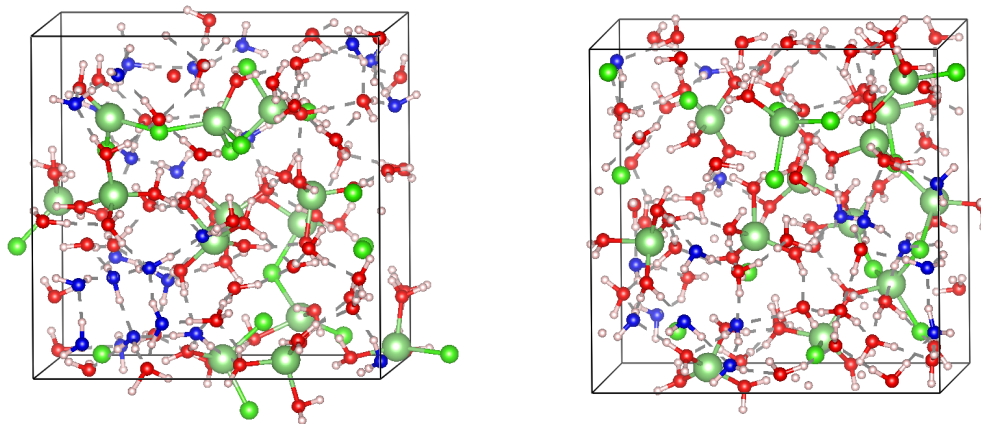


Figure 3.3: DDAA water molecules under two different temperature.

Upon lowering the temperature, the populations of DA, DDA, and free water decrease linearly while the proportions of DDAA and DAA rise. Aqua ligands of Li^+ are members of the DA or DDA populations (Fig. 2d), and the reduction of these populations indicates dehydration of Li^+ ions' hydration shells. The water molecules that are released from Li^+ ions are then free to accept H from two neighboring water molecules and contribute to the increase of DAA and DDAA configurations (reminiscent of low-density liquid water)[23].

The AIMD simulation results indicate the notion of dehydration from the Li^+ ions' coordination shells at low temperatures. From room temperature to $-78\text{ }^\circ\text{C}$, the hydration number of Li^+ decreases by ~ 0.8 in its first coordination shell (3.4). It also reveal that as water vacates the Li^+ hydration shells it is replaced by Cl^- ions, thus leading to the formation of partially hydrated Li-Cl complexes such as $[\text{LiCl}_x(\text{OH}_2)_{4-x}]_{1-x}$. From 20 to $-78\text{ }^\circ\text{C}$, the average number of Cl^- coordinating Li^+ increases from 0.2 to ~ 1 at 2.5 \AA ,

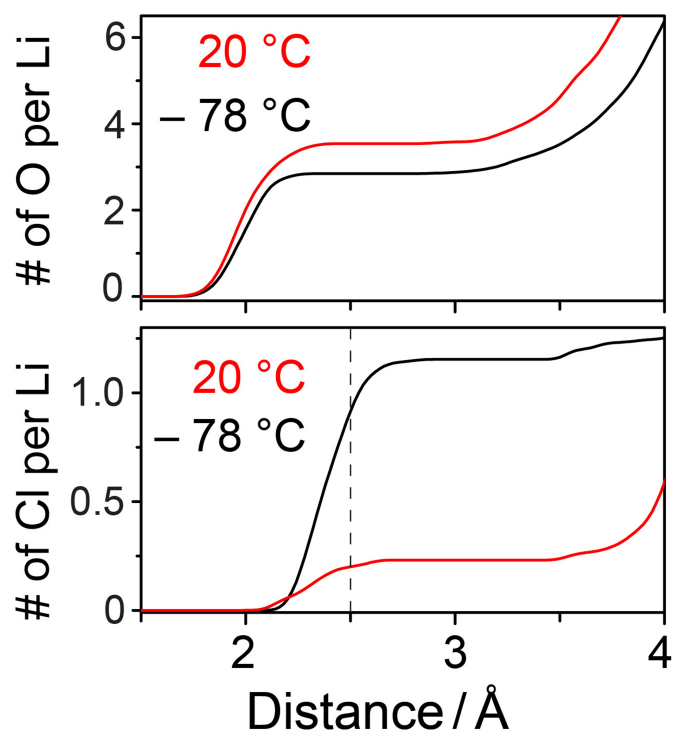


Figure 3.4: (g) and (h) The average numbers of water's O coordinating Li^+ and Cl^- coordinating Li^+ as a function of distance revealed in AIMD simulations of 25 wt.% LiCl .

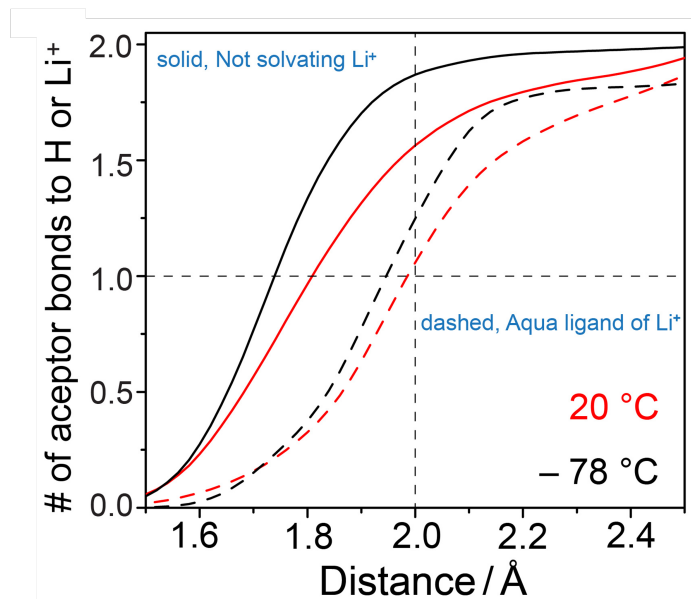


Figure 3.5: Number of acceptor bonds to H of water or Li^+ as a function of distance from AIMD simulations. Solid lines are about the oxygen of water molecules that are not aqua ligands, and dashed lines denote the oxygen of water molecules that are aqua ligands of Li^+ .

a typical ion-pair distance (3.4). Interestingly, in these complex ions the rarely observed bending mode of water becomes more polarizable (mainly due to adjacent Li^+).

The evidence of Li^+ and Cl^- pairing up at low temperatures accompanied by an increase in the occurrence of $\text{H}\cdots\text{Cl}^-$ bonding would seem paradoxical. The number of Cl^- ions in the solution is fixed, and coordinating to Li^+ ions inhibits coordination by H. Our AIMD simulations resolve this paradox. Fig. 2k shows the average distributions of O and H around Cl^- ions at two temperatures. Cooling from room temperature to $-78\text{ }^\circ\text{C}$, the average O count in the region from 2.5–3.25 Å from the Cl^- ions drops by 0.45, consistent with Cl^- pairing with Li^+ . However, while there are fewer O atoms surrounding Cl^- , the H atoms bonded to those O atoms move closer to the Cl^- (solid lines in Figure 3.6). At $-78\text{ }^\circ\text{C}$,

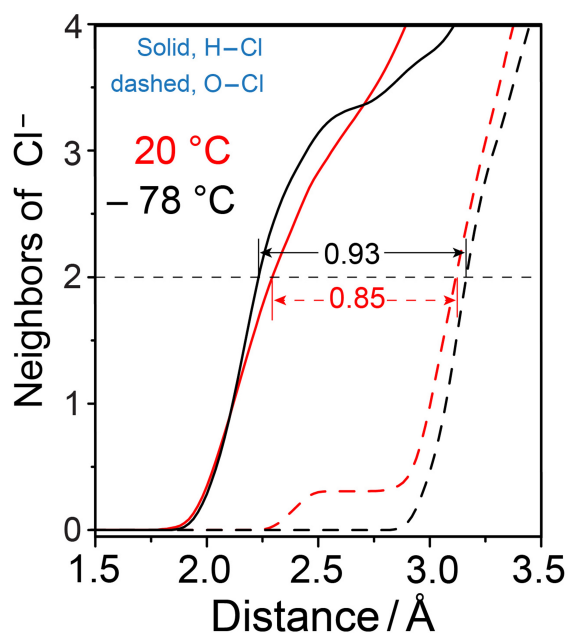


Figure 3.6: The average number of O per Cl^- (dashed lines) and H coordinating Cl^- (solid lines) as a function of distance retrieved from AIMD simulations of 25 wt.% LiCl.

the difference between the distance from Cl^- to the first two surrounding O atoms and the distance to the same number of H is 0.93 \AA —close to the length of an O–H covalent bond. This indicates that, upon cooling, water molecules rotate to straighten and strengthen the O–H \cdots Cl $^-$ bond.

At room temperature, water serves primarily as aqua ligands of Li^+ ; at $-78 \text{ }^\circ\text{C}$, water clusters and $\text{Li}^+\text{--Cl}^-$ pairs form, and chloride-binding water molecules become more prevalent. The most significant local structure change is found. Then how does such a new chemical environment widen water’s stability window? In what way does it improve battery performance? Conventionally, the enhancement of water’s inertness is reasoned by the strengthened water’s O–H bond.

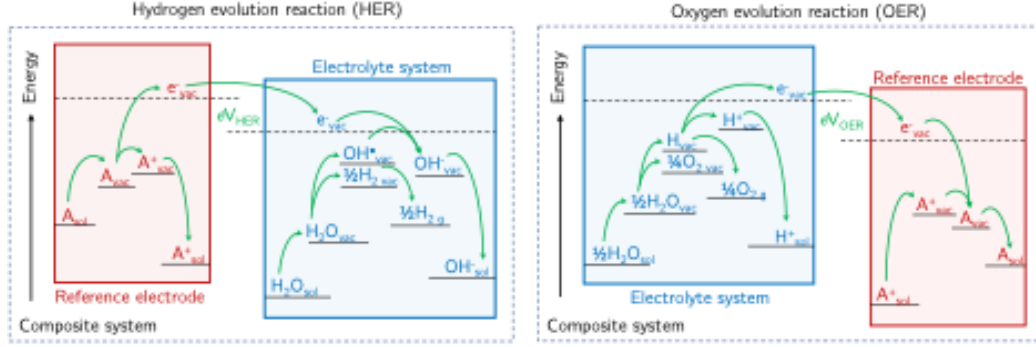


Figure 3.7: Key steps in the hydrogen evolution reaction (HER, left) and oxygen evolution reaction (OER, right).

While the stability window is related to the covalent bond strength of water, the chemical environment for HER and OER products could play a more important role. Thus, we must understand the detailed reaction energetics of HER and OER (Figure 3.7).

For the HER, the products are $\frac{1}{2}H_2 + OH^-$. Considering the thermodynamic cycle occurred on a cathode[43], V_{HER} is derived as:

$$V_{HER} = -\frac{1}{e^-}(\Delta G_{(sol,H_2O)} - \Delta G_{(sol,OH^-)} - \frac{1}{2}\mu_{H_2} + E_{H_2O}) + V'_A \quad (3.3)$$

$$V_{HER} = -\frac{1}{e^-}(\Delta G_{(sol,H_2O)} - \Delta G_{(sol,OH^-)} - \frac{1}{2}\mu_{H_2}) + V_A \quad (3.4)$$

Here $\Delta G_{(sol,H_2O)}$ and $\Delta G_{(sol,OH^-)}$ are the energies for solvating a single molecule of H₂O or OH⁻, respectively, and E_{H_2O} represents the bond enthalpy in vacuum. The term μ_{H_2} is the chemical potential of the evolved H₂ gas, e^- is the electron charge (which is negative), and V'_A is a constant which is independent from the temperature and conditions in the electrolyte. The upper equation encompasses the contributions from the solvation energies, and the O–H bond enthalpy, E_{H_2O} , of a water molecule in isolation. However, the

vibrational modes of water are so stiff that these modes are frozen out at room temperature, and thus the change in the thermal contribution to the bond enthalpy between 20 °C and -78 °C is less than 10^{-5} eV, and thus E_{H_2O} can be considered a constant. Therefore, the HER potential can be simplified as the second equation, where $V_A = V'_A - \frac{1}{e^-}E_{H_2O}$. On the other hand, for OER, the products are $\frac{1}{2}O_2 + 2H^+$. Following a similar Gedanken reaction path as the HER, V_{OER} is derived as:

$$V_{OER} = -\frac{1}{e^-} \left(\frac{1}{2} \Delta G_{(sol, H_2O)} - \Delta G_{(sol, H^+)} - \frac{1}{4} \mu_{O_2} + \frac{1}{2} E_{H_2O} \right) + V'_B \quad (3.5)$$

$$V_{OER} = -\frac{1}{e^-} \left(\frac{1}{2} \Delta G_{(sol, H_2O)} - \Delta G_{(sol, H^+)} - \frac{1}{4} \mu_{O_2} \right) + V_B \quad (3.6)$$

wherein $\Delta G_{(sol, H^+)}$ is the solvation energy of H^+ , μ_{O_2} is the O_2 chemical potential, and V'_B and V_B are constants. Overall, the chemical stability window of water at low temperatures can thus be written as:

$$\Delta V_{H_2O} = V_{OER} - V_{HER} = \frac{1}{e^-} \left[\left(\frac{3}{2} \Delta G_{sol, H_2O} - \Delta G_{sol, H^+} - \Delta G_{sol, OH^-} \right) - \left(\frac{1}{4} \mu_{O_2} - \frac{1}{2} \mu_{H_2} \right) \right] + V_c \quad (3.7)$$

where $V_c = V_B - V_A \approx 17.1V$ is the electrochemical stability of a water molecule in vacuum, and our model predicted that ΔV_{H_2O} is 1.55 V, which is close to the experimental value of 1.44 V at room temperature. This substantiates the validity of our theoretical modeling to examine individual contributions from fundamental processes like the solvation of pertinent water, proton, and hydroxide species. Although μ_{O_2} and μ_{H_2} vary with temperature, they contribute only $\sim 0.1V$ to the widening of ΔV_{H_2O} at -78 °C.

The solvation energies $\Delta G_{(sol,H_2O)}$ and $\Delta G_{(sol,OH^-)}$, and $\Delta G_{(sol,H^+)}$ depend on both temperature and local structure of the electrolyte. At standard temperature and pressure, $\Delta G_{(sol,H_2O)}$, $\Delta G_{(sol,OH^-)}$, and $\Delta G_{(sol,H^+)}$ in pure water are -0.27 , -4.55 , and -11.37 eV/molecule, respectively[56]. Given its small magnitude, even a large fractional change in $\Delta G_{(sol,H_2O)}$ could not reasonably supply the $\sim 1.4V$ increase to ΔV_{H_2O} measured at -78 °C. The energy for solvation of water in 25 wt.% LiCl is larger than in pure water as around half of the water molecules are coordinating Li+, but there remains more scope for reduction in $\Delta G_{(sol,OH^-)}$ and $\Delta G_{(sol,H^+)}$. This leaves solvation of OH⁻ and H⁺ becoming less favorable as the structure of the electrolyte reconfigures at low temperatures as the major cause for HER and OER suppression.

At -78 °C, the 25 wt.% LiCl solution forms more ice-like DDAA water clusters that intrinsically makes it unfavorable to solvate ions (H⁺ and OH⁻) compared to that at room temperature. This local structure transformation leads to less negative (in eV unit) $\Delta G_{(sol,OH^-)}$ and $\Delta G_{(sol,H^+)}$, significantly reducing the HER onset potential and increasing the OER potential.

To validate our new understanding, we performed additional AIMD simulations to study the energetics of removing H₂O or introducing H⁺/OH⁻ into the simulated LiCl solution at 20 and -78 °C. In addition to the simulating approach described in Method Chapter, the energy of a lone OH⁻, which cannot be directly inserted into our simulation box due to the hydroxide ion's large size, is modeled by the difference of energy of system with a missing H⁺ minus energy of the system with a missing H₂O.

Averaged energy:	LT (eV)	RT (eV)	RT-LT (eV)	Surrogate
Additional H+	-0.04	-5.9	-5.8	$\Delta G_{(sol,H^+)}$
H+ Removal(non-Li+ H2O)	4.5	2.0	-2.5	N/A
Non-Li+ H2O Removal	1.1	3.8	2.7	$-\Delta G_{(sol,H_2O)}$
OH^- insertion	-3.1	-8.3	-5.2	$\Delta G_{(sol,OH^-)}$

Table 3.1: Energy change resulted from ion insertion. LT represents low/cryogenic temperature set up while RT means room temperature. the last column quantities represent a surrogate for the thermodynamic quantities. They indicate the trends in the thermodynamic variable for which they are surrogate with cooling, but not their accurate values.

Because of the restricted box size, these calculations do not give the true solvation energy and they result in a significant compositional change in the electrolyte. However, we can view these energies listed in Table 3.1 as surrogates for the relevant terms that reveal the trend in solvation energies with the structural changes that arise from cooling the electrolyte. As the temperature is lowered, the energy required to remove a water from the box is reduced which would narrow the electrochemical stability window. However, there is a larger change in the energy to introduce H^+ or OH^- into the electrolyte which more than offsets the reduction in the water solvation, thus leading to an overall widening of the electrochemical stability window. Upon cooling, the solvation of water becomes less favorable as fewer water molecules coordinate to Li^+ . This would narrow the electrolytes' electrochemical stability window; however, this is more than offset by a stronger decrease in the solvation energy of H^+ and OH^- leading to an overall increase in stability. The results suggest that at low temperatures, 25 wt.% LiCl is a less capable solvent for H^+ and OH^- , which explains why the HER and OER are inhibited at low temperatures.

In summary, our results provide a more general and robust mechanism to explain the shifted HER and OER potentials of aqueous electrolytes and the widened electrochemical stability window from the energetic point of view, particularly for the solvation energies

of OH⁻ and H⁺, the side product of HER and OER reactions, across a broad temperature range. The synergistic AIMD results revealed the formation of more DDAA structures and a departure of aqua ligands from Li⁺ ions to allow Li⁺ to pair with chloride ions, in conjunction with the enhanced H⁺...Cl⁻ hydrogen bonding. Besides offering a new theory instead of a simply strengthened O-H bond argument, this unveiled chemical environment can effectively lower the solvation energies of OH⁻ and H⁺ ions and energetically suppresses both HER and OER, thus enhancing water's electrochemical stability—an outcome that enables the encouraging low-temperature performance in aqueous Li-ion batteries using 25 wt.% LiCl as the electrolyte.

Chapter 4

Percolation in Vacancy Network in Prussian Blue Analogs

4.1 Background

Prussian Blue and its Analog compounds (PBAs) have an open framework structure which has generated great interest as potential battery electrodes. The open structure enables transport and hosting of a batteries working ion with counter charges accommodated by reduction of cations in the PBA framework. It was recently shown that PBAs could be used as the electrode in batteries using protons as their working ions. In this case protons can be transported into the electrode through the crystal water inside the PBA using the Grotthuss diffusion mechanism enabling extremely fast charging and discharging of the battery. The outstanding rate performance of these batteries implies that there must exist contiguous chains of water in the PBA structures along which protons are transported.

Here we present the results of simulations that reveal that water uptake into vacancies leads to correlated vacancy arrangements that can provide such contiguous pathways.

With the general chemical formula as $M_x[M'(CN)_6]_z \cdot n H_2O$, PBAs usually adopt a phase with FCC structure (see Figure 4.1) when in a low alkalinity environment. The high compatibility of water in PBAs comes from the microporous structure as well even larger pore size caused by vacancies with considerable amount of both ligands and zeolitic water in the lattice. Each vacancy is proven enlarging a micropore with an effective diameter of approximately 8.5 Å that exceeds the distance between neighboring M' sites ($\frac{a}{\sqrt{2}} \approx 7.2$ Å). Hence, neighbored vacancies would provide larger micropores or even bulk microporosity for water molecules to form big clusters or water droplets. Apart from adsorbed water on surfaces, within PBA frame, water can be separated into two types – 1) zeolitic water in interstitial site; and 2) coordinated water that chemically coordinated with M ions. While in each unit cell, eight subunit cells are available for zeolitic water or alkaline ions to accommodate, the number and location of coordinated water are affected by $M(CN)_6$ vacancy fraction and distribution.

4.2 Methods

4.2.1 Monte Carlo Simulation

Vacant sites are presumed to be random in PBA frame with uncontrollable distribution, which makes the existence of percolation hardly predictable, therefore, the probability was modeled with Monte Carlo simulation at multiple vacancy fraction, to find the

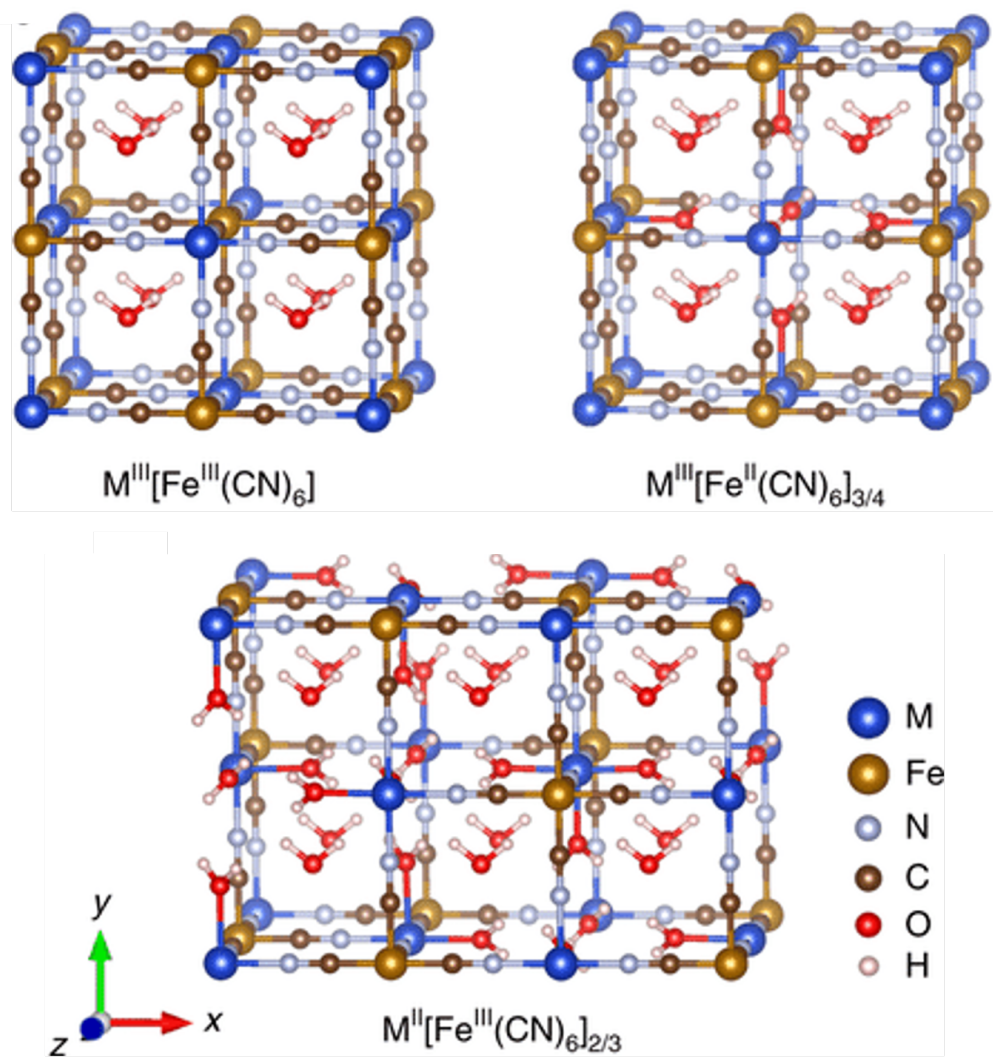


Figure 4.1: Example structure of PBA and PBA with vacancy.

appropriate number that offers more charge carrier transportation pipes. Starting with the initial structure without any defects, $M(CN)_6$ units were randomly removed to vacate spherical pores till reaching a given percentage to generate PBAs with various vacancy configurations.

At every step, one of the sites which originally has one $M(CN)_6$ residing in will randomly chosen with equal probability. And the selected site would be replaced with vacancy, or simpler speaking, the $M(CN)_6$ would be removed. If the chosen site is already a vacancy, then another site would be chosen following the same rule, and a swap would be taken action, that the previous one would become non-vacancy and the other being vacated if the the second site is not a vacancy yet. Otherwise, the configuration maintains unchanged at current step. The process above is repeated multiple times until vacancy fraction requirement is met.

4.2.2 Vacancy Connectivity Analysis

A cavity neighboring to another can be considered connected, implying a path for charge transportation being opened, based on which, the connectivity of a given PBAs atomic structure can be analyzed by graph theory with vacancies considered vertices and the 'tunnel' being edges. As defined, a system with most vacancies linked to a dominant graph is a high-connectivity one. At the meantime, less isolated components in the created graph representing a PBA frame leads to higher freedom for charge carriers to reside or diffuse.

One of the goals is to find the bottom line of vacancy occupation that guarantees the existence of percolation path. Thus, one hundred samples are created with random

insertion at each vacancy ratio respectively for average percolation probability calculation. For each new model generated from random vacancy insertion, the pipework would be analyzed with Depth First Algorithm to efficiently identify all subgraphs within the given system. All subgraphs with at least one node as vacancy on the box surface are identified accessible. And those involving vacancies on a pair of opposite sides of the cubic box can provide paths for percolation which plays the most important role in charge transportation. The box size was set as 100x100x100 unit cell to keep the randomness of arrangement and the robustness of the result while balancing with analysis resource limitation.

4.3 Results and Discussion

To discuss resulted transport path for proton hopping, the H-bonding network is still requested on top of vacancy connectivity analysis without promises on contiguous water chain directly from neighboring vacancies. Water in frame model were initialized with GCMC simulation associated with LAMMPS. The force field parameters for this specific MD process are calibrated by pure water model at room temperature with density as $997.77\text{kg}/\text{m}^3$ and average hydrogen bonding about 3.94. While all the frame atoms frozen during the whole simulation, insertion and deletion of TIP4 water molecules are performed under NVT ensemble to find the capacity and low energy state of water in PBAs.

To identify potential pathway, the connectivity of water network analysis and transferring action simulation are done to verify an open path in between two neighbored vacancies, where the acceptors and doners were defined with Vuilleumier and Borgis' definition. With the well-defined bonding threshold, a directed graph would be created for a

similar analysis as for vacancy connectivity. As shown in 4.4, a connected water chain can be found from one vacancy to another that is next to it for most of the cases. RAPTOR, as a Multi-scale Reactive Molecular Dynamics method, is designed to track chemical reactions in condensed phase. In our simulation, the reactions are defined with two types: one from Eigen cation to Zundel water and another the other way around. A RAPTOR simulation is accomplished by initializing one of the water molecules to a hydronium cation and relaxed under NVT.

The assumption of accessible vacancy guarantees proton transport to that area solely with Grotthus mechanism has been proven correct under common circumstances. From Figure 4.2(d), there are two examples of RAPTOR result, an inserted proton is transferred from dark blue labeled water to light blue ones. Random assignments were performed on 3 different vacancy confederations, and the conclusion claiming proton travels from outside of the box to accessible vacancies and from one vacancy to the others connected are further proven to be correct with RAPTOR simulations.

GCMC outcome gives an extra information that water capacity is not fixed to a number only decided by the fraction of vacancy (Table 4.1). In addition to the 6 extra water theoretically more caused by missing component, the extra large pore space and surface area actually slightly provides more site for water insertion. At the meanwhile, not only the net size differs with arrangement, the resulted connectivity can differentiate. In the third case we can clearly see, even though the water number is not significantly larger than the other two topology. However the connectivity increase is obvious. The path for proton transfer is shortened to half the length, which can potentially improve the battery efficient to another

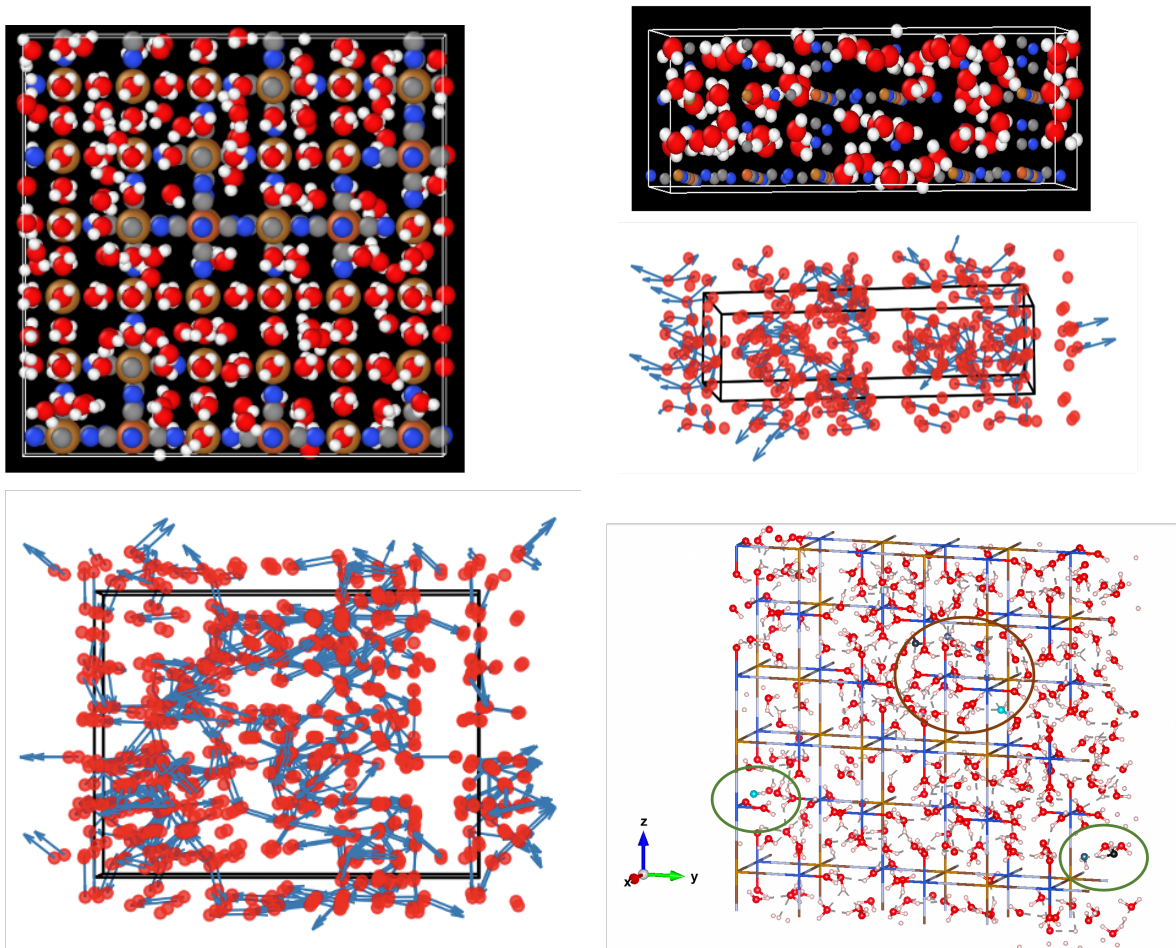


Figure 4.2: Two GCMC example results of two simulation box size: 4.2(a) and 4.2(c) for a $3 \times 3 \times 1$ system and 4.2(b) for smaller case which is $3 \times 1 \times 1$. Combined with the atomic structure, we can clearly find a path in the directed graph of water network topology from one vacancy to another if they are next to each other for most of the cases. 4.2(d) shows two transporting path found after proton insertion with RAPTOR simulation.

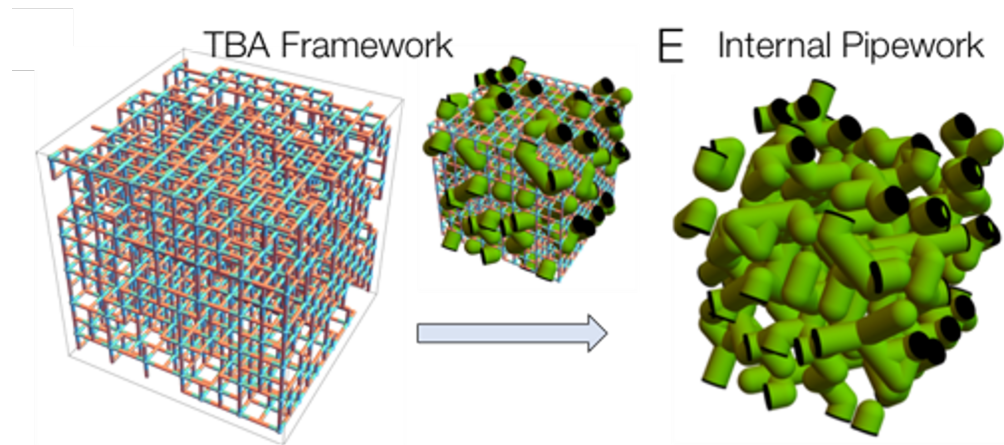


Figure 4.3: PBAs frame and pipe configuration.

level. Therefore, a cluster of vacancies are desired, which means the connectivity of vacancy network is wanted.

The aim is converted and simplified to a discussion of vacancy connectivity. The randomness of configuration is hardly maintained under periodic condition, therefore, a larger scale is highly required. However, even skipping atomic simulation, the graph itself can be supremely large to be analyzed, for example in Figure 4.3. Eventually, the cost and accuracy are balanced with percolation threshold, with the scale change making less significant change on results (Figure 4.4(a), 4.4(b)). As seen, at reasonable scale, we can confidentially say that percolation always appear while vacancy fraction larger than 20.20%. At the meanwhile, the inaccessible, who belongs isolated component in the graph are less than 20%. It is safe to say, that as long as the vacancy configuration is fully random and without any external or manual intervening, when the vacancy fraction exceeds 20.2%, the conductivity of proton in PBAs without alkaline ion insertion can be considered relatively good. The further change on vacancy number won't play the most important role on battery

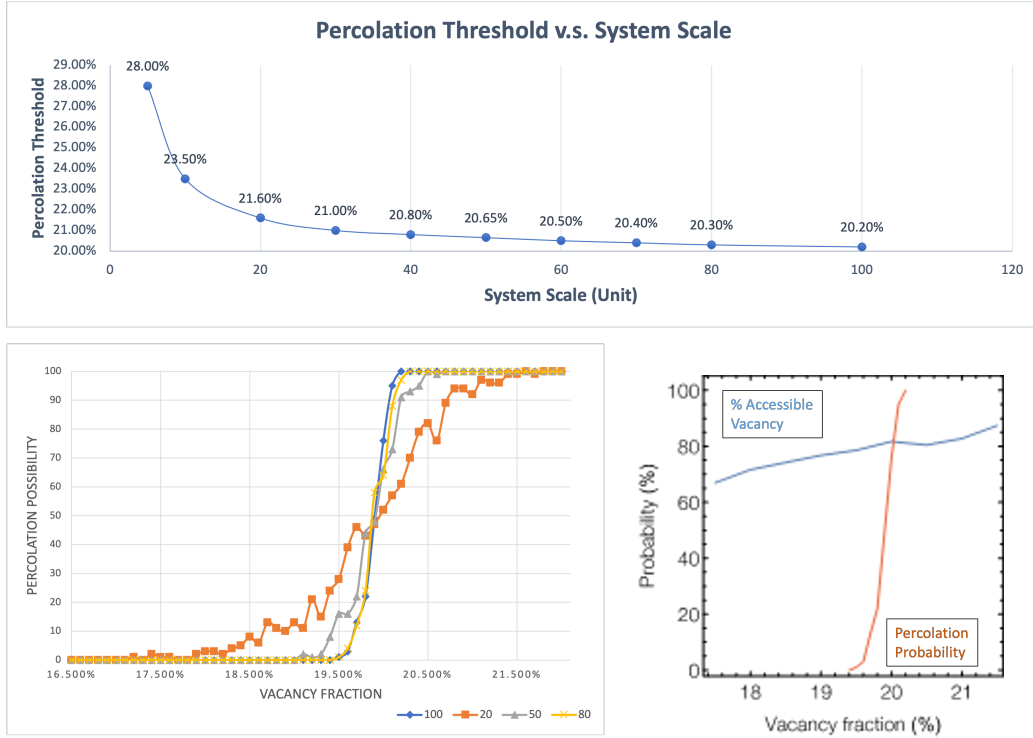


Figure 4.4: Vacancy connectivity analysis. 4.4(a) and 4.4(b) showing percolation thresholds at different system scale and percolation probability change with vacancy fraction. 4.4(c) shows that accessible vacancies change much smoother than percolation probability

performance optimization, but instead the arrangement is more worthy to be discussed. If future technology allows controlling vacancy correlating positions, battery charging speed is expected to be accelerated.

	Water Capacity	H-bonding per water	Chain length
Case 1	290	1.314	20
Case 2	294	1.327	21
Case 3	300	1.44	9

Table 4.1: Water and hydrogen bond number comparison of three vacancy topology with same number of vacancies.

Chapter 5

Efficiently Forecasting Organic Ligands Mechanical Properties From Atomic Structure via Deep Learning

Great potential for gas separation or storage has been found in Metal-Organic-Frameworks(MOFs) for their high flexibility of linker under pressure change.[7, 13] The search and design of metal organic frameworks and other coordination polymers highly rely on efficient methods discussing the properties of ligand molecules from their structure or bonding topology because of the high variety. Otherwise, the motion simulation of the large amount of molecules could be time-consuming given the complexity of organic molecule structure and high diversity in combination of building units and functional groups.[12].

To introduce machine learning as the predictor, with proven capability finding hidden correlation effectively, an encoding solution of molecular structure quantitatively is required, so that it may be used as inputs to the molecules behaviors forecast model. We will present the suitability of four different molecule encoding schemes for the prediction of two different classes of mechanical properties. The work demonstrates the possibility to predict, or forecast, the mechanical response ligands with a high degree of accuracy, and that there is not one best structure encoding which is universally optimal for forecasting all mechanical properties.

5.1 Background

Cheminformatics in chemical or bioactivity prediction of molecules from atomic-level structure have had a significant impact on many aspects of chemistry, from the discovery of new pharmaceuticals[30, 15] to the design of catalyst,[63] and elucidating reaction mechanisms.[48] Data-driven approaches eliminate the need to conduct a broad experimental survey of molecules' properties by replacing with forecasting based on the known data of materials structurally or chemically similar to the research subjects. Artificial intelligence has been introduced to learn quantitative structure–activity relationships (QSAR) with high correlation factors in the condition that lack theoretical understanding on the underpinning mechanisms[36] to accomplish and accelerate high-throughput screening on molecules with desired properties to save the costs of long simulations using realistic structural models or experiments considering the huge number of candidates.[3]

Success on molecular property forecasting hinges on possessing a suitable knowledge base to choose the algorithm model, and more importantly an encoding scheme able to quantify the structural similarity of molecules, and discriminating the structures active in the structure-property mechanisms underpinning the property(s) of interest. For many applications, the mechanisms may not be known *a priori* and so it is desirable for the encoding scheme to be general and agnostic in its recording of structure without missing key factors while symbolizing. From the previous researches on many chemical or bioactive properties, quantifying similarity of molecules through structural descriptor has been proven sufficient, for example, fingerprint catalogs all the unique chemical moieties exist in one molecule.[11, 16, 62, 27] However, predicting mechanical properties, such as bending, twisting, vibrating and mechano-isomerization, of a molecule poses a bigger challenge which depend on the local chemistry and bonding topology of moieties within a molecule, as is the case for chemical activity, but also depend strongly on the *location* of these moieties within the molecule’s overall geometry.

The work is motivated by the desire to accelerate design of new flexible MOFs by providing the ability to predict the likely mechanical properties of a candidate ligand molecule without computational overhead on detailed atomic simulations. We have developed an approach for automated design of molecules in which candidate molecules are created by starting with a seed molecule and descending a decision tree for the application of a series of molecule editing rules.[32] By formalizing the molecule design-space as a decision tree our approach makes the task of molecule design amenable to advanced tree searching algorithms. These computationally designed molecules could be screened by per-

forming atomic simulations to evaluate each candidate’s properties or performance for a particular application. However, the design space for molecules is astronomically large — encompassing 1×10^{15} unique molecules by only the tenth level of the design tree. It is not tractable to simulate every candidate in such a vast pool of molecules, for that only a deforming process along one direction would cost hours with molecular dynamic simulation. To screen any meaningful fraction of the candidate pool requires methods that can rapidly *forecast* the properties of a candidate molecule from descriptors of its structure.

Four approaches encoding molecular structure were implemented and compared, and their suitability as machine learning algorithms input were discussed in two cases forecasting different mechanical behavior of tethered ligand molecules — molecules that could be for example the linker of mechanically active metal-organic frameworks (MOFs). The four approaches to describe molecular structure that being used are: i) geometric moments of the molecule, ii) the point cloud of atomic coordinates, iii) a voxelized representation of the atomic coordinates, and iv) the molecules’ SMILES grammar. This is the first time for approach 1 and 2 examined for their potential utility in chemical informatics predictions, and while the latter two approaches are well established molecular descriptors they have not been applied to forecasting mechanical behavior. To test these structure-encoding schemes’ ability capturing the structure-property physics relevant for mechanical behavior, two different mechanical properties are considered based on different aspects of a molecule’s structure: stiffness in axial loading, and the ability of a molecule to mechanically isomerize. The former depends predominantly on global geometry, while the latter depends on the presence of local structures that function mechanically as hinges or revolute joints.

5.2 Methods

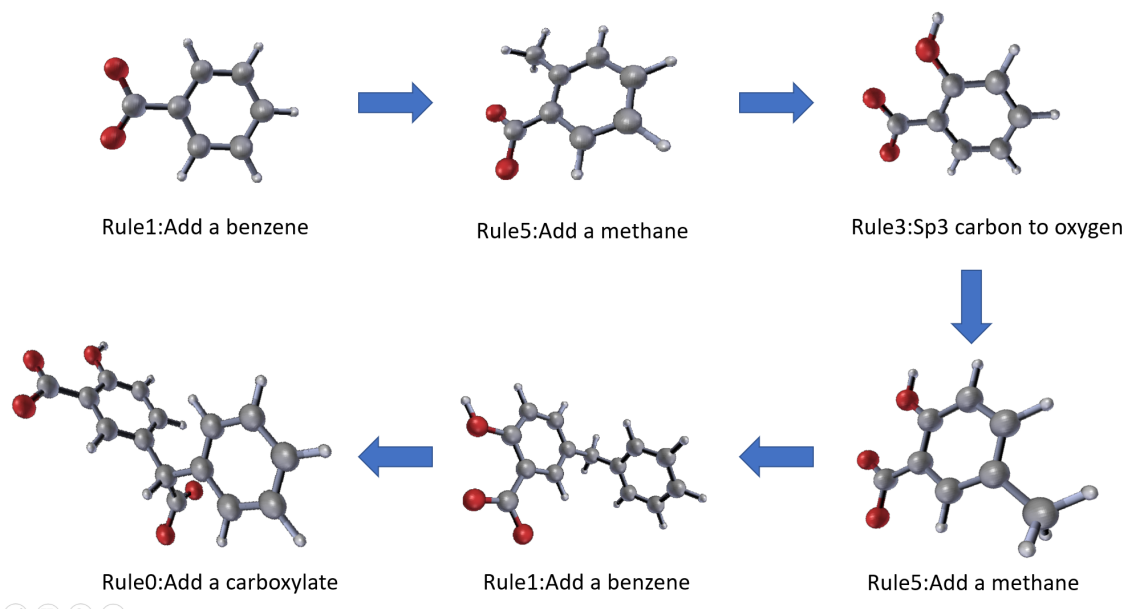


Figure 5.1: A example sequence of design steps for automated generation of a linker molecule: Starting from the carboxylate seed molecule editing rules are selected at random and applied to the molecule. Between each editing step the incomplete linker is structurally relaxed, and the list of editing rules that can be applied to the new partial molecule are determined. In the final editing step a second carboxylate group is added that acts as the candidate linker's second connection point in a MOF.

Database in this work is generated with the automated design scheme (shown schematically in Fig. 5.1) starting with one carboxylate as seed and finished by adding the second carboxylate with at maximum 6 rules applied, with which a dataset of 171,391 ditopic ligand molecules is generated with molecule size ranges from 9 to 57 atoms. "Linkers" designed with this process are not necessarily compatible in a MOF in this research to maintain high variety in the database as the purpose is to identify if any ligand a good candidate for flexible MOF. That is, that the backbone of the linkers can be curved, and the carboxylate ends need not be co-linear or aligned. The rationale for this is that our

goal is to test descriptors of structure, and we therefore seek to maximize the structural diversity in our data set.

To computationally characterize elastic properties of these ligand molecules, molecular dynamics simulations of the molecules were performed undergoing quasistatic axial deformation; deforming along the backbone axis after structure optimization, in which, interatomic bond interaction were modeled with a harmonic-style equation, pair interaction with standard 12/6 Lennard-Jones potential equation, and the angles, dihedrals and impropers in Fourier style with the parameters from Universal Force Field (UFF).[47, 44]

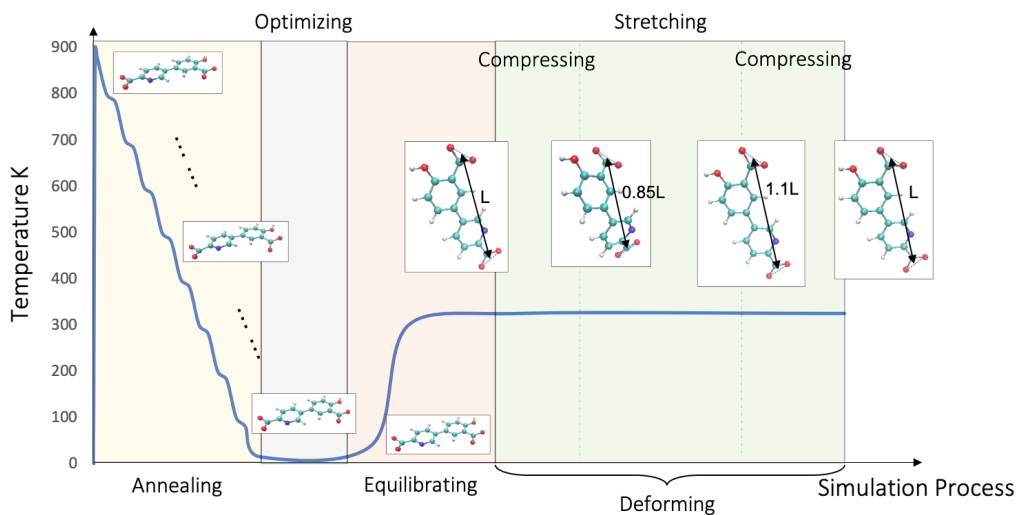


Figure 5.2: The simulation process. The whole experiment is separated to four parts: annealing, energy minimizing, equilibration, and deformation.

As in Figure 5.2, the structural relaxation was obtained starting with an energy minimizing relaxation, and a following annealing step in the canonical ensemble (NVT) with temperature rapidly increasing to 900 K (with bond breaking not permitted) and then going through a series of cooling steps followed by a 50-picosecond dwell until the system

reached 5 K at which point its structure was relaxed using the FIRE algorithm. [5] Given that deformation process would be simulated at 300 K, the entire system then would be heated back to 300 K and held there for a further 50 ps to ensure the system was well equilibrated. This rather rigorous and involved structural relaxation was found to be essential for the reproducibility of the deformation process to follow. To monitor stiffness during crushing and stretching, ligands are deformed at low speed which is $0.005\text{\AA}/ps$ by increasing and decreasing distance between two carboxylates in the range from 85% to 110% of original molecule length. Ligands are firstly compressed, then stretched, and finally recovered to original length as shown in Figure 5.2. Forces parallel to the deforming direction is the component required to accomplish deformation, which would be used to further calculate stiffness. Given that these ligands are going to be coordinated to metal clusters and build frameworks, their stiffness is defined as force change over strain percentage instead of length change itself.

As seen (Figure 5.3), stiffness of one ligand is not remaining the same while the shape changed. Based on the stiffness trend during whole deforming, molecular behavior is featured with 4 factors to discuss elastic properties for compressing and stretching process respectively, which are the stiffness of compressed state (Stiffness 2) and tensile (Stiffness 1) status, stiffness at original length, and whether mechano-isomerization could be found during the deformation. In advance of the calculation, the noise of real-time force-strain curve was reduced with univariate spline reducing noise with automatic assigned parameters ensuring with different length and numbers of atoms, all data were post-processed appropriately without losing information, despite their different levels of thermal noise.

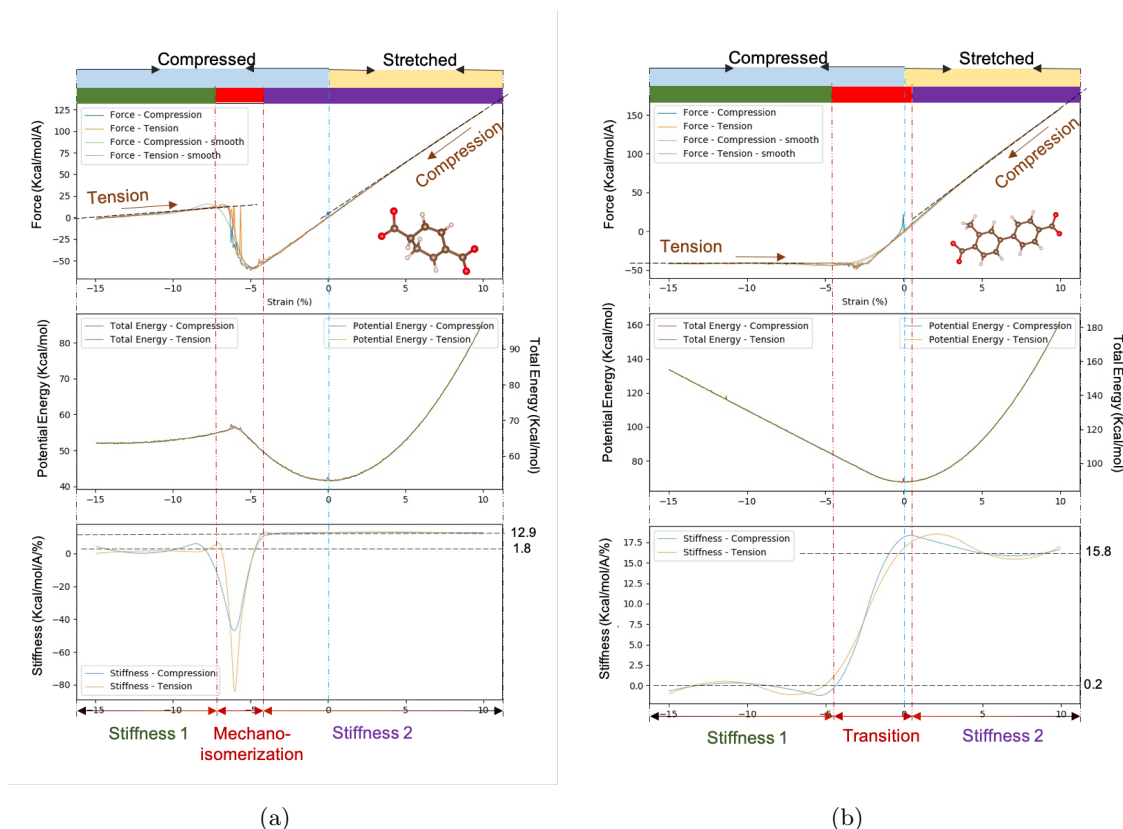


Figure 5.3: Example molecules, 5.3(a) molecule with mechano-isomerization which could be found by obvious drop or impulse of stiffness and 5.3(b) a typical molecule that geometry and as well stiffness changing smoothly during deformation, in the 3rd depth with 3 rules applied. Three subplots from top are the force-, energy-, and stiffness/length-strain curves respectively. Stiffness curve is the derivative of force curve.

Energy-strain curve reflects that the optimization was done successfully finding lower-energy status. Besides, the paralleled total energy and potential energy curves imply that these deformations are slow enough that the temperature and kinetic energy is not fluctuating significantly in each step. From the stiffness changes, most of designed ligands were found with the deformation process composed by two sections with an significant difference of stiffness, whose average are calculated as stiffness 1 and 2. In the case with negative stiffness, the mechano-isomerization phenomena would be tagged with their depth

and width of the period if exists. Regarding hysteresis, which is barely happening and hard to identify, with the combination of two stiffness and mechano-isomerization phenomena describing the behaviour in deforming process molecules can be distinguished and decided whether a flexible candidate.

Characterized molecules with their mechanical factors are split into three subsets, 120,000 in training set, 30,000 in validation set and the rest as test set. Neural network has been used in predicting molecule activities that could meet the theoretical results or atomic level simulation with acceptable accuracy.[40, 49] Therefore, deep neural network is chosen to be trained to forecast the mechanical factors as network output respectively, by different structural descriptors as input vectors. To guarantee the representation approaches being useful for the purpose of making predictions about the molecules' properties, the four structure encoding scheme are chosen following the requirements: i) unique and unambiguous, ii) universal, iii) compact, iv) computationally efficient to form and parse, and v) systematic and property relevant. A good descriptor should be one able to encode the structure of all hypothetical molecules with every molecules' description in the encoding being unique with only one possible description for each molecule, and at the meanwhile requiring minimal digital memory to store, and of fixed or predictable size. A low computationally cost to generate the description of structure is also a key feature, and it must be easy to parse and process — although it is not necessary to revert to the structure of the molecule from the description. Finally, and most importantly the encoding must vary systematically with structure, and the encoding must encapsulate and resolve the aspects of structure that are active in the structure-property mechanisms relevant to the behavior of interest.

To meet the requirements, the first descriptor attempted is geometric moments. Higher order moments are considered possible related to turbulence[9], and according to the definition, mass scattering in the space and shape information of the objects even for discrete system like single molecules are highly relevant to moments, but not only for rigid bodies. we noticed that the effect of atoms to elastic property, for involving deformation—one of mechanical mechanisms but not just inertia, which implies the factors used in calculating should highly possible be more than mass, for example nuclear charge, and atomic radius. Instead of choosing masses as weight to generate overall moments, geometric moments as known as image moments for 2-D images are defined with multi-channels and 1 as weight for each atom.[34]

$$\hat{M}_{(e,l_1\dots l_n)} = \sum_i^{all\ atoms} \frac{\prod_{j=1}^n x_{i_j}}{N(e)} \quad (5.1)$$

Here n is the order of moment with each of l_s is one number from $\{1,2,3\}$, and x_1, x_2, x_3 are the re-oriented coordinates of atom. e represents element species, and N is the number of atoms of this specie. A multi-channel solution makes the contribution to properties of different elements demonstrated by deep learning but not based on theoretical assumption.

Structure could be better described with more details by more and higher-order moments. Considering the high-order of moments causes exponentially growing distribution ranges, geometric moments are modified to keep the value and variance lower with equation below.

$$M_{l_1\dots l_n} = sgn(\hat{M}_{l_1\dots l_n}) \times \sqrt[n]{|\hat{M}_{l_1\dots l_n}|} \quad (5.2)$$

To dismiss the variance to any rigid transformation and for higher property-relevance with each dimension delivering similar physical interpretation for all molecules, the inputs are re-oriented canonically before moment feature extraction following the way in support information. Molecules vibrating caused by thermal motion leads to inaccuracy of moments calculation with atom position fluctuating, which is discussed by moments fluctuating variance during an equilibrium process (in Supporting Information). With the order of magnitude 2 to 5 smaller to the variance of whole data-set, the accuracy could be justified high enough representing molecules with a complex of several orders of moments, even though reconstruction from moments is hard to accomplish. To determine the best number of order in this circumstances, up to 5th orders were compared with both general moments and geometric moments.

Point cloud is simply using atoms as a group of points as a descriptor with each point represented by the atom coordinates and their species. To be precise, point cloud formats as a set of atoms $\{A_i | i = 1, \dots, n\}$, where for each atom A_i is a tuple (P_i, E_i) representing its position $P_i = (x_i, y_i, z_i)$ and element species E_i , modeled as a one-hot vector $(I(A_i, 'H'), I(A_i, 'C'), I(A_i, 'O'), I(A_i, 'N'))$ of 4 dimension. The $I(A_i, 'X')$ is indicator function to identity different element species for each atom A_i . The similar concept has been used to train neural network to identify specific items from 3-D images.[45] For molecules, it could also effectively differentiate between molecular structures. Unlike the voxelized representation or geometric moments who need more view from different orientation to accomplish a better representation, a set of discrete coordinates and their element information is less sensitive to parameters. According to Qi et al.'s work, it has also been specified that

its data structure requires point cloud being unsequenced that the order of points should not influence the network.

Voxelization is one of the other common methods quantifying 3-d system in computational vision. We implemented the descriptor by atom position based on work from Kuzminykh et al., who found a solution involving interaction with voxelization[27]. Coordinates are oriented with the same way as for moments. Following, the 3-D space is grid into small units with element size of $a \times a \times a$.

$$X(x, y, z) = \begin{cases} 1 & \text{if voxel contains an atom} \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

Afterwards, to capture more interaction and to avoid high sparsity, a Gaussian blurring with wave transform is used for a continues map.

$$\hat{X}(x, y, z) = \sum_{\delta_x=-k_0}^{k_0} \sum_{\delta_y=-k_0}^{k_0} \sum_{\delta_z=-k_0}^{k_0} X(x+\delta_x, y+\delta_y, z+\delta_z) k_{wave}(\delta_x, \delta_y, \delta_z; c(x+\delta_x, y+\delta_y, z+\delta_z)) \quad (5.4)$$

$c = c(x,y,z)$ returns the element in the voxel containing (x,y,z). If their is no atom locating at this position, then return 0. And k_0 is the maximum cut-off of all elements in the system.

The Gaussian and wave transform kernel are as

$$k_{\text{Gauss}}(x, y, z; c) = \begin{cases} \exp\left(-\frac{x^2 + y^2 + z^2}{2\sigma_c^2}\right) & \text{if } \sqrt{x^2 + y^2 + z^2} \leq k_c \\ 0 & \text{else} \end{cases} \quad (5.5)$$

$$k_{\text{wave}}(x, y, z; c) = k_{\text{Gauss}}(x, y, z; c) \cos 2\pi\omega_c \sqrt{x^2 + y^2 + z^2}, \quad \omega_c = \frac{1}{\sigma_c} \quad (5.6)$$

The Gaussian kernel cut-off k_c and σ_c were set relative to their elements' *Van Der Waals* radius.

$$k_c = r_c \times k, \sigma_c = \sigma \times r_c \quad (5.7)$$

To balance the size of input and resolution of representation for a better learning efficiency and effect with: i) Lower sparsity; ii) Capability to accurately identifying molecules; and iii) shorter length of representing vector, the element size of voxel is defined as 0.5Å and the factors for kernel function as $k = 3$ and $\sigma = 0.6$. The detailed discussion could be found in Supporting Information.

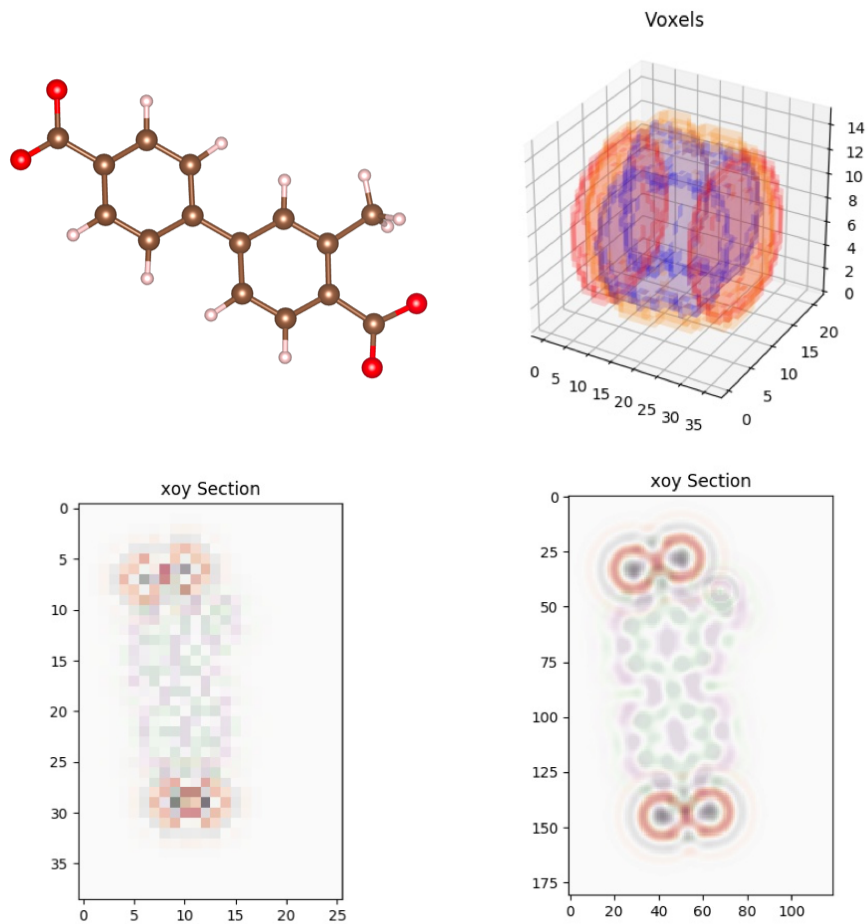


Figure 5.4: Voxelization example with $a = 0.5\text{\AA}$, $\sigma = 0.6$. 5.4(a) shows the molecule geometry; 5.4(b) is the voxelized space, and 5.4(c) is XOY surface section of voxelized space. 5.4(d) is a high resolution section image of XOY surface.

SMILES (Simplified Molecular-Input Line-Entry System) is a widely describing chemical bonding that are used to interpret molecular structure with primary bonds in molecules for predicting chemical reactivity[19]. The encoded one-hot vectors from SMILES as the actual descriptor is generated following Kusner et al.'s work instead of directly using strings. SMILES strings are generated following specific grammars[26, 59], therefore the formation of a string could be parsed to a parse tree showing how grammars work. This tree gives a sequence of production rules which could be converted into one-hot indicator vectors to be used as the descriptor.

One-hot-encoding of SMILES has been proven a good chemical property-predicting representation[14], which makes it benchmark for the comparison of all four descriptors. The independence from the 3d structure of molecules could lead to high exactness that the descriptor would not be affected by condition changes and is irrelevant from the energy-minimizing but only their chemical formula. But at the meantime, it could cause the lack of direct information about weak interaction and secondary bonds such as *Van Der Waals* force or non-covalent interaction. Therefore, other descriptors related from the positions of every atoms and moieties are highly potential more accurate prediction on mechanical behavior.

5.3 Results and Discussion

Performance of four descriptors are discussed by training a binary classifier to predict the four mechanical property factors: i) the existence of mechano-isomerization in compression process, ii) stiffness 1 or 2 respectively, and iii) stiffness of fully relaxed

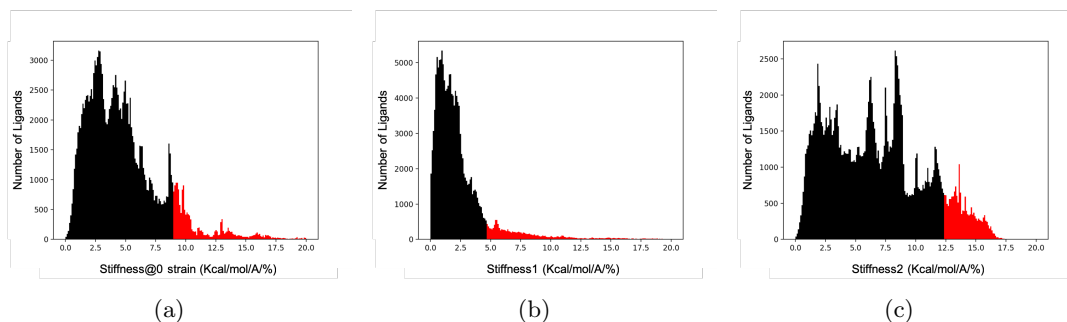


Figure 5.5: Histogram of 5.5(a) stiffness at 0 strain, 5.5(b) stiffness 1, and 5.5(c) stiffness2. Red is molecules with high stiffness and black represents flexible molecules.

molecule, in which, molecules with top 10% stiffness are highly stiff and the rest are defined flexible, therefore, the low/high stiffness ratio is 9: 1 in the dataset (Figure5.5). The number of molecules with and without mechano-isomerization in dataset is 150,813 and 20,577, respectively. The performance of binary classification is evaluated by area under the curve (AUC) of precision recall curve. Higher value of $AUC \in [0, 1]$ indicates higher quality of the forecasting[38], which tells the classifier is successfully picking out molecules with high stiffness or appears mechano-isomerization correctly with less false negatives.

With all four models, performance are all convincingly well that they all could be potential descriptors predicting elastic properties with relatively high AUC value (Table 5.1). Taking forecasting stiffness at 0 strain status by moments as an example, false predictions are found mostly with stiffness value around the threshold, which is acceptable given the data cannot be clearly classified into groups. The stiffness variance was calculated by starting deformation at different equilibrium states of same molecule, and results shows the stiffness are not returned as a exact number with all starting point but with average variance about $0.0257 Kcal \cdot mol^{-1} \text{\AA}^{-1} \%^{-1}$, in which case, a few mistaken predicting especially

for data points around demarcation is reasonable. According to the simulation, stiffness variance is found increasing while the molecule is too stiff, and high stiffness is always seen accompanied by short distance between two carboxlates comparing to the whole molecule size, in which case, deformed status would be less stable, and causes higher uncertainty.

To exclude the obvious high-stiffness molecules, who could be easily identified by their carboxlates distance, and to ensure classifiers are not only works by locating two carboxlates, we further cleansed the data-set by excluding data points whose backbone is too short or force is too biased from backbone, that reduces the training, validation and testing set sizes to 56,000, 14,000, and 10,000 respectively. The method automatically identifying atoms in backbone could be found in Supporting Information.

Descriptor	AUC			
	mechano-isomerization	Stiffness @ 0 strain (non-cleasning data-set)	Stiffness 1	Stiffness 2
Moments	0.836	0.834(0.963)	0.664	0.905
Voxelization	0.705	0.763(0.930)	0.638	0.811
Point Cloud	0.784	0.898(0.979)	0.743	0.963
SMILES	0.947	0.771(0.939)	0.522	0.795

Table 5.1: Prediction AUC

As shown in Table 5.1, the significantly high AUC evinces the assumption about the effect from carboxylates, therefore, the performance on descriptors was discussed without these exemptions.

It is notable that AUC for stiffness 1 are slightly lower, which is caused by the distribution of data for the high concentration of data points combined with high uncertainty instead of the chosen algorithms. Despite that, the descriptors are making satisfying forecasting. Moments' good performance for all properties proved its property-correlation as believed. Yet both being geometric representation, voxelization is shown not as good as the other choices while point cloud being significantly better, which shows, in this case, less pre-processing on structural data and relatively lower feature space dimension could make a better descriptor.

The three other descriptors performed much better than or at least as well as SMILES-based encoding, which has been widely used in chemoinformatics researches, except for predicting mechano-isomerization. The success of SMILES brings the possibility that mechano-isomerization might be caused by specific components or neighbour in one molecule, for example, the conformational isomerism of molecule components could be the reason for the phenomenon during the deformation.

5.4 Conclusion

In summary, we compared how different descriptors perform for predicting elastic properties that proved their potential for further uses in other mechanical activity forecasting. Though not as good as in forecasting stiffness while relaxed, the success in predicting

mechano-isomerization and other two stiffness at different status showed possibility that using characterization of one snapshot to learn behaviours of a whole process. Point cloud is found the best in comparison and the deformed structures' behaviours are proved possible to be foreseen with characteristics of fully relaxed structure. By expanding our experiments to larger data set and improving the descriptors, hopefully, it will be possible to predict complicated mechanical properties of more complex mechanisms or larger molecules, which could help understand how structures determine or influence molecule behavior and looking for potential materials.

5.5 Appendix

5.5.1 Molecule Orientation

To make sure the representation invariant to any rigid transformation, the inputs are re-oriented canonically before moment feature extraction. At the same time, every dimension in the descriptor should deliver similar physical interpretation after oriented following same method. To ensure more property-relevant factors included in the moment representation, molecules are re-located with having the mass center as zero point and setting x-axis along force direction (crossing two carboxylates' center). Considering most atoms on backbone are near the surface defined by the mass of center and two oxygen atoms of carboxylate, the XOY plane are the average of two planes determined by two ends of ligands(Figure 5.6).

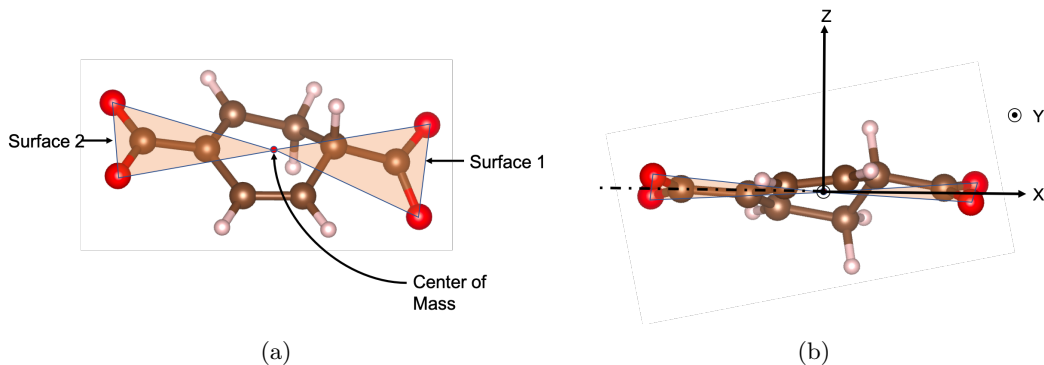


Figure 5.6: Orientation for moments calculation. 5.6(a) the two surface determining the XOY plane, and 5.6(b) the re-oriented system.

5.5.2 Moment Variance

We discussed the noise of moments with from various status of the same molecule during an equilibrium period at room temperature, the temperature of deforming simulation. By multi-measuring during equilibrium, the average uncertainty of one molecule could be analyzed, with whose error is assumed normally distributed. Taking M_{C111} , M_{C2333} , M_{H11121} , and M_{N3333} as example, the average uncertainty of one molecule's equilibrium as about 2.73×10^{-4} , 9.37×10^{-3} , 1.29×10^{-4} , and 3.93×10^{-3} respectively. Even though there is instability of labelled moments in our work, its small scale is convincing enough being an descriptor in these data set. The magnitude order of average moments variance in the period of equilibrium are mostly between -2 to -5 , with very few exemptions as -1 in $\{M_{l_1 \dots l_n} | l_1 = l_2 = \dots = l_n\}$. Comparing to data set variance of each moment, whose order of magnitude are about in the range of $(0.27, \infty)$, the accuracy, with uncertainty's order of magnitude 2 to 5 smaller, could be justified high enough identifying molecules with a complex of several orders of moments.

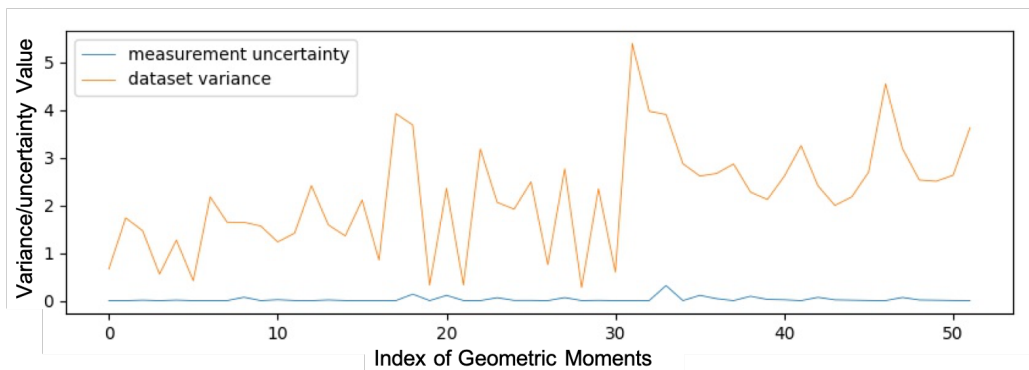


Figure 5.7: Comparison of carbon channel geometric moments variance of dataset and uncertainty during equilibrium. Up to 5th order moments without the 1st are shown in the figure.

5.5.3 Performance Comparison of Different Order of Moments

The pre-test (Table 5.2) on different order of moments shows that 4-channel with up to 2nd order moments gives the best performance on our dataset. 1-channel moments are the traditional ones with atoms all weighted with their masses and considered as a whole, while the 4-channel ones are as described previously classified by element species. In general, 4-channel moments perform better than traditional ones, but without obvious difference.

The small and lowering gap between them tells that atom mass does play a significant role in molecular elastic properties. Theoretically, with more orders of moments involved, the more information is delivered to learning network, in which case the prediction is supposed to be more accurate. However, given that we are working on relatively small molecules, the fact that low-order moments could be well enough for an accurate forecasting is reasonable because of low complexity of molecular structures. The slight decrease of AUC could be the consequence of higher uncertainty that high order moments brings in. On the

contrary, the improved performance of 1-channel moments descriptors implies the effect of thermal motion of every atoms is partially dismissed by others.

Moments Order		AUC
up to 2 nd	1-channel	0.9368
	4-channel	0.9692
up to 3 rd	1-channel	0.9465
	4-channel	0.9657
up to 5 th	1-channel	0.9487
	4-channel	0.9643

Table 5.2: Stiffness Prediction AUC

5.5.4 Voxelization Parameters

Wave transform parameter σ is compared in Table 5.3. By comparing their performance in interaction representing, we chose $\sigma = 0.6$.

As in Table 5.4, histogram shows that the number of voxels containing invalid information, which conveys value approx zero, in the space increases significantly while higher resolution gives a clearer image of the molecule structure. Although the smaller voxel we choose, the lower the blurriness of image would be. With the fact that high sparsity of descriptor might lower learning accuracy because of over-fit and to reduce learning expense as well, especially for that the size of training set is only 120,000 data points, there needs a compromise with reconstruction image quality.

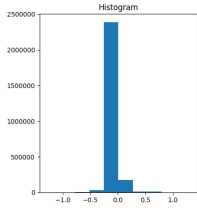
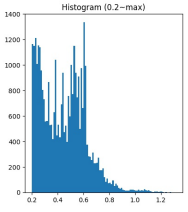
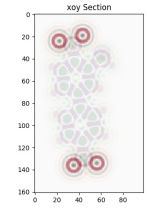
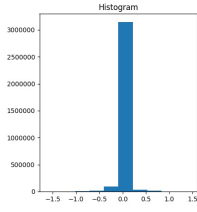
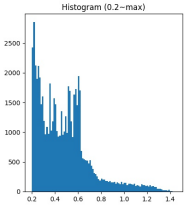
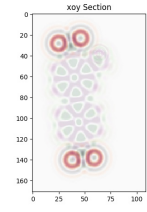
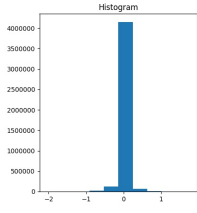
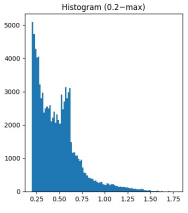
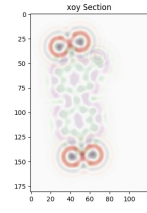
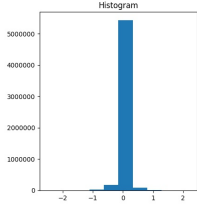
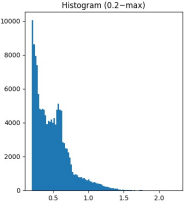
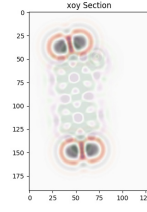
Parameters	Histogram / Histogram in range 0.2 \rightarrow <i>max</i> / XOY section		
$a = 0.1\text{\AA}, \sigma = 0.4$			
$a = 0.1\text{\AA}, \sigma = 0.5$			
$a = 0.1\text{\AA}, \sigma = 0.6$			
$a = 0.1\text{\AA}, \sigma = 0.7$			

Table 5.3: Comparison of voxelization performance with σ changing

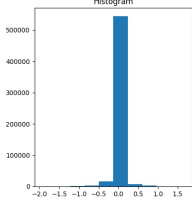
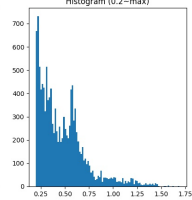
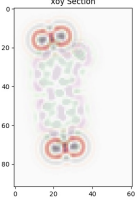
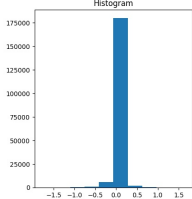
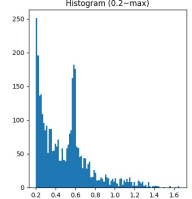
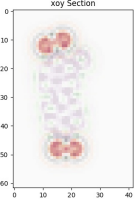
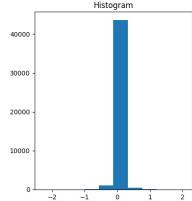
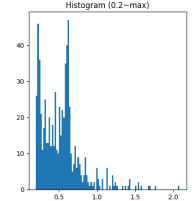
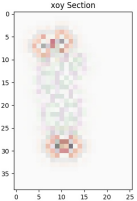
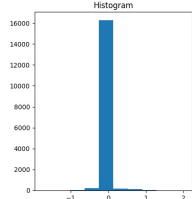
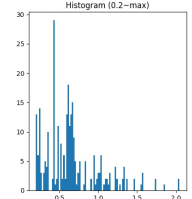
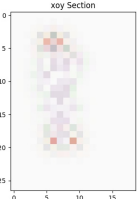
Parameters	Histogram / Histogram in range 0.2 \rightarrow <i>max</i> / XOY section		
$a = 0.2\text{\AA}, \sigma = 0.6$			
$a = 0.3\text{\AA}, \sigma = 0.6$			
$a = 0.5\text{\AA}, \sigma = 0.6$			
$a = 0.7\text{\AA}, \sigma = 0.6$			

Table 5.4: Comparison of voxelization performance with different voxel size

5.5.5 Identifying Backbone

Molecules are represented as 2D graphs for this purpose with atoms as nodes and bonds as edges. The process starts with the shortest path from one carboxylate to another. All the atoms on the shortest path or paths are supposed to be in the backbone. To make sure the rings are complete and not ignored in the first step. The edges in backbone are one-by-one broken. If the two carboxylate are still connected, then the new shortest path

should be also in the backbone. For the final step, all nodes only one edge away from any of nodes in the backbone are counted in.

Chapter 6

Conclusions

The research above are mostly based on MD and AIMD simulated atomic modeling followed by topology analysis. Several different techniques were used in this thesis to prove the efficiency and convenience of interpreting atomic structure understanding. In first two chapters, we found the possibility for deeper understanding of a phenomenon by discussing the atomic modeling result in addition to geometric factor analysis. And in the following two chapters, the benefits of structural representation were revealed that further advanced algorithms i.e. graph theory and machine learning can be introduced into atomic level materials research with a proper way quantitatively describing a material, which leads to higher potential of computational materials science.

Although there has not been a universal solution on atomic structure understanding or representing, the success explaining several mechanism/phenomenon in different fields indicated that saving resources and time by structural interpretation is achievable. Moreover, not only the new perspective looking at atomic model provides a new way to ratio-

nalizing experiment, the understanding achieved by geometrical analysis offers insights for materials design. When a certain application or property is settled, it is possible to find the best material without tons of experiments.

Bibliography

- [1] K. M. Abraham. How comparable are sodium-ion batteries to lithium-ion counterparts? *ACS Energy Letters*, 5(11):3544–3547, 2020.
- [2] Y. Baskin and L. Meyer. Lattice constants of graphite at low temperatures. *Phys. Rev.*, 100:544–544, Oct 1955.
- [3] Jörg Behler. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angewandte Chemie International Edition*, 56(42):12828–12840, 2017.
- [4] N Bernstein, J R Kermode, and G Csányi. Hybrid atomistic simulation methods for materials systems. *Reports on Progress in Physics*, 72(2):026501, jan 2009.
- [5] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, Oct 2006.
- [6] Xinlong Chen, Yuheng Zheng, Wenjian Liu, Can Zhang, Sa Li, and Ju Li. High-performance sodium-ion batteries with a hard carbon anode: transition from the half-cell to full-cell perspective. *Nanoscale*, 11:22196–22205, 2019.

- [7] Rosario M. P. Colodrero, Pascual Olivera-Pastor, Enrique R. Losilla, Miguel A. G. Aranda, Laura Leon-Reina, Maria Papadaki, Alistair C. McKinlay, Russell E. Morris, Konstantinos D. Demadis, and Aurelio Cabeza. Multifunctional lanthanum tetrakisphosphonates: Flexible, ultramicroporous and proton-conducting hybrid frameworks. *Dalton Trans.*, 41:4045–4051, 2012.
- [8] Ali Darwiche, Cyril Marino, Moulay T. Sougrati, Bernard Fraisse, Lorenzo Stievano, and Laure Monconduit. Better cycling performances of bulk sb in na-ion batteries compared to li-ion systems: An unexpected electrochemical mechanism. *Journal of the American Chemical Society*, 134(51):20805–20811, 2012. PMID: 23194439.
- [9] T. Dudok de Wit. Can high-order moments be meaningfully estimated from experimental turbulence measurements? *Phys. Rev. E*, 70:055302, Nov 2004.
- [10] Bruce Dunn, Haresh Kamath, and Jean-Marie Tarascon. Electrical energy storage for the grid: A battery of choices. *Science*, 334(6058):928–935, 2011.
- [11] Felix A. Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S. Schoenholz, George E. Dahl, Oriol Vinyals, Steven Kearnes, Patrick F. Riley, and O. Anatole von Lilienfeld. Prediction errors of molecular machine learning models lower than hybrid dft error. *Journal of Chemical Theory and Computation*, 13(11):5255–5264, 2017. PMID: 28926232.

- [12] Michael Fernandez, Peter G. Boyd, Thomas D. Daff, Mohammad Zein Aghaji, and Tom K. Woo. Rapid and accurate machine learning recognition of high performing metal organic frameworks for co2 capture. *The Journal of Physical Chemistry Letters*, 5(17):3056–3060, 2014. PMID: 26278259.
- [13] Gérard Férey and Christian Serre. Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences. *Chem. Soc. Rev.*, 38:1380–1399, 2009.
- [14] Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *arXiv e-prints*, page arXiv:1610.02415, Oct 2016.
- [15] Mike M Hann and Tudor I Oprea. Pursuing the leadlikeness concept in pharmaceutical research. *Current Opinion in Chemical Biology*, 8(3):255 – 263, 2004.
- [16] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and non-locality in chemical space. *The Journal of Physical Chemistry Letters*, 6(12):2326–2331, 2015. PMID: 26113956.

- [17] John Holoubek, Haodon Liu, Zhaohui Wu, Yijie Yin, Xing Xing, Guorui Cai, Sicen Yu, Hongyao Zhou, Tod A. Pascal, Zheng Chen, and Ping Liu. Tailoring electrolyte solvation for li metal batteries cycled at ultra-low temperature. *Nature Energy*, 6:303–313, 2021.
- [18] Bao-Hua Hou, Ying-Ying Wang, Dao-Sheng Liu, Zhen-Yi Gu, Xi Feng, Haosen Fan, Tuofeng Zhang, Changli Lü, and Xing-Long Wu. N-doped carbon-coated ni_{1.8}co_{1.2}se₄ nanoaggregates encapsulated in n-doped carbon nanoboxes as advanced anode with outstanding high-rate and low-temperature performance for sodium-ion half/full batteries. *Advanced Functional Materials*, 28(47):1805444, 2018.
- [19] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv e-prints*, page arXiv:1802.04364, Feb 2018.
- [20] Bradley A. Johnson and Ralph E. White. Characterization of commercially available lithium-ion batteries. *Journal of Power Sources*, 70(1):48 – 54, 1998.
- [21] William L. Jorgensen, Jayaraman Chandrasekhar, and Jeffrey D. Madura. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [22] Stephen E. Kesler, Paul W. Gruber, Pablo A. Medina, Gregory A. Keoleian, Mark P. Everson, and Timothy J. Wallington. Global lithium resources: Relative importance of pegmatite, brine and other deposits. *Ore Geology Reviews*, 48:55 – 69, 2012.

- [23] Kyung Hwan Kim, Alexander Späh, Harshad Pathak, Fivos Perakis, Daniel Mariedahl, Katrin Amann-Winkel, Jonas A. Sellberg, Jae Hyuk Lee, Sangsoo Kim, Jaehyun Park, Ki Hyun Nam, Tetsuo Katayama, and Anders Nilsson. Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science*, 358(6370):1589–1593, 2017.
- [24] Shinichi Komaba, Wataru Murata, Toru Ishikawa, Naoaki Yabuuchi, Tomoaki Ozeki, Tetsuri Nakayama, Atsushi Ogata, Kazuma Gotoh, and Kazuya Fujiwara. Electrochemical na insertion and solid electrolyte interphase for hard-carbon electrodes and application to na-ion batteries. *Advanced Functional Materials*, 21(20):3859–3867, 2011.
- [25] R Komanduri and L M Raff. A review on the molecular dynamics simulation of machining at the atomic scale. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 215(12):1639–1672, 2001.
- [26] Matt J. Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar Variational Autoencoder. *arXiv e-prints*, page arXiv:1703.01925, Mar 2017.
- [27] Denis Kuzminykh, Daniil Polykovskiy, Artur Kadurin, Alexander Zhebrak, Ivan Baskov, Sergey Nikolenko, Rim Shayakhmetov, and Alex Zhavoronkov. 3d molecular representations based on the wave transform for convolutional neural networks. *Molecular Pharmaceutics*, 15(10):4378–4385, 2018. PMID: 29473756.
- [28] Languang Lu, Xuebing Han, Jianqiu Li, Jianfeng Hua, and Minggao Ouyang. A review on the key issues for lithium-ion battery management in electric vehicles. *Journal of Power Sources*, 226:272 – 288, 2013.

- [29] Yuhao Lu, Long Wang, Jinguang Cheng, and John B. Goodenough. Prussian blue: a new framework of electrode materials for sodium batteries. *Chem. Commun.*, 48:6544–6546, 2012.
- [30] Alessandro Lusci, Gianluca Pollastri, and Pierre Baldi. Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *Journal of Chemical Information and Modeling*, 53(7):1563–1575, 2013. PMID: 23795551.
- [31] Christina M. MacLaughlin. Innovations in lithium ion battery technologies: A conversation with will west, nancy dudney, and andrew westover. *ACS Energy Letters*, 0(0):786–788, 0.
- [32] Charles Manion, Ryan Arlitt, Matthew I. Campbell, Irem Tumer, Rob Stone, and P. Alex Greaney. Automated design of flexible linkers. *Dalton Trans.*, 45:4338–4345, 2016.
- [33] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. A package for building initial configurations for molecular dynamics simulations. *J. Phys. Chem. Lett.*, 30(13):92157–2164, 2009.
- [34] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, February 1962.
- [35] Osamu Mishima and H. Eugene Stanley. The relationship between liquid, supercooled and glassy water. *Nature*, 396:329–335, 1998.

- [36] John B. O. Mitchell. Machine learning methods in chemoinformatics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(5):468–481, 2014.
- [37] Christophe Monnin, Michel Dubois, Nicolas Papaiconomou, and Jean-Pierre Simonin. Thermodynamics of the $\text{liCl} + \text{H}_2\text{O}$ system. *Journal of Chemical & Engineering Data*, 47(6):1331–1336, 2002.
- [38] J. Myerson, L. Green, and M. Warusawitharana. Area under the curve as a measure of discounting. *Phys. Rev. E*, 76:235–243, 2001.
- [39] Miki Nagaoa, Cédric Pittelouda, Takashi Kamiyamab, Toshiya Otomob, Keiji Itohc, Toshiharu Fukunagac, Kuniaki Tatsumid, and Ryoji Kannoa. Structure characterization and lithiation mechanism of nongraphitized carbon for lithium secondary batteries. *J. Electrochem. Soc.*, 153:A914–A919, 2006.
- [40] Suresh Kondati Natarajan and Jörg Behler. Neural network molecular dynamics simulations of solid–liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys.*, 18:28704–28725, 2016.
- [41] Roman Pallach, Julian Keupp, Kai Terlinden, Louis Frenzel-Beyme, Marvin Kloß, Andrea Machalica, Julia Kotschy, Suresh K. Vasa, Philip A. Chater, Christian Sternemann, Michael T. Wharmby, Rasmus Linser, Rochus Schmid, and Sebastian Henke. Frustrated flexibility in metal-organic frameworks. *Nature Communications*, 12(1):4097, July 2021.

- [42] Verónica Palomares, Montse Casas-Cabanas, Elizabeth Castillo-Martínez, Man H. Han, and Teófilo Rojo. Update on na-based battery materials. a growing research path. *Energy Environ. Sci.*, 6:2312–2337, 2013.
- [43] Pekka Peljo and Hubert H. Girault. Electrochemical potential window of battery electrolytes: the homo–lumo misconception. *Energy Environ. Sci.*, 11:2306–2309, 2018.
- [44] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1 – 19, 1995.
- [45] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation, 2016.
- [46] Yuruo Qi, Yaxiang Lu, Feixiang Ding, Qiangqiang Zhang, Hong Li, Xuejie Huang, Liquan Chen, and Yong-Sheng Hu. Slope-dominated carbon anode with high specific capacity and superior rate capability for high safety na-ion batteries. *Angewandte Chemie International Edition*, 58(13):4361–4365, 2019.
- [47] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.
- [48] Hans Martin Senn and Walter Thiel. *QM/MM Methods for Biological Systems*, pages 173–290. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

- [49] Khosrow Shakouri, Jörg Behler, Jörg Meyer, and Geert-Jan Kroes. Accurate neural network description of surface phonons in reactive gas–surface dynamics: N₂ + ru(0001). *The Journal of Physical Chemistry Letters*, 8(10):2131–2136, 2017. PMID: 28441867.
- [50] D. A. Stevens and J. R. Dahn. High capacity anode materials for rechargeable sodium-ion batteries. *J. Electrochem. Soc.*, 147(4):1271–1273, 2000.
- [51] Ning Sun, Huan Liu, and Bin Xu. Facile synthesis of high performance hard carbon anode materials for sodium ion batteries. *J. Mater. Chem. A*, 3:20560–20566, 2015.
- [52] Liumin Suo, Oleg Borodin, Tao Gao, Marco Olguin, Janet Ho, Xiulin Fan, Chao Luo, Chunsheng Wang, and Kang Xu. *water-in-salt*; electrolyte enables high-voltage aqueous lithium-ion chemistries. *Science*, 350(6263):938–943, 2015.
- [53] Deng T, Ji X, Zou L, Chiekezi O, Cao L, Fan X, Adebisi TR, Chang HJ, Wang H, Li B, Li X, Wang C, Reed D, Zhang JG, Sprenkle VL, Wang C, and Lu X. 17(3):269–277, 2022.
- [54] Jean-Marie Tarascon. Is lithium the new gold? *Nature Chemistry*, 2:510, 2010.
- [55] K. Tatsumi, J. Conard, M. Nakahara, S. Menu, P. Lauginie, Y. Sawada, and Z. Ogumi. ⁷Li nmr studies on a lithiated non-graphitizable carbon fibre at low temperatures. *Chem. Commun.*, pages 687–688, 1997.
- [56] G. J. Tawa, I. A. Topol, S. K. Burt, R. A. Caldwell, and A. A. Rashin. Calculation of the aqueous solvation free energy of the proton. *The Journal of Chemical Physics*, 109(12):4852–4863, 1998.

- [57] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 271:108171, 2022.
- [58] Ke Wang, Yaobin Xu, Yuan Li, Vinayak Dravid, Jinsong Wu, and Ying Huang. Sodium storage in hard carbon with curved graphene platelets as the basic structural units. *J. Mater. Chem. A*, 7:3327–3335, 2019.
- [59] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.
- [60] Martin Winter, Jürgen O. Besenhard, Michael E. Spahr, and Petr Novák. Insertion electrode materials for rechargeable lithium batteries. *Advanced Materials*, 10(10):725–763, 1998.
- [61] Chongyin Yang, Ji Chen, Xiao Ji, Travis P. Pollard, Xujie Lü, Cheng-Jun Sun, Singyuk Hou, Qi Liu, Cunming Liu, Tingting Qing, Yingqi Wang, Oleg Borodin, Yang Ren, Kang Xu, and Chunsheng Wang. Aqueous li-ion battery enabled by halogen conversion–intercalation chemistry in graphite. *Nature*, 569:245–250, 2019.
- [62] Kun Yao, John E. Herr, Seth N. Brown, and John Parkhill. Intrinsic bond energies from a bonds-in-molecules neural network. *The Journal of Physical Chemistry Letters*, 8(12):2689–2694, 2017. PMID: 28573865.

- [63] Andrew F. Zahrt, Jeremy J. Henle, Brennan T. Rose, Yang Wang, William T. Darrow, and Scott E. Denmark. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science*, 363(6424), 2019.