Synergy of classical and quantum computational methods to investigate the properties of microporous materials

Thèse N° 7584

Présentée le 25 octobre 2019

à la Faculté des sciences de base
Laboratoire de simulation moléculaire
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Daniele ONGARI

Acceptée sur proposition du jury

Prof. A. Züttel, président du jury
Prof. B. Smit, directeur de thèse
Prof. L. Gagliardi, rapporteuse
Prof. J. Hutter, rapporteur
Prof. N. Marzari, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

> The most serious drawback of Hartree's method is probably the neglect of interchange effect. […] Fock has considered this question and has given equations which may be numerically solved by methods similar to Harthree's, but which include interchange. So far no applications have been made of these, but several computations are in progress.
>
> — L. Pauling and E. B. Wilson,
> *Introduction to Quantum Mechanics*, 1935

Acknowledgements

“That’s what you will stare at, for 10 hours per day, if I convince you to do my same job”. Every year, at the exercise session of the course CH-452, this was my comment after introducing the UNIX environment. Students reacted differently to my provocation, someone with a chuckle, someone looking outside the window and wondering if that class was the best place to stay in a Wednesday afternoon, when autumn is still warm, and someone else staring at the projected screen, still hypnotized by the blinking cursor of the command line. However, even if the most of the acknowledgements that I had to include in my manuscripts were dedicated to computers, as they are the closest companions in our daily routines, I can finally take the occasion to thanks all the people that motivated me in these very short four years. I would like to thank my supervisor for giving me the opportunity of starting this adventure, and giving me the space and the support to find my direction and exciting projects to work on. Thanks especially for always indicate me new methods and techniques that I had very much fun to learn and apply to my systems. Also, thanks for the patience in helping me to improve my drafts, and teaching me how to organize the concepts and the results, to make them ready to be communicated. I would like to thank the members of my group as a source of scientific inspiration. Also, for all the adventures in Amsterdam, Berkeley, Milan, on glaciers, skis, hikes, with spying satellite dishes, “via ferratas”, black nose sheep, etc. etc. etc.

During these 4 years I had the opportunity to assist the course CH-452 (Computational methods in molecular quantum mechanics) for three years in a row, which was for me a very motivating experience. I would like to thank the teacher for being patient with me, involving me so much in the preparation of the course, always guiding me, and giving me the opportunity of sharing the passion that I have for my work. I’m a bit sorry for the students that enrolled for my first time, it was quite tough the first year, but I’m happy that I could have a second and a third chance to improve, and to be a little more self-confident about my teaching skills.

I need to mention my master thesis supervisor who introduced me to this wonderful world of molecular simulation. My life would have been quite different if six years ago an oral exam of kinetics hadn’t gone better than expected, I hadn’t thought “why not to ask him a project for my master?”, and he hadn’t gave me a very stimulating project, involving both DFT and classical molecular mechanics. Also, I need to give credit to the professor of the last course of my master, who, at the end of his last class, told us: “You have learned so many things in the chemical engineering program. Now, you can never let anyone tell you that there is something you can not do!". It was just a sentence, but it came back to my mind many times.

I would also like to give credit to my friends who could manage to get (almost) rid of me for (at least) four years. It seems like yesterday when I was announcing that I was moving to "Louisiana". They gave me the advice to share my craziness only little by little with my new colleagues, in order to get them used to it. I don't think I was always able to follow this precious advice! I'm especially grateful to those friends who, knowing that we had to spend less time together, made that time more valuable.

Also, I think you can understand that for an Italian couple, seeing their only child moving out of home, before even being thirty, it is a very suffering situation. However, they motivated me so that this transition happened very smoothly, and they ended up being proud of my decision.

Finally, I would like to thank that girl that, with questionable taste, chose to have such a strange guy on her side: now that we mentioned each other in our theses, ... *scripta manent*!

Domodossola, 7 July 2019

Abstract

The aim of this thesis is to explore the power and the limits of classical and quantum molecular modelling, for the investigation of the adsorption properties of microporous crystalline materials. The materials analyzed are metal organic frameworks (MOFs) and covalent organic frameworks (COFs), two classes of frameworks that in the last two decades attracted the interest of the scientific community due to their limitless possibility of tunability and their remarkable properties. The four articles that are gathered in this document describe the advances in (1) modelling the interaction of polar molecules with MOFs' open metal sites, (2) computing the micorpore volume that can be occupied by gas molecules, (3) benchmarking the different protocols that are used to estimate partial charges in the frameworks and (4) constructing a full workflow to evaluate the performance of COFs for carbon capture and storage, from just their crystal structure.

Riassunto

Lo scopo di questa tesi é di esplorare i poteri e i limiti della modellazione molecolare classica e quantistica, nell'investigazione delle proprietà di assorbimento di materiali microporosi cristallini. I materiali analizzati sono *metal organic frameworks* (MOFs) e *covalent organic frameworks* (COFs), due classi di strutture che negli ultimi due decenni hanno attiratto l'interesse della comunità scientifica per la loro illimitata possibilità di modulazione e le loro eccezionali proprietà. I quattro articoli che sono riuniti in questo documento descrivono i progressi nel (1) modellare le interazioni tra molecole polari e i siti metallici aperti dei MOFs, (2) calcolare il volume dei micropori che può essere occupato dalle molecole gassose, (3) confrontare differenti protocolli usati nella stima delle cariche parziali di queste strutture e (4) costruire un intero flusso di lavoro per valutare la performance dei COFs per la cattura e il sequestro di anidride carbonica, partendo unicamente dalla loro struttura cristallina.

Contents

Acknowledgements **v**

Abstract (English/Italiano) **vii**

Introduction **1**

Chapters **5**

1 Origin of the interaction between CO_2 and Cu paddle-wheels **7**

1.1 Abstract 7
1.2 Introduction 7
1.3 Computational methods 10
1.4 Results and discussion 12
1.4.1 Comparison of simulated and experimental isotherms in HKUST-1 . . . 12
1.4.2 Interactions computed in the cluster models 14
1.4.3 Multireference calculations 16
1.4.4 Correction of the force field 18
1.4.5 Investigation of the "double" open metal site interaction in Cu-TDPAT . 20
1.5 Conclusions 22

2 Accurate Characterization of the Pore Volume **23**

2.1 Abstract 24
2.2 Introduction 24
2.3 Methods 25
2.3.1 Experimental Measurement of the Pore Volume 25
2.3.2 Computational Methods to Assess the Pore Volume from the Unit Cell . 26
2.3.3 Accessible versus non-accessible channels 30
2.3.4 Algoritm to compute the occupiable pore volume 30
2.4 Computational versus experimental pore volumes 32
2.5 Software and parameters 32
2.6 Results and discussion 33
2.6.1 3D Model for the Full Characterization of the Pore Volume 33
2.6.2 Comparison of Different Pore Volume Definitions with Experimental Data for HKUST-1 34

2.6.3 The Helium Void Fraction 36
2.6.4 CoRE MOF screeening 37
2.6.5 Comparison With Experimental Data for 10 MOFs 40
2.7 Conclusions 42

3 Evaluating charge equilibration methods 43
3.1 Abstract 43
3.2 Introduction 44
3.3 Theoretical Aspects 45
3.3.1 Charge equilibration (Qeq) and periodic charge equilibration (PQeq) methods 45
3.3.2 Modifications to the Qeq method 49
3.3.3 Split charge equilibration methods (SQE) 52
3.4 Applications to MOFs 53
3.4.1 Periodic Qeq 53
3.4.2 Extended Qeq 54
3.4.3 Formal Charge and Ionizing Qeq 55
3.4.4 MOF electrostatic potential optimized Qeq and EQeq+C 55
3.4.5 Split Charge Equilibration MEPO 56
3.4.6 Other methods 57
3.4.7 Which one is the most reliable method to compute partial charges in MOFs? 57
3.5 Computational Details 58
3.5.1 Programs to compute Qeq and DDEC charges 58
3.5.2 Ionization energies and radii 58
3.5.3 Adsorption calculations 61
3.6 Results and discussion 61
3.6.1 Analysis of the charges obtained from DDEC 61
3.6.2 Analysis of the charges obtained with different methods, charge centers and parameters 64
3.6.3 Analysis of the adsorption results 71
3.7 Conclusions 76

4 Building a consistent and reproducible database of COFs 79
4.1 Abstract 79
4.2 Introduction 80
4.3 Building of the database 83
4.3.1 Structure labels 84
4.3.2 Data cleaning 84
4.4 Results and discussion 87
4.4.1 Analysis of experimental structures 87
4.4.2 Cell optimization 88
4.4.3 CO_2 separation performances 93

4.5 Conclusions 98
4.6 Methods 99
4.6.1 DFT calculations 99
4.6.2 Partial charges 100
4.6.3 Geometry-based descriptors 100
4.6.4 Parasitic energy evaluation 100

Conclusions and future work **101**

Bibliography **105**

List of publications **127**

Curriculum Vitae **129**

Introduction

Finding the best material and operating conditions for a given process or application is one of the ultimate goals in materials science and chemical engineering. To accomplish this task, scientists are required to combine expertise and intuition to sort out the most promising candidates. Also, they need a lot of patience and perseverance to test these, in a trial and error approach. One may think about Thomas Edison and his numerous attempts before coming up with a proper filament for the light bulb, or the two German scientists, Franz Fischer and Hans Tropsch, testing all the possible catalysts and reaction conditions, and looking for the proper combination that could turn syngas (i.e., a mixture of CO and H2) into a synthetic fuel, making the Germany of the 1920s energetically independent.

Nowadays, computers offer new ways to achieve such a task. First, they allow one to easily share the structure of newly synthesized materials, making possible the creation of databases that contain millions of these entries that can be screened for any target application. Second, they provide a considerable computing power, which allows one to deal with all these materials. The problem is conceptually simple: given a target application, evaluate and rank all known materials, and pick the best. However, computers limit themselves to repeating a precise sequence of operations, and, compared to the early inventors of the last centuries, computers have unbeatable patience and perseverance but an obvious lack of intuition. In order to make this ranking possible, a computational scientist has to understand the key properties that are important for the target application, how these relate to the chemistry of the materials, and then code this information in a computer. Currently, we are able to quantitatively describe how the fundamental components of these materials, the atoms, behave. The solution of the Schrödinger's equation provides the essential description on how these atoms combine and interact with each other. This is, however, an enormously demanding task, and, as always in science, when we want to apply a theory we can not avoid considering a certain amount of approximations.

Starting from the description of the atomic behaviour, to ending up predicting the cost of removing a kilogram of carbon dioxide from a coal power plant's flue gas using a specific material for the adsorption, is for sure an impressive and exciting challenge. This PhD work deals with the approximations that one needs to make in order to cover this leap. The CO_2 adsorption was the main topic of investigation in the published results of my PhD. However, the fundamental understanding of the laws that govern the interactions between molecules and microporous materials allowed me to go beyond the investigation of CO_2 separation and apply similar principles to study noble gases, catalysis, and even a device for the detection of THC (the active principle of marijuana) from human breath.

The research was focused on microporous materials, i.e., materials that contain pores whose size is comparable to the size of gas molecules. According to the definition given by the International Union of Pure and Applied Chemistry (IUPAC), micropores are defined as *not exceeding about 2 nm*. In particular, there is another requirement that informed the selection of materials for our investigations: we considered only materials that are crystalline. Crystalline materials can be represented by a primitive unit (i.e., the unit cell). Since for most of the cases, it is reasonable to neglect the phenomena acting on the surface, and focus only on the bulk, then the unit cell with periodic boundary conditions is a representative model for the whole material. Therefore, understanding how a gas molecule behaves inside the unit cell at the atomic level, allows us to understand the macroscopic adsorption performances of the material.

Speaking about the length scale, the unit cells of these microporous crystals are typically bigger than the units cells of, for example, metals or metal oxides, as they need to contain the description of the pores. The majority of them have lengths between 10 and 50 Angstroms and contain 50 to 1000 atoms. These dimensions are prohibitive for higher accuracy wave-function methods such as the “gold standard” coupled-cluster methods. Density Functional Theory (DFT) methods should also be used with care: input settings should be wisely calibrated and one can not expect to have the same high accuracy as when modelling small molecules or crystals that have a few atoms per unit cell.

Moreover, there is not only a problem of length scale but also a problem of time scale. Since we want to model the equilibrium properties of these materials at finite temperature and pressure, this means integrating over all the possible configurations that the atomic system can explore at the given temperature and pressure conditions. This is essentially a sampling problem, and statistical thermodynamics provides the proper theory to consider this aspect. However, even exploiting efficient sampling methods, we need to explore a very wide configurational phase space. One may think, for example, about all the positions that a CO_2 molecule can take inside a small box, and how this number of configurations exponentially increases when more CO_2 molecules are added. Even exploiting efficient sampling methods, we are required to compute the energy for millions of geometric configurations. This is again a calculation that we can not achieve by brute force using DFT. We need to introduce some approximations to be able to perform an energy evaluation in a fraction of a second. Here comes into play the fundamental role of empirical potentials (also referred to as “force fields”) to cheaply describe the interactions between the atoms in the system. This big shortcut in time comes from very radical approximations: even if it is impossible to aim for the “perfect” force field, it is an essential requirement that the force field can at least qualitatively describe the important phenomena that contribute to the ranking in the performance for a desired application. Finally, also with force field based methods one can meet sampling issues when the density of the adsorbed molecules in the pores becomes liquid-like and the micropores are saturated. Combining the cost of exponentially more interactions, with exponentially more possible configurations that comes from an increased number of particles, enhanced sampling methods are necessary to obtain a reliable description of the system. These methods by themselves require tailor-made implementations depending on the case, which are still difficult to automate for the screening of thousands of materials.

Given this perspective, we can now anticipate the path that links together the four chapters of this thesis. In the first chapter, “Origin of the Strong Interaction between Polar Molecules and Copper(II) Paddle-Wheels in Metal Organic Frameworks”, we investigated the accuracy and the parametrization of force field methods to model the interaction between the adsorbent HKUST-1 and CO_2. The reliability of these force fields can be evaluated by comparing the outcome of the model with experimental results, or by comparing directly the energies with the values from higher accuracy (dispersion corrected) DFT methods. In this system, we found a discrepancy between these two evaluation protocols, which we motivated by the evidence that DFT also fails to correctly model the investigated interactions. We therefore needed to employ a more expensive wave-function based method, to be able describe the experimental evidence, and tune the force field.

The second chapter, “Accurate characterization of the pore volume in microporous crystalline materials”, aims to develop a novel protocol to compute geometrically the amount of void space that a reference probe, e.g., nitrogen, can occupy. This is a fundamental descriptor that can be compared with experimental measurement: it can reveal if the model of our unit cell is reliable, especially when assessing the gas loading at saturation. Before going to more advanced molecular simulation, one had better check this comparison and, in case of mismatch, investigate whether the experimental framework collapsed, contains residual solvent, or has a non-negligible percentage of defects.

The chapter “Evaluating charge equilibration methods to generate electrostatic fields in microporous materials” aims for an extensive benchmark of empirical methods that are commonly employed to assign partial charges to the atoms of frameworks. These partial charges are fundamental to describe the interactions with polyatomic molecules such as CO_2, as they allow one to compute the Coulombic contribution of the interactions that is mainly responsible of the selectivity toward a different gas molecule. Point charges can be obtained from a partitioning of the electron density as computed from DFT, but charge equilibration (QEq) methods are considerably faster and an attractive choice when considering thousands of materials. We concluded, however, that the latest variants of these Qeq methods were arguably too optimistic in ensuring their reliability.

In contrast to the study described in the first chapter, which is based on the investigation of just one material, the studies described in the second and third chapters only led to clear conclusions when thousands of materials were compared with consistent metrics. The latter results evidently can not be computed with the same accuracy that is achievable when investigating only one material. Therefore, these must be computed with a well defined and consistent protocol that allows for a fair analysis of trends. This awareness motivated us to adopt AiiDA (Automated Interactive Infrastructure and Database for Computational Science) in our research. Explained in a few words, AiiDA is a "program to manage programs" that allows the researcher to automate workflows combining different simulation packages. The added value of using AiiDA is the advantage of having a common platform to collaborate and share these workflows, as well as its data-tracking system, which allows for full reproducibility of the calculations. In the final chapter of my PhD, titled "Building a consistent and reproducible database for adsorption evaluation in covalent organic frameworks", I present my attempt to find a balance between accuracy and efficiency for the analysis of over 300 synthesized and reported covalent organic frameworks. In this work, we automated the full workflow that goes from creating the unit cell to the assessment of the CO_2 separation performance, tying together the DFT optimization, the calculation of partial charges, the calculation of CO_2 and N_2 isotherms, and, finally, the process modelling the pressure swing adsorption.

Chapters

1 Origin of the Strong Interaction between Polar Molecules and Copper(II) Paddle-Wheels in Metal Organic Frameworks[1]

1.1 Abstract

The copper paddle-wheel is the building unit of many metal organic frameworks. Because of the ability of the copper cations to attract polar molecules, copper paddle-wheels are promising for carbon dioxide adsorption and separation. They have therefore been studied extensively, both experimentally and computationally. In this work we investigate the copper-CO_2 interaction in HKUST-1 and in two different cluster models of HKUST-1: mono-copper Cu(formate)$_2$ and di-copper Cu_2(formate)$_4$. We show that density functional theory methods severely underestimate the interaction energy between copper paddle-wheels and CO_2, even including corrections for the dispersion forces. In contrast, a multireference wave function followed by perturbation theory to second order, using the CASPT2 method, correctly describes this interaction. Restricted open shell Møller-Plesset 2 method (ROS-MP2, equivalent to (2,2) CASPT2) was also found to be adequate in describing the system and was used to develop a novel force field. Our parametrization is able to predict the experimental CO_2 adsorption isotherms in HKUST-1, and it is shown to be transferable to other copper paddle-wheel systems.

1.2 Introduction

Metal organic frameworks (MOFs) are a class of three-dimensional nanoporous materials composed of metal nodes connected by organic ligands. The oriented coordination bond between these two components is responsible for the structure of the crystal. The possibility of combining different metals with different ligands provides a large variety of MOF structures. More than ten thousand structures have already been synthesized, [1] but this is only a small fraction of the hundreds of thousands of structures that have been predicted computationally.[2]

---

[1] Preprint version of the article: Daniele Ongari, Davide Tiana, Samuel J. Stoneburner, Laura Gagliardi and Berend Smit, *The Journal of Physical Chemistry C*, 121(28), 15135-15144, 2017, https://doi.org/10.1021/acs.jpcc.7b02302. D.O. performed all the force field, DFT and MP2 calculations and tuned the force field to reproduce the MP2 potential.

MOFs have attracted considerable attention in the past decade for various applications, including gas adsorption and storage,[3] gas separation,[4] fuel production,[5] chemical sensing,[6] and catalysis.[7]

Computational modeling is extensively used to investigate the properties of synthesized materials for a given application, and to predict the performance of hypothetical structures. In the case of gas adsorption, the quality of the model directly derives from the accuracy with which one can describe the microscopic interactions between the guest molecules and the framework. Density functional theory (DFT) calculations are routinely used for this purpose.[8, 9, 10] However, weak interactions, due to dispersion forces arising from electron correlation, are poorly described by standard DFT methods. Corrections need to be introduced for this purpose (see the recent review of Grimme *et al.*[11] and references therein). Alternatively, post Hartree-Fock methods can be employed to evaluate interaction energies with high accuracy. However, because of the unfavorable scaling with the size of the system, they can hardly be used directly to compute interaction energies in MOFs, whose unit cells typically contains hundreds of atoms.[10]

This work focuses on the interaction between the carbon dioxide molecule and the copper (II) paddle-wheel, which is a metal organic structure composed of two copper cations connected to four carboxylates anions in a square planar coordination geometry. The smallest example of this structure is the $Cu_2(formate)_4$ molecule (Figure 1.1, left).

The copper paddle-wheel is the building unit of many MOFs, including HKUST-1 ($Cu_3(BTC)_2$) as shown in Figure 1.1, right. The structure of HKUST-1 presents three pores (Figure 1.2, left) and several characteristic adsorption sites for CO_2 (Figure 1.2, right). The biggest pore is characterized by the presence of twelve open metal sites (OMSs) i.e. unsaturated copper cations which are obtained after solvent removal and which are able to attract polar molecules through electrostatic interaction.

HKUST-1 is one of the earlier reported MOFs.[12] It is among the best performers for natural gas storage,[13] and it has also attracted interest for gas separation[14, 15, 16] and heterogeneous catalysis.[17, 18, 19] Because of its popularity, many experimental data are available for this framework. Wu *et al.*[20] conducted *in situ* neutron diffraction studies for CO_2 adsorption in HKUST-1 that show that at low loading of CO_2 and low temperature (20 K) the open metal site is the strongest adsorption site because it is the only one to be occupied at a 1:1 CO_2:Cu ratio of loading. They were also able to rank the strength of the secondary sites by increasing the amount of CO_2 and observing the filling in each site: small pore windows sites and center sites are the second and the third, respectively, and large pore corner sites are the fourth in

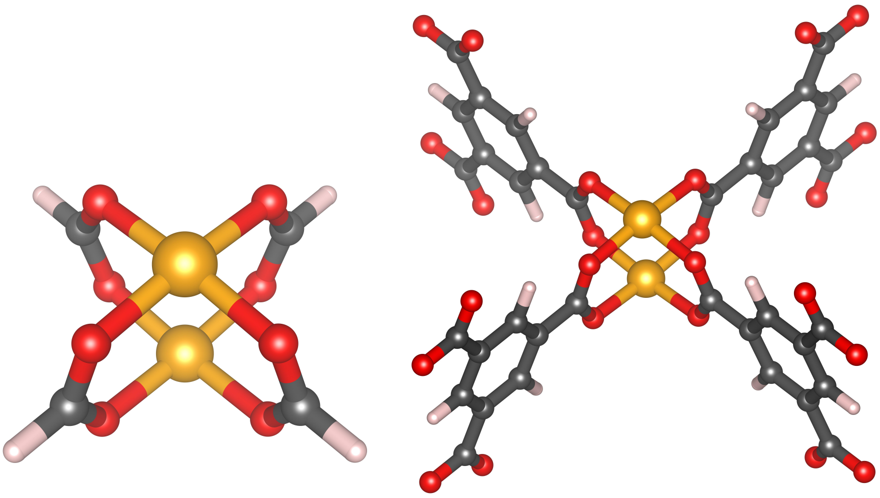


Figure 1.1 – Copper paddle-wheel structure is composed of two coppers atoms bridged though four di-carboxilate anion. Cu_2(formate)$_4$ (left) represents the simplest paddle-wheel geometry possible. Di-copper benzil-1,2,3-trimetylcarboxylate, Cu_2(BTC)$_4$ (right), is the building unit of HKUST-1 framework: each BTC has three caboxylate groups that allow the creation of a three-dimensional network.

terms of order of filling and therefore interaction energy strength.

In a recent work, Grajciar *et al.*[21] showed that DFT dispersion corrected methods, e.g. Grimme's pairwise correction for dispersions (D2[22] and D3[23]) and van der Waals density functionals (vdW-DF[24] and vdW-DF2[25]), underestimate the strength of the open metal site and are not able to reproduce the experimental adsorption data obtained by Wu *et al.*

The van der Waals density functional methods, in particular, were used previously by our group to compute the CO_2 binding energy in MOF-74 for different metals[8, 26, 27, 28] and to parametrize the associated force field [29, 30]. A good agreement with experiment was always observed, giving rise to the question of why the same ab-initio methods are not able to model correctly the CO_2 interaction with the open metal site in a copper paddle-wheel framework. This underestimation of the interactions in HKUST-1 motivated Grajciar *et al.* to employ a DFT - Coupled Clusters corrected (DFT/CC) method[31] to study this system and obtain a tailor-made correction for the CO_2 interaction with HKUST-1. In DFT/CC, the error associated with the PBE density functional is corrected by a term dependent on the pairwise distance between the CO_2 atoms and the atoms of the framework. This term was estimated from the difference between the DFT and the CCSD(T) computed one-dimensional potential energy curves of CO_2 interacting with some other reference molecules, i.e. H_2, benzene, CO_2, and Cu(formate)$_2$.

It is known that a copper-copper magnetic interaction is present in HKUST-1, [32] and consequently the correlation between the electrons of the two coppers can affect the interaction with the CO_2. Because of this, we investigated the legitimacy of transferring the DFT error for CO_2 interaction from the mono-copper system Cu(formate)$_2$ to the di-copper paddle-wheel

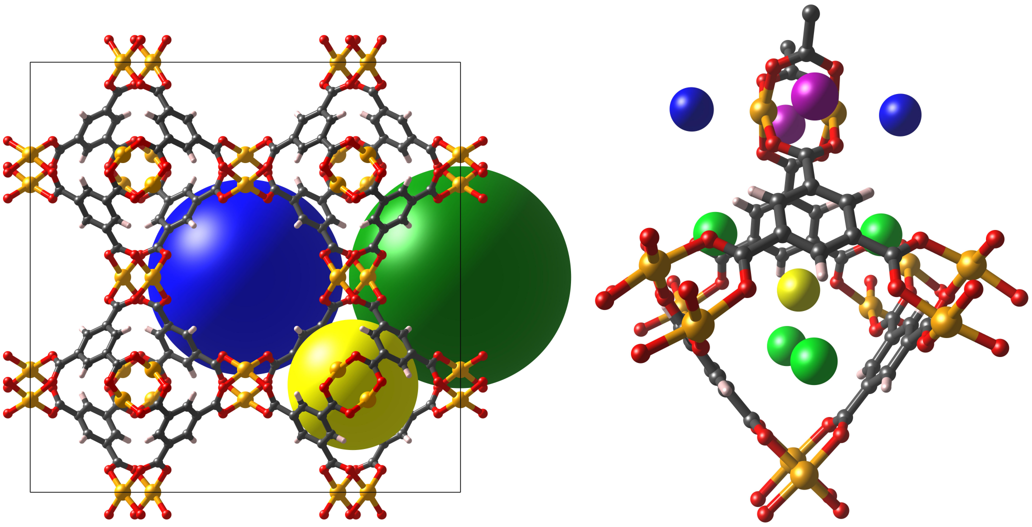


Figure 1.2 – Three different pores in HKUST-1 (left): big pore (blue), medium pore with open metal sites (green), small pore (yellow). Characteristic sites of adsorption for CO_2 (right): open metal site (blue), small pore window (green), small pore center (yellow) and large pore corner (purple).

structure (and to the HKUST-1 framework) by using multireference wave function methods. These methods are critical for accurately modeling systems with a relevant magnetic coupling such as the Cu paddle-wheel.[33, 34, 35] Accordingly, we explored in this work the adequacy of different quantum methods for describing the electronic structure of the system and the interaction between the metal cation and carbon dioxide.

Furthermore, we used our insights to develop a classical force field that is able to accurately describe the Cu paddle-wheel interaction with CO_2 and model the adsorption in MOFs containing this building unit. It was estimated[36] that among 4764 three-dimensional MOF structures from the Cambridge Structural Database[37] (as refined in the CoRE MOF database),[38] 4.2% of them contains the Cu paddle-wheel, and another 3.5% contains the paddle-wheel motif formed by other cations. Cu paddle-wheels are a recurrent building unit among the different MOFs, and with a reliable and transferable force field it would be possible to also screen these frameworks and identify their performances for CO_2 adsorption.

1.3 Computational methods

The periodic calculations were performed using the Perdew-Burke-Ernzerhof GGA method PBEsol[39] to optimize the framework, and the second version of van der Waals dispersion corrected density functional vdW-DF2[25] to compute the interactions. The plane wave Quantum Espresso 5.4 package[40] was employed. We adopted the projector augmented wave (PAW) method[41, 42] with a cutoff energy of 60 Ry for the wave function and 300 Ry for the electron density. Due to the dimension of the unit cell of HKUST-1 a Γ-point sampling of the

Brillouin zone integration was used, with a smearing occupation of 0.02 Ry.

For the cluster calculations, geometry optimizations were performed using the unrestricted M06-L/cc-pVDZ [43] level of theory and subsequent single point energy difference calculations were performed using restricted open shell MP2,[44] (ROS-MP2) and unrestricted M06-L and M06[43]. The Gaussian-09 package[45] was employed. We tested the convergence of the basis set using cc-pVDZ, AUG-cc-pVDZ, cc-pVTZ and AUG-cc-pVTZ[46, 47, 48, 49] A spin multiplicity of three was used to model the magnetic state of the copper paddle-wheel clusters. To account for the error in computing the interaction due to the basis set superposition, the counterpoise method by Boys and Bernardi was employed[50]. For the ROS-MP2 calculations the frozen orbitals are the 1s for C and O, and 1s, 2s, 2p, 3s and 3p for Cu.

Multireference calculations were performed on the cluster models using the complete active space self-consistent field method (CASSCF) [51] followed by second-order perturbation theory (CASPT2) [52] using Molcas 8.2.[53] All CASSCF/CASPT2 calculations were performed without symmetry. Relativistic basis sets of atomic natural orbital type (ANO-RCC) [54] were employed for all the atoms. To explore basis set convergence, three different basis sets of increasing size were tested. The first one, BS1, is of double-ζ quality plus polarization; the second one, BS2, is of triple-ζ quality plus polarization on Cu, O, C atoms and double-ζ quality plus polarization on H atoms; the third one, BS3, is of quadruple-ζ quality plus polarization for Cu and CO_2, triple-ζ quality plus polarization on the remaining C and O atoms, and double-ζ quality plus polarization on H atoms. Scalar relativistic effects were included using the Douglas-Kroll-Hess Hamiltonian.[55] The computational cost arising from the two-electron integrals was drastically reduced by employing the Cholesky decomposition technique.[56] The decomposition threshold was chosen to be 10^{-4}, as this should correspond to an accuracy in total energies of the order of mHartree or higher. In the CASPT2 calculations, in order to prevent possible intruder states, an imaginary shift of 0.1 au was added to the zero-order Hamiltonian. The default IPEA shift of 0.25 au was used. The default choices of the program were employed for freezing orbitals, resulting in the 1s orbitals of C and O being frozen, along with the 1s, 2s, 2p, and 3s orbitals of Cu.

For the cluster models the interaction energy between the framework and the CO_2 molecule was computed as the difference between the energy of the super-system, the framework plus CO_2, and the energies of the two isolated components, namely CO_2 and the framework.

The Raspa 2.0 package [57] was employed for the force field calculations. In all the simulations TraPPE[58] Lennard-Jones parameters and charges were used to model CO_2-CO_2 interactions, while different sets of parameters were used to model the framework-CO_2 interaction, as discussed within the results. The details of the simulations are provided in the Supporting Information.

1.4 Results and discussion

1.4.1 Comparison of simulated and experimental isotherms in HKUST-1

The CO_2 isotherms computed with the standard force field i.e. UFF[59], DREIDING[60] and TraPPE[58], are found to be in strong disagreement with the experimental data in the range of pressure from zero to one bar. Figure 1.3 shows the simulated isotherms, computed using the Grand Canonical Monte Carlo (GCMC) technique with different sets of parameters for the dispersion forces and the corresponding experimental isotherms.

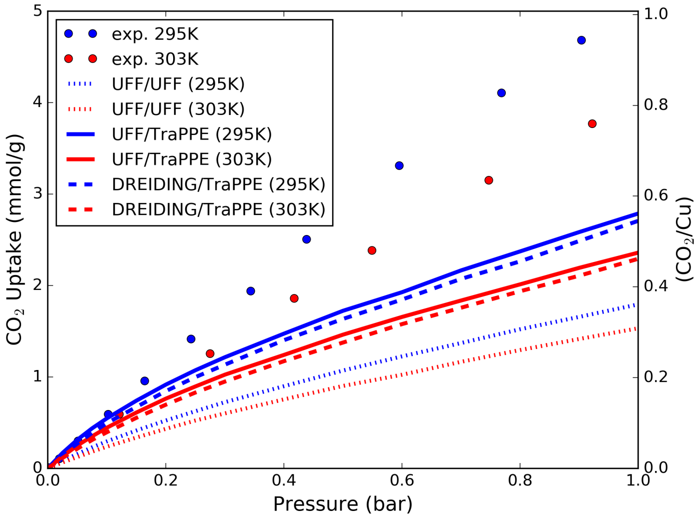


Figure 1.3 – Comparison of experimental (295K,[14] 303K[31]) and simulated adsorption isotherms. TraPPE[58] Lennard-Jones parameters and charges are used for CO_2-CO_2 interactions. To compute the dispersion forces acting between CO_2 guest molecules and the crystal, three commonly used approaches are compared. First we used Lennard-Jones parameters from UFF[59] (Lorentz-Berthelot mixing rules). Then we used UFF/TraPPE and DREIDING/TraPPE parameters[60] (notation: $FF_{framework}/FF_{adsorbate}$). The point charges for the framework atoms are extracted from a PBEsol DFT calculation using the REPEAT scheme;[61] in the Supporting Information we reported the charges' values, and we compared them with the values obtained by using Bader's method.[62] The framework is assumed to be rigid in all the simulations.

All the simulations underestimate the uptake of CO_2, which means that the force field underestimates the adsorbate-host interactions. The force field interaction energies for specific sites are compared to those obtained by DFT calculations in Table 1.1. The binding energy for each site, corresponding to the optimized position of a CO_2 molecule in the open metal site, in the small pore window site, and in the small pore center site, are reported.

From neutron diffraction *in situ* experiments by Wu *et al.* we know that OMSs are the first filled sites, then windows and cage sites get populated by CO_2. This observation proves that OMSs have the strongest binding energy. Despite the fact that UFF/UFF, UFF/TraPPE and DREIDING/TraPPE force field are giving similar results to the vdW-DF2 method, and this could in principle validate the force fields, we clearly see from Table 1.1 that in all four of

Table 1.1 – Interaction energy (kJ/mol) between CO_2 and HKUST-1, for different adsorption sites: the open metal site in the apical position of copper paddle-wheel, the window and the center of small octahedral pores. Force field and periodic DFT calculations are compared. Results obtained with PBEsol are also reported to appreciate the contribute of the dispersion corrections introduced by the vdW-DF2 non local functional.

Method	**Open metal**	**Window**	**Center**
FF (UFF/UFF)	-19.3	-25.7	-26.3
FF (UFF/TraPPE)	-19.0	-27.5	-29.0
FF (DREIDING/TraPPE)	-19.4	-27.2	-28.5
DFT (vdW-DF2)	-22.1	-30.2	-26.3
DFT (PBEsol)	-12.1	-6.7	-0.8
DFT/CC (Grajciar *et al.*[31])	-28.2	-23.1	-23.2

these cases the OMS is predicted to be the weakest site. As a consequence, these standard methods erroneously predict that the OMS is poorly occupied, as its interaction energies with CO_2 are $\sim 4k_bT$ and $\sim 60k_bT$ weaker than other sites at 303K and 20K, respectively. Standard force fields are known to incorrectly model the strong interaction of adsorbate molecules with OMSs in MOFs,[63] but vdW-DF2 is also showing the same problem in the case of copper paddle-wheel, while it was found to model accurately the open metal site interaction with CO_2 for other MOFs.[26, 29, 30]

There are different assumptions in these calculations that may not hold for this system, therefore the interaction energy between carbon dioxide and the copper atom in HKUST-1 was also computed using other approaches. We considered the introduction of the Hubbard correction[64] to model the d orbitals of copper, because it was shown to influence the CO_2 interaction with the OMS in MOF-74.[65, 26] The value of U=3.8 eV, which can reproduce the experimental oxidation energy of copper,[66] was used. Also different versions of the van der Waals density functional were compared to vdw-DF2 method, i.e. vdW-DF[67] and revised-vdW-DF2.[68] In all the cases the geometry of CO_2 was optimized keeping the framework rigid, as obtained from the PBEsol calculation. The results are reported in Table 1.2. No significant deviations in the interaction energy were found, the only slightly increased value being obtained with vdW-DF, which is known to systematically overestimate dispersion interactions.[69]

Table 1.2 – CO_2 open metal site interaction energies in HKUST-1 computed with different dispersion corrected DFT methods.

Method	**Open metal site interaction**
vdW-DF	-24.9 kJ/mol
vdW-DF2	-22.1 kJ/mol
vdW-DF2+U	-21.4 kJ/mol
vdW-DF2-rev	-20.2 kJ/mol

Finally, the rigid framework assumption was neglected, performing a full optimization of the framework's atoms with the adsorbed molecule in the OMS, using vdW-DF2. No significant deviation in the binding energy was found: -1.6 kJ/mol of difference from the rigid calculation.

Moreover, we noticed an exaggerated distortion of the copper paddle-wheel structure which has not been reported experimentally, suggesting the inadequacy of the vdW-DF2 method to optimize the crystal geometry. The rigidity of the adsorbent was therefore assumed as reasonable.

1.4.2 Interactions computed in the cluster models

To understand why vdW-DF2 method underestimates the CO_2-Cu interaction in HKUST-1, we analysed two smaller representative clusters, Cu(formate)$_2$ and Cu$_2$(formate)$_4$. The interaction energy with carbon dioxide was scanned at different distances by keeping the CO_2 molecule perpendicular to the CuO_4 plane, as shown in Figure 1.4.

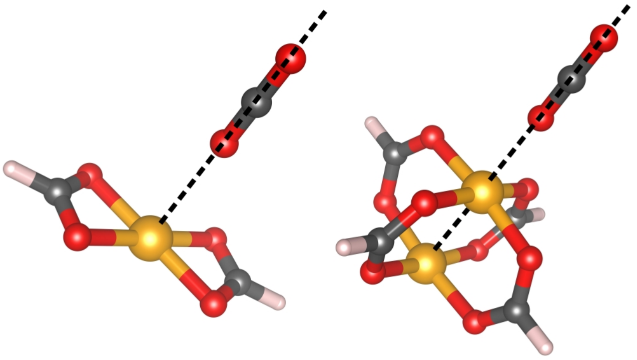


Figure 1.4 – Path representation of linear scans of CO_2 interacting with Cu(formate)$_2$ (left) and Cu$_2$(formate)$_4$ (right). The dotted line, along which the CO_2 molecule is displaced, is perpendicular to the CuO_4 plane.

This configuration, referred here as "linear", was chosen to decrease the number of degrees of freedom for the CO_2 position to just one i.e. the copper-oxygen distance in the axial direction. This configuration also minimizes all the pairwise contributions of the interaction but the copper-oxygen one, which is the one vdW-DF2 is failing to model properly. Within HKUST-1, the optimal linear configuration corresponds to a distance of 2.65 Å and a binding energy of -13.4 kJ/mol, computed using vdW-DF2. Figure 1.5 and Figure 1.6 show the interaction energy of CO_2 as a function of the Cu-O distance computed with different methods in Cu(formate)$_2$ and Cu$_2$(formate)$_4$ respectively.

The inspection of the energy profiles reported in Figure 1.5 for Cu(formate)$_2$, shows that the CO_2-copper binding energies differ within 4 kJ/mol among the various methods, ranging between -8.0 kJ/mol (vdW-DF2) and -13.1 kJ/mol (MP2). The minimum energy distance for vdW-DF2 is longer than with the other methods, 2.9 Å instead of 2.5-2.6 Å. The M06 and M06-L functionals produce similar energy profiles. Hence, the inclusion of the semi-local contribution with Hartree-Fock exchange present in M06 has a minor effect. It is also interesting to note the overall good agreement with the UFF force field. The attraction computed by the force field is mainly due to the coulombic (REPEAT-TraPPE) interaction, with only a small

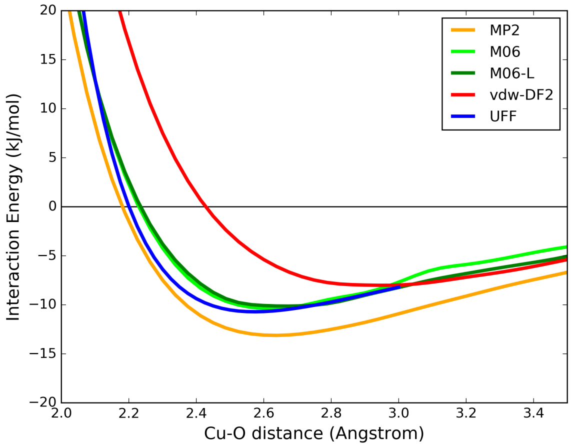


Figure 1.5 – Interaction energy profile for the CO_2-Cu(formate)$_2$ linear scan: the interaction energy is plotted as a function of the distance between the copper atom and the CO_2 molecule's oxygen.

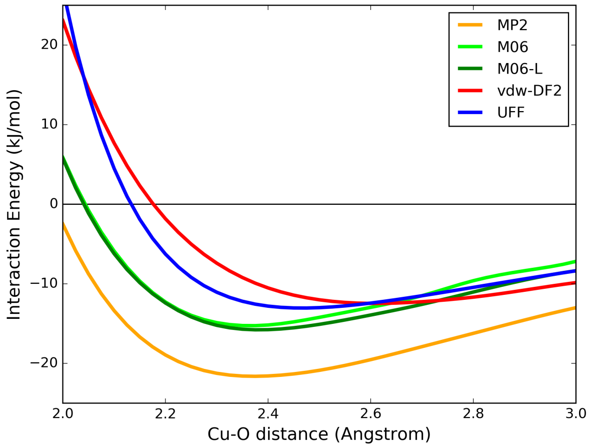


Figure 1.6 – Interaction energy profile for the CO_2-Cu_2(formate)$_4$ linear scan: the interaction energy is plotted as a function of the distance between the CO_2 molecule's oxygen and the closest copper.

influence of dispersion forces: the electrostatic contributions represent 96% of the interaction at the optimal distance of 2.50 Å.

In the Cu_2(formate)$_4$ case, vdW-DF2, M06 and M06-L underestimate the interaction energy compared with ROS-MP2 by 9.1 kJ/mol, 6.4 kJ/mol and 5.8 kJ/mol, respectively. Moreover, if compared to the mono-copper system, the ROS-MP2 calculation leads to a binding energy which is 8.5 kJ/mol more stable in this di-copper system.

In a second series of calculations we optimized the position of the CO_2 molecule, keeping the Cu_2(formate)$_4$ cluster rigid. The CO_2 molecule creates an angle with the copper-copper line of 109 (vdW-DF2) to 116 (M06-L) due to both the interaction of the lone pair of CO_2 oxygen with the copper and the partially positive CO_2 carbon with partially negative oxygen from the paddle-wheel. This optimized configuration is referred here as "tilted" position,

Table 1.3 – Energy of interaction (kJ/mol) between Cu_2(formate)$_4$ and CO_2 in linear and tilted conformation. For all the calculations that employ gaussian basis functions, the energies obtained without counterpoise correction are reported in parentheses. ROS-MP2 calculations without augmented basis function are included to show the variability due to their exclusion in computing interactions.[70] ROS-MP2/ANO-RCC calculations are also compared with CASPT2 results in Section 3.3: for consistency we used the same basis set as BS2, with triple-ζ quality plus polarization on Cu, O, C atoms and double-ζ quality plus polarization on H atoms.

Method	**linear CO_2 Inter. Energy (kJ/mol)**	**Cu-O distance (Å)**	**tilted CO_2 Inter. Energy (kJ/mol)**	**Cu-O distance (Å)**	**Cu-O-O angle (deg)**
FF(UFF/UFF)	-13.0	2.5	-14.3	2.5	127.4°
ROS-MP2/cc-pVTZ	-18.2 (-24.5)	2.4	-22.9 (-31.3)	M06-L opt	M06-L opt
ROS-MP2/ANO-RCC(BS2)	-20.4 (-38.0)	2.4	-24.8 (-43.3)	M06-L opt	M06-L opt
ROS-MP2/aug-cc-pVTZ	-21.6 (-27.1)	2.4	-27.2 (-33.1)	M06-L opt	M06-L opt
M06/aug-cc-pVTZ	-15.2 (-17.7)	2.4	-21.9 (-25.3)	2.4	114.5°
M06-L/aug-cc-pVTZ	-15.8 (-18.6)	2.4	-23.3 (-26.0)	2.4	115.9°
vdW-DF2/cutoff=60Ry	-12.5	2.6	-18.4	2.6	109.9°

because of the CO_2 inclination with respect to the the CuO_4 plane. The interaction energies between Cu_2(formate)$_4$ and the linear and tilted configurations of CO_2 computed with different methods are reported in Table 1.3.

Based on quantum calculations, the tilted conformation binding energy is ca. 5.5-7.5 kJ/mol larger than the linear conformation binding energy. The force field model, based on pairwise interactions, underestimates this difference at only 1.3 kJ/mol.

Finally, we tested the possible additive effect on the CO_2 binding energy by adding a second CO_2 molecule bonded symmetrically on the other copper of Cu_2(formate)$_4$. The binding energies computed for this system don't show any significant deviation (-21.0 kJ/mol and -26.9 for the linear and tilted conformations respectively) and therefore any additive effect can reasonably be neglected.

To summarize, vdW-DF2 underestimates the CO_2-Cu_2(formate)$_4$ binding energy by 9.1 and 8.8 kJ/mol, respectively for the linear and tilted configurations, if compared to the ROS-MP2/aug-cc-pVTZ calculations. Considering the ROS-MP2 results, we are now able to improve our model for HKUST-1 and similar copper paddle-wheel MOFs.

1.4.3 Multireference calculations

To have more insight into the interaction between CO_2 and Cu_2(formate)$_4$, we performed wave-function based multireference complete active space calculations, followed by second order perturbation theory.

A variety of active spaces were explored, including an active space with two electrons in two orbitals (2,2) and one with ten electrons in ten orbitals (10,10). The (2,2) CASSCF calculation is equivalent to a restricted open-shell (ROS)-HF calculation, while the (2,2) CASPT2 calculation is equivalent to the ROS-MP2 calculation. Notice that a singlet CASSCF (2,2) active space indeed corresponds to a multireference calculation in the sense that it generates a wave function that is the combination of two configuration state functions (or Slater determinants). Both the singlet and triplet lowest spin states were explored. In all cases the singlet state is the ground state and it lies 3 kJ/mol lower than the triplet state. This result is in good agreement with the experimental values obtained for MOF-11: 3.4 and 5.3 kJ/mol respectively for the water bound and the anhydrous structure.[71] It is also in good agreement with the 3.2 kJ/mol Maurice et al. calculated with DDCI3 on a similar system, copper acetate monohydrate.[72]

In the following we will discuss the energetics and electronic structure configurations of the singlet. However, as discussed above and also in the literature,[33, 34, 35] it is reasonable to expect that the open-shell singlet and the triplet potential energy surfaces have a parallel shape. The singlet state is a linear combination of two electronic configurations with 50% weight each (See Table S1). The first configuration corresponds to orbital MO_1 doubly occupied (MO_1^2, Figure 1.7a) and the second to orbital MO_2 doubly occupied (MO_2^2, Figure 1.7b). In the (2,2) calculations these orbitals are the only ones included in the active space. They have an average occupation number of about 1 each (because each of them has only a 50% probability of being doubly occupied). In the (10,10) calculation, the other orbitals included in the active space are π and π^* orbitals on the O and C atoms of the paddles. They have occupations of 2 and 0, respectively, within each pair. Additional details regarding the active space orbitals, including visual plots of the (10,10) orbitals, are presented in the Supporting Information.

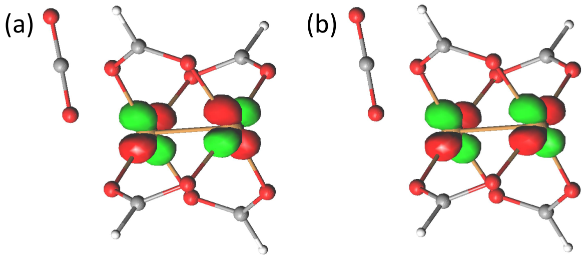


Figure 1.7 – The two molecular orbitals MO_1 (a) and MO_2 (b), in the tilted di-copper system at equilibrium, with their occupation number in parentheses. In the linear system they look similar. Their occupation number is 1. They correspond to an overall configuration of 0.51 MO_1^2 + 0.49 MO_2^2.

The binding energies are reported in Table 1.4. In the di-copper system, the binding energy of CO_2 to Cu is significantly larger than in the mono-copper case, as already discussed in Section 3.2. This behaviour can be explained by inspection of the electronic configuration of the Cu_2 system. The two Cu atoms are close enough to have electronic communication and the overall wave function is a superposition of two electronic configurations. A multiconfigurational method is therefore needed to correctly describe this system in the singlet ground state. The mono-copper system, on the other hand, has a single configuration, which is reasonably well described by MP2. The triplet state of the Cu_2 system is also single-configurational.

The interaction energies for the singlet and triplet states are very similar (within 1 kJ/mol) and only the singlet energies are reported in Table 4. Our results show that an active space of (2,2)

Table 1.4 – CASPT2 interaction energies (kJ/mol) between Cu_2(formate)$_4$ and CO_2 in linear and tilted conformations for different active spaces and different basis sets for the singlet ground state. The distance between CO_2 and copper is 2.4 Å for both the linear and the tilted conformations. Values include counterpoise correction. Values without counterpoise correction are in parentheses.

configuration	Active Space	BS1	BS2	BS3
linear	(2,2)	-15.0 (-43.2)	-18.7 (-33.5)	-20.2 (-31.7)
linear	(10,10)	-14.8 (-46.6)	-18.6 (-36.8)	-20.1 (-35.0)
tilted	(2,2)	-17.7 (-49.3)	-23.5 (-40.0)	-25.8 (-39.6)
tilted	(10,10)	-15.8 (-51.1)	-21.8 (-41.7)	-23.9 (-41.1)

followed by PT2, equivalent to ROS-MP2, is sufficient to describe the binding of this system, as the binding energy does not change by more than 2 kJ/mol when increasing the active space to (10,10).

Basis set effects were explored for the CASPT2 calculations. Table 1.4 shows that going from BS1 to BS2 the uncorrected binding energy decreases by about 10 kJ/mol, while it remains almost unchanged going from BS2 to BS3. The counterpoise-corrected binding energies change by 3-6 kJ/mol going from BS1 to BS2, while again undergoing little change when going from BS2 to BS3. The CASPT2 results with the (2,2) active space reported in Table 1.4 should be compared to the ROS-MP2/ANO-RCC (BS2) results reported in Table 1.3. The only difference between these two sets of results is that those in Table 1.3 are obtained for the triplet, while those in Table 1.4 are obtained for the open-shell singlet and with unfrozen 3p orbitals for Cu. The two sets of values including counterpoise corrections differ by less than 2 kJ/mol, and more generally the most accurate CASPT2/BS3 energies agrees well with the ROS-MP2/aug-cc-pVTZ values, especially in the linear conformation (difference of 1.5 kJ/mol).

1.4.4 Correction of the force field

In order to model properly the interaction of the carbon dioxide with the open metal site in a classical force field, we needed to correct the potential energy curve based on our first principle calculations. The most representative path for different CO_2-Cu distances is the one where the energy is mainly influenced by the interaction with the cation rather than the interaction with other atoms of the cluster (or framework). Hence, we fitted the linear CO_2-Cu_2(formate)$_4$ curve obtained with the ROS-MP2/aug-cc-pVTZ method to obtain the new parameters for the force field. Only the Cu-O van der Waals potential was tuned, while keeping the standard UFF parameters for all other atoms pairs and REPEAT (PBEsol derived) point charges to model electrostatic interactions. For the Cu-O interaction, a Buckingham potential was adopted to correctly represent the repulsion at short distance and an r^{-8} attractive term was added to account for the stabilization observed in the ROS-MP2 calculations. The details about the fitting and the coefficient for the Cu-O potential are reported in Supporting Information. The optimal CO_2 interaction with Cu_2(formate)$_4$, which corresponds to the tilted conformation, computed with the fitted force field parameters has a value of -23.2 kJ/mol. This result is consistent with the UFF difference between the linear and tilted configurations of -1.3 kJ/mol. We notice that by applying this relatively simple but effective correction, obtained without modifying the pairwise interaction with other atoms and without introducing a specific contribution based on the Cu-CO_2 angle, the minimum interaction energy obtained for Cu_2(formate)$_4$ is in fair agreement with the ROS-MP2 result of -27.2 kJ/mol.

Finally, we replicated the GCMC simulations in HKUST-1, using our UFF parameters with the corrected Cu-O potential. The comparison with experimental date is reported in Figure 1.8. The simulations are still slightly underestimating the measured uptake, and this is reflecting the previously mentioned underestimation of ca. 4 kJ/mol for the interaction energy in the optimal tilted configuration. However, the assumptions made for the force field are sufficient to obtain a good representation of the uptake around ambient temperature, and a significant improvement with respect to the standard force fields.

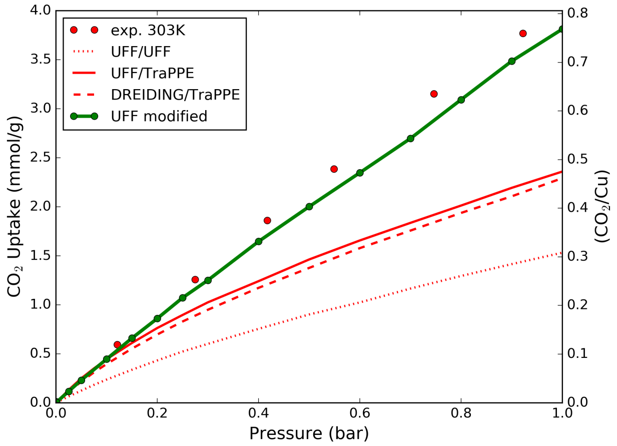


Figure 1.8 – Comparison between the experimental[31] and simulated isotherms for CO_2 inside HKUST-1 at 303K. The modified UFF forcefield is obtained by fitting the Cu-O potential on ROS-MP2 calculations.

The minimum energy of interaction computed with our new force field in the three main adsorption sites of HKUST-1, i.e. open metal, small pore window, and small pore center sites, are now ranked correctly: -27.3 kJ/mol, -26.8 kJ/mol, and -26.8 kJ/mol respectively, and the OMS stability is not underestimated any longer compared to the *in situ* experimental results.

As a starting point for our correction, we used UFF/UFF mixed parameters instead of UFF/TraPPE or DREIDING/TraPPE, because from the simulated isotherm (Figure 1.8) we can observe that these last force fields are already predicting the experimental uptake at very low pressure (below 0.1 bar), even if the open metal site interaction is strongly underestimated. This is an artifact due to a fortuitous error cancellation with the overestimation of the interaction in other sites, i.e. the small pores centers (see Table 1.1), which are already saturated at 0.82 mmol/g, as clearly shown by the deviation of the simulated isotherm from the experimental one. Therefore, employing the conventionally used UFF/UFF, UFF/TraPPE or DREIDING/TraPPE mixed parameters to describe the guest-host interaction in an analysis of the site occupancy would lead to wrong conclusions, i.e. that in HKUST-1 the open metal sites are very poorly occupied at low uptake.[73] Ulterior comparisons with experimental data is provided in the Supporting Information: CO_2 uptake at higher pressure and different temperatures[74], and the heat of desorption in function of the uptake.[31, 75, 76]

1.4.5 Investigation of the "double" open metal site interaction in Cu-TDPAT

To further test the reliability of our force field, we investigated another interesting copper paddle-wheel metal organic framework, Cu-TDPAT, firstly synthesized by Li *et al.*[77] The crystalline structure is characterized by the presence of strong adsorption sites for CO_2, where both oxygens of the guest molecule are attracted to two different copper cations (Figure 1.9), leading to an interaction energy which is roughly double with respect to the conventional single open metal site of copper paddle-wheel.

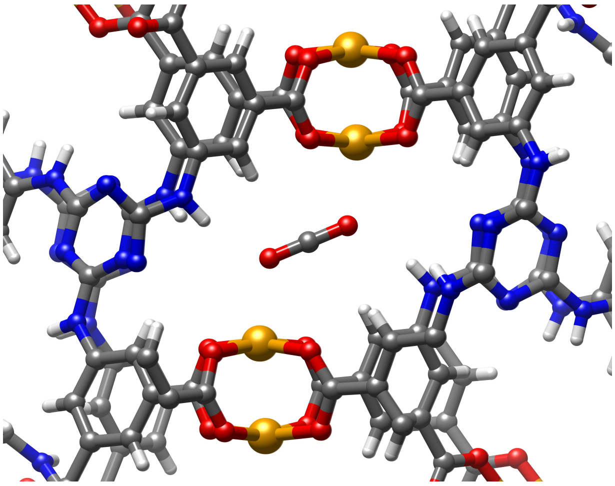


Figure 1.9 – CO_2 molecule adsorbed in the double open metal site of Cu-TDPAT.

Because of this reason, Cu-TDPAT is one of the top performing MOFs for both gravimetric and volumetric CO_2 uptake at ambient pressure.[78] The conventional unit cell of Cu-TDPAT contains 48 copper cations: 24 of them compose 12 double open metal sites while the remaining 24 atoms compose 24 single open metal sites, with a conformation very similar to the OMS of HKUST-1. Due to the large dimension of the unit cell (960 atoms), the crystal is too big to perform a DFT calculation with an accuracy comparable with our previous calculation on HKUST-1. Consequently, we employed the Extended Charge Equilibration (EQeq) method[79] to compute the partial charges of the framework. This method is able to self-consistently compute point charges for MOFs, with results very similar to the charges obtained by fitting the electrostatic potential from a quantum calculation, e.g. REPEAT. HKUST-1 itself was successfully tested in the original paper presenting the EQeq method.[79] Compared to the quantum electrostatic potential fitting, this method is drastically faster (a few minutes instead of hours for HKUST-1) and is applicable to a unit cell containing a large number of atoms, which is practically forbidden to DFT calculations. The result obtained for the copper paddle-wheel is consistent with our PBEsol calculation in HKUST-1. Using the EQeq method the average point charges for Cu-TDPAT are 0.905 and -0.398 for the copper and the carboxylic oxygen respectively, versus 0.914 and -0.57 for HKUST-1 computed using REPEAT. With the new set of parameters, we compared the results of the GCMC simulations to experimental data (Figure 1.10).

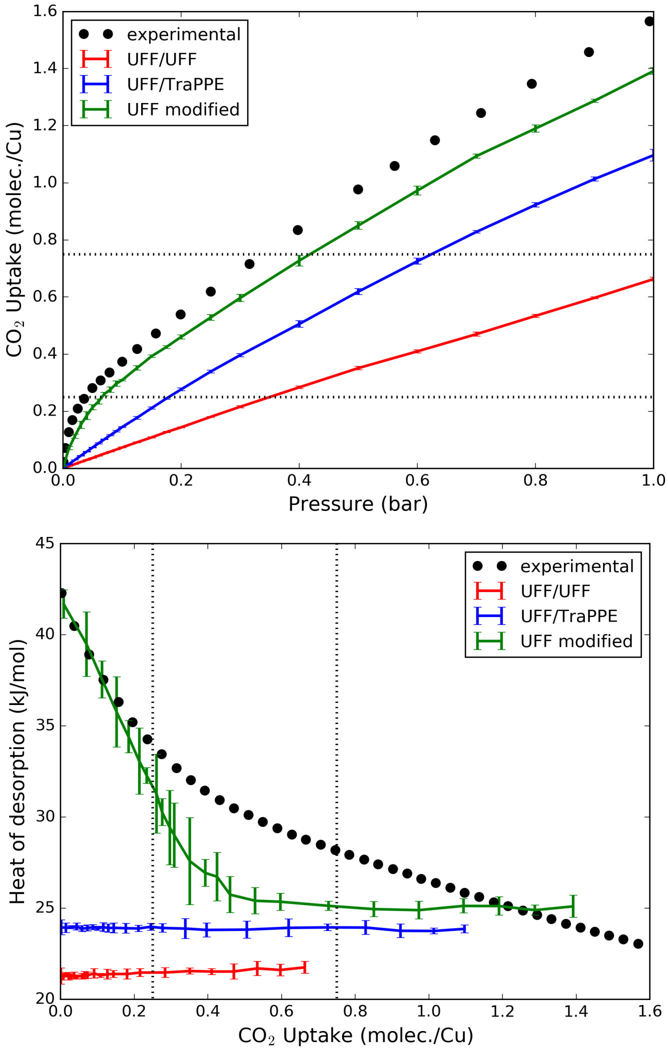


Figure 1.10 – Comparison of experimental[77] and simulated adsorption of CO_2 in Cu-TDPAT at 298K using different set of parameters. The force field developed in this method is reported as "UFF modified", while UFF/UFF and UFF/TraPPE are conventionally used standard sets of parameters. In both plots the uptake is converted to CO_2 molecules per copper ratio, and the equivalence to the number of double open metal sites (0.25 CO_2/Cu) and the number of total open metal sites (0.75 CO_2/Cu) is highlighted with a dotted line. The experimental heat of desorption (black dots, right picture) has been computed through the Virial-Langmuir method while the simulated values (colored lines) are computed from the guest molecules number fluctuation in the GCMC simulation.

This comparison shows a good agreement, as for HKUST-1, which gives us some confidence in the transferability of our force field to model CO_2 adsorption. Moreover, it becomes more evident how UFF/UFF and UFF/TraPPE parametrizations do not capture the strong interaction between CO_2's oxygen and copper.

1.5 Conclusions

In this work we have shown that the Cu-Cu interaction in copper paddle-wheel systems is the reason why DFT methods, even when they include dispersion corrections, systematically underestimate the interaction between CO_2 and copper paddle-wheel motif. Our calculations confirm the presence of the copper-copper coupling, influencing the attraction of the CO_2, and suggest that the mono copper cluster Cu(formate)$_2$ is not a realistic model to describe this interaction.

One thus needs an electronic structure theory that properly describes the Cu-Cu interaction, such as the ROS-MP2 wave function. We show that if this interaction is included in our calculations, the prediction of the binding energies is in better agreement with the experimental data. To justify the choice of ROS-MP2 method, which is equivalent to a (2,2) CASPT2 calculation, we performed a number of multireference calculations over a variety of active spaces, basis sets, and spin states. We concluded that the ROS-MP2 level of theory is good enough to model the Cu_2(formate)$_4$-CO_2 interaction.

Using the ROS-MP2 results, we reparametrized the UFF pairwise potential to correctly model the interaction of CO_2 with the open metal site in HKUST-1, which was severely underestimated by conventional force fields. The results obtained from our new force field agree with experimental isotherms as well as with *in situ* PXRD studies, which found the open metal site to be the strongest adsorption site for CO_2 rather than the small pore center site. The correction proposed in this work acts in proximity of the open metal site, and this local character of the correction terms allows us to transfer the same parameters to other MOFs containing the copper paddle-wheel motif.

To test this transferability, we employed our force field to model the CO_2 interaction with the "double" open metal sites of Cu-TDPAT framework, and we showed a significant improvement with respect to conventional UFF parameters in this case as well. The modified set of parameters proposed makes it possible now to accurately describe the adsorption behavior for this class of MOFs. In this study we have focused on CO_2, but similar effects can be expected for other polar molecules.

2 Accurate Characterization of the Pore Volume in Microporous Crystalline Materials[1]

December 26, 2017 Volume 33, Number 51
pubs.acs.org/Langmuir
LANGMUIR
The ACS journal of fundamental interface science
Computing the Pore Volume in Microporous Crystals
ACS Publications
Most Trusted. Most Cited. Most Read.
www.acs.org

---

[1] Preprint version of the article: Daniele Ongari, Peter G. Boyd, Senja Barthel, Matthew Witman, Maciej Haranczyk and Berend Smit, *Langmuir*, 33(51), 14529-14538., 2017, https://doi.org/10.1021/acs.langmuir.7b01682. D.O. did the literature review on the different methods to compute the pore volume, improved the *probe-occupiable* method from an initial idea of P.B., implemented the algorithm in Zeo++ and performed the calculations.

2.1 Abstract

Pore volume is one of the main properties for the characterization of microporous crystals. It is experimentally measurable and it can also be obtained from the refined unit cell by a number of computational techniques. In this work we assess the accuracy and the discrepancies between the different computational methods which are commonly used for this purpose, i.e, geometric, helium and probe center pore volume, by studying a database of more than 5000 frameworks. We developed a new technique to fully characterize the internal void of a microporous material and to compute the probe accessible and occupiable pore volume. This value can be directly compared with experimentally measured pore volumes from nitrogen isotherms.

2.2 Introduction

The internal void volume is an important characteristic of microporous materials, as it will determine its permeability to guest molecules, the adsorption capacity, and many other properties that can be engineered for the industrial applications that involve the use of these material, such as gas separation,[4] gas storage,[3] catalysis,[80] or drug delivery.[81] The field of microporous materials used to be dominated by zeolites, but recently new classes of microporous materials have been published. Examples include Metal Organic Frameworks (MOFs),[82] Covalent Organic Frameworks (COFs),[83] Zeolitic Imidazolate Frameworks (ZIFs),[84] Porous Polymer Networks (PPNs),[85] etc. For each of these classes a large number of different materials can be obtained by combining different ligands and nodes, leading to millions of frameworks, each with different topology, pore shape, and chemistry. For example, at present over ten thousand MOFs and related porous materials have been synthesized,[1] and large databases of computationally hypothesized structures are rapidly expanding.[2] All the main applications for porous materials involve the adsorption of guest molecules in the pores. For this reason it is of critical importance to correctly characterize the pore volumes of these materials as this is the first, and often the only, step to characterize a material.

Therefore, the internal void volume of a porous material can be determined computationally from the crystal structure.[86, 87] This theoretical value of the pore volume can be compared with the experimental pore volume derived, for example, from the nitrogen uptake at low temperature.[88] The comparison of the two values can give some insight on the characteristics of the synthesized crystal. For example, if the experimentally measured void volume is smaller than the computed one, this can be symptomatic of an incomplete desolvation (solvent molecules still trapped inside the pore), limited permeability at the surface, or defects in the crystal. In addition, deviations of theoretical pore volume from the experimental one can also indicate that the reported crystal structure is a poor representation of the actual material.

In this paper we review a number of different methods employed to compute the void fraction.[86, 87, 89, 90, 91] We show that, because of the different assumptions, each method computes a (slightly) different portion of the volume. For some particular cases these differences can be large and, more importantly, the theoretical pore volume cannot be compared with the experimental pore volume. One of the reasons for these differences is that the definition of pore volume depends on the type of probe that is used to measure it. To address this issue we introduce and compute the "probe accessible occupiable volume". It represent the internal free space of the material where a spherical probe can have access and that it can occupy. We will highlight why this measurement can be meaningfully compared with experimental data. To illustrate the need and the importance of this concept of probe occupiable volume, we introduce a simple but representative model of a microporous material to test our algorithm. Then we investigate the discrepancies in the values of the volume as computed by different methods for a set of more than 5000 three-dimensional MOFs from the Cambridge Structure Database (as refined in the CoRE MOF database[38]). Finally we demonstrate some of the practical consequences by considering a sample of ten structures, for which we can directly compare the computed pore volume with available experimental data.

2.3 Methods

2.3.1 Experimental Measurement of the Pore Volume

The internal free volume of a microporous material can be experimentally measured by determining the maximum loading of a gas in the pores of the material. Nitrogen is commonly used for this purpose because of its small size and it is weakly interacting with the framework. In addition, its normal boiling point is sufficiently low (77K) that condensation at the exterior of the pores is avoided before the full saturation inside the pores. The pore volume is obtained under the assumption of validity of the Gurvich rule:[92, 13] the density of the saturated nitrogen in the pores is assumed equal to its liquid density regardless to the shape of the internal void network and, because of the weak interactions, regardless to the chemistry of the framework. The pore volume (ν_{pore}) and the void fraction (θ) are computed from:

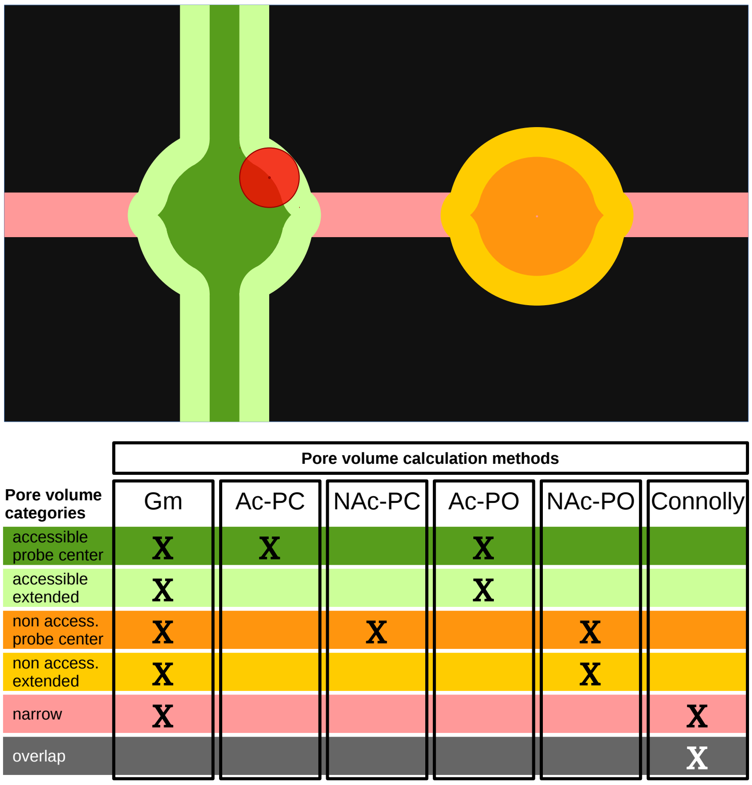
$$\nu_{\text{pore}} = \frac{n_{N_2}^{\text{ads,sat}}}{\rho_{N_2}^{\text{liq}}} \tag{2.1}$$

$$\theta = \nu_{\text{pore}} \cdot \rho_{\text{cryst}} \tag{2.2}$$

where ν_{pore} is commonly expressed in cm^3 per gram of crystal, $n_{N_2}^{\text{ads,sat}}$ is the specific amount of nitrogen adsorbed (grams of nitrogen per gram of crystal), $\rho_{N_2}^{\text{liq}}$ and ρ_{cryst} are the densities of the liquid nitrogen (0.808 g/cm_{liq}^3) and of the material respectively. The commonly used protocol to determine the pore volume involves measuring the nitrogen uptake just before it starts to condensate outside the material, i.e., 0.9 P/P_0,[13] with P_0 being the atmospheric pressure. To compare this pore volume with a theoretical value obtained from the crystal structure it is important to realize that this experimentally measured value is not considering all the small interstices between the atoms where the nitrogen molecule can not fit, nor the non-accessible pores, i.e., the pores connected only by channels too narrow for a nitrogen molecule to enter.

2.3.2 Computational Methods to Assess the Pore Volume from the Unit Cell

To compute the pore volume of a microporous crystal from the knowledge of the atomic structure of the unit cell, there are a number of different methods that are currently employed.[86, 87, 89, 90, 91] Each one computes slightly different portions of the full internal volume, as showed in Figure 2.1.



	Pore volume calculation methods					
Pore volume categories	Gm	Ac-PC	NAc-PC	Ac-PO	NAc-PO	Connolly
accessible probe center	X	X		X		
accessible extended	X			X		
non access. probe center	X		X		X	
non access. extended	X				X	
narrow	X					X
overlap						X

Figure 2.1 – Qualitative two-dimensional model of the unit cell of a microporous material, permeable to a spherical probe (red). Each color corresponds to a different category of volume. In the table, the color coding is explained and it summarized which portions of the volume are considered for each method: geometric pore volume (**Gm**), accessible and non-accessible probe center pore volume (**Ac-PC**, **NAc-PC**), accessible and non-accessible probe occupiable pore volume (**Ac-PO**, **NAc-PO**) and solvent free Connolloy volume.

Here we propose a list of precise definitions to distinguish the volume computed with each method. For all these definitions, the pore volume can be further characterized either as accessible (**Ac**, part of an accessible network) or non-accessible (**NAc**, isolated pocket).

- Geometric pore volume (**Gm**). Defined as all the volume of the unit cell which is not overlapping with the atoms of the crystal. In Figure 2.1 this is the non-black area.
- Probe center pore volume (**PC**). Defined as the volume that the center of a spherical probe can occupy. In Figure 2.1 this is the sum of the dark green area (for pores that are accessible from the outside) and dark orange area (for pores that are non-accessible from the outside).
- Helium pore volume (**He**). In the definition of **PC** volume we assume hard-sphere interactions between the probe atoms and the atoms of the pore. In the definition of the helium pore volume these hard-core interactions are replaced by a more realistic intermolecular potential, which makes this volume depended on the temperature assumed for the calculation. In Figure 2.1 the **He** volume is represented by the same colors as the **PC** volume (the dark green and dark orange).
- Probe occupiable pore volume (**PO**). This is a definition which we introduce here to ensure that the theoretical pore volume matches the pore volume obtained experimentally from the nitrogen isotherms. The experimental definition assumes that we can take the bulk density of the gas and computes the volume from the number of adsorbed gas molecules per unit volume. This volume, however, has no notion of atoms and should be defined as the entire volume **enclosing** all the adsorbed gas atoms. Therefore, if Figure 2.1 this volume has to include the light green (for accessible pores) and light orange areas (for non-accessible pores) in addition to the the dark green and dark orange areas. If we have a system with large pores the difference between the **GM** and **PO** volume is small, but for micropores, however, this difference can be significant.

These pore volumes can be multiplied for the density of the material to be converted to the corresponding void fractions. The frameworks are assumed rigid, i.e., considering the atoms frozen in their crystallographic positions.

For the geometric pore volume **(Gm)** we assume that the atoms can be approximated as spheres with a conventional radius, depending on the atom type and which represent their electron cloud, i.e., the van der Waal (vdW) radius. The analytical calculation of the **Gm** pore volume needs to consider all the many-body overlaps between the atoms. Consequently, the most efficient solution to obtain the geometric pore volume is to perform a Monte Carlo test. A number of points, randomly displaced in the unit cell or taken on a 3D grid, are evaluated: if a point is overlapping with an atom, i.e., the distance of the point with that atom is less than its vdW radius, then a value of zero is assigned to that point. A value of one is assigned otherwise. Therefore, the **Gm** void fraction θ_{Gm} of the crystal from N sample points is obtained as:

$$\theta_{\mathrm{Gm}} = \frac{\sum_1^N \mathrm{value}}{N} \tag{2.3}$$

Consequently the geometric pore volume can be obtained multiplying the void fraction by the density of the framework (Eq. 2.2). In this measurement the volume inside the large pores is summed together with all the small interstices in the framework, which are too narrow to be effectively occupied by a guest molecule. Hence, the value computed in this way will be always an upper bound for the volume that a probe can effectively access.

The probe center pore volume **(PC)**, often named simply "pore volume"[90, 87] considers the shape of the probe used for the measurement, conventionally spheres with a radius of 1.32 Å for helium and 1.86 Å for nitrogen.[38, 93] In this definition it is important to recall that even the nitrogen molecule is treated as a spherical probe, as shown in Figure 2.2.

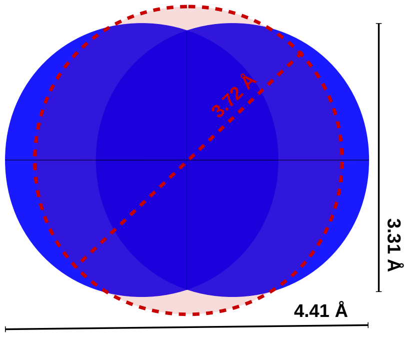


Figure 2.2 – N_2 spherical model of radius 1.86 Å (3.72 Å diameter) compared to the van der Waals representation of the same molecule (using the Lennard-Jones sigma value and the N-N distance of the TraPPE model)[58]

For this calculation, the same Monte Carlo test is performed but this time the radius of the framework's atoms is taken as the sum of the atomic radius plus the probe radius. The obtained void fraction is then representing the portion of the volume which is occupiable by the centers of the probe (Figure 2.3).

It is also important to note that the **PC** pore volume for a probe of zero radius, corresponds to the **Gm** pore volume.

(He) A third solution is to compute the helium pore volume. Similarly to the **Gm** pore volume, a collection of sampling points are considered, but instead of assigning a value of 0 or 1 depending on the overlap with atoms, this time the Boltzmann factor (BF), related to the insertion of an helium atom, is computed.

$$\mathrm{BF} = \exp\left(\frac{-E_{\mathrm{int}}}{K_b T}\right) \tag{2.4}$$

E_{int} is the energy of interaction of the helium atom with the atoms of the framework, as computed using, for example, the Lennard-Jones potential (see SI). Similarly to the previous cases, the void fraction θ_{He} (and therefore the pore volume) is computed as the average over

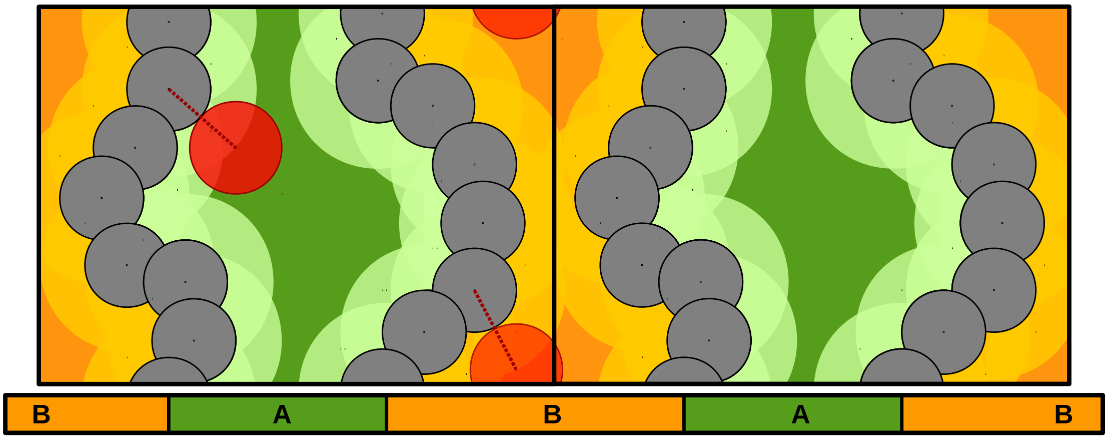


Figure 2.3 – Two-dimensional example of the probe center pore volume calculation. The periodic unit cell is duplicated in the y directon. The radius of the framework's atoms (black) is expanded by the radius of the red probe (light green and light orange). The remaining area is what we define as probe center pore volume (dark green and dark orange). The framework is composed by two channels: channel A (green) which is accessible and channel B which is non-accessible (orange). Channel B is too narrow for the probe to pass from one side to the other, and can be referred as an isolated pocket.

all the sample points:

$$\theta_{\mathrm{He}} = \frac{\sum_1^N \mathrm{BF}}{N} \tag{2.5}$$

It is worth noting that this measurement is influenced by the force field and the temperature used. It is therefore important to use a consistent choice to compare different sets of results.[90] We need to stress that the **He** void fraction, in the way it is measured, does not correspond to the amount of helium that can saturate in the pores. The physical meaning of the **He** void fraction is linked to the probability of a single helium atom to be adsorbed in the framework at a certain temperature, which is chosen to be 298K by convention.[90]

(PO) At this point it is important to recall that none of the previously summarized methods to compute the pore volume exactly matches with the pore volume we obtain from the nitrogen isotherms. To arrive at a definition of pore volume that can be directly compared to experiments we introduce the probe occupiable pore volume, and we propose an algorithm to compute it. We use the term "occupiable" to define the portion of the space that can be spanned by the probe, that should not be confused with the term "accessible" (**Ac**), which defines the pores where the probe can have access.

2.3.3 Accessible versus non-accessible channels

In these Monte Carlo simulations we are probing a number of points within the unit cell to measure the void fraction (and so the pore volume) of the bulk material. However, it is also important to know if the detected free space is accessible from the outside, i.e., if a cavity forms a multi-dimensional network where a guest molecule can enter at the solid/gas interface and diffuse. The same analysis allows to detect if a solvent molecule is able to exit the pores, and a synthesized crystal can be effectively desolvated.

This concept of accessibility is obviously related to the size of the molecule, represented as a spherical probe, that we are interested to evaluate. Once we computed the **PC** volume, we can further categorize this internal space as accessible (**Ac-PC**) or non-accessible (**NAc-PC**) by considering whether it composes or not a multi-dimensional network along the periodic boundaries. This is illustrated in the two-dimensional example of Figure 2.3: the central channel (A) is accessible to the probe while the other one (B) is not, because the **PC** pore volume doesn't form a continuous path. The accessibility test can be performed by doing a percolation analysis along the edges obtained from the Voronoi decomposition[94] or analysing a grid of points.[95, 96]

The same concept can be applied to compute the **Ac-PO** (as presented in the next section) or the **Ac-He** pore volume. In the second case, one needs to first assume an energy cutoff for the helium-framework interactions, which defines the regions that are diffusively inaccessible on an experimental time scale (e.g., 15 k_bT). Then, one must consider the regions of the volumes where the interaction energy is lower than the cutoff to perform a percolation analysis.[97] For what concerns the **Gm** volume, the calculation considers a dimensionless probe and therefore we do not have any practical interest in analysing its accessibility.

2.3.4 Algoritm to compute the occupiable pore volume

In this section we propose an algorithm to obtain the experimental pore volume from our definition of the accessible and occupiable volume (**Ac-PO**) and in general to fully characterize the internal volume of a microporous material.

1. Let us consider a set of N sample points, randomly selected within the unit cell.
2. For each point we compute its distance to the framework's atoms: if this distance is smaller than the atomic radius, the sample point is categorized as "overlap", if it is larger than the sum of the atomic and the probe radius, it is categorized as **PC**. If the second condition is verified, we compute the distance δ between the point and the surface of the **PC** volume, defined as:

$$\delta = d - r_{\text{probe}} - r_{\text{atom}} \tag{2.6}$$

with d being the distance to the closest atom, r_{probe} the radius of the spherical probe and r_{atom} the vdW radius of the closest atoms of the framework (Figure 4). In addition, we use a percolation algorithm[87] to further classify the sample point as **Ac-PC** or **NAc-PC**.

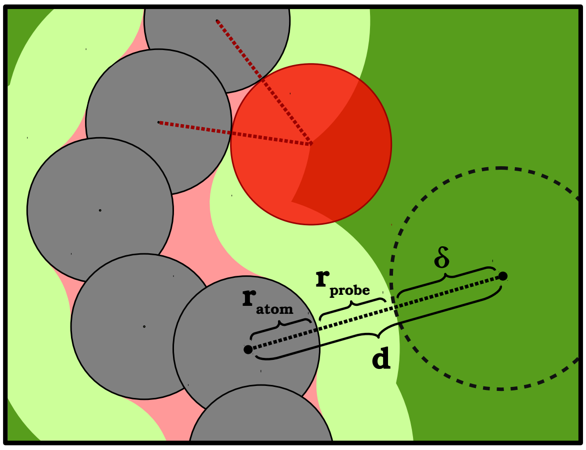


Figure 2.4 – The calculation of δ from eq.(2.6) is showed in a two-dimensional model. The color coding is consistent with the one reported in Figure 2.1: the atoms (which correspond to the overlap volume) in black, the probe in red, the accessible probe center volume in dark-green, the accessible extended volume in light-green and the narrow volume in pink.

3. For every sample point left and every **Ac-PC** marked point, the algorithm computes the distance between the two points and it tests if the uncategorised points is closer to the **Ac-PC** surface than the probe radius.

$$\text{distance} < \delta + r_{\text{probe}} \tag{2.7}$$

   If the inequality is verified, the uncategorised point will be considered as part of the now defined "accessible extended volume" (light green in Figure 2.1). The inclusion of δ improves the speed and the accuracy of the algorithm, because in this way also the internal points of the **Ac-PC** volume are giving some information on the position of its surface.

4. The same test is performed for the **NAc-PC** points: in case of success uncategorised points will be marked as belonging to the "non-accessible extended volume" (light orange in Figure 2.1).

5. If none of the previous test is true, it means that the sample point belongs to what we define as "narrow volume" (pink in Figure 2.1).

It follows that the **PO** volume is given by the summation of the probe center and the extended volume. Figure 2.1 presents all the different categories of volume with color coding for an illustrative two-dimensional model.

With this definitions, we marked as "narrow" the whole volume that can not be touched by the probe because it is hindered by the framework. It can be the case of a narrow channel (pink, Figure 2.1) or the small interstices between the atoms of the crystal (pink, Figure 2.4). Moreover, the overlap volume summed to the narrow volume, gives what is commonly defined in biochemistry as the "solvent free volume" computed analytically by Connolly[91] (Figure 2.1).

2.4 Computational versus experimental pore volumes

Now that we have fully characterized the pore volume inside a microporous framework, we can couple the result with experimental measurements. Under the assumption of Gurvich rule the experimental 77K nitrogen's pore volume can be compared with the **Ac-PO** pore volume computed from the unit cell, using a spherical N_2 probe. The nitrogen's **NAc-PO** pore volume, could also be measured experimentally with smaller probing molecules, e.g., hydrogen or helium, or with Positron Annihilation Lifetime Spettroscopy (PALS).[98] The measurements with these techniques is not as frequently used. If we would like to compare our theoretical calculations with these experiments we need to use an appropriate probe for the calculation.

We stress once more that for these methods the thermal vibrations of the atomic positions are not taken into account. Moreover, for the **Ac-PO** calculation we use hard-sphere potentials for which the effective volume does not depend on temperature. These assumptions holds for the experimental conditions (i.e., 77K for nitrogen adsorption). However, we do assume that the crystal structure does not change upon adsorption of nitrogen (e,g, swelling or ration of ligands). For cases where the diameter of the channel is very similar to the diameter of the probe, further investigations are needed,[99] because a small distortion of the framework or a different choice of the parameters can drastically change the amount of **Ac** and **NAc** volume detected.

2.5 Software and parameters

In this section we illustrare how the different pore volumes are determined in the different software packages that compute pore volumes.

The Poreblazer package[86] computes the **Gm** and **He** pore volume using sample points laying on a grid with 0.2Å bin size.

The Zeo++ package[87] gives the **Gm** and the **PC** volume, the first one being obtained by setting the radius of the spherical probe to zero. In this software the number of sample points specified in the input is randomly displaced in the unit cell.

The PLATON package[89] computes the **PO** volume using a grid of points. Points belonging to the **PC** pore volume are first detected, and then their neighbour points are considered.

Contrary to Zeo++, this software does not distinguish between **Ac** and **NAc** volume. Also, one should pay attention to the terminology: in this software the authors define as "accessible" volume what here we define as "occupiable" volume.

The Raspa package,[100] (which is mainly used for for Monte Carlo and molecular dynamics simulations), provides the **He** pore volume considering a specified number of sample points in random positions of the unit cell.

The algorithm we proposed in this work to compute the **Ac-PO** volume and fully characterize the internal pore volume has been implemented as an extension of Zeo++.[87]

In our calculations, the **He** volume is computed at 298K (25°C), which are the typical conditions of most previous calculations.[90] We used the Lennard Jones potential to describe the dispersion interactions, applying the Lorentz-Berthelod mixing rules and considering a cutoff distance of 12.8Å, beyond that the potential is set to zero. Parameters for the framework and for helium were taken from the Universal Force Field (UFF)[59] and from Hirschfelder,[101] respectively. Concerning the "hard spheres" calculations, (**Gm**, **PC** and **PO**) and for all the softwares (Poreblazer, Zeo++ and PLATON), the Lennard-Jones sigma values from UFF were used as the diameter of the framework atoms, to be consistent with the **He** calculations. A kinetic radius of 1.86 Å was considered for nitrogen.[93]

2.6 Results and discussion

2.6.1 3D Model for the Full Characterization of the Pore Volume

To illustrate the difference between the various approaches we applied our algorithm on a three-dimensional model which is able to represent qualitatively the characteristics of a microporous material, inspired by the two-dimensional example reported in Figure 2.1. The model has one accessible pore and one non-accessible pore, with a narrow channel (i.e., with a diameter smaller than the probe's) connecting the two. The model is built with a large number of spheres laying on a grid to represent the framework, and leaving free space that corresponds to pores and channels (Figure 2.5, upper).

In this simplified model of a porous framework we can really distinguish between all the different categories of internal volume listed in Figure 2.1: the result from the analysis with 500,000 sample points is shown in Figure 2.5 (lower) using the same color coding for the points.

To assess the convergence of the method we run our algorithm for different numbers of sample points (Figure 2.6). One can see that the **Ac-PO** calculation is converged to the 0.1% of the void fraction with 10 points per cubic angstrom. Within our algorithm, to measure the **PO** void fraction, we need first to accurately locate the surface of the **PC** volume and expand this volume by the length of the probe radius. To minimize the error associated to

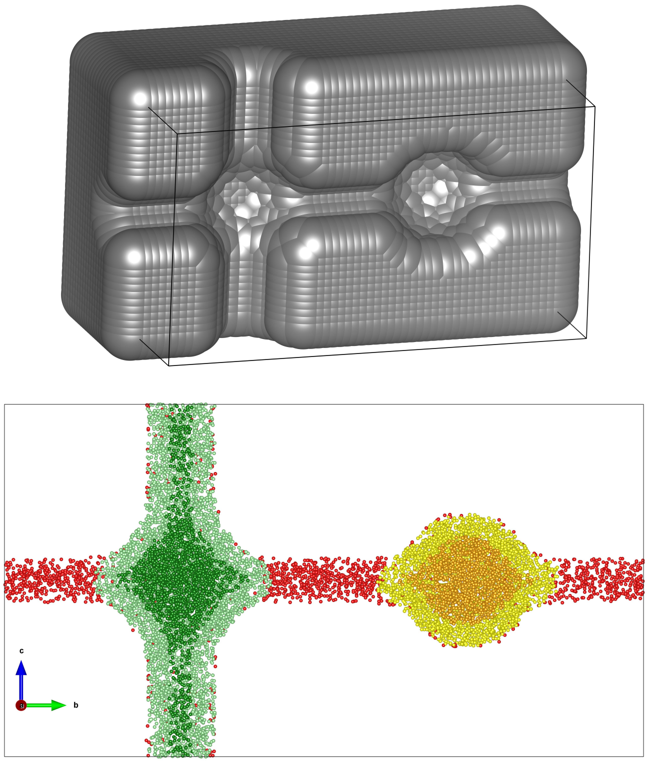


Figure 2.5 – Section of the 3D model (top) and analysis of the pore volume using 500,000 points (bottom). For the color coding one has to refer to Figure 2.1, overlap points were omitted. The probe has a diameter of 2 Å. The diameter of the accessible and the narrow channels are 3 Å and 1.5 Å, respectively.

a poorly sampled **PC** surface, one should increase the number of of sample points, albeit with a significant computational cost. Nevertheless, in real frameworks we can consider as reasonable to use a convergence within the 1% of the void fraction, to compare the calculated values with experimental data.

2.6.2 Comparison of Different Pore Volume Definitions with Experimental Data for HKUST-1

The triclinic unit cell structure of HKUST-1 (CSD code: FIQCEN) was considered to compute the void fraction with the different methods, and compare the values. Water solvent molecules were removed from the original deposited structure.[38] No **NAc** volume was detected. The resulting void fraction and computational time is reported as a function of the number of samples per cubic Angstrom that were used for the calculation (Table 2.1).

We use as experimental value for the void fraction 0.678,[13] which is the highest value we could find in the literature for the desolvated crystal. Lower values were reported in literature, from 0.590 to 0.660.[102, 103, 104, 105, 75] The computed **Ac-PO** void fraction converges to a

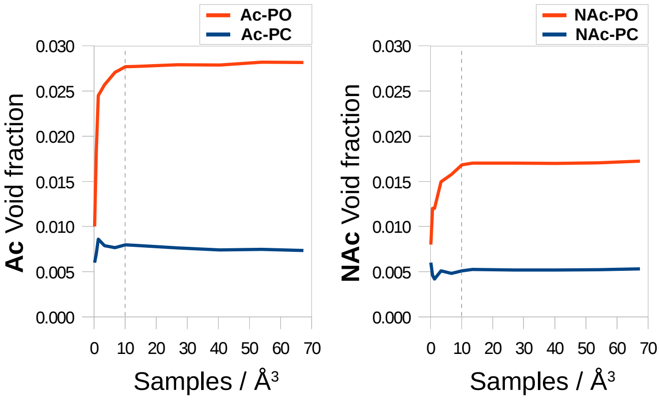


Figure 2.6 – Assessment of the different volume fractions in the three-dimentional model of Figure 2.5, for an increasing number of sample points per cubic angstrom. One can note how different are the values for the **PC** and the **PO** void fraction. Numerical data provided in Table S2.

Table 2.1 – Assessment of the **Ac-PC** and **Ac-PO** void fraction in the FIQCEN structure (HKUST-1 with triclinic unit cell), for an increasing number N of sample points. The cpu time refers to a 3.60 GHz Intel processor: for this framework the time needed for the calculation is proportional to ~ $N^{1.5}$.

Ac-PC	Ac-PO	N	N/Å^3	cpu time (s)
0.240	0.606	4571	1	2
0.250	0.649	22851	5	9
0.250	0.656	45702	10	28
0.249	0.658	68553	15	55
0.250	0.660	91404	20	90
0.248	0.663	228510	50	475
0.248	0.664	342765	75	998
0.249	0.665	457020	100	1702

value which is close to the experimental result while the **Ac-PC** void fraction is significantly smaller. The **PO** void fraction computed with the CALC_SOLV routine in PLATON is 0.654: this result was obtained in 165 seconds with a minimum grid spacing (0.14 Å). These settings gives 365 samples per cubic Angstrom and it is the most accurate sampling that the program can manage before crashing. The **Gm** void fraction of 0.708 reasonably agrees with the **Ac-PO** value, meaning that the percentage of narrow volume is negligible. On the other hand, the **He** calculation gives a value of 0.764 which overestimates the experimental void fraction. It is surprising to note that using a different parametrization for the Lennard-Jones interactions, i.e., UFF's[59] instead of Hirschfelder's[101] parameters for helium, we obtain an **He** void fraction of 0.947, which disagrees with the experimental and the **Ac-PO** values. This evidence motivated a deeper analysis of the physical an mathematical meaning of the **He** calculation.

2.6.3 The Helium Void Fraction

The **He** calculation is very commonly used to compute the void fraction.[106, 107] As we demonstrated in the previous section, its value depend strongly on the force field parameters used to model the helium-framework interactions, and it can lie far off the experimental value. Therefore, we analyse the underlying mathematical reason for this variability. First, we study the case of an helium atom interacting with a carbon atom, using the Hirschfelder-UFF parameters to represent their interaction at different distances. The potential and the Boltzmann factor (BF) for at different He-C distances is shown in Figure 2.7.

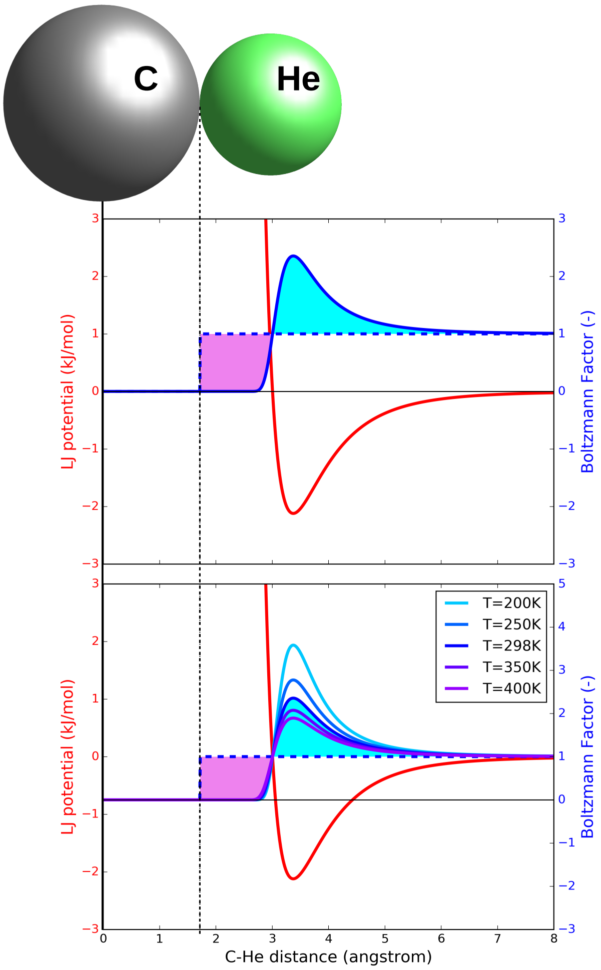


Figure 2.7 – One dimensional representation of the Lennard-Jonnes potential and the associated Boltzmann factor in function of the C-He distance (system shown on the top). The Boltzmann factor (blue solid line) function is compared to the factor associated to the occupability of the space i.e., one everywhere outside the carbon's van der Waals radius (blue dashed line). In the bottom figure the sensitivity of the Boltzmann factor to the arbitrary value of the temperature is investigated. Notice that doubling or halving the temperature corresponds to respectively halving or doubling the value of epsilon for the Lennard-Jones interaction.

We can now compare the **He** calculation to the **Gm** calculation (in this bi-atomic model the

Gm and **Ac-PO** volumes are equivalent). For the **He** calculation the BF is the value assigned for every He-C distance while for the **Gm** calculation we assign a value of zero for a He-C distance inferior to the carbon's radius (equivalent to half the Lennard-Jones's sigma for carbon) and a value of one elsewhere (see the dashed blue line in Figure 2.7). Therefore, the void fraction is the integration of these values over the entire volume considered. The **He** and the **Gm** coincide exactly in the case when the two integrals are equal, i.e., when there is a match between the cyan and purple areas in Figure 2.7. The BF depends on the set of parameters used and on the temperature assumed in the calculation. Indeed, the common choice of the temperature of 298K is just a convention, and its variation can drastically affect the **He** calculation, as shown in Figure 2.7. Moreover, the **He** void fraction is not strictly restricted to be smaller than one, since also the BF can take values larger than one, especially for framework's atom with a large epsilon Lennard-Jones parameter. In UFF, for example, the epsilon value for aluminium, silicon and phosphorus is ca. 5, 4 and 2.5 times the carbon's value, which may give unrealistic contributions larger than one for part of the pores.

To see for which types of pores the **Gm** and **He** void fractions shows the largest differences, we extended our analysis to cylindrical and spherical pores and a reticular structure. We modelled the framework with a smeared continuous distribution of carbon atoms. The details are reported in the Supporting Information. Figure 2.8 shows the comparison between **Gm** and **He** void fraction in these models. We observe for all the three pore shapes that the **Gm** void fraction is greater than the **He** void fraction for small pores, while for bigger pores the **He** void fraction becomes greater. This is due to the fact that for smaller pores the BF for helium is always less than one, because of the unfavourable interaction between the particle and the framework. For bigger pores the BF can assume values larger that one, and in such cases the **He** void fraction systematically overestimates the experimental void fraction. A similar trend for the **He** vs **Gm** curve is observed for the three types of pores in Figure 2.8, with the main difference being the value of the intersection with the bisector, which is therefore dependent on the geometry of the pore.

The **Ac-PO** volume is expected to be similar to the **Gm**, with the notable difference that it collapses to zero for small pores, i.e., for:

$$L < 2\,(r_{\text{probe}} + r_{\text{atom}}) \tag{2.8}$$

2.6.4 CoRE MOF screeening

Our model calculations shows that the differences between the **He** and the **Gm** void fractions are not negligible and can be interesting to see how these model calculations compare with the void fractions for the experimental MOF structures. A set of 5109 MOF structures was investigated from the CoRE MOF database: 4764 frameworks were modified by the authors (solvent removal, and other adjustment described in the paper)[38] and the remaining 345

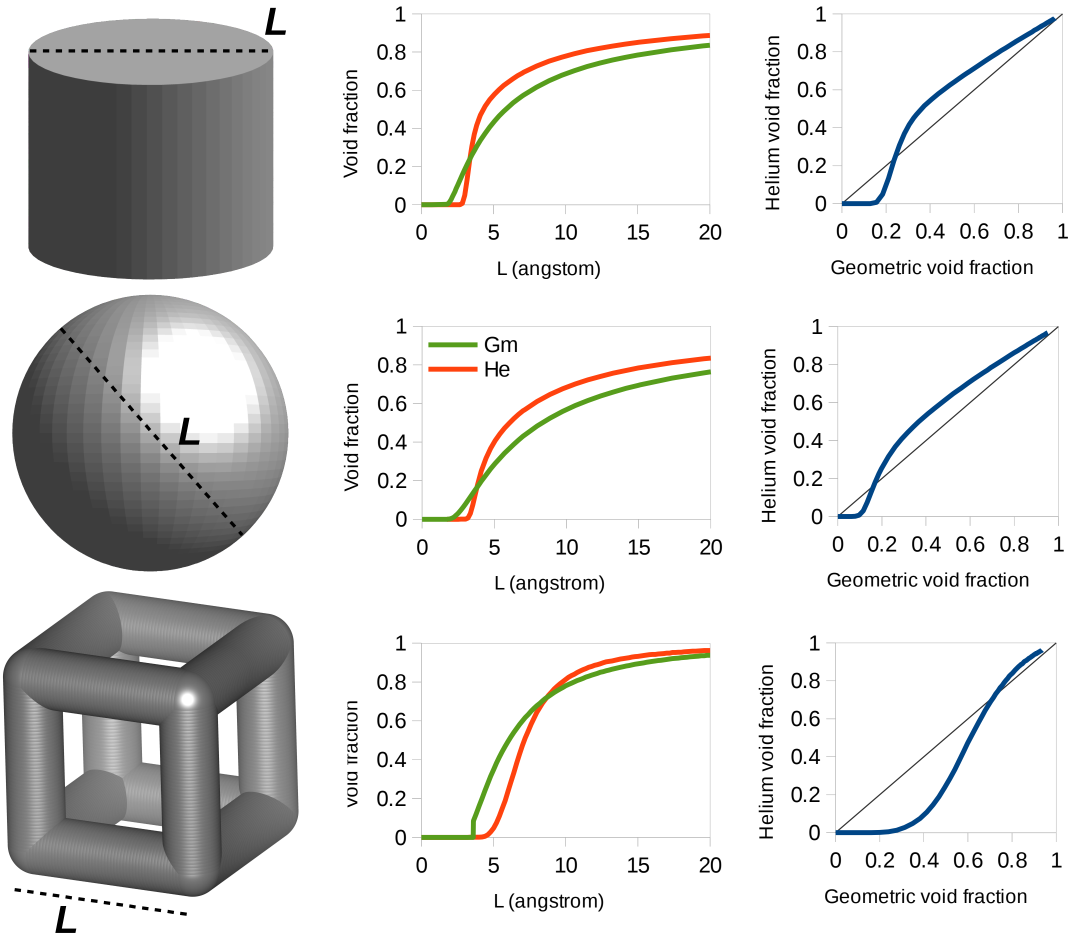


Figure 2.8 – Geometry of the pore for the different three dimensional models, comparison of the **Gm** (green) and **He** (red) void fraction versus the characteristic length *L* of the pore and direct comparisons of the two values of the void fraction for pores with different dimensions.

frameworks were downloaded directly from the Cambridge Structure Database,[37] without any further manipulation. The results of computing the **He** and **Gm** void fraction for these structures are shown in Figure 2.9.

The average trend is mostly similar to the reticular model presented in the previous section. One can notice that for many material the void fraction computed using the **He** method is higher than the **Gm** void fraction, when the **Gm** method should compute an upper bound value for the void fraction. The most extreme example for this overestimation, is the structure LOFZUB[108]: this frameworks contains aluminium and phosphorus, which have a particularly high Lennard-Jones epsilon. On the other side, a few frameworks appear to have the opposite trend, showing a moderate **Gm** void fraction but a lower **He** void (highlighted in yellow in Figure 2.9. Interestingly, all of them have a similar chemistry, i.e., the ligands of these structures are based on C≡N and C≡C bonds. These kind of ligands are particularly thin and simple, resulting in weaker dispersion forces which explains the low **He** void fraction.

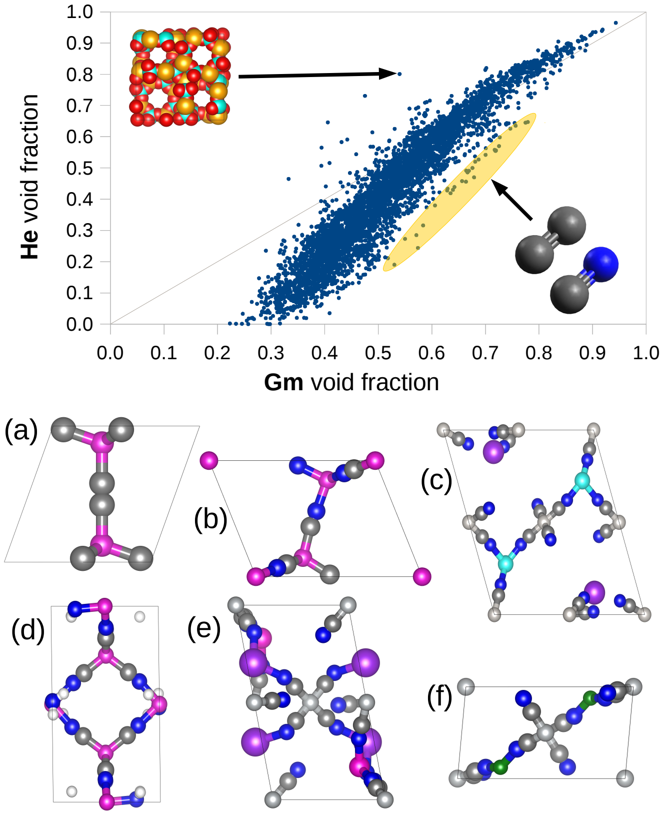


Figure 2.9 – Comparison of the **Gm** and the **He** void fraction for the materials from the CoRE MOF database.[38] Some out-trend materials are highlighted. LOFZUB has a remarkably higher value for the **He** void fraction compared to the **Gm** measurement (upper inset, Al: orange, P: cyan) and O: red) . The opposite out-trend is noticed for the structures highlighted in yellow, characterized by C≡N and C≡C based ligands (C: gray, N: blue). Some of these structures are showed: PIYZAZ (a), KECRAL10 (b), YEQRER (c), YARYEV (d), YEQRER (e), EBEMEF (f).

Using our algorithm we computed the **Ac-PO** void fraction for all the frameworks considering a probe of 1.86 Å (N_2) and using 100,000 sample points. The results are compared with the **Gm** and **He** void fractions in Figure 2.10.

As expected, the value of the **Ac-PO** void fraction is always smaller than the **Gm** void fraction. This behaviour is more pronounced for very dense materials, were the atoms of the framework create many small interstices (narrow volume) that are excluded for the calculation of the **Ac-PO** void fraction. Also, for many structures the void fraction collapse to zero, meaning that under the assumption of a rigid framework, these crystals are completely impermeable to the probing sphere. The material labelled SETPEO is a prominent example: the 0.71 geometric void fraction of this material can be decomposed to a 7% narrow volume and 64% of volume non-accessible to the nitrogen probe (the 29% remained, is the volume occupied by the atoms). For this material we can expect, if not a complete impermeability, a slow diffusion of nitrogen inside the activated crystal. Moreover, methanol is used as solvent for the synthesis and given the size of methanol, we can expect the impossibility of a complete desolvation, as effectively

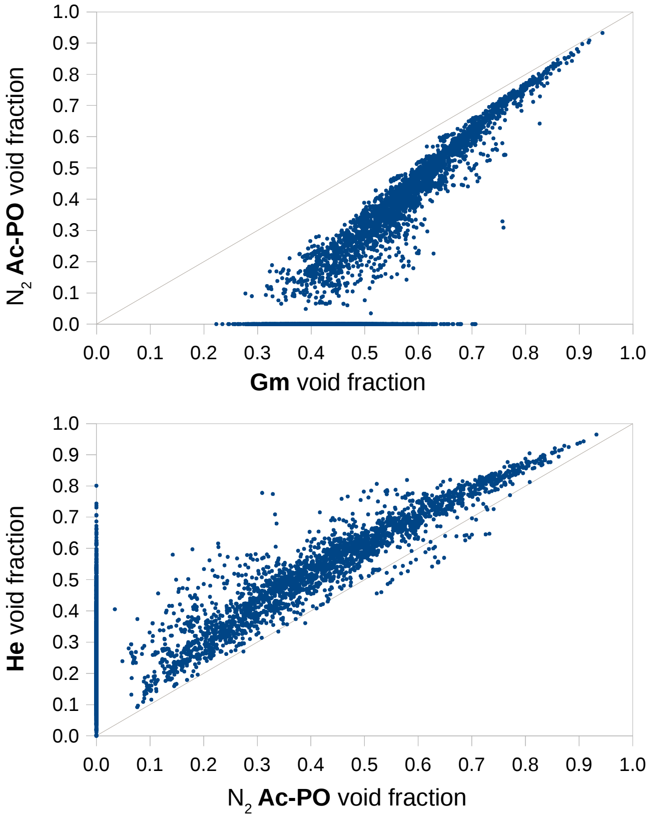


Figure 2.10 – (upper) The **Gm** void fraction is compared with **Ac-PO** void fraction for a 1.86 Å nitrogen probe. For the materials non permeable to the probe, the **Ac-PO** void fraction collapses to zero. The (lower) **He** and the **Ac-PO** void fractions are compared.

reported.[109]

If we compare the **He** and the nitrogen **Ac-PO** volume fraction, it is interesting to note the systematic overestimation of the pore volume which affects the **He** method. There are three reasons for this: the helium probe is smaller, the non-accessible volume is not excluded and, most important the possibility of the BF to be higher than one. The structures with opposite trend, where the void fraction is underestimated by the **He** method, are again the ones characterized by C≡N and C≡C ligands (shown in Figure 2.9) .

2.6.5 Comparison With Experimental Data for 10 MOFs

We studied in detail 10 different MOFs (including HKUST-1) to obtain some insights in the practical consequences of the differences in pore volume that are computed by the different methods and their agreement with experimental data[110, 111, 112, 13, 113, 114, 115, 116, 117, 118]. All the frameworks investigated have accessible channels for nitrogen and no **NAc** pore volume was detected. Figure 2.11 shows that the **PO** method leads to the best agreement among the different methods. These results emphasize that the value for the **PC** pore volume (sometimes simply defined as "pore volume") leads to a significant underestimation of the experimental pore volume. Another consideration is that for these ten structures the total **Gm** void fraction is close to the **PO** void fraction, meaning that in these samples the narrow volume is a negligible percentage of the **Gm** pore volume.

The **He** void fraction is close to the experimental value if we use Hirschfelder's Lennard-Jones parameters for helium, noticing however a systematic but relatively small overestimation. Nevertheless, the same calculation employing the He parameters from UFF is showing a much larger overestimation of the void fraction, even with non-physical values greater than one for SNU-30 and UTSA-62.

In four materials the experimental volume is more than 10% lower than the computed value (PCN-46, SNU-30, UTSA-34, and UTSA-64). We attribute this difference to some incomplete desolvation or pore shrinking after the removal of the solvent. At this point it is important to note that the computational pore volume are based on structures from the CoRE MOF database in which solvent molecules are removed computationally keeping the rest of the crystal structure unchanged.[38] In some cases this procedure is unrealistic and the most evident example is SNU-30 where the computed void fraction is eight times the measured value. The authors of its synthesis already reported a big discrepancy between the experimental and the computed surface area, which was attributed to the shrinking of the evacuated pores.

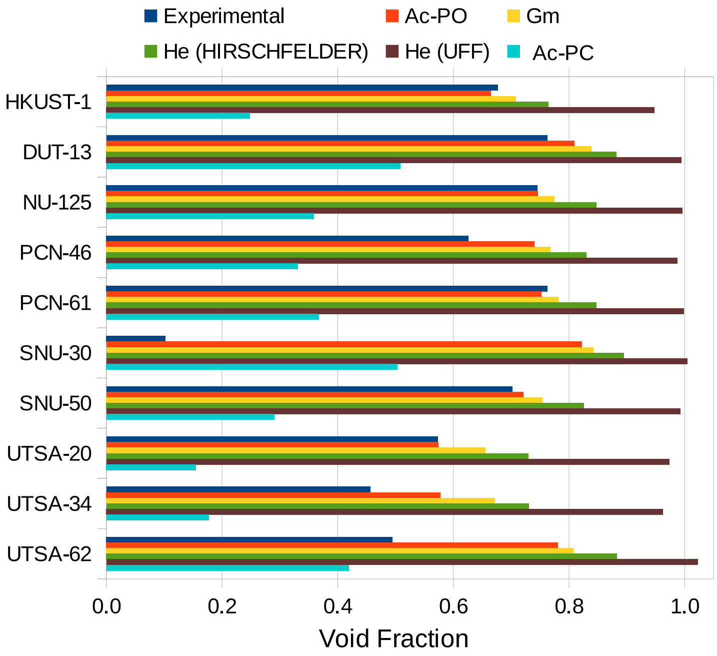


Figure 2.11 – The void fraction as computed with different methods is shown and compared with experimental data. The structures were computationally desolvated as reported in the CoRE MOF database.[38] A list with the references for the experimental values is provided in the SI.

2.7 Conclusions

In the present work we compared different methods that are used to to compute the pore volume of a crystalline microporous material from its crystal structure. We show that these methods use different definitions of the pore volume and we show that in particular for micropores these differences can be quite significant. These volumes are referred in this work using a consistent nomenclature, i.e., the geometric (**Gm**), the helium (**He**), the probe center (**PC**), and the probe occupiable (**PO**) methods. For the last two it is meaningful to further identify the volume as accessible (**Ac**) or non-accessible (**NAc**).

The main conclusion of this work is that the accessible probe occupiable (**Ac-PO**) pore volume gives the closest representation of the experimentally measured pore volumes for all types of pores. The other methods show systematic deviations. The geometric (**Gm**) calculation lead to a value for the pore volume which is an upper limit for this quantity, while the probe center (**PC**) calculation considerably underestimate the experimental value. The helium (**He**) void fraction was shown to be very dependent on the parameters and on the reference temperature assumed for the calculation.

In addition, we have presented a novel algorithm which is an extension of the freely available Zeo++ code, to fully characterize the internal volume of a crystal and extract its accessible probe occupiable (**Ac-PO**) pore volume. The algorithm is taking into account both the solvent accessibility and occupability of the internal pore cavity, and therefore its result can be meaningfully compared with the measurement of the pore volume, as obtained from the nitrogen uptake. The comparison between the experimental data and the accessible probe occupiable (**Ac-PO**) void fraction, allows to detect discrepancies due to low crystallinity, poor desolvation and pore shrinking in the real material.

3 Evaluating charge equilibration methods to generate electrostatic fields in nanoporous materials[1]

3.1 Abstract

Charge equilibration (Qeq) methods can estimate the electrostatic potential of molecules and periodic frameworks by assigning point charges to each atom, using only a small fraction of the resources needed to compute density functional (DFT)-derived charges. This makes possible, for example, the computational screening of thousands of microporous structures to assess their performance for the adsorption of polar molecules. Recently, different variants of the original Qeq scheme were proposed to improve the quality of the computed point charges. One focus of this research was to improve the gas adsorption predictions in Metal Organic Frameworks (MOFs), for which many different structures are available. In this work, we review the evolution of the method from the original Qeq scheme, understanding the role of the different modifications on the final output. We evaluated the result of combining different protocols and set of parameters, by comparing the Qeq charges with high quality DFT-derived DDEC charges for 2338 MOF structures. We focused on the systematic errors that are attributable to specific atom types to quantify the final precision that one can expect from Qeq methods in the context of gas adsorption where the electrostatic potential plays a significant role, namely CO_2 and H_2S adsorption. In conclusion, both the type of algorithm and the input parameters have a large impact on the resulting charges and we draw some guidelines to help the user to choose the proper combination of the two for obtaining a meaningful set of charges. We show that, considering this set of MOFs, the improvements in the newer Qeq variants are not evident in general, but only in the presence of certain atom types, such as alkali metals.

---

[1] Preprint version of the article:: Daniele Ongari, Peter G. Boyd, Ozge Kadioglu, Amber K. Mace, Seda Keskin and Berend Smit, *Journal of chemical theory and computation*, 15(1), 382-401, 2018, https://doi.org/10.1021/acs.jctc.8b00669. D.O. did the literature review about the different Qeq variants and performed all the calculations except for the comparison of the electrostatic potentials as computed from a grid.

3.2 Introduction

Nanoporous crystals have attracted the interest of the scientific community for a variety of applications, ranging from catalysis[119, 120] to gas separation[121] and storage.[122, 123] Widely investigated classes of these materials are, for example, zeolites and, more recently, Metal Organic Frameworks (MOFs), for which large number of new structures are reported every year.[124] In addition to these experimental structures, hypothetical frameworks are generated in-silico,[125, 126] to be screened and possibly suggested as interesting materials on which synthetic chemists should focus their attention. Therefore, simulations (e.g., molecular dynamics or Monte Carlo) represent an important tool to efficiently assess the performance of thousands of different materials for a range of specific applications.

There are three main inputs that one must provide to set up the molecular simulation: the geometry of the system, a set of parameters for dispersion (Van der Waals) interactions, and a set of partial charges. For the dispersive interaction, we can rely on a generic force field such as UFF[127] or Dreiding[128] to get a reasonable estimate of the interactions. However, for all but the simplest adsorbates (e.g., noble gasses or methane) the dispersive interactions need to be supplemented with the Coulombic (e.g., ionic, dipololar, quadrupolar, ...) interactions. The most reliable approach is to obtain these charges from a converged electronic structure calculation, which gives the ab initio electronic density, e.g., using wave function (WF) or density functional theory (DFT) methods, and the electrostatic potential. From these, one can derive different sets of point partial charges. The most popular schemes quantify the partial charges of the atom from a partition of the electron density (e.g., Mulliken,[129] Hirshfeld,[130] iterative Hirshfeld[131] Bader,[132]) or by fitting the electrostatic potential around the atoms (e.g., RESP,[133] CHELPG,[134] REPEAT[135]), or from both. An example of the last is the DDEC method, where the two types of information are included in the charge optimization functional.[136, 137, 138]

Obtaining the point charges from quantum calculations requires a relevant amount of computational time to perform the electronic structure calculation first and the post-process fitting later. For porous materials, with hundreds of atoms per unit cell, this typically requires hours on multiple compute cores. Such a calculation becomes very expensive, if not prohibitive, when investigating several thousands of different structures. This motivated the development of approximated methods that can empirically compute partial charges much faster, i.e., in less than a minute running on a single CPU. Most of these methods are grouped under the name of "charge equilibration methods (Qeq)", with the most popular algorithm being proposed by Rappé and Goddard in 1991.[139] The major differences between the different variants of this scheme can be summarized in four main concepts, that will be discussed in detail in the next section:

- The choice of the atomic parameters

- The center and the order of the Taylor expansion of the energy as a function of the charges.

- The analytic form to compute the pairwise interaction between the atoms with respect to the geometry of the system.
- The inclusion of further parameters to characterize each bond type.

Notice that, when a new Qeq variants is proposed, the authors usually suggest modifications that belong to more than one category.

For the screening of libraries of MOFs, the Qeq methods attracted a lot of interest due to their ability to quickly compute partial charges for thousands of materials and to screen their performances for gas uptake and separation.[140, 126] However, at present there is little known about the accuracy and transferability of the different methods for a large range of diverse MOFs. In this work we make a detailed comparison of the different methods. In particular, we investigate the influence of the chosen set of parameters on the resulting partial charges and the subsequent adsorption calculations. For this purpose, the uptake of CO_2 is computed and compared for 2338 MOFs, containing a total of 59 atomic elements. These structure are included in the Computational Ready Experimental (CoRE) MOF database,[141] and their partial charge was computed from the DFT electron density, following the DDEC scheme.[136, 142] Structures containing lanthanides were excluded due to the difficulty of computing or obtaining experimental parameters needed for certain Qeq variants. In this work, the heat of adsorption at zero loading is compared for CO_2 and H_2S as test cases. What we report here is the largest comparison of Qeq methods available in literature, with the aim of recognizing symptomatic problems and assess the accuracy that one can expect when using these methods for computing adsorption properties.

3.3 Theoretical Aspects

3.3.1 Charge equilibration (Qeq) and periodic charge equilibration (PQeq) methods

Before we discuss the results of our calculations we review the different charge equilibration methods. All the variants that will be mentioned are summarized in Table 3.1.

The charge equilibration (Qeq) method[139] allows to compute the partial charges for the atoms in a molecule, by using as input its geometry and three important properties related to the isolated atoms. The first is the ionization potential, i.e., the energy needed to remove the outer valence electron, the second is the electron affinity, i.e., the energy difference related to the injection of an extra electron, the last is the atomic radius. These quantities can be obtained from experimental measurements and/or computed ab initio. The Qeq method is based on Sanderson's concept of electronegativity equalization, postulating that two or more atoms combining within a molecule gets their electronegativity equalized.[152, 153] Therefore,

Table 3.1 – Summary of all the Qeq methods described in this paper.

Method	Full name	Description	Parameters	Ref.
EEM	Electronegativity Equalization method	Fitted $\Delta\chi_A$ and ΔJ_{AA}	fitted	[143]
Qeq	Charge equilibration	nS Slater-type overlap, corrections for H	GMP	[139]
PQeq	Periodic Qeq	Extension to periodic systems (Ewald summation)	GMP	[144]
SCQeq	Self Consistent Qeq	Fourth order Taylor expansion	fitted	[145]
EQeq	Extended Qeq	Charge center for the Taylor exp. selectable in the input	exp'tal	[146]
FC-Qeq	Formal Charge Qeq	Taylor exp. centered in the formal charge	CCSD(T)	[147]
I-Qeq	Ionizing Qeq	Taylor exp. centered in the partial charge, iterative	CCSD(T)	[147]
MEPO-Qeq	MOF Electrost. Potential Optimized Qeq	Qeq parameters fitted to reproduce charges in MOFs	fitted	[148]
EQeq+C	EQeq corrected	Extra parameters are added to EQeq and fitted	fitted	[149]
SQE	Split Charge Equilibration	Split charge formalism	fitted	[150]
SQE-MEPO	Split Charge Equilibration MEPO	SQE parameters fitted to reproduce charges in MOFs	fitted	[151]

if we assume that the atomic ionization potential and electron affinity of an isolated atoms are similar to the ones of the same atom type bonded inside a molecule or a crystal, we can derive its partial charge.

To understand the inner working of this method, we start by expanding the energy of an isolated atom $\widetilde{A}$, as a second order Taylor expansion related to its charge $Q_{\widetilde{A}}$ and centered on its neutral reference point[154]:

$$E_{\widetilde{A}}(Q_{\widetilde{A}}) = E_{\widetilde{A}}(0) + Q_{\widetilde{A}}\left(\frac{\partial E_{\widetilde{A}}}{\partial Q_{\widetilde{A}}}\right)_{Q=0} + \frac{1}{2}Q_{\widetilde{A}}^2\left(\frac{\partial^2 E_{\widetilde{A}}}{\partial Q_{\widetilde{A}}^2}\right)_{Q=0} \tag{3.1}$$

By definition, the energies related to the removal and the addition of one electron starting from the neutral state of the atom, are given by the ionization potential (IP_0) and electron affinity (EA_0):

$$\mathrm{IP}_0 = E_{\widetilde{A}}(+1) - E_{\widetilde{A}}(0) \tag{3.2}$$

$$\mathrm{EA}_0 = E_{\widetilde{A}}(0) - E_{\widetilde{A}}(-1) \tag{3.3}$$

Substitution in equation 3.1 gives:

$$\left(\frac{\partial E}{\partial Q}\right)_{\widetilde{A}} = \frac{1}{2}(\mathrm{IP}_0 + \mathrm{EA}_0) = \chi_A^0 \tag{3.4}$$

$$\left(\frac{\partial^2 E}{\partial Q^2}\right)_{\widetilde{A}} = \mathrm{IP}_0 - \mathrm{EA}_0 = J_{AA}^0 \tag{3.5}$$

χ_A^0, is commonly defined as *electronegativity*. The difference between IP_0 and EA_0, named J_{AA}^0, is identified in the first approximation as the electron repulsion in the outer atomic orbital, and referred as *idempotential* (or *self-Coulomb interaction*). J^0_{AA} is also known as *atomic hardness.*[155] The superscript ^{0} for χ_A and J_{AA}, and the subscript $_0$ for EA and IP, indicates that the reference state is the neutral ($Q = 0$) atom. Equation 3.1 can now be rewritten as:

$$E_{\tilde{A}}(Q_{\tilde{A}}) = E_{\tilde{A}}(0) + \chi^0_A Q_{\tilde{A}} + \frac{1}{2} J^0_{AA} Q^2_{\tilde{A}} \tag{3.6}$$

A similar expression for molecules and periodic crystals can be obtained by keeping the atomic values for χ^0_A and J^0_{AA} and by adding an additional term to account the pairwise interaction between the atoms:

$$E(Q_1, Q_2, ..., Q_N) = \sum_{A=1}^{N} \left(E_{\tilde{A}} + \chi^0_A Q_A + \frac{1}{2} J^0_{AA} Q^2_A + \sum_{B>A}^{N} J_{AB} Q_A Q_B \right) \tag{3.7}$$

where A and B are two atoms in the molecule, and J_{AB} is the function that describes their pairwise interaction. The charge equilibration scheme assumes that the charge distribution is such that the electric energy given by equation 3.6 is minimized with respect to the charge distribution (Q_1, ... Q_N). We define the partial derivatives of the energy with respect to the charge Q_A, as χ_A:

$$\chi_A(Q_1, Q_2, ..., Q_N) = \frac{\partial E}{\partial Q_A} = \chi^0_A + J^0_{AA} Q_A + \sum_{B \neq A} J_{AB} Q_B \tag{3.8}$$

The minimum energy is found if:

$$\chi_1 = \chi_2 = ... = \chi_N \tag{3.9}$$

which, together with the constrain on the total charge

$$Q_{tot} = \sum_{A=1}^{N} Q_A \tag{3.10}$$

gives a system of N equations that one need to solve to obtain the molecular partial charges Q_i. This minimization resembles a typical thermodynamic equilibrium condition, hence the partial derivatives in equation 3.8 are often referred to as *atomic-scale chemical potential*, and the entire scheme as *charge equilibration.*

In the original Qeq method, the charge of the atoms was allowed to vary within the possible occupations of the valence shell of electron. As for the analytic form of J_{AB}, the Coulombic potential

$$J_{AB}(R) = \frac{1}{4\pi\epsilon R} \tag{3.11}$$

is assumed only for a large distance R between two atoms. ϵ is the relative dielectric constant, considered unitary (as in vacuum) in Qeq. Equation 3.11 can give unrealistically large values for J_{AB}, as the $1/R$ term explodes when two atoms are close, as in the case when they share a covalent bond. As a consequence, this term will dominate over the others in equation 3.8 and the minimum energy, if still existing, will be found for partial charges with very high absolute values. This problem is known as *infinite charge separation*[146] (or *polarization catastrophe*)[156] and gives non-physical high value partial charges. To ensure that $J_{AB}(R)$ converges for small values of R a shielding is needed, which physically arises from the overlap of the electron densities and can be computed ab initio from the Coulomb integral between the Slater-type densities of neighbouring atoms. Therefore, Qeq still considers the bonded atoms like isolated atoms that are pulled close. To simplify this calculation, Rappé assumed the electron densities to be spherically symmetric for all the atoms, i.e., as normalized nS Slater densities in the form[157]

$$\rho_{n\xi}^{Slater}(r) = N_n r^{n-1} e^{-\xi r} \tag{3.12}$$

where N_n is the normalization constant, n the valence shell and the exponent ξ is computed from the characteristic size of each atom, as:

$$\xi_A = \lambda(2n+1)/(2r_{\widetilde{A}}) \tag{3.13}$$

Here, the crystal covalent radius $r_{\widetilde{A}}$ is a specific property of the atom and the scaling factor λ was estimated as $\lambda = 0.5$ for the whole periodic table.[139] The Coulomb integral, for short distances R, is therefore computed as:

$$J_{AB}(R) = \int\int \rho_A(r_A)\frac{1}{R}\rho_B(r_B)dV_A dV_B \tag{3.14}$$

Note that the damping of the $J_{AB}(R)$ term for low distance was the main novelty of the Qeq method over earlier similar schemes, grouped under the name of *Electronegativity Equalization Method* (EEM).[158, 143, 159, 160] In this scheme the atomic-scale chemical potential reads as:

$$\chi_A^{EEM} = (\chi_A^0 + \Delta\chi_A) + (J_{AA}^0 + \Delta J_{AA})Q_A + \sum_{B\neq A}\frac{Q_B}{R} \tag{3.15}$$

where the two extra parameters $\Delta\chi_A$ and ΔJ_{AA} needs to be fitted for a training set of molecules to match the partial charges from ab initio calculations. This is another difference with the Qeq method, where the input parameters comes only from the properties of isolated atoms (except for H, as we will see in the next paragraph), and therefore no training is required.

In its original implementation, the EA_0 experimental values for hydrogen was found to lead to non-physical partial charges. It is not surprising, because the addition of one electron to the hydrogen atom gives a free H^- ion, which is more stable than a negatively charged H inside a molecule. Therefore, the atomic derived EA_0 is overestimated if used for Qeq. To fix this problem the authors proposed a charge dependent idempotential and Slater density exponent for hydrogen:

$$\xi_H(Q_H) = \xi_H^0 + Q_H \tag{3.16}$$

$$J_{HH}(Q_H) = \left(1 + \frac{Q_H}{\xi_H^0}\right) J_{HH}^0 \tag{3.17}$$

where ξ_H^0 is computed from the standard procedure of equation 3.13. χ_H^0 and the J_{HH}^0 parameters were fitted to reproduce the experimental partial charge of five small molecules (HF, H_2O, NH_3, CH_4 and LiH): the experimental value of χ_H^0=7.17 eV was reduced to 4.53 eV. With this correction for hydrogen atom, the calculation of Q_H becomes iterative, starting from the initial guess of null partial charge.

The charge equilibration method was successively extended for periodic systems (PQeq) by using the Ewald summation[161] to ensure the convergence of the Coulomb term in an infinite periodic system.[144] The solution of the system of linear equations need an outer level of iterations, where an initial set of charges is assumed and updated at each step until reaching the convergence.

3.3.2 Modifications to the Qeq method

Several modifications of the Qeq method were proposed over the years, to fix specific problems or to improve the physical description of the system. Oda and Hirono, for example, claimed that the two-center Coulombic integrals of nS Slater-type densities, used by Rappé, gives imprecise values of the energy, for small interatomic distance.[162] Hence, they tested five different empirical formulation for the $J_{AB}(R)$ term and they verify that, among these, the DasGupta-Huzinaga approximation[163],

$$J_{AB}(R) = \frac{1}{R + \frac{1}{\frac{J_{AA}}{2} e^{kR} + \frac{J_{BB}}{2} e^{kR}}} \tag{3.18}$$

with the Klondike parameter, k, chosen equal to 0.4 for all the atoms, gives the best agreement with the ESP-fitted charges computed with HF/6-31G**.

Another interesting modification is the inclusion of the third and the fourth order terms of the Taylor expansion in equation 3.6.[164] The motivation was given by the non-physical charges computed with Qeq for the Ag_5Li_5 cluster, where a partial charge bigger than +3 was obtained for Li. Since the set of equations becomes nonlinear, the solution needs to be achieved through an iterative procedure. The method was therefore named *Self-Consistent charge equilibration* (SCQeq) and it has two nested iterative loops: one for the hydrogens and one for solving the system of equations. The coefficient for the first two terms, i.e., χ_A^0 and J_{AA}^0, as well as the coefficients for the third and fourth term were computed by fitting ab initio results. The parameters were obtained for 6 metals (Li, Na, K, Cu, Ag and Au) and successively, Oda and Takahashi[145] extended the same approach to organic molecules.

In their Extended charge equilibration (EQeq) method, Wilmer et al.[165, 146] suggested that using a different "charge center" than the neutral one, avoids extrapolations in the Taylor expansion, while still considering the truncation beyond second order. They suggested, as an educated guess, to use the formal oxidation state as the charge center for coordinated metals (e.g., in a metallic complex or in MOFs) or metals forming ionic bonds. Therefore, we can generally define the ionization potential and the electron affinity of the ion, rewriting equation 3.2:

$$\mathrm{IP}_n = E_{\tilde{A}}(n+1) - E_{\tilde{A}}(n) \tag{3.19}$$

$$\mathrm{EA}_n = E_{\tilde{A}}(n) - E_{\tilde{A}}(n-1) = \mathrm{IP}_{n+1} \tag{3.20}$$

From these two the electronegativities and idempotential are directly computed from equation 3.4 and 3.5. Figure 3.1 show for three atom types, the change in the potential due to the choice of different charge centers.

This protocol implies that the same atom type can be treated differently depending on the choice of the charge center. Taking copper as an example, if one considers the oxidation number as the charge center, EQeq will treat this atom type differently if Cu(I) or Cu(II). Therefore, the choice of the charge center is an extra input that the user has to provide. The assumption that Wilmer et al. made in presenting their EQeq method applied to MOFs, was to change the center of the Taylor expansion to the formal oxidation number for metallic cations only: atoms such as nitrogen and oxigen, which typically have a negative formal charge, are still treated with an expansion centered in their neutral state. These assumptions are unavoidable given the lack of experimental data for -2, and lower negative ionization energies, due to the practical impossibility of injecting more than one electron in an isolated atom.[167]

For the development of EQeq, Wilmer and coauthors compared different analytic forms of $J_{AB}(R)$ and chose the one that leads to the best agreement with CHELPG charges, i.e.,

$$J_{AB}(R) = \frac{K}{R} - e^{-(J_{mix}R/K)^2}\left(\frac{K}{R} - J_{mix} + \frac{J_{mix}^2}{K}R\right) \tag{3.21}$$

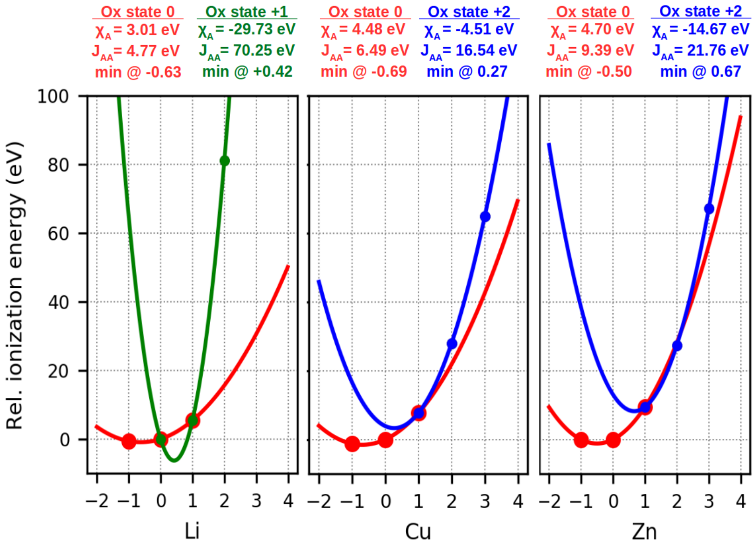


Figure 3.1 – Comparison of the potential $\chi^n_A Q_{\tilde{A}} + \frac{1}{2} J^n_{AA} Q^2_{\tilde{A}}$ when the charge center n is chosen to be zero (as in Qeq, red lines) or equal to the common oxidation number (green and blue lines). The experimental relative energies for the ions (points) are shown for each formal charge (x axis) and the solid line represent the potential centered in different charge center.[166, 167] Neutral (red) and +1 (green) expansions are shown for lithium. Neutral (red) and +2 (blue) expansions are shown for Cu and Zn.

where $K = e^2/4\pi\epsilon$, with e being the charge of the electron, and $J_{mix} = \sqrt{J^0_{AA} J^0_{BB}}$.[165] The need for reiteration was removed for the hydrogens, by assigning an effective electron affinity of $\mathrm{EA}^H_{eff} = +2$ eV (the measured value is -0.754 eV). This correspond to fit the $E(Q)$ potential for hydrogen with a quadratic expression instead of cubic one (as in Qeq) and causes instability issues that were solved by increasing the relative dielectric constant ϵ to ϵ_{eff}=1.67. Wilmer et al. also warned that higher values for ϵ may be required to model high-density system such as alloys and non-porous solids. In summary, two ad-hoc parameters, EA^H_{eff} and ϵ_{eff} are introduced in the EQeq protocol. The final formulation of the EQeq energy of the system as a function of the partial charge, to be compared with equation 3.7, is:

$$E_Q(Q_1, Q_2, ..., Q_N) = \sum_{A=1}^{N} \left(E_{\tilde{A}^{cc}}(Q^{cc}_A) + \chi^{cc}_A (Q_A - Q^{cc}_A) + \frac{1}{2} J^{cc}_{AA} (Q_A - Q^{cc}_A)^2 + E^{coul}_A + E^{damp}_A \right) \tag{3.22}$$

In this equation Q^{cc}_A is the input charge center for atom A (an integer number), E^{coul}_A is the sum of the Coulombic interactions and E^{damp}_A is equal to the sum of all the $J_{AB} Q_A Q_B$ terms, with J_{AB} from eq. (3.21). In the way EQeq treats the hydrogen atoms and the Ewald summation, the system of equation can be solved directly without the need of reiterations.

A further modification of the Qeq method, in the direction of shifting the charge centers for the Taylor expansion, was proposed later by Wells.[147] Both positive (oxidated) and negative (reduced) ionization states of the atoms are considered as charge centers. To allow fractional formal charges (e.g., +0.5 charges on N in a ZnN_4 coordination) the ionized value of the atomic electronegativity χ_A^{iz} is expressed as a linear interpolation:

$$\chi_A^{iz} = \left(\frac{dE_A}{dQ_A}\right)_{iz} = [1-(Q_A^{iz}-\gamma)]\left(\frac{dE_A}{dQ_A}\right)_{\gamma} + (Q_A^{iz}-\gamma)\left(\frac{dE_A}{dQ_A}\right)_{\gamma+1} \tag{3.23}$$

where Q_A^{iz} is the formal charge of the atom and γ corresponds to the value of Q_A^{iz} rounded down to the nearest integer. As an example, a +1.25 oxidated atomic electronegativity $\chi_A^{iz=+1.25}$ will be obtained by 75% of $\chi_A^{iz=+1}$ plus 25% of $\chi_A^{iz=+2}$. The algorithm still requires the formal charges for the atoms (their total sum has to be null) as input, which are assigned in a preceding routine of the code, by considering the connectivity in the framework. This method was named *Formal Charge equilibration method* (FC-Qeq). A variant was also proposed, where the input formal charges is not required but obtained in a self-consistent fashion from the computed partial charges. This took the name of *Ionizing charge equilibration method* (I-Qeq) and it is more computationally expensive, due to the inner iterative process: for each atom, χ_A^{iz} and J_{AA}^{iz} are updated from Q_A at every step. To remedy the lack of experimental data, the ionization energies were computed ab initio, using the coupled-cluster CCSD(T) method for all the elements for which the aug-cc-pvqz basis set is available (H-Ar, Sc-Kr)[168, 169, 170]. As for hydrogen, the ionization energies were computed for the H_2 molecule instead of the isolated H atom, giving a value of 2.62eV for the EA, in close agreement with the effective value of 2 eV adopted by Wilmer in his screening study.[126] The DasGupta–Huzinaga approximation[163] (equation 3.18) was used to compute the $J_{AB}(R)$ term, and the Ewald summation was used to compute the long range Coulombic interaction. The method proposed by Wells is apparently more rigorous from a mathematical standpoint but as this method relies on single reference coupled cluster calculations using standard basis set, it is questionable whether such a method provides reliable values for the ionization energies of isolated atoms (see the section "Ionization energies and radii").

3.3.3 Split charge equilibration methods (SQE)

All the methods described so far are based on the hypothesis that the intrinsic properties of the atoms, to lose and gain electronic charge, are transferable from isolated atoms to molecular systems and crystals. Other methods introduce further parameters that are specific for each bond type, to better characterize the atoms in molecules from their connectivity. This class of methods is called *Split-Charge Equilibration* (SQE) and was generalized in 2006 by Nistor et al.[150] based on previous models.[171, 172] Here, the energy of the system is not function of the atomic charges Q_A itself, but it is a function of the charge flown between two connected atoms A and C, defined as q_{AC}. Therefore, Q_A and q_{AC} are related by

$$Q_A = \sum_C q_{AC} \tag{3.24}$$

where the summation is over all the atoms connected to A. In neutral systems, the split charges has to satisfy the antisymmetry condition, $q_{AC} = -q_{CA}$. We can now plug equation 3.24 into the expression for $E_Q(Q)$ (equation 3.7) and compute the cross terms. Simplifying these cross terms, one can derive:[150, 151]

$$E_q(\mathbf{q}) = \frac{1}{2}\sum_A \sum_C \left(\chi_{AC}^{bond} q_{AC} + \frac{1}{2} J_{AC}^{bond} q_{AC}^2 \right) + E_Q(\mathbf{Q}) \tag{3.25}$$

where $E_Q(\mathbf{Q})$ is from Qeq, equation 3.7. From this reformulation, it is clear that the SQE introduces for each specific bond two extra parameters, χ_{AC}^{bond} and J_{AC}^{bond}, that needs to be fitted. It is important to note that the SQE method is based on a more accurate physical description of the charges distribution in a molecular system. With respect to Qeq, the SQE model therefore allows for a better modelling of the dielectric properties and polarizability in the system.[150, 173, 174, 175]

3.4 Applications to MOFs

3.4.1 Periodic Qeq

In 2012, Sholl and co-workers[140] were the first to screen a large amount (~500) of MOFs using PQeq, looking for materials with high CO_2/N_2 selectivity. The atomic parameters for χ_A^0 and J_{AA}^0 were obtained using the Generalized Mulliken-Pauling (GMP) method,[127] i.e., the same scheme as originally used by Rappé.

The charges computed with PQeq were validated by comparing the Henry's constants for CO_2 adsorption in four MOFs (IRMOF-1, ZIF-8, ZIF-90, and Zn(nicotinate)$_2$) where, as a reference, the electrostatic potentials inside the frameworks were computed directly from DFT. Also, the results were compared with non-charged systems, where the CO_2-framework interactions are computed with the same Lennard-Jones potential but charges are set to zero. The benchmark for just four materials seems quite limited, and in one case (Zn(nicotinate)$_2$) the Henry's coefficient computed without charges charges gets even closer to the result obtained from DFT-derived electrostatic potential than when using PQeq charges. Despite this, PQeq could provide the same ranking as the DFT-derived simulations, for this four test structures. The PQeq calculation successfully converged for 489 of the 500 structures, CO_2/N_2 selectivity at infinity dilution was computed from the Henry coefficients of the two gas molecules. For further analysis they selected, from the group of materials with a selectivity larger than 100, six MOFs with a large difference between the PQeq and non-charged results and other five MOFs that experimentally were proven to be stable after activation (i.e., solvent removal). For these eleven structures the selectivity was compared with the one obtained with DDEC (electron density derived) charges system. In all cases, the PQeq was shown to do at least better than the non-charged model when compared to the selectivity obtained by using DDEC charges, but still none of the two methods, PQeq and non-charged system, gave the same ranking as DDEC for the selectivities of these 11 materials.

3.4.2 Extended Qeq

The EQeq algorithm by Wilmer et al., was specifically designed to improve the description of the charge on the metallic nodes of MOFs even if its use can be extended to other molecules and materials, just requiring a reasonable assumption for the charge center. The new method was validated for 12 common MOFs.[146] The discrepancy between the charges computed with DFT-based methods and the charges computed with EQeq method was shown to be significantly less than Qeq in five out of twelve cases, and comparable for the other cases. As a further test, the CO_2 adsorption (gravimetric uptake at 198 K and 0.1 bar) that was computed with the different charges was compared to the experimental values for these 12 MOFs. To do this, the Spearman's correlation coefficient (see SI) was used to estimate the ability of the different protocols to rank the materials according to their CO_2 uptake. Keeping the experimental ranking as a reference, the authors concluded that EQeq can provide reliable charges despite the low computational cost and the simplicity of the implementation: the Spearman's coefficient obtained was 0.727, while the calculations with Qeq charged MOFs led to a correlation of only 0.35.

The predictive power of EQeq was successively utilized by Wilmer et al.[126] to assign partial atomic charges for a set of more than 137,000 hypothetical MOFs, and these charges were used in a subsequent screening study for CO_2 and N_2 adsorption. The value for the effective dielectric constant ϵ_{eff} was increased from 1.67 to 2.0 for all the MOFs. This increase in the dielectric constants weakens the Coulombic interaction and more structures could converge to physical partial charges, but it also leads to artificially lower partial charges. The amount of data collected in a consistent way for such a large set of frameworks, allowed the authors to draw some important considerations on the relation between the structure (e.g., pore volume, surface area, channel, and pore diameters) and the performance of these MOFs for CO_2 capture. An interesting conclusion regarding the contribution of the partial charges, was that MOFs with F and Cl functional groups were identified as potentially well performing for this application due to their polar nature.

More recently, Li et al.[176] screened 2932 MOFs from the CoRE database[142] for CO_2 capture under humid conditions, comparing the results for the CO_2 selectivity over H_2O, using the DDEC or the EQeq charges to model the Coulombic interactions with the adsorbates. They found that, from the 15 materials with the highest CO_2 selectivity with EQeq modelling, only 8 of them are confirmed to be selective when using the more accurate DDEC charges. The remaining seven MOFs are therefore false positives. Also, they highlighted seven additional structures that show a high CO_2 selectivity with DDEC charges (comparable to the top 15 found) where the calculations with EQeq charges underestimates the selectivity. It would be interesting, however, to rationalize these differences in the context of the present study. Unfortunately, Li et al. have not fully documented all the EQeq parameters for us to reproduce their results. For example, we miss information on the ionization potentials, the value used for the effective dielectric constant, and the charge centers used for the Taylor expansion.

3.4.3 Formal Charge and Ionizing Qeq

The FC-Qeq and I-Qeq variants, proposed by Wells, were tested for 24 MOFs.[147] The DFT-derived electrostatic potential was compared with the one computed from Qeq, EQeq, FC-Qeq and I-Qeq charges. Based on the relative root mean square error, EQeq and FC-Qeq are shown to perform significantly better than PQeq, but I-Qeq was found to be the best performing method among the four. This is particularly encouraging since the I-Qeq method, without the need of input formal partial charges, can be effectively used for obtaining charges for a large number of different MOFs. Little is reported on convergence problems for these methods. Moreover, to extend the use of the I-Qeq method for all the MOFs, the ionization parameters need to be computed also for heavier metals, for which aug-cc-pvqz basis sets are not available.

3.4.4 MOF electrostatic potential optimized Qeq and EQeq+C

A step further in the direction of modelling charges in MOFs was made by the group of Tom Woo. Using the original version of the PQeq algorithm, as implemented in the GULP package,[177] they fitted the atomic values χ_A^0 and J_{AA}^0 over a training set, to reproduce the DFT/PBE[178] electrostatic potential inside the framework. For this procedure, named as *MOF electrostatic-potential-optimized charge* (MEPO-Qeq), a training set of 543 hypothetical MOFs was employed and the new parameters were validated on a second set of 693 hypothetical MOFs. These MOFs were built in-silico, by combining 52 different ligands, 4 common metallic nodes (Zn_4O, Zn_2-paddlewheel, Cu_2-paddlewheel and V_2O_2) and modifying the ligand to include 17 different functional groups. In MEPO-Qeq, the parameters for a total of 10 atom types were fitted, while the parameters for hydrogen were kept the same as in Qeq: the large number of hydrogens on the internal surfaces of MOFs, would lead to instabilities in the fitting procedure. To test the new method, the uptake and the heat of adsorption for CO_2 was compared between PQeq (with GMP parameters), EQeq, MEPO-Qeq, and non-charged systems. The reference is the DFT-derived REPEAT charged system. Considering the validation set, the authors showed that MEPO-Qeq give a better agreement than Qeq and EQeq. In addition, for that set of frameworks, these two methods lead to worse agreement to the REPEAT calculation than the simulations without charges. The authors insist on the fact that most of the materials where Qeq and EQeq is significantly overestimating the value of the partial charges (and consequently the CO_2 uptake) contain F and Cl functional groups. This is an important point as exactly for MOFs with F and Cl functional groups Wilmer and Snurr observed exceptionally high selectivities.[126] Qeq and EQeq method assign the same null charge center to Cl and F, potentially leading to similar partial charges on these atom types. According to the authors, the conclusion that Wilmer et al. draw on the performance of MOFs with these functional groups seems to result from an artefact in the EQeq calculation, and in our work we aim for a deeper investigation on this issue.

The MEPO-Qeq method has the strong limitation of being transferable only to MOFs with similar structures. To give two examples, the MEPO-Qeq parameters are not able to compute partial charges that correctly describe the electrostatics in the materials, in the cases of zeolitic imidazolate frameworks (ZIFs), which are based of a different Zn-based secondary building unit, and MIL-100, having vanadium open metal sites. This is an important warning to avoid meaningless extrapolation for MOFs with a very different topology with respect to the training set used. In this case a new fitting should be performed.

Qiao et al.[179] used MEPO-Qeq to obtain the charge of ~5,000 MOFs from the CoRE MOF database,[141] to investigate CO_2/N_2 and CO_2/CH_4 separation. The reliability of transferring the fitted parameter to different topologies that were not included in the MEPO training set is therefore questionable and should be further investigated.

A similar procedure was published by J. Schrier and co-authors.[149] In their EQeq+C method they introduced a correction to the EQeq scheme inspired by the Charge Model 5 (CM5) model.[180] Instead of tuning the IP_n and the EA_n parameters directly, a new parameter for each atom type was introduced. The new method was applied to 17 amine-templated metal oxides and to the 12 MOFs Wilmer already tested in his EQeq paper.[146] When using these 12 MOFs for both the training and the validation, the authors could achieve a significant improvement in the correlation with the REPEAT charges. While they could lower the mean absolute deviation by a 34-68% for the most of the frameworks, the mean absolute deviation for ZIF-8 increased by 54%. They suggest that a better fit for this material could be achieved if more ZIFs were included in the training set.

The effectiveness of these methods based on fitting the input parameters is shown to be very dependent to the similarity between the training and the test sets. In his work Verstraelen[174] analysed the limits of these approaches involving parameter's calibration, suggesting some useful guidelines. However, in the case of a very diverse set of materials (like the CoRE MOF database that we want to consider) the calibration became less effective and one has to rely on the parameters measured, or computed, for the isolated atoms.

3.4.5 Split Charge Equilibration MEPO

Woo et al. re-parametrized the coefficients for the SQE method[151] analogously to MEPO-Qeq. More than a thousand of frameworks (MOFs and porous polymer networks, PPNs) were split into a train and a validation set. Compared to MEPO-Qeq, many more parameters need to be considered: for SQE-MEPO, 91 parameters were fitted considering 17 different atom types and the bonds present in structures of the adopted training set, while, for the same set of structures, only 34 parameters would be sufficient for a Qeq method.

The reparametrization was shown to outperform MEPO-Qeq when comparing the CO_2 uptake and heat of adsorption to a system with REPEAT charges. However, one should consider that the MEPO-Qeq parameters used in the comparison were not re-fitted for the new training set and the parameters for the missing atom types were taken from GMP.

3.4.6 Other methods

It is worth mentioning two other methods that were used to obtain partial charges in MOFs without the need of computing their electron density. The first is the connectivity-based atom contribution (CBAC) method[181] which assumes the transferability of DFT-derived CHELPG charges computed for molecular cluster, to atoms with the same bonded neighbors. The second is the recent molecular building block-based (MBBB) method[182] in which the partial charges are computed separately for the ligands and the metallic nodes, properly capped into molecular clusters, and transferred to similar MOFs with different topologies and metal/ligand combinations. The MBBB charges were shown to reproduce considerably better the DFT-derived electrostatic potential than EQeq and CBAC methods. These methods require an extensive library of fragments: the CBAC was tested for 43 structures using a total 35 atom types The MBBB was parametrized for only 5 inorganic nodular, 6 organic nodular and 13 connecting building blocks. Therefore, these methods are not immediately ready to be used for large screening of MOFs with diverse chemistry and topology. Moreover, the MBBB method, is clearly designed for building and characterizing hypothetical structures from scratch but it still needs to be integrated with a building block recognition protocol for managing general structures.

3.4.7 Which one is the most reliable method to compute partial charges in MOFs?

Considering all the variants that have been proposed for the Qeq method one may ask which one is the *best* method to obtain the partial charges in a set of diverse materials such as Metal Organic Frameworks, to be used, for example, in the assessment of gas adsorption properties. In this context, we would like to define as "*best*" the scheme that reproduces the experimental data. However, these Qeq methods are aimed to be a computational efficient protocol to reproduce the charge distribution as obtained by a more accurate method, such as DDEC that, among other DFT-derived methods, is specifically designed to generate partial charges that reproduce the electrostatic potential and ensure chemically meaningful values. Therefore, in this context, we define the *best* to be the method that assigns partial charges which are in close agreement with DDEC charges. This allow us to compare the point charges directly (due to the chemical meaning) and the electrostatic interactions (since they reproduce the electrostatic potential). DDEC charges have already been computed by Nazarian et al. for 2894 experimentally reported MOF structures,[142] and they will be used here as a reference for our benchmarks. The validation set in our work is considerably larger than the small sets typically used before, i.e. 15, 12, 24 and 693 MOFs for PQeq,[140] EQeq,[146] I-Qeq,[147] and MEPO-Qeq,[148] respectively, aiming for a more complete picture of the accuracy and the weaknesses of these methods.

We only considered for the benchmark off-the-shelf methods, i.e., the ones that do not require additional fitting parameters. We focus on those methods where the parameters are obtained from isolated atoms, i.e., PQeq, EQeq, FC-Qeq and I-Qeq. The only exception we included is

MEPO-Qeq, with the aim of assessing the transferability of the parameters specifically fitted on MOFs. Moreover, we tested for each method, different sets of parameters, i.e., derived from GMP, experiments and CCSD(T). This will give us some insights about the improvement in a new Qeq variant, whether it is due to the modifications in the algorithm or to the choice of different parameters.

3.5 Computational Details

3.5.1 Programs to compute Qeq and DDEC charges

The variants of the Qeq method that are compared in this work are: the original version by Rappé (PQeq), MEPO-Qeq, EQeq, FC-Qeq and I-Qeq. For the first two we used a modified version of the General Utility Lattice Program (GULP)[177] named 'egulp'[183], which can take as input the parameters for the electronegativity and the idempotential. EQeq charges were computed using the program released by Wilmer in his paper.[146] As for FC-Qeq and I-Qeq, the program provided by Wells was adopted,[147] but all the input parameters were recomputed in this work (see the next section). Some considerations and benchmarks about the speed of different softwares are reported in the Supporting Information. As an example, the calculation for IRMOF-1 (conventional cell, 424 atoms) on a 2.6 GHz CPU, took 6, 7, 10 and 27 second respectively for FC-Qeq, I-Qeq, PQeq and EQeq. The DDEC charges were computed by Nazarian et al.[142] using the DDEC3 scheme[137] as implemented in the January 2014 version of the code. The PBE functional[178] was used to compute the electronic density for the charge fitting.

3.5.2 Ionization energies and radii

In all the Qeq variants we compared in this work, the user has to provide a set of isolated atom ionization energies, which can be measured experimentally or computed ab initio. From these, IP_n, EA_n, χ_A^0 and J_{AA}^0 are calculated.

The ionization energies can be computed ab initio using an accurate method such as the coupled cluster CCSD(T). To ensure consistency of our ionization energies we recomputed the energies for all the ions from -5 to +5 charge, using the Gaussian09[184] quantum code. The protocols was extended to all the atom types for which the basis set is available, i.e., H-Ar/Sc-Kr (34 atoms) for the aug-cc-pvqz[168, 169, 170] basis set and H-La/Hf-Rn (72 atoms) for the def2qzvpp[185] basis set. The first basis set, aug-cc-pvqz, has the advantage of including diffuse basis functions to better represent the broad electron density in anions. On the other side, the def2qzvpp basis set has been extended to include heavier atom, which are commonly found in MOFs, and therefore can be used to parametrize the Qeq methods for a larger number of material. However, in def2qzvpp the inclusion of diffuse function has been limited only to a few atoms, because these smooth gaussian functions often introduce more numerical instabilities in the convergence of the electronic structure or they lead to worse results.[186]

For consistency we decided not to use diffuse functions for the def2qzvpp basis set.

Since the most favourable spin configuration of an ionized state is not known a priori, for every atom and ionization state the energy was computed at different multiplicities: up to 11 for atoms with an even number of electron and 12 for atoms with odd electrons. Finally, only the multiplicity with the lower CCSD(T) energy was considered. The lowest energy multiplicity is reported in Tables S1 and S2. The ionization energies are reported in Tables S3 and S4, compared with experimental values in Tables S5 and S6. There is a noticeable discrepancy between the experimental and the ab initio results: on one hand one can argue that the measurement are subject to the experimental error, on the other hand single reference coupled cluster calculations with standard basis sets (the protocol adopted by Wells for his FC-Qeq and I-Qeq methods[147]) are not suitable to compute the energy of ions with a moderate positive or negative charge. As for the basis set limitations, we don't use the frozen core approximation in our CCSD(T) calculations (notice that def2 basis set is using effective core potentials for atoms heavier than Kr), but still the core basis function are not well designed for a contraction or expansion in highly charged ions.

If we just consider the IP_0 for the different atom types, we see a good agreement with experiments: the largest discrepancies are attributable to transition metals, but also for heavier atoms when using def2qzvpp. Considering the EA_0 (equivalent to IP_{-1}) for the H-Cl/Sc-Br atoms, and excluding all the noble gasses which show in a very large deviation in the EA_0, the mean absolute deviation to experimental values is 0.33 eV for aug-cc-pvqz and 0.49 eV for def2qzvpp. Note also that for certain atom types, e.g., Zn and other metals where the outer orbitals are fully occupied (Mg, Mn, Cd and Hg), the EA_0 is negative, the exact value is not reported from measurements[167] and it is taken as zero.[146]. In these cases the injection of one extra electron is not energetically favourable, and negative values found in coupled cluster calculations are artefacts due to the forced localization imposed by the gaussian basis set.

In Table S7 the results from the CCSD(T) are compared between the two basis set for the atoms H-Ar/Sc-Kr. The positive ionization states generally show a good agreement for most of the atoms, while for the negative states the results show a systematic deviation, with aug-cc-pvqz predicting in almost all the cases an higher ionization potential. This is reasonable because aug-cc-pvqz, due to its diffuse functions, can better accommodate extra electrons, leading to more stable negative ions. Again, this strong basis set dependence can be attributed to the artefact of forcing the extra electrons to stay close to the nucleus when using localized basis functions.

Table 3.2 compares the EA_0 and the IP_0 for the five most recurrent atoms (excluding hydrogen where effective parameters are used in the Qeq methods).

One can note a large deviation for copper between experimental and coupled cluster values, where it is not clear if the discrepancy comes from the experimental error or some approximations in the calculation (e.g., a strong static correlation). We will show that different sets of parameters often lead to very different partial charges, and consequently the choice of one set

Table 3.2 – Comparison of the EA and IP for the most recurrent atoms in MOFs. Energies are expressed in eV.

atom	**EA_0 CC/aug**	**EA_0 CC/def2**	**EA_0 exp**	**EA_0 GMP**	**IP_0 CC/aug**	**IP_0 CC/def2**	**IP_0 exp**	**IP_0 GMP**
C	1.25	1 .09	1.26	0.28	11.24	11.23	11.26	10.41
N	-0.23	-0.60	-0.07	1.02	14.53	14.51	14.53	12.78
O	1.40	1.08	1.46	2.06	13.53	13.50	13.62	15.42
Cu	3.01	2.89	1.24	-0.02	5.63	5.69	7.73	8.42
Zn	-0.52	-0.86	<0.00	0.82	9.19	9.16	9.39	9.39

of parameters over another can be as influential as the choice of the Qeq method itself.

Table 3.2 also reports the EA_0 and IP_0 parameters derived from GMP's electronegativity and idempotential (equations 3.4 and 3.5). These values are listed in the Open Babel package[187] and some of them were published in the 1991's Qeq paper[139] while the parameters for other atoms remain unpublished but were used to derive the atomic properties for the whole periodic table in the UFF force field.[127] These GMP parameters are referred as *generalized Mulliken-Pauling electronegativities and idempotential* by Rappé and Goddard and came from experimental "state-averaged" ionization potentials and electron affinities to mitigate spin state/exchange effects, but a detailed description of the protocol never appeared in print.[188] Notice that for the EA_0, the GMP parameters show significant deviations if compared with both experimental and coupled cluster values.

There is also some confusion in the literature about the values that are effectively implemented in the different programs that are used to compute the Qeq charges. As pointed out by Kadantsev[148] for the Qeq implementation in the GULP package,[177] the parameters for copper, differ from the original ones (GMP). Only Cu's and Ce's parameter are different. The values used for Cu in GULP are 2.48 eV and 4.98 eV, for the EA_0 and IP_0, respectively, while the corresponding GMP values are -0.02 eV and 8.42 eV. Notice that the GULP values are closer to the coupled cluster parameters than the GMP ones. The reason of this discrepancy is unclear and leads to partial charges in worse agreement with the DFT-derived ones.[148] However, both Wilmer and Wells used these parameters to compare Qeq, EQeq, FC-Qeq and I-Qeq,[146, 147] and therefore their conclusions need to be revised.

For FC-Qeq and I-Qeq methods, the radius for every ionization state is needed. This value was computed as the mean HF/def2qzvpp electron density $\rho_i(r)$, i.e.:

$$\langle r_i \rangle = \frac{\int r\rho_i(r)\, dr}{\int \rho_i(r)\, dr} \tag{3.26}$$

The radii computed using this protocol are reported in Table S10 for atoms types up to Radon, and for ions in the range of charges from -5 to +5.

3.5.3 Adsorption calculations

The RASPA 2.0 molecular simulation software[189] was employed to compute the adsorption properties of the frameworks at using the different sets of charges. A Lennard-Jones 12-6 potential was used to reproduce the dispersion forces. Parameters from UFF[127] were adopted for the frameworks' atoms and TraPPE force field was employed to model the adsorbed molecules, CO_2[190] and H_2S (4-3 model)[191] Frameworks and gas molecules are assumed to be rigid upon adsorption. Mixed Lennard-Jones coefficients are obtained according to the Lorentz-Berthelot combining rules, with a truncated cutoff of 14 Angstrom. Coulombic interactions were calculated adopting the Ewald summation scheme.[192] The CO_2 uptake was computed running 10,000 GCMC[193] cycles (5,000 for equilibration plus 5,000 for production) at the industrially relevant conditions for flue gasses, i.e., 298 K and 0.2 bar. The fugacity of CO_2 at these conditions was computed using the Peng-Robinson equation of state.[194] The insertion of CO_2 and H_2S was probed according to the Widom's test particle method to estimate the Henry's coefficient and the heat of adsorption at infinite dilution at 298K. For each molecule, the interaction energy was computed for 100,000 random positions inside the framework.

3.6 Results and discussion

3.6.1 Analysis of the charges obtained from DDEC

To assess the ability of the different Qeq variants to reproduce the partial charges in MOFs, we employed as a reference 2894 frameworks, for which Nazarian and co-workers computed the DFT derived DDEC charges.[142] These MOFs are extracted from the Cambridge Structural Database (CSD) and the solvent molecules have been removed computationally to allow for adsorption studies.[141] Out of the initial set of 4519 structures, for ca. one third of the frameworks the electronic structure calculation did not converge because of the large size of the unit cell or other issues, and for these MOFs the DDEC charges were not reported. From this set, we considered only the materials for which the def2QVZPP basis set is available, i.e., up to Rn and excluding the rare earth metals, Ce-Lu, for which also experimental EA_0 are not reported.[167] This gives us 2338 MOFs that we used in the analysis for the present work.

We set the stage by analyzing the different atom types that are represented in this study. Figure 3.2 (upper) shows, for each atom type (excluding H, C, N, O), the number of MOFs that contain it and the count of the total number of that atom type present in the set. To give an example, 74 MOFs contain F and 194 contain S, but there are more F atoms present in MOFs than S (1618 versus 1533, see Figure 3.2, upper). The count of atoms is important because, when comparing charges, deviations for atoms that are more frequent in the set will contribute more on the total mean standard deviation. Figure 3.2 (lower) shows, the average partial charge, the standard deviation and the minimum/maximum DDEC charge for this set of MOFs.

In our reference set the most recurrent metals, are the transition metals of the first row, from Mn to Zn, and also Cd. For these, the average partial charge is close to +1, but in certain

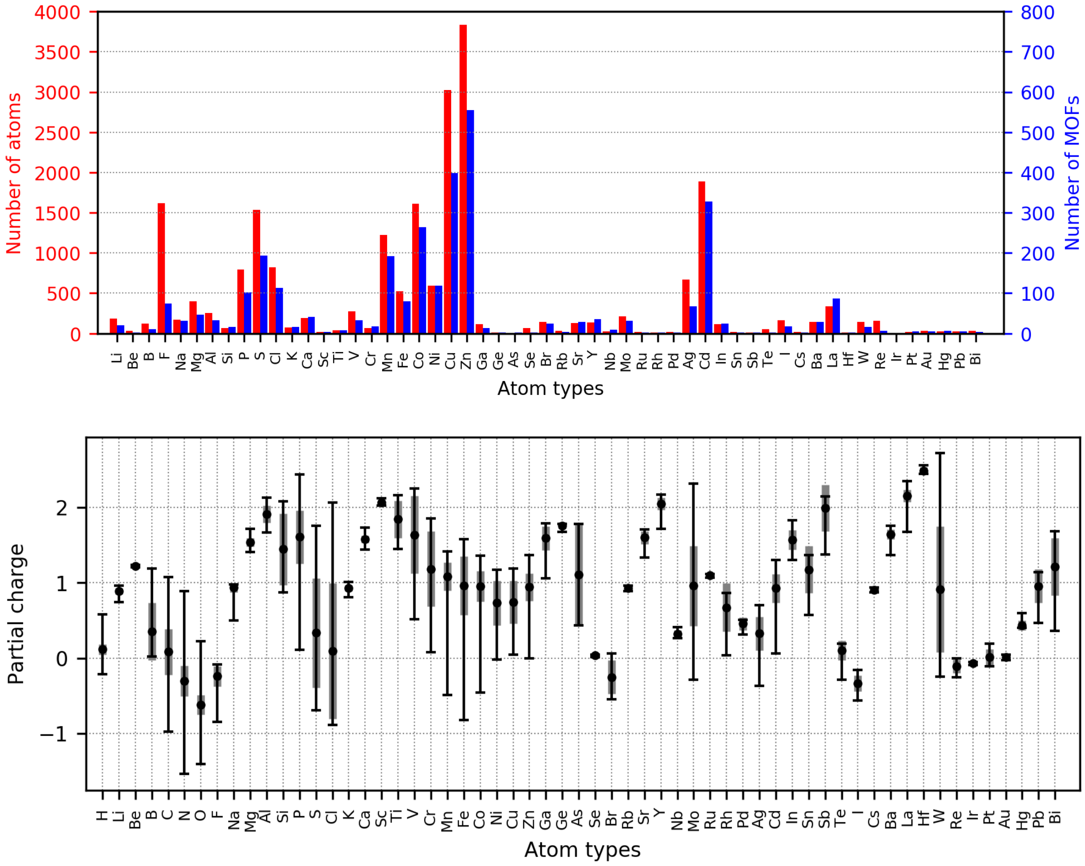


Figure 3.2 – The upper histogram shows the frequency of the different atom types in the set of 2338 MOFs we considered in this works. They are counted as total number of atoms (red bars) and number of MOFs containing that specific atom type (blue bars). H, C, N and O are excluded from the graph: their counts are 102’028, 14’4025, 28’123, and 53’796, respectively. Also, atoms types that are not present in the set are not showed in the figure. The lower graph show the maximum, minimum and average DDEC charge for every atom type, considering the 2338 MOFs dataset. Grey bars show the standard deviation. We use electron charge as unit charge.

cases they can take also a negative DDEC charge. It is important to remember that while the DDEC charges are fitted to reproduce the electrostatic potential inside the pores, they are also based on the electron density of the framework. Therefore, they are shown[136] to be less sensitive to the problem of non-physical charges on "buried atoms". As was pointed out from the work of Verstraelen et al.,[174] point charges exclusively derived from the ab initio electrostatic potential (e.g., using CHELPG or REPEAT schemes), should not be compared with Qeq charges, or worse, be used to fit the input parameters for Qeq methods, because they can take up extrapolated non-physical values. Therefore, for the case of DDEC charges, the negative charges are possibly due to the local environment of the metal instead of a bad fitting. It is interesting to notice, for example that all five structures where Fe has a negative DDEC partial charge (see Table 3.3) share the same chemical environment, with Fe coordinated to eight CN ligands with an octahedral geometry.

Table 3.3 – MOFs containing negatively charged Fe atoms, as computed with the DDEC method. For all the other 75 MOFs that contain Fe atoms, the charges on these are positive.

MOF	Fe partial charge
GEHSAN	-0.82
HIFTUM	-0.51
INIQUR	-0.47
OTOROF	-0.29
XULCIR	-0.49

Here, we are interested in comparing directly the partial charges obtained from different methods, and to assess how these different charges affect the typical experimental properties that can be predicted if these charges are used in molecular simulations. We focus on the adsorption of gas molecules with a partial charge, for which we use CO_2 as example.

To exclude non-porous structures, we considered a spherical probe with a diameter of 3.05 Angstrom (size of the oxygen for CO_2 in TraPPE's force field) to estimate that 77 structures over 2338 have zero probe occupiable pore volume,[195] meaning that they are non-porous. We excluded these structures from our adsorption analysis, and for the remaining ones we didn't block the inaccessible pockets in the adsorption calculations, because the aim of this study is to probe the electrostatic potential inside the pores. However, when simulations are compared with experiments, the inaccessible pockets, i.e., pores where the openings are too small for the molecule to diffuse inside, should be blocked to obtain a consistent estimation of the uptake.[189] Moreover, one should also verify that the MOF can be effectively activated (i.e., the coordinated solvent can be removed applying vacuum) and the framework retains its structure after desolvation.

To illustrate the importance of charges we compare the heat of adsorption (Figure 3.3a) and volumetric uptake (Figure 3.3b) in the different MOFs as computed with the Lennard-Jones potential and DDEC charges with the results in which the charge has been set to zero. The results show that both the heats of adsorption and the uptake are, on average, underestimated if Coulomb interactions between partial charges are not considered.

The material with the highest volumetric CO_2 uptake is VODSEM (316.6 $\mathrm{cm}^3_{STP}/\mathrm{cm}^3$) and the MOF with the lowest heat of adsorption is ICOYIK (-121.1 kJ/mol). Both contain La atoms with partial charges in the range of 2.14-2.24 electrons (among the highest, comparing Figure 3.2), and a favourable geometry that allows CO_2 to be bound from both the oxygen atoms.

In this study we are combining point charges with the dispersion potential obtained by mixing UFF and TraPPE parameters. These parameters were derived with very different procedures and philosophies and, they are widely adopted for screening studies,[179, 126, 176] assuming that their combination is a good guess for the framework-adsorbate interaction energies. However, in MOFs that are identified from the screening as particularly interesting, it is a common practice to derive tailor-made parameters for the host-adsorbate interactions from ab initio calculations.[196] Deviations are expected in MOFs where unsaturated metal centers

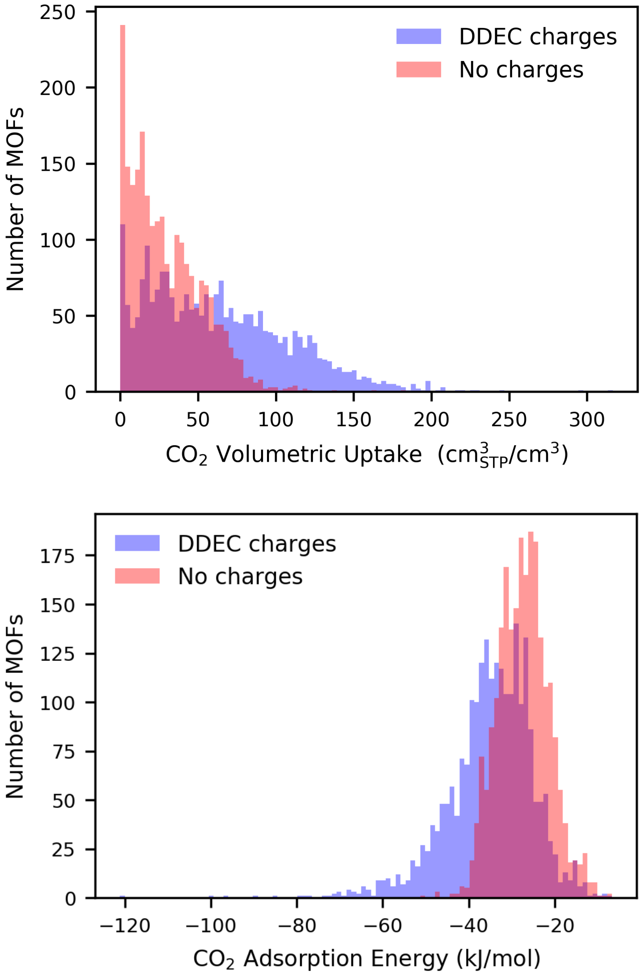


Figure 3.3 – Distribution of the considered MOFs by CO$_2$ volumetric uptake and adsorption energy, as computed from a GCMC simulation using DDEC charges and non-charged frameworks. It is evident how the atomic charges are influential for determining both properties.

are present.[197, 198, 199] We limit ourself to use standard force fields as a way to observe the variability related to different set of partial charges. Further comments on the charges assigned to CO$_2$ are reported in the Supporting Information.

3.6.2 Analysis of the charges obtained with different methods, charge centers and parameters

Comparing different charge center and parameters for EQeq

We computed the charge with the EQeq method using two settings: first, imposing zero charge center for all the atoms, versus using the formal oxidation states for transition metals as a reasonable guess suggested by Wilmer et al., and second, employing the ab initio computed set of parameters for the ionization energies, versus employing the experimental ones. Let us call the four combination EQeq/zero/exp, EQeq/zero/def2, EQeq/ox/exp, and EQeq/ox/def2. To set the charge centers for the whole periodic table, we assigned to all metals the lower common oxidation state.[200] However, the experimental values for the ionization energy are not available for all cases, especially for high oxidation states. In these cases we lowered the input charge center to the highest computable with available data, assuming a minor change in the resulting partial charges. As for non-metals, we assigned zero charge center according as suggested by Wilmer et al. The list of the input formal charges is reported in Table S11. Hydrogen was always treated using the effective parameters fitted by Wilmer et al., and an effective value of 1.67 was used for the relative dielectric constant ϵ_{eff}.

The EQeq code[146] was modified to address a convergence issue of the charges with the number of unit cells. Considering for example DOTSOV02 (HKUST-1), the charges on Cu change from 0.68-0.80 in the 1x1x1 calculation, to a value of 0.90 in the 3x3x3 calculation. This problem arises from the lack of spherical cutoff in the Ewald summation. The EQeq program, by default, expands the input structure to a 5x5x5 structure for the calculation of the Coulombic interactions. After some testing to verify the convergence of the output charges, we fixed this problem by increasing the default expansion of the unit cell to 13x13x13. The time for the calculation of DOTSOV02 significantly increased from 4 to 58 seconds (see SI).

Figure 3.4 shows the comparison of the EQeq charges with the DDEC charges for some representative atom types in the 2338 MOFs of the set: C, N, O, Cl, Cu, and Zn. We first focus on C (Figure 3.4a): if the EQeq would be in perfect agreement with DDEC, all the points would collapse on the dashed line. We see that for most of the structures there is a good agreement for all methods, but we also observe clusters of points that are far from the diagonal. Detailed inspection of these structures shows that these involve atoms with a similar bonding connectivity where EQeq and DDEC give discrepant prediction of the charge. For C, one can notice several "spikes", meaning that the DDEC gives a well-defined charge but the QEq method returns a random non-physical charge. The corresponding structures typically represent a specific carbon type environment. Let's take as an example the carbons that have a DDEC charge of 0.72 and a EQeq/zero/exp charge higher than two. This is the red spike on the top right of Figure 3.4a. For all the cases (EGELUY, EHALOP, SABVOH, and WAYMIU structures) these are carboxilic carbons coordinated to Al through bridging oxygens. In these structures the carbon just reflects a problem with the partial charge of Al, which takes non-physical values (higher than 10 electrons) when using the EQeq/zero/exp protocol. The same problem remains in EQeq/zero/def2, since some blue points are detectable in the same peak and therefore we can conclude that the proper charge center on Al needs to be specified in order to have a reliable result for these structure: indeed, no yellow or blue markers are present in this peak of Figure 3.4a. One can note these peaks also for other recurrent atoms such as N and O (Figure 3.4b and c).

To obtain physical charges, the second and the third term of the atomic-scale chemical potential (equation 3.8) need to be consistent, such that the hardness matrix is *positive definite*

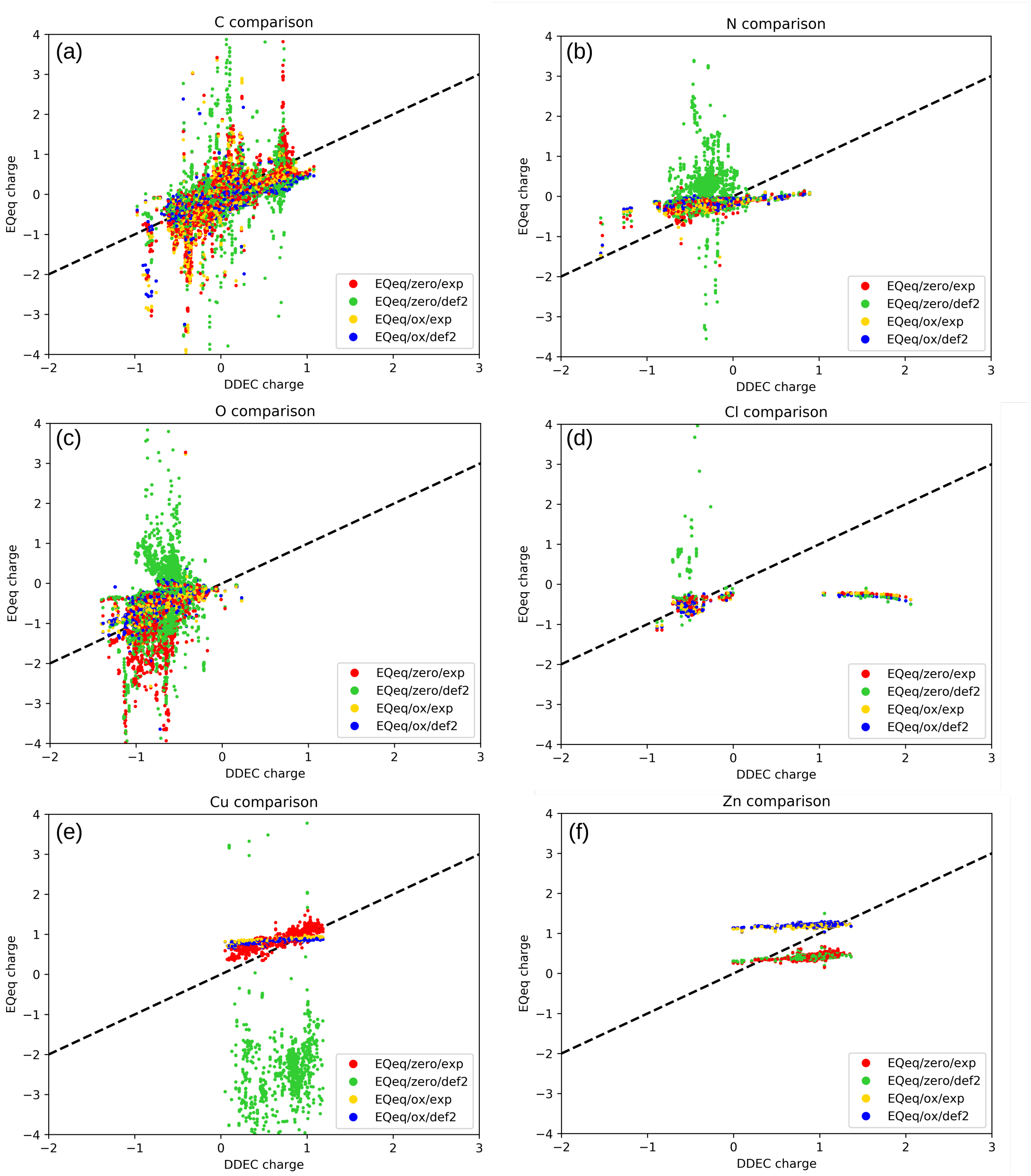


Figure 3.4 – Comparison of DDEC charges with EQeq charges computed with different settings and parameters: using zero or the common oxidation state as charge center and experimental or CCSD(T)/def2qzvpp parameters.

and a minimum for the energy (equations 3.7 for Qeq and 3.22 for EQeq) exists.[201, 174] This is an essential condition that one has to remember when attempting the training of these parameters, e.g., to reproduce a set of ab initio computed partial charge. However, in this study we use experimental and coupled cluster computed electronegativity and hardness parameters, and therefore this condition is not explicitly imposed, resulting in non-physical computed charges when certain atom types and type of bonds are present in the structure.

For chlorine (Fig. 3.4d), we observe an interesting feature: an horizontal series of points in the lower right of the graph, representing Cl atoms that are predicted to be positive by DDEC method but negative by all the EQeq calculations. All these cases correspond to the Cl of a perchlorate anion (ClO_4^-). These perchlorate molecules are, in fact, not part of the structure but charged solvent molecules. The EQeq method is not computing correctly their partial charges, independently to the chosen parameters. These structures, where the ClO_4^- solvent was not completely removed, are listed in Table S12.

For N, Cu and Zn, we observe that certain EQeq protocols give similar charges for all the structures, resulting in an horizontal line (Figures 3.4b, 3.4e, and 3.4f) indicating that the EQeq charge is less sensitive to the environment than the reference DDEC charges. For different choices of the charge center (see Figure 3.1) the $\chi_A^n Q_{\tilde{A}} + \frac{1}{2} J_{AA}^n Q_{\tilde{A}}^2$ parabola can be sharper, hindering more the partial charge on that atom, or smoother, allowing for a larger influence from the environment. This is especially evident in the case of Cu (Fig. 3.4e), when using experimental parameters: using the +2 charge center (EQeq/ox/exp) all the charges are narrowly centered in the 0.88±0.06 value. On the other hand, EQeq/zero/exp charges are more correlated to the DDEC charges, showing that with these parameters the charge of Cu is more flexible and sensitive to the environment. However, when using the zero/def2 settings the charges on Cu diverge to non-physical values. In this case, it is evident the large sensitivity of the charge on the choice of different Cu parameters: extreme care should therefore be paid on the parameters choice for this atom type, being the second most common metal in MOFs after Zn. Because of this reason, we preferred to use experimental values for the ionization energies in the comparison with other methods (PQeq, FC-Qeq and I-Qeq), as they are more robust.

Finally, we note that for Zn the experimental and CCSD(T)/def2, are giving very similar results. The distributions of the Qeq charges are quite narrow: 0.44±0.06 for zero/exp, 0.43±0.04 for zero/def2, 1.21±0.03 for ox/exp, 1.22±0.02 for ox/def2. The use of 0 or +2 charge center result in just a shift of 0.77 in the partial charge.

Another interesting comparison can be made on the alkali metals. The charge on these systematically diverges when the null charge center is adopted (Figure 3.5). An analogous result is obtained for K, Rb, and Cs. For alkali metals, the Taylor expansion centered in the zero or the first ionization states are very different (see Figure 3.1 for Li) and the χ_A^0 and J_{AA}^0 parameters are not able to reproduce the proper partial charges in the framework. Even if for some atom types (e.g., Cu) it was not obvious from these results if the zero or the formal oxidation state should be used as charge center, in the case of alkali metals the choice seems to be mandatory. Alternatively, one should use a higher order Taylor expansion, like in the work of Zhang et al.[164] Indeed, their work was motivated by the non-physical Qeq charges observed for AgLi cluster, where it is now clear that the problem is related to the presence of alkali metals.

Knowing the range of value for partial charges as computed from DDEC method, we will

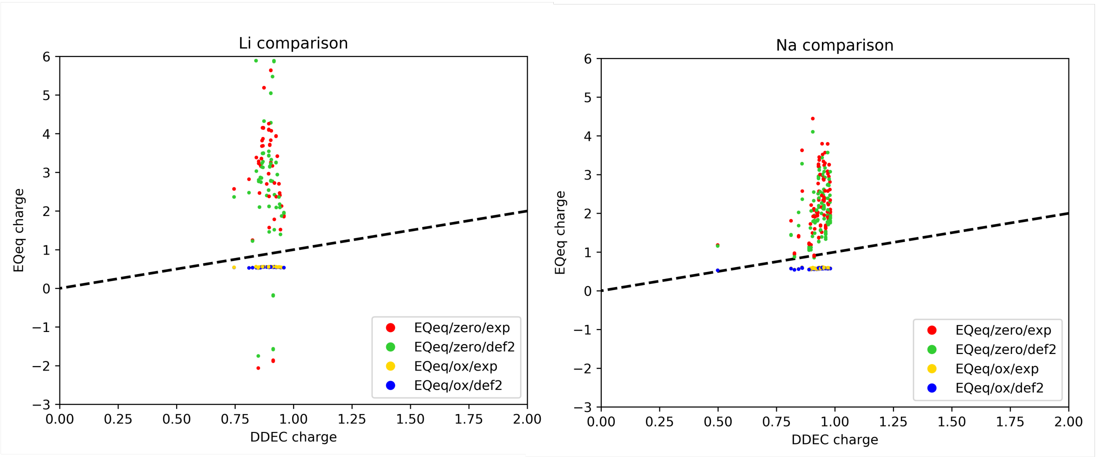


Figure 3.5 – Comparison of DDEC charges for alkali metals Li and Na. EQeq charges are computed with different settings and parameters: using zero or the common oxidation state as reference and experimental or CCSD(T)/def2qzvpp parameters.

impose, from now on, an upper limit of +3 and a lower limit of −2 for the partial charges. Frameworks with any charge outside this interval will be considered non-physical and discarded as if the method did not converge, to avoid the inclusion of these values in the statistics. Table 3.4 reports the mean absolute deviation for every method compared to DDEC, together with the number of invalid (i.e., discarded) outputs over 2338 frameworks.

Table 3.4 – Comparison of partial charges assuming DDEC charges as reference. "Invalid" structures are the ones for which the method did not converge or gave as output at least one charge outside the −2 to +3 range. For MEPO-Qeq all the structures that contain non-parametrized atoms are considered as invalid. The mean absolute deviation (MAD) is computed by comparing the charges of all the atoms belonging to valid structures.

Method	**Param.**	**Invalid**	**MAD**
EQeq	zero, exp	119	0.144
EQeq	zero, def2	564	0.167
EQeq	ox, exp	46	0.131
EQeq	ox, def2	30	0.148
FC-Qeq	def2	104	0.184
I-Qeq	def2	716	0.123
FC-Qeq	exp+def2	95	0.175
I-Qeq	exp+def2	214	0.118
PQeq	GMP	14	0.125
PQeq	exp	92	0.231
PQeq	MEPO fit	1566	0.165

Just considering the EQeq results, one can note how the choice of both the charge center and the ionization parameters is heavily affecting the final result. The robustness of the method is lower when using the neutral charge center: only 94.9% and 75.9% of the structures gave valid results with experimental and def2 parameters, respectively, versus a 98.0-98.7% when using the common oxidation states. Considering the mean absolute deviation, in both cases the experimental parameters lead to a better agreement with the DDEC charges than the CC/def2 parameters. Therefore, the choice of using experimental parameters and the common oxidation state, consistently to what Wilmer et al. suggested, seem the best combination for this method. In the paper by Nazarian et al. [142] a null charge center was used for many atoms for computing the EQeq charges to be compared with DDEC charges. This led to a poor agreement between the two results (see Figure 2 in Ref. [142]), which is especially evident for alkali metals.

Comparing different Qeq methods and parameters

We continue our benchmark, considering other Qeq variants with different sets of parameters. For FC-Qeq and I-Qeq we used both the ionization energies computed using the CCSD(T)/def2qzvpp method and the experimental ones. In the second case, the missing parameters (to have all the values for the ionized states from -5 to +5, as the methods require) were included from the ab initio values. Table 3.4 shows that for FC-Qeq and I-Qeq many structures did not converge or gave non-physical charges. I-Qeq outperforms EQeq/ox/exp, resulting in a mean absolute deviation as low as 0.118 when the experimental values are employed. However, we also have to take into account that with I-Qeq/exp the 9.2% of the structures are invalid: in particular for 99 the iterative routine did not converge and for 115 the partial charges went outside the boundary of -2/+3.

For I-Qeq, as we already noted with EQeq, CC/def2 parameters are responsible for many Cu charges to diverge. On the other hand, using the experimental values we obtained reliable charges for almost all the cases (Figure 3.6).

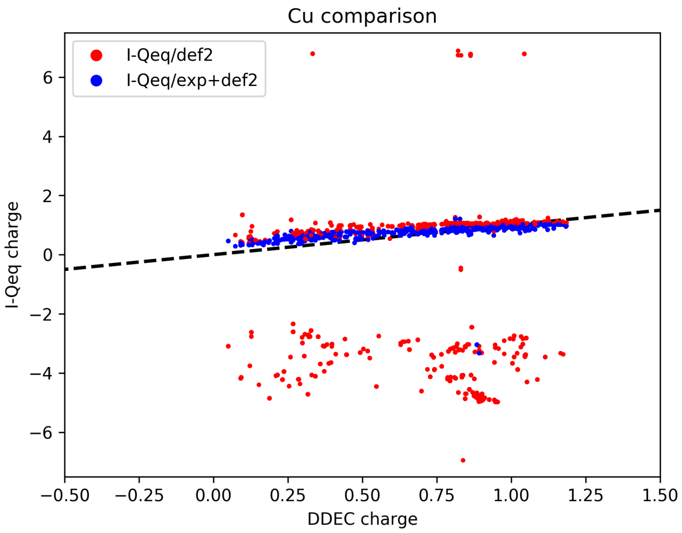


Figure 3.6 – The partial charges on Cu atoms are compared between DDEC and I-Qeq. These last were obtained using ab initio and experimental ionization energies as input. Charges from non-converged calculations are not shown.

Copper is a recurrent atom type in this set of MOFs, and therefore the choice of the set of parameter is important to judge the performance of the method. For example, the numerous HKUST-1 structures that are present in our set of MOFs (38 DOTSOV variants) failed to converge the I-Qeq calculation with CC/def2 parameters. The key problem is the low relative energy associated to the +1 ionization energy of copper, as we reported in Table 3.2. The reason why this problem did not emerge in the I-Qeq paper (where HKUST-1 is included in the validation set) is because the author tacitly assumed for the +1 ion an higher spin state for Cu (triplet) for which the IP gets closer to the experimental value and gives a robust convergence. However, this high spin state is less favourable than the singlet spin state (for both the basis sets) and this choice is not consistent with the declared assumption of considering the lower spins state. Other atom types for which the ab initio parameters give diverging I-Qeq charges for most of the structures, are: Mn, Ba, and La. In all these cases, the experimental ionization energies, expanded with CC/def2 only for the missing data, lead in general to a more robust convergence of the I-Qeq method.

For the PQeq method, we adopted three sets of electronegativities and idempotentials: the parameters from GMP (PQeq/GMP), the ones computed from experimental values (PQeq/exp) and the values fitted through the MEPO procedure (MEPO-Qeq). If we compare the results from PQeq/GMP and PQeq/exp (Table 3.4), it is surprising how different the mean absolute deviation is when using one set of parameters instead of another, showing once more the sensitivity of these methods on the parameters. As for the MEPO-Qeq method, we stress again that one should use this protocol with care: not only the applicability of this method is limited to a smaller set of atom types, resulting in a total of 772 structures over 2338, but also it should be restricted to frameworks having a topology (intended as metal coordination) which is similar to the ones in the training set, e.g, Zn and Cu paddlewheels. In this analysis we are extrapolating the results for a wide class of different coordination environments to test how consistent the computed partial charges are. From Table 3.4 one can note that the calculations converged for all the structures (the 1566 marked as "invalid" are the ones that contain non fitted atom types), but the mean absolute deviation is 0.165. This value is higher than using the PQeq/GMP method, despite the fact that it is evaluated over a smaller subset of atoms for which the parameters have been fitted. It seems evident, therefore, how the approach of fitting the effective electronegativy and idempotential is only limited to be transferred to framework which are very similar to the ones in the training set.

Figure 3.7 shows the deviations in the charges computed with the four protocols that gave the lowest mean absolute deviations. Even if I-Qeq/exp+def2 and PQeq/GMP have a similar mean absolute deviations (0.123 and 0.125) the former has a very peaked distribution of the error close to zero but also has many outliers, while the latter has a broader distribution but fewer outliers.

In Figure 3.8 the comparison of the partial charges computed with the same 4 Qeq methods are plotted versus the DDEC charges, for all the atoms of 2338 MOFs. Some common features can be highlighted. One is the horizontal line of values in the lower right, which was already

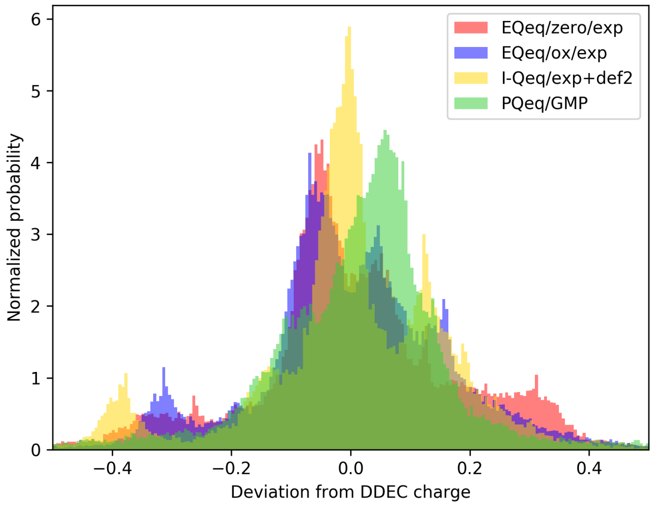


Figure 3.7 – Normalized histogram of the errors in the Qeq charges, considering DDEC as reference.

explained to be referred to perchlorite anions (Cl atoms are shown in magenta). Also, consistent with the Li charges in Figure 3.1, we see in both EQeq/zero and PQeq/GMP, a vertical series of point at ca. +1 DDEC charge, which correspond to the non-physical charges of alkali atoms when modelled with the potential centered at zero oxidation (green markers in Fig. 3.8). Notice that most of the structure containing alkali metals made the I-Qeq/exp+def2 calculation diverge and therefore only a few green markers are shown. Another interesting systematic deviation is the negative tail in the EQeq results at ca. -0.3 DDEC charges: these correspond to the charges on carbon (cyan in Fig. 3.8), and for all the cases where this deviation occur, there is a bond with a nitrogen involved.

From Figure 3.8d we can expect that using PQeq charges, despite the low mean absolute deviation, we will have a lower CO_2 adsorption in MOFs. In fact, the positively charged atom in the range +1/+2, that are the main attractive sites for CO_2 are systematically underestimated by PQeq. On the other side, PQeq overestimate the positive charge for alkali metals, but also B, Ga and In (red markers in Fig. 3.8). These three atoms belong to the 13th group of the periodic table and similarly to alkali have a single electron in the outer orbital, a p-orbital in this case.

3.6.3 Analysis of the adsorption results

To assess the impact of different set of charges on the adsorption properties that are commonly computed with molecular simulations, we considered 8 different sets of charges. Mixed UFF and TraPPE parameters were used to model the dispersion interactions in all cases. Charges are computed using these protocols (summarized in Table 3.5): (1) EQeq with common oxidation states and experimental parameters,[166, 167] (2) FC-Qeq and (3) I-Qeq using for both exp+CC/def2 ionization energies, (4) PQeq with GMP paramenters, (5) PQeq with

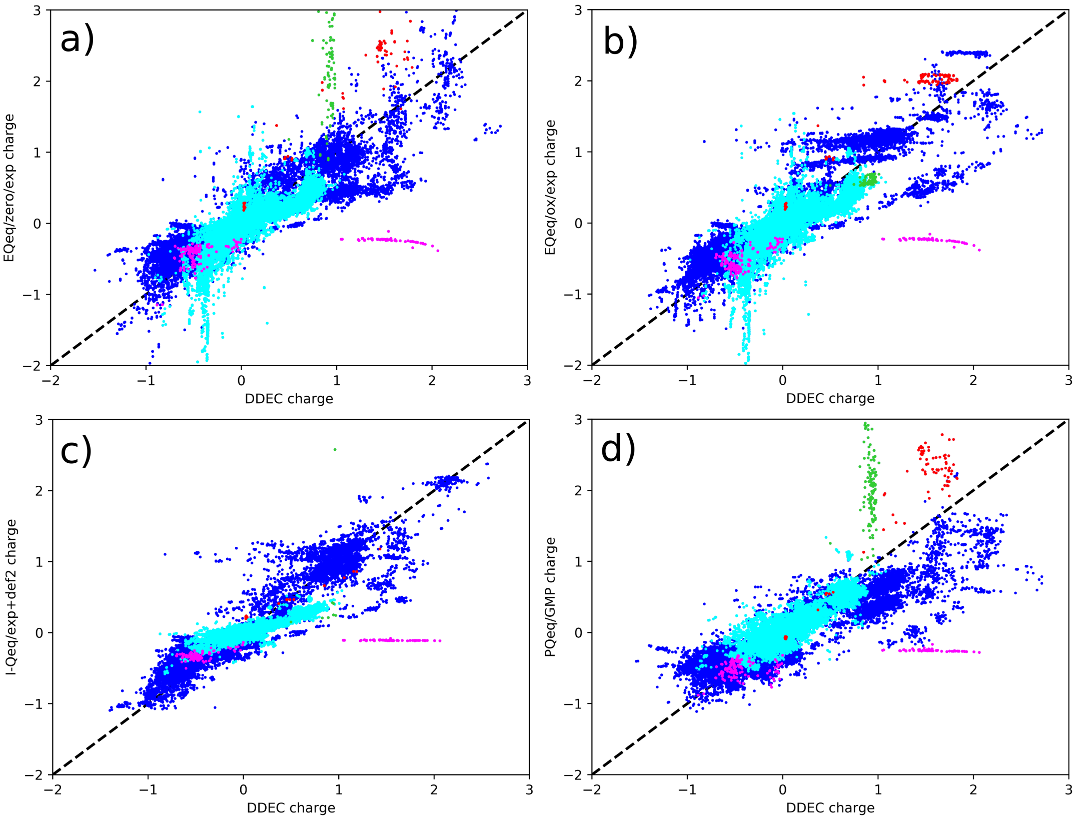


Figure 3.8 – Direct comparison of DDEC partial charges with (a) EQeq/zero/exp, (b) EQeq/ox/exp, (c) I-Qeq/exp+def2 and (d) PQeq/GMP. Color coding is used for different atom types: carbon (cyan), chlorine (magenta), alkali metals (i.e., Li, Na, K, Rb, Cs, in green), B, Ga and In (red). Charges for calculations that did not converge are not shown.

experimental parameters,[166, 167] and (6) MEPO-Qeq. We added also (7) a set of charges, labelled as "AVG-Q", where for every atomic element its partial charge is the average DDEC charge over the set (see Figure 3.2), slightly shifted to maintain the neutrality of the cell. Finally, (8) a set of null charges for every atom (NO-Q) was considered. To compute statistics, we took as reference the results of the simulations obtained with DDEC charges. As for previous comparisons, we discarded all the structures that did not converge or have charges outside the -2 to +3 range, and the ones with non-zero probe occupiable pore volume. All the other structures are included in the comparison.

Figure 3.9 and 3.10 show the CO_2 heat of adsorption and volumetric uptake computed using partial charges from the eight protocols and compared with DDEC charged systems. Table 3.6 and 3.7 report, for the same quantities, the mean absolute deviation, mean signed deviation, Pearson and Spearman coefficient.

We can start commenting that the heat of adsorption and the volumetric uptake give both the same ranking for the performance of the different methods: the lowest mean absolute

Table 3.5 – Summary of the 8 Qeq protocols for which adsorption properties are assessed in this study.

Method	**Notes**
(1) EQeq/ox/exp	Experimental[166, 167] χ_A^n and J_{AA}^n
(2) FC-Qeq/exp+def2	Experimental[166, 167] ionization energies are used, integrated with CC/def2 computed energies when missing.
(3) I-Qeq/exp+def2	Same as for FC-Qeq.
(4) PQeq/GMP	Generalized Mulliken-Pauling χ_A^0 and J_{AA}^0
(5) PQeq/exp	Experimental[166, 167] χ_A^0 and J_{AA}^0
(6) MEPO-Qeq	χ_A and J_{AA} fitted for MOFs[148]
(7) AVG-Q	Atomic averaged DDEC charges from the CoRE dataset[142]
(8) NO-Q	No Coulombic interactions considered

Table 3.6 – Comparison of the CO_2 heat of adsorption (kJ/mol) at infinite dilution. Mean absolute deviation (MAD), mean signed deviation (MSD), Pearson and Spearman coefficient are shown, with the reference being the value computed with DDEC charges.

Method	**MAD**	**MSD**	**Pearson**	**Spearman**
(1) EQeq/ox/exp	4.649	-0.755	0.779	0.791
(2) FC-Qeq/exp+def2	5.933	-2.644	0.710	0.725
(3) I-Qeq/exp+def2	3.258	1.878	0.889	0.881
(4) PQeq/GMP	4.879	1.612	0.636	0.798
(5) PQeq/exp	5.651	0.061	0.591	0.722
(6) MEPO-Qeq	2.798	2.403	0.878	0.888
(7) AVG-Q	13.456	-12.574	0.558	0.549
(8) NO-Q	7.867	7.687	0.473	0.562

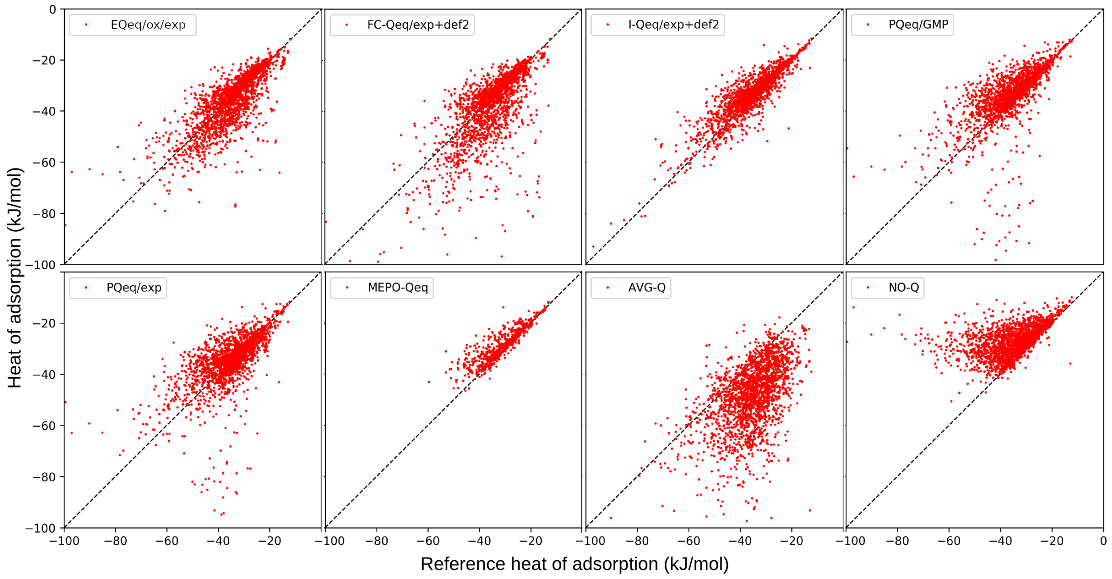


Figure 3.9 – Comparison of the CO_2 heat of adsorption (kJ/mol) at infinite dilution. Reference calculation are computed using DDEC partial charges.

deviation is obtained with MEPO-Qeq, then I-Qeq < EQeq ~ PQeq/GMP < PQeq/exp ~ FC-Qeq << NO-Q << AVG-Q. A similar trend is drawn by the Pearson and Spearman coefficients. Comparing together the mean absolute and signed deviations one can highlight a systematic

Table 3.7 – Comparison of the CO_2 volumetric uptake (cm$^3_{STP}$/cm^3) from GCMC calculations at 298 K and 0.2 bar. Mean absolute deviation (MAD), mean signed deviation (MSD), Pearson and Spearman coefficient are shown, with the reference being the value computed with DDEC charges.

Method	**MAD**	**MSD**	**Pearson**	**Spearman**
(1) EQeq/ox/exp	17.181	2.115	0.808	0.828
(2) FC-Qeq/exp+def2	22.608	7.896	0.689	0.754
(3) I-Qeq/exp+def2	13.611	-8.107	0.880	0.901
(4) PQeq/GMP	18.578	-9.530	0.743	0.802
(5) PQeq/exp	21.493	-2.328	0.692	0.734
(6) MEPO-Qeq	12.868	-11.682	0.841	0.898
(7) AVG-Q	60.144	53.115	0.431	0.425
(8) NO-Q	36.034	-35.007	0.446	0.511

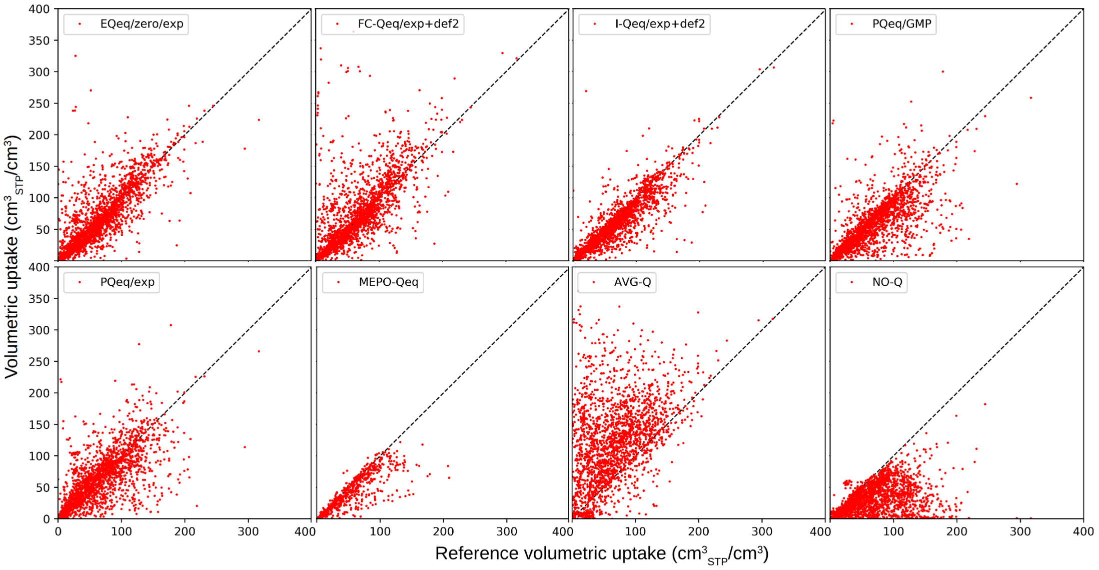


Figure 3.10 – Comparison of the CO_2 volumetric uptake (cm$^3_{STP}$/cm^3) from GCMC calculations at 298 K and 0.2 bar. Reference calculation are computed using DDEC partial charges.

deviation from the reference set of values. Indeed, these values are similar when the Qeq method lead to a systematic overestimation of the adsorption property (e.g. in the case of AVG-Q) and they are opposite when there is a systematic underestimation (e.g., for MEPO-Qeq and, as expectable, NO-Q). Using average charges (AVG-Q protocol) leads to the highest mean absolute deviation and a systematic overestimation of the adsorption properties, meaning that such a simplistic approach is too coarse for screening calculations. FC-Qeq performance also lead to a relatively high mean absolute deviation, even if it was shown to perform similarly to EQeq for a limited set of MOFs.[147] The performance of the FC-Qeq method is very much biased by the choice of the input formal charge, which is done internally by an initialization routine that evaluates the connectivity of atoms in the framework. Possibly, this part of the code needs to be further improved and tested for a more diverse set of structures.

Regarding the influence of the parameters on the final results, one can notice from Tables 3.6

and 3.7 that depending on the set of electronegativity and idempotential used, the PQeq/GMP method performs similarly to EQeq or considerably worse when using experimentally measured values (PQeq/exp). Hence, the experimental set of parameters is a good choice for I-Qeq but not for PQeq.

To evaluate how sensitive is the comparison to the utilized the probe, Table 3.8 shows the heats of adsorption for H_2S.

Table 3.8 – Comparison of the H_2S heat of adsorption (kJ/mol) at infinite dilution. Mean absolute deviation (MAD), mean signed deviation (MSD), Pearson and Spearman coefficient are shown, with the reference being the value computed with DDEC charges.

Method	MAD	MSD	Pearson	Spearman
(1) EQeq/ox/exp	4.165	0.453	0.756	0.773
(2) FC-Qeq/exp+def2	4.713	-0.943	0.678	0.725
(3) I-Qeq/exp+def2	3.122	2.221	0.868	0.882
(4) PQeq/GMP	4.105	1.072	0.651	0.818
(5) PQeq/exp	4.497	0.344	0.615	0.779
(6) MEPO-Qeq	2.667	2.535	0.899	0.901
(7) AVG-Q	20.776	-20.162	0.425	0.398
(8) NO-Q	6.635	6.622	0.672	0.692

The ranking according to the mean absolute deviation is again very similar as for CO_2, with PQeq/GMP's mean absolute deviation being, in this case, slightly lower than EQeq. Therefore, referring to different adsorption properties, ranking parameter or probing adsorbates, we note that the best methods are MEPO-Qeq and I-Qeq which are also the ones that are applied on the smallest subset. In our set, only 772 MOFs contain the 11 atom types that have been reparametrized for MEPO-Qeq, and 24 of them were further excluded because they were non-porous. When using this smaller set of structures for a fair comparison with the other methods, we obtained the results in Table 3.9. The mean absolute deviation for the subset of MOFs which are porous and valid for I-Qeq, 2064 in total, are also reported in Table 3.9.

Table 3.9 – Comparison of the CO_2 heat of adsorption (kJ/mol) at infinite dilution. The smaller subsets of MEPO-Qeq- and I-Qeq-valid calculations are considered. Mean absolute deviation (MAD) are shown, with the reference being the value computed with DDEC charges.

Method	MAD MEPO-Qeq-valid	MAD I-Qeq-valid
(1) EQeq/ox/exp	3.644	4.662
(2) FC-Qeq/exp+def2	4.696	5.848
(3) I-Qeq/exp+def2	2.713	3.258
(4) PQeq/GMP	2.206	4.126
(5) PQeq/exp	3.751	4.792
(6) MEPO-Qeq	2.798	2.842
(7) AVG-Q	13.993	13.324
(8) NO-Q	4.565	7.840

For the subset of 748 structures, PQeq/GMP is surprisingly outperforming MEPO-Qeq and all the other methods. I-Qeq has also a comparable but lower mean absolute deviation than

MEPO-Qeq and finally EQeq has a mean absolute deviation of 0.85 kJ/mol higher than MEPO-Qeq. On the other hand, when using the subset of 2064 MOFs the ranking of the methods remains more or less the same. Notice also that, in this case, a Qeq methods performs worse than the simulations without charges: this is the case of FC-Qeq when comparing the 748 "MEPO-valid" MOFs. All this illustrates the delicate comparison of different method to assess which one is performing better for an arbitrarily selected set of structures. We also have to consider that in the paper of Kadantsev et al.[148], the accuracy they could get from applying MEPO-Qeq to a validation set of 693 MOFs, all having a consistent metal coordination, was as high as 0.98 for both the Pearson and Spearman coefficients for CO_2's heat of adsorption. It is clear therefore that MEPO-Qeq is not a good choice for MOFs that are not similar to the ones in the training set that was used for the reparametrization, giving worse result than PQeq itself.

The results of our work allow us to comment on the possible overestimation of the charge on Cl and F functional groups in EQeq, which is corrected in MEPO-Qeq, as claimed by Kadantsev et al.[148] Figure S1 shows the charge comparison for this two atom types for MEPO-Qeq and the four different settings for the EQeq protocol. We can see how for Cl an F atoms with a weak DDEC partial charge, which in most cases correspond to benzene functionalization, the EQeq methods are overestimating the magnitude of the charge, while MEPO-Qeq is slightly underestimating it. Moreover, for strongly charged cases this method is unable to predict a proper partial charges. As for EQeq, a good match is found for Cl (excluding for the zero/def2 settings that we already showed to lead to inaccurate results) with some dependence on the choice of the charge center on the metal. However, for F atom types, large discrepancies are observed.

As a final benchmark, we compared directly the electrostatic potential as computed on a grid with a 0.2 Å spacing. Only the pore volume of the materials is considered, i.e., the volume that can be spanned by the center of a spherical probe with 1.5 Å radius. The diameters for the frameworks' atoms are taken as the Lennard-Jones's sigma from UFF. Table 3.10 reports the average mean absolute deviation for the different Qeq methods, considering different subsets of structures. The rankings of this analysis are comparable to the ones estimated from the adsorption calculations, with MEPO-Qeq outperforming the other methods when considering the full set of MOFs, but showing a similar deviation to I-Qeq and PQeq/GMP when the same small subset of MEPO-valid structures is adopted.

3.7 Conclusions

In this work, we assessed the performance of the different charge equilibration (Qeq) methods with a variety of different input parameters, over a set of 2338 MOFs for which the DFT-derived DDEC partial charges were available in the literature. These methods are usually validated over a restricted set of structures, and then used for the screening of thousands of structures to predict their adsorption properties and identify the best performing materials for a specific

Table 3.10 – Comparison of the electrostatic potential as computed on a 3D grid, for different sets of point charges. Three subsets with all, MEPO-Qeq-valid and I-Qeq-valid structures are considered. For every structure, the mean absolute deviation (MAD) with respect to the DDEC electrostatic potential is computed over the gridpoints of the pore volume, and the average MAD from all the calculations is reported in meV. The number of porous structures considered for each averaging is showed in brackets: non-porous structures were excluded.

Method	MAD all	MAD I-Qeq-valid	MAD MEPO-Qeq-valid
(1) EQeq/ox/exp	10.69 (2123)	10.73 (1941)	9.61 (726)
(2) FC-Qeq/exp+def2	9.83 (2080)	9.62 (1923)	8.05 (723)
(3) I-Qeq/exp+def2	6.07 (1976)	6.07 (1976)	5.14 (709)
(4) PQeq/GMP	8.65 (2148)	7.71 (1974)	7.02 (727)
(5) PQeq/exp	11.90 (2082)	11.22 (1970)	11.19 (727)
(6) MEPO-Qeq	6.13 (727)	6.09 (709)	6.13 (727)
(7) AVG-Q	27.96 (2160)	28.22 (1976)	26.59 (727)
(8) NO-Q	10.88 (2160)	10.40 (1976)	8.52 (727)

application. Assuming DDEC point charges as a reference, we assessed the discrepancy to the set of charges computed using different Qeq variants. Also, we quantified the deviations we observed when using these charges for computing common adsorption properties. In our benchmark study we show how the different methods suffer from very specific problems, in many cases related to a certain category of atom types, and in other cases to the choice of the parameters.

We showed that the second order Taylor expansion centered in the neutral state is not a good approximation for the $E(Q_A)$ potential of alkali metals, and therefore the standard PQeq method should not be used for these elements, while for other atom types the choice of which reference ionization state to use as charge center is questionable. Moreover, we have shown that the results are very sensitive to the choice of the parameters for the ionization energies (i.e., IP_n and EA_n). Therefore, we recommend to always specify the set of values employed in detail, to ensure the reproducibility of the study. Ab initio ionization energies calculated with coupled cluster methods guarantee consistent and reproducible set of values for most of the periodic table, but still suffer from the dependence of the basis set employed. We therefore suggest using experimental ionization energies for EQeq, FC-Qeq and I-Qeq, and GMP parameters for Qeq. These combinations of methods and parameters ensure a more robust convergence of the algorithms and more reliable results. The alternative (e.g., MEPO-Qeq) is to obtain the parameters using a training set of structures for which the electrostatic potential is obtained from higher level DFT calculation. However, we showed that if these parameters are used to compute charges in frameworks that are topologically different to the training set, the results are actually worse than when using the original Generalized Mulliken-Pauling (GMP) parameters derived from isolated atoms. Despite the discrepancies in the electrostatic potential obtained from Qeq or DDEC charges, the use of the Qeq partial charges generally leads to better agreement than non-charged or average-charged systems, based on all the descriptors considered in this study: mean absolute deviation, Pearson and Spearman coefficients. However, when considering this large set of 2338 MOFs, our results did not provide the type of evidence one would like to see to confirm a clear improvement in the accuracy of the Qeq methods over the years.

For more details on the utilization of these Qeq methods the user is suggested to visit the Material Cloud Archive (https://archive.materialscloud.org/2018.0017/v2) where all the inputs parameters and the results are provided for sake of reproducibility. Also, we provide the comparison of the different Qeq methods with DDEC charges, grouped by atom type, to allow for the recognition of symptomatic problems directly related to a specific combination of settings and atoms.

4 Building a consistent and reproducible database for adsorption evaluation in Covalent-Organic Frameworks[1]

4.1 Abstract

We present a workflow that traces the path from the bulk structure of a crystalline material to assessing its performance in carbon capture from coal's post-combustion flue gases. This workflow is applied to a database of 324 covalent-organic frameworks (COFs) reported in the literature, to characterize their CO_2 adsorption properties using the following steps: (1) optimization of the crystal structure (atomic positions and cell) using density functional theory, (2) fitting atomic point charges based on the electron density, (3) characterizing the pore geometry of the structures before and after optimization, (4) computing carbon dioxide and nitrogen isotherms using Grand canonical Monte Carlo simulations with an empirical interaction potential and, finally, (5) assessing the CO_2 parasitic energy via process modelling.

The full workflow has been encoded in the Automated Interactive Infrastructure and Database for Computational Science (AiiDA). Both the workflow and the automatically generated provenance graph of our calculations are made available on the Materials Cloud, allowing peers to inspect every input parameter and result along the workflow, download structures and files at intermediate stages and start their research right from where this work has left off. In particular, our set of CURATED (Clean, Uniform and Refined with Automatic Tracking from Experimental Database) COFs, having optimized geometry and high quality DFT-derived point charges, are available for further investigations of gas adsorption properties. We plan to update the database as new COFs are being reported.

---

[1] Preprint version of the article: Daniele Ongari, Aliaksandr V. Yakutovich, Leopold Talirz and Berend Smit, *ACS Central Science* 2018, *in press*, https://doi.org/10.1021/acscentsci.9b00619. D.O. conducted the search of experimental structures from the literature and did the data cleaning. He proposed the structure of the workflows, designed them in collaboration with A.V.Y. and L.T. and tested them. He performed all the calculations.

4.2 Introduction

Covalent organic frameworks (COFs) are a class of emerging materials composed of covalently bonded organic residues.[202] While classical covalent crystals like diamond, graphene, or carbon nitride could also be considered as COFs, this nomenclature highlights the concept of reticular design, where the precursor molecules, termed *ligands* or *linkers*, are rationally combined into a framework structure. The ancestors of COFs are metal organic frameworks (MOFs), where ligands form coordinate bonds with metal centers. In COFs, however, ligands are directly connected, forming the topology of the structure.

The first two-dimensional microporous COFs, named COF-1 and COF-5, were synthesized by condensation reactions of phenyl diboronic acid in the group of Yaghi.[203] Two years later, the same group employed a similar condensation reaction to obtain the first three-dimensional COFs.[204] Up to date hundreds of different COFs have been reported by different research groups.[205, 206] This new class of materials attracted the interest of the scientific community for applications ranging from gas adsorption,[207] to catalysis,[208] and photocatalysis,[209] setting the stage for a possible commercial use in the near future.[210] In addition to the synthesized COFs, new hypothetical COF structures have been designed *in silico*, [211, 212, 213, 214] and investigated to highlight possible promising frameworks that experimental chemists could target in their synthesis. The total number of these hypothetical structures exceeds the 560,000.

Similarly to MOFs, we see a rapid increase in the number of novel COFs that are being reported. Often these materials are synthesized with a particular application in mind and experimental data are often limited to this application; it is simply not practical to experimentally test a novel material for all possible applications. For such studies a computational screening to identify the most promising candidates is often the most efficient first step. Essential in these computational studies is that all experimental structures are curated in, what is referred to in the literature as "Computation-Ready Experimental" (CoRE) structures.[141, 124] The importance of this curation step cannot be emphasized enough, and, as we will argue in this work, that, in particular for COFs, the lack of a systematic approach makes it impossible to reliably compare structures for different applications.

The first public database to present the "CoRE" concept is the CoRE-MOF database, which collects over 5,000 structures.[141] This database has been used by many different groups for a large range of studies. The importance of such a database is that all structures are reported in a uniform format (e.g., Crystallographic Information Files, CIF, with P1 symmetry). Indeed, the possibility of easily accessing and computationally analyzing these frameworks, allows to screen thousands of different materials, compare their performance for specific applications, and find correlations between geometric properties and performances.[215] An important point is that the CoRE-MOF database not only involved a conversion to a uniform format but also includes a minimal curation, that mainly involves the removal of solvent molecules from the pores of the framework. Solvent molecules are often present in the reported structure of

MOFs, and they they can be (partially) removed. The idea of the CoRE-MOF database is to arrive at a solvent free material which serves as a reference for computational modelling. The difficulties of simply removing the solvent molecules from the structure, becomes clear if one realizes that for some MOFs the solvent molecules are essential, without solvent the framework collapses. Another important issue is that from the X-ray data one does not have sufficient information of the position of the H-atoms. These are then often added manually or using different software packages, where different protocols would put them on slightly different positions. In addition, depending of the quality of the X-ray data there can be uncertainty in the position of the atoms. The net result of all these small difference is that these screening studies are of limited use, and potential source of artifacts, due to the lack of uniformity between the investigated structures.

Contrary to MOFs, the atomic coordinates of synthesized COFs are not deposited in commonly available standardized databases. Therefore, it is already difficult to precisely estimate how many have been reported so far, and an extensive literature review is necessary to collect these structures: they must be extracted from experimental papers, checked and converted into a common format for storing atomic coordinates. Recently, Tong et al.[216] published a collection of 280 structures, which were parsed from the experimental literature, and that we used as a starting point to build the database of frameworks of the present study.

It is interesting to note that only a very small fraction of these COFs have been reported in the popular Cambridge Structural Database[217] (CSD). To be accepted into the CSD, a structure has to be determined directly *from X-ray and neutron diffraction studies, and from powder studies using a constrained refinement.*[218] This is the case of the recently synthesized single crystal COF-300, COF-303, LZU-79, and LZU-111.[219] In the most of the cases though, the poor long-range crystallinity makes it impossible to resolve the structure directly. The fact that from an experimental point of view it is difficult to obtain a sufficiently reliable crystal structures, requires experimental groups to increasingly rely on computational techniques to make a model of the material. These proposed structures are subsequently validated by comparison with experimental properties, such as NMR and/or “DFT Pore-size distribution” fitted from gas isotherms.[203] Part of this work was motivated by the notion that for these types of computational structure creation from scratch, there are no standards available. Depending on the methodology, it is easy to optimize towards different meta-stable states, different approaches may use using different bond lengths, different positions of hydrogen atoms, etc. In addition, these computation often involve geometry optimization which are performed with very different procedures (e.g., using different *ab initio* density functional theory methods or empirical force fields) that are often poorly documented. A direct use of these structures for molecular simulations therefore becomes a source of artifacts; these seemingly small differences in the way a structure is determined, can lead to significant differences in the prediction of, for example, the thermodynamic properties of these materials.

In this work, we make a first step towards a systematic approach, which allow for a standardization. The challenge we aim to address is to develop a protocol that not only allows such a standardization, i.e., all materials are optimized identically and using a reproducible protocol, but also allow for systematic improvements and additions. Our approach aims to address on one hand the needs of synthetic chemists with limited knowledge of computational methods, but that would like to use state of the art methods in an automated protocol to standardize the geometry of their just-synthesized crystal. On the other hand we provide a tool to the computational community to collectively improve these optimization protocols.

In this work, we consider a total of 324 COFs reported experimentally from 2005 to 2018 (we refer to the Materials Cloud, see Associated Content, for the references to the original structures), cleaned from solvent molecules, and with no partial occupations or disorder in the structures. We propose a systematic approach to organize and store these frameworks in a curated database, to track eventual corrections to the structures or additions to the database, and to link each entry back to its original source. The atomic coordinates and the cell dimensions have been optimized using density functional theory (DFT) to give a coherent geometry to all the frameworks. In addition, we have analyzed the results from DFT calculations to spot and correct errors in the structure and to achieve a robust and automated procedure to perform the DFT modelling and the optimization. DFT-derived partial charges of the structures are computed and provided for molecular simulations.

Whereas the optimization of a single structure can be seen as a routine DFT calculation, the optimization of over 300 hundred structures, including both 2- and 3-dimensional topologies is particularly challenging. For a single structure, one can develop a recipe that is tailor-made and optimized for a given material. For such a larger number of structures, the challenge is to build a workchain that is sufficiently general and robust such that any COF structures can be optimized. Moreover, given the number of calculations, one cannot afford to manually check all the outputs and ensure that all the calculations proceeded correctly. Therefore, it is important for the workchain not only to print out just the important values that the user need to monitor, but also to automatically act when a known problem is encountered. Our aim is to shift the attention from the single calculation to the inner logic of the workchain that, performing multiple calculations and tests, can chose automatically the most efficient path that brings to the solution (in this case the optimized geometry), and which can be extended to identify more problems and improve its robustness.

The important additional complication is that our workchain requires the interaction of multiple codes. Here we show the advantage of employing a supervision tool that can automatically execute these codes. For this purpose we used the Automated Interactive Infrastructure and Database for Computational Science (AiiDA),[220] that allows to represent our simulation protocol (the “workchain”) in a form of python script and execute it in an automated manner. It also give a common language for different groups to collaborate in extending and improving the routines we designed.

A novel aspect of this work is not only the development of this workchain, but also the full integration between AiiDA and the Materials Cloud. The importance of this integration becomes clear once groups start using our refined structures. For example, for a 5% of the structures our workflow did not succeed in obtaining a converged structure because of inconsistencies or errors in the deposited crystal structure. These errors range from missing hydrogen, incorrect labeling of atoms, wrong cell dimensions, etc. The accesses to the Materials Cloud allows any user to inspect all corrections we have made on the original structure, the outcome of the geometry optimization, and the computed partial charges to be used for molecular simulations. In addition, the reader can find on the Materials Cloud the detailed workchain, i.e., the logic and the input/output files for all programs that have been used to compute and characterize the final structures. These workchains can be downloaded, allowing any researcher to reproduce our results, or extend our calculations while maintaining their complete providence. We feel that it is important to create a precedent on the level of transparency and reproducibility that can be obtained.

The other important point we aim to highlight is the potential for future extensions of the entries in our COF database: the number of novel COFs that get published each is still increasing, novel structures can be added easily, allowing different groups to collectively contribute to extending the number of structures with uniformly refined geometry.

Therefore, in this work we go beyond the "CoRE" approach, presenting the "CURATED" COF database, where the focus is on the full tracking and reproducibility of the operation that are needed to make the database (possibly) error-free, consistent, uniform and easily upgradable.

Finally, optimizing >300 structures also gives important insights in the state of the art of COF synthesis, and in the applicability of the different programs that were used in our workchain. To show an example usage of the structure we obtained, with refined geometry and partial charges, we screened the COFs in our database for their potential application in capturing carbon dioxide from coal's post-combustion gasses.

4.3 Building of the database

We aim to build a database of COFs structures reported from the literature and with refined geometry. In the following, we will refer to these structures as *CURATED* COFs, an acronym that stands for: *Clean, Uniform and Refined with Automatic Tracking from Experimental Database.* The database includes one set of CURATED COF structures that contains the original frameworks as reported in the literature, with uniform orientations of the layers and minor corrections in case of typos, chemical errors or solvent removal ("*CURATED from literature*", 324 COFs), and another set containing the structures that succeeded the DFT optimization ("*CURATED DFT-optimized*", 308 COFs). The whole design of the CURATED COFs repository was conceived to easily correct for possible errors and to be upgraded with more recently reported frameworks.

4.3.1 Structure labels

In the CURATED database, covalent organic frameworks are labelled using 7-character strings (see Figure 4.1) that encode information about their structure (charge state and dimensionality) as well as their provenance (publication year, article ID and structure ID).

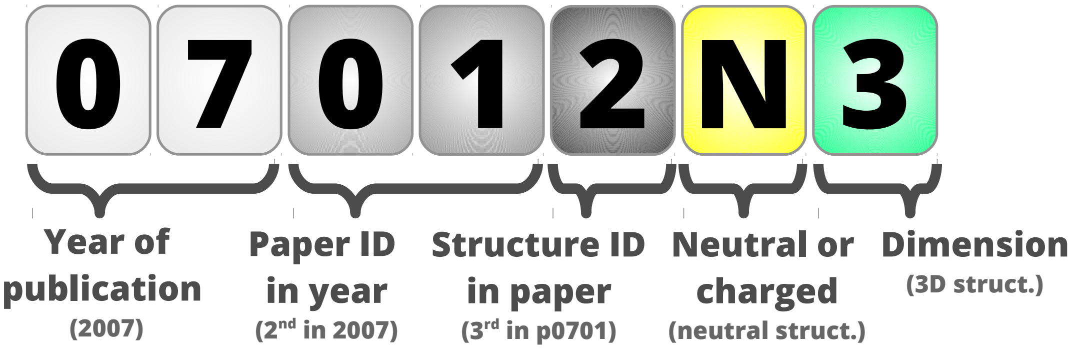


Figure 4.1 – Structures labels used throughout this work. The structure ID is a progressive counter starting from zero and following the order of insertion into the database (not the chronological order of publication). The charge state of a framework may be neutral ('N') or charged ('C', needing floating counterions to balance its charge) The first four integers are therefor and index for the reference paper, e.g., p0701, where "p" stands for "paper". Note that the counting for the IDs starts form zero.

Therefore, simply looking at a structure's label one can realize in one glance that, for example, COFs 07010N3, 07011N3, 07012N3 and 07013N3 are all 3D, do not contain charge-balancing counterions, and were published in the same paper of 2007 (p0701).[204]

4.3.2 Data cleaning

In this work, we argue that an important requirement of our database is that at the source of the initial structures as well as all the modifications that these structures underwent during the data cleaning process need to be fully tracked. Modification are necessary to ensure a uniform formatting of the structures and to remove errors in the structure that were detected after a visual inspection or a suspect failure in the DFT optimization.

Our database contains 324 CURATED structures from literature. The most of them comes from the previous collection done by Tong et al., in particular, we focused on the second version of their CoRE database.[216, 221] In the three versions of their database, Tong et al. collected 187,[222] 280,[216] and, only very recently, a third version of the database contains 309 COF structures.[223] The most of these structures were manually parsed from the PDF documents in the Supporting Information of the relative papers, and *some COFs reported without atomic coordinate files are constructed following the experimental information provided in the corresponding synthetic studies.*[222] As mentioned in the introduction, only a few of the known COFs are deposited in the CSD database. Despite the care of the authors in this necessary manual parsing, errors are unavoidable. Indeed, we had to apply ca. 20 corrections to the structures originally provided in the database of Tong et al. As our parsing was also done manually, and therefore equally prone to errors, we used the Git version-control system to permanently keep track of these modifications.[224] Our repository is built from the 280 structures reported in the second version of Tong's et al. database.[216]

It is important to realize that most of the errors were found by warnings in the convergence of the DFT optimization, and subsequently manually corrected. It was therefore important to monitor this optimization step and double-check the input structure of problematic runs. Problems in the initial structure are typically: missing hydrogen atoms, incorrect elements, highly unfavourable protonation states, overlapping atoms (originated by the parsing of partial occupations), and incorrect unit-cell dimensions.

To extend the database, we added new structures from the literature research by querying papers that contain the keywords 'covalent organic framework' or 'COF' in the title. For the most of the new structures, we had to manually parse the coordinates, and apply a few modifications: solvent removal, manual fix of partial occupations, unit cell's realignment for 2D COFs, or correction of typos in the original publication.

It is instructive to discuss a few examples of this detection and correction of typos. In three cases the DFT calculation failed and from visual inspection we observed unphysical bond lengths. The problem was fixed by keeping the same fractional coordinates for the atoms and rescaling the unit cell dimension to obtain physical C-C bonds in benzene rings (see COFs 18021N3, 18121N3, and 18122N3).[225, 226] We see this as an example of a reasonable but ad-hoc fix, which might be challenged in the future. In another case, a significant change in the unit cell was found after optimization, and this was caused by a manual error in the parsing from the SI: the COF with ID 120 in the database of Tong et al.[216], was incorrectly obtained by mixing the unit cell of HAT-NTBA-COF with the atomic coordinates of HAT-NTBCA-COF.[227] The mistake has been corrected in the present database, and the two structures are now reported as 17061N2 and 17060N2, respectively.

Geometry and cell optimization

The 85% of the COFs in this database are layered structures. To have consistent representation of all these layered COF structures, the unit cells were chosen to contain two layers for all COFs, and these layers were placed perpendicular to the z axis (c dimension). The stacking of two-dimensional COFs is important for the evaluation of adsorption properties, and having only one layer would not allow the geometry to find its energy minimum in a possible AB stacking during the optimization. Out of 274 2D COFs, 253 single-layers structures were found, and expanded assuming perfectly eclipsed stacking. We would like to emphasize that one also could opt for 3 or more layers, but which would make the optimization even more CPU intensive.

For the DFT optimization, it is important to mention that the unit cells of several structures in the database have more than 1000 atoms and volumes beyond 100.000 Å^3. Such large volumes and number of atoms pose a challenge for performing density function theory in a high-throughput context, informing our choice for the DFT code. Therefore, our choice of the CP2K package is motivated by its efficient DFT implementation, that exploits mixed Gaussian and Plane-Waves (GPW) method based on pseudopotentials, and the efficient Orbital Transformation (OT) method for the optimization of the wave function. Since this method is not suited for structures with vanishing DFT gap, in case of final band gap smaller than 0.1 eV (a total of only 44 structures) the calculation was flagged, and different settings were used for the workchain, applying diagonalization and smearing (see Methods section for the details).

For the unit cell and geometry optimization we designed within AiiDA an advanced workchain that ensure a more robust convergence by performing a “three-stages” optimization. First, a preliminary cell optimization with fixed unit cell’s angle is run for maximum 20 steps. Then 100 steps of flexible cell ab-initio molecular dynamics (AIMD) were performed to give a “shake” to the structure and escape from metastable or higher energy local minima. At the end, a final cell optimization is performed without constraints on the unit cell’s angles. All the details on the parameters can be found in the reported workchain on Materials Cloud and are summarized in the Methods section.

The second step, the AIMD stage was found to be particularly necessary for the optimization of 2-dimensional COFs. We report in the Supporting Information the comparison with a standard direct cell optimization protocol, where we typically observe that, without the AIMD, the geometries gets stuck in a more symmetric state with higher energy, typically the perfectly eclipsed AA configuration for 2D COFs.

In the way we designed and applied this workchain, it can almost be seen as a chemical sanity check: if a structure does not pass this stage it most likely violates the collective knowledge of quantum chemistry. This either implies DFT to have some severe inaccuracies in modeling the crystal, or, what we found in most of the cases, chemical inconsistencies in the initial structure that are needed to be fixed manually before the structure can be optimized.

Geometric properties and partial charge assignment

To characterize the structures before and after the cell optimization, a number of geometric descriptors were evaluated, e.g., pore surface, pore volume, largest free sphere, pore size distribution, etc.

DFT-derived partial charges (DDEC protocol) are computed for the optimized structures, utilizing the electron density that was already computed in the previous optimization stage. These partial charges are already a valuable result of our study as it allows for the modelling of the interactions with polar adsorbates such as CO_2 or H_2O.

Adsorption calculations and parasitic energy

As an illustration of the use of these COFs, we screened the materials for their performance to separate CO_2 from flue gasses in a case study in which these materials are use to carbon capture from a coal fired power plant followed by geological sequestration of the captured CO_2. In our model, we mimic a temperature-pressure swing process to separate CO_2 from coal post-combustion flue gasses, and its compression to 150 bar for underground storage. The aim is to find the materials that gives the lowest parasitic energy. This parasitic energy is defined as the loss of electricity production caused by separation and compression of one kg of CO_2.[228] To evaluate this energy we need to predict the pure component isotherms of CO_2 and N_2 in the COFs at 300K, as evaluated from Grand canonical Monte Carlo (GCMC) simulations.[229] We have developed an AiiDA workchain to compute the minimal CO_2 parasitic energy from the optimized structures, provided with DDEC charges. The full workflow is sketched in Figure 4.2.

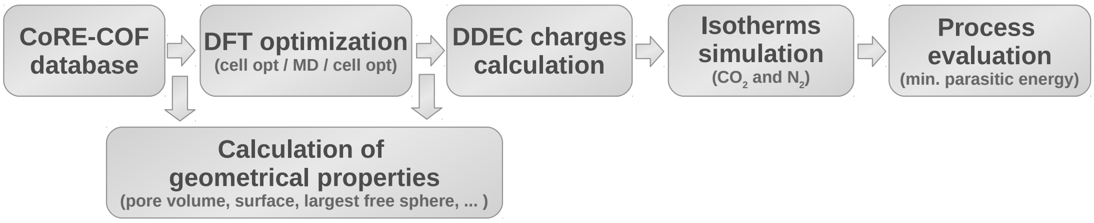


Figure 4.2 – Block diagram of the workflow used in this work.

4.4 Results and discussion

4.4.1 Analysis of experimental structures

The database of experimental COFs we are considering contains 324 structures, with 310 being neutral and the remaining 14 being charged: in this second case counterions are necessary to maintain the neutrality of the system, and these charged molecules have been kept in the pore as reported in the reference paper. The counterions found in these structures are dimethilamonium, tetrafluoroborate and single atom ions, e.g., Li^+, NA^+, F^+, Cl^+, Br^+, and I^+. Figure 4.3 reports the variety of elements in these crystals, distinguishing between 2D/3D and neutral/charged frameworks. One can note that these COFs contain a number of transition metals, e.g., Co, Ni, Cu, and Zn. However, unlike MOFs, these metal are not part of the connection nodes, but they are embedded in ligands, typically in porphyrines and phthalocyanines.

An important point that we need to discuss is the stacking of 2D COFs, i.e., the arrangement of neighbouring layers, sticking together because of non-covalent interactions. Reviewing the experimental literature about the synthesis and characterization of COFs,[203, 230, 205] we found that it is common to consider just two types of stacking: the eclipsed AA stacking, where the different layers are perfectly superimposed and the staggered AB stacking, where

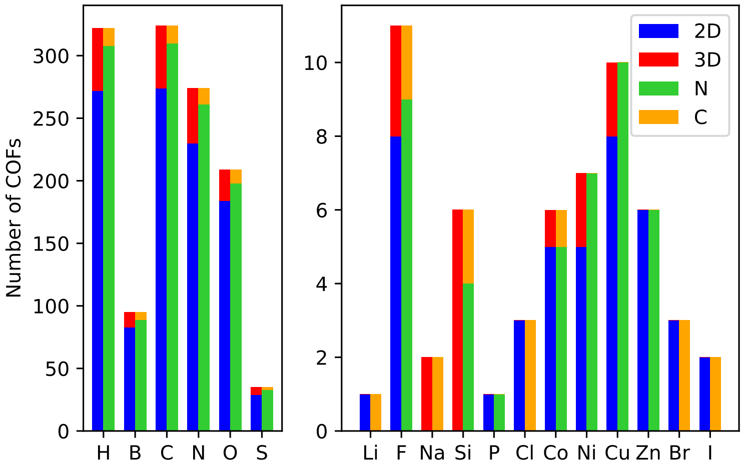


Figure 4.3 – Distribution of elements for the 324 COFs actually present in the database. COFs are also distinguished into 3D or 2D, and neutral ("N") or charged ("C", that need floating counterions).

there is an alternation of even and odd layers with an offset in the xy plane. Figure 4.4 shows as example these two configuration for COF-1 and COF-5.

Lukose et al. found that the minimum energy configuration for these COFs has a slightly tilted stacking.[231] Instead of the perfectly aligned AA and AB, COF-1 and COF-5 were found to have its minimum energy when the layers have an offset of ca. 1.4 Å. This misalignment was shown to be hard to distinguish from the experimental PXDR pattern and one has better to rely on computational methods. Recognizing the correct stacking is indeed a crucial aspect for many applications, and particularly for gas adsorption, where shifted or staggered configuration results in a shrinking of the channel diameter, possibly blocking the channels to the diffusion of gas molecules.

Analyzing the collected 2D COFs we noted that in only few cases (21) two layers are reported in the unit cell, and it is often not clear the method that was used for the generation of that particularly stacking configuration over the all possible ones. For the remaining 253 structures only one layer was reported, assuming AA stacking. Therefore, by consistently considering two layers in the conventional unit cell for these 2D COFs, we allow the geometry optimization to explore both the "serrated" configuration (i.e., where odd layers are slightly shifted) and the "inclined" configurations (i.e., where there is a constant offset of the layers, resulting in tilted unit cells). Note that the serrate configuration can not be obtained when only one layer is considered.

4.4.2 Cell optimization

We performed the cell optimization for the 310 COFs that do not contain counterions in their pores (i.e., labelled as "N", neutral). As we reported in the previous section, the most of the

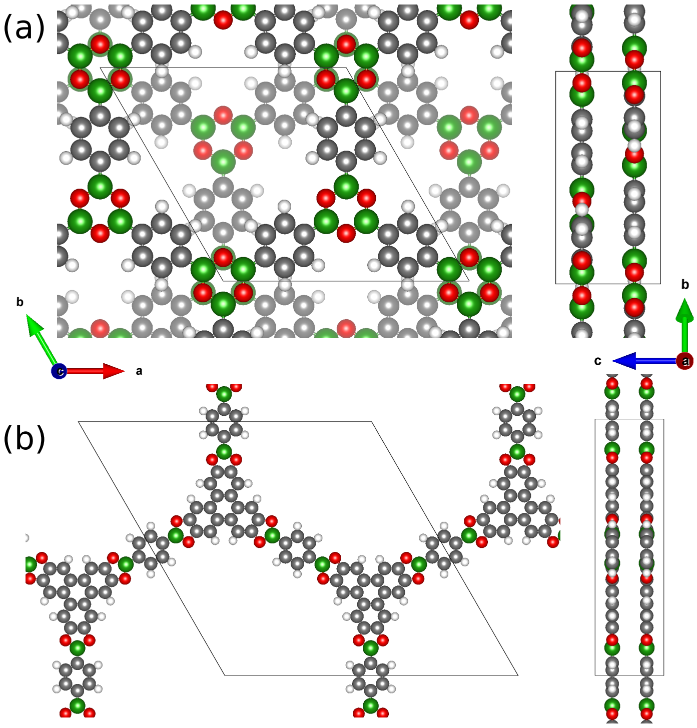


Figure 4.4 – The AA and AB stacking for 05000N2 (a) and 05001N2 (b), as reported in the experimental reference,[203] are compared. Conventional names are COF-1 and COF-5, respectively. In that report, the assessment of the presumed stacking was done by generating for both COFs the AA configuration and a reasonable AB configurations, and comparing the computed XRD with the experimental PXRD.

problem in the DFT optimization routine was solved after a careful inspection and fix of the initial structures. The problems that could be effectively attributed to the DFT implementation are related to only two structures: 18081N2 and 18082N2. These two both have cobalt and they need for electron smearing in the SCF: even testing different input parameters (e.g, lower mixing alpha, spin state, etc.) did not lead to a successful optimization. Considering that cobalt's pseudopotential is designed for 17 valence electrons (the largest number of electrons in the set we used), this is known to be a challenging element for the SCF diagonalization, and some effort is needed to design a more effective protocol for this particular element. However, AiiDA allowed to fully track the problem and report the issue to the CP2K developers.

Given that more than 99% of the structures in the database converged, our three-stages DFT optimization routine is suitable for high-throughput calculations. For comparison, in a previous attempt of a systematic DFT-based geometry optimization of 2,612 MOFs, only 879 (the 33.7%) successfully converged.[232] In that case the unit cell dimension was not optimized and no check was included for the band gap.

Figure 4.5 shows the evolution of DFT energy over our three-stages structure optimization workchain for two exemplary structures: 05000N2 (COF-1) and 05001N2 (COF-5).

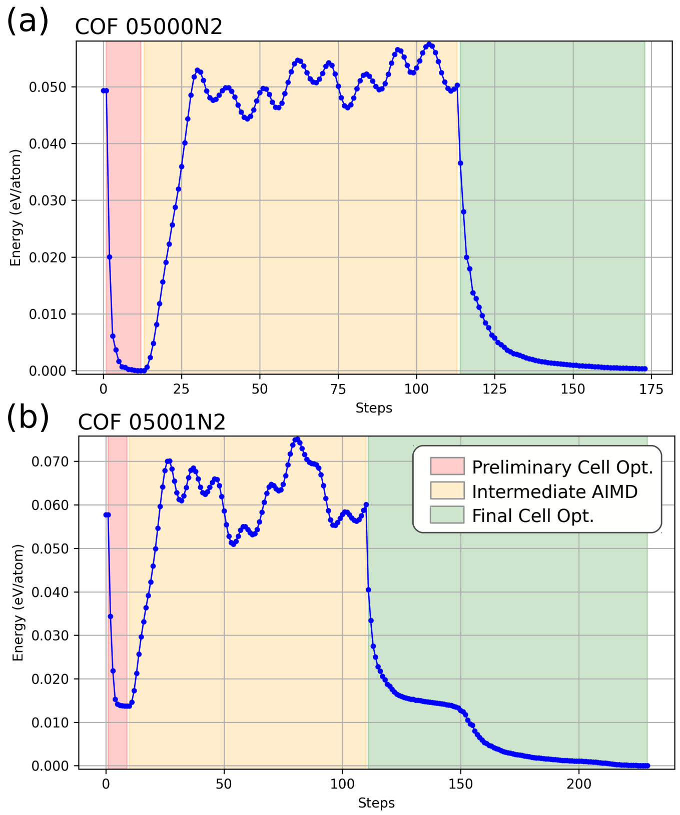


Figure 4.5 – Energy profile during the optimization workchain for COFs (a) 05000N2 and (b) 05001N2. Conventional names are COF-1 and COF-5, and the starting configurations of the layers, as reported,[203] are staggered and eclipsed, respectively. The three colored regions of the plot correspond to the first cell optimization with constrained cell's angles and 20 steps maximum (red), the NPT AIMD at 400 K / 1 bar for 50 fs i.e. 100 steps, (yellow) and the final cell optimization without constrains (green). The energy is shifted to assign a 0 value to the minimum.

As one can note that for the COF 05000N2 structure[203], we already find the minimum in the first stage and there is no apparent need for stages two and three. For COF 05001N2, however, the MD "shaking" of the structure results in it finding a more stable minimum that is 11.8 meV/atom more stable. This final optimization results from a shifting of the perfectly eclipsed layers in the reported structure, coherently with the work of Lukose et al.[231] In the Supporting Information, we show that, without the intermediate AIMD stage, the structure gets stuck in the first minimum, due to the high initial symmetry where layers are perfectly eclipsed.

To evaluate for the entire COF database the relative importance of the two cell optimization stages, we plotted in the histograms of Figure 4.6 the difference in energy in the first cell optimization (with constrained angles) and the further drop in the energy for the final cell optimization, showing also the contribution of dispersion interactions.

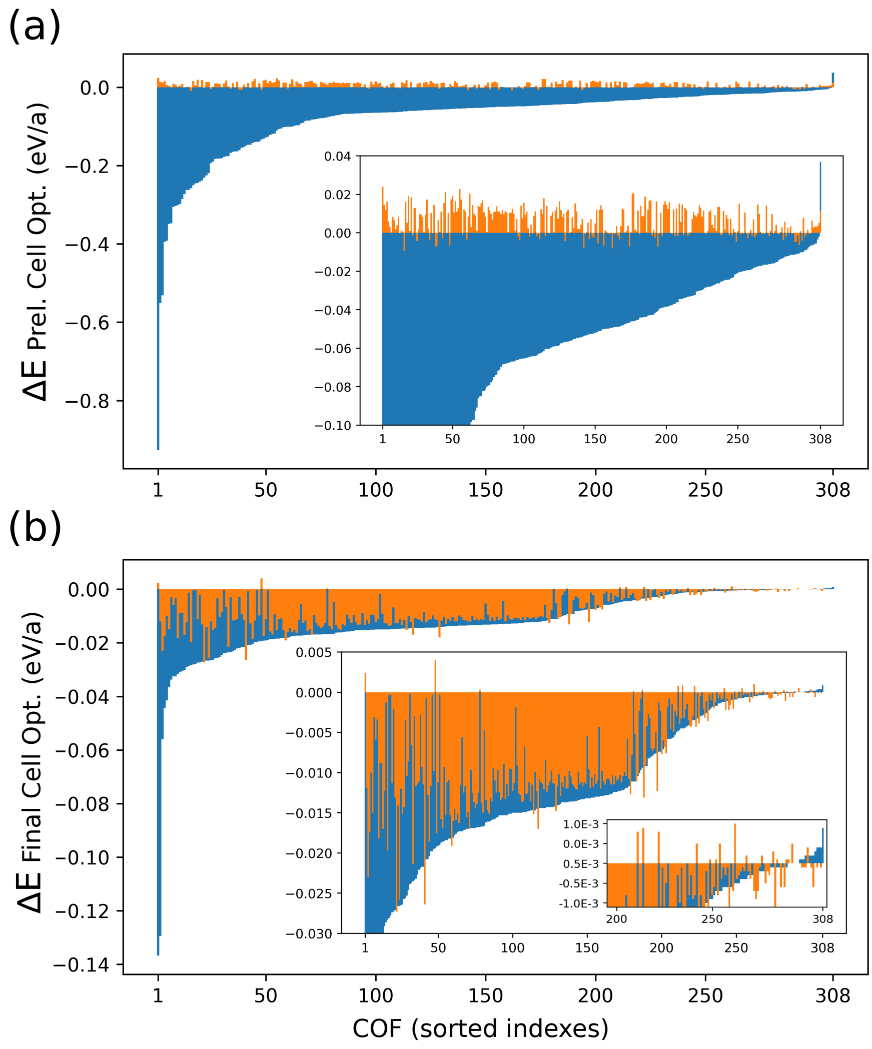


Figure 4.6 – Comparison of the difference in the energy in the first (a) and third (b) cell optimization stages. These two stages are also referred in the text as “preliminary” and “final”, respectively. Difference in energies (blue) are sorted for the 308 COFs in the histogram, showing also the relative contribution given by dispersion interactions (orange).

We observe that in most cases dispersion interactions dominate the final cell optimization. In the first cell optimization, where the typical changes in energy are larger by a factor of 10 (c.f., compare the y axis scaling in Figure 4.6), we mainly observe from the optimization trajectories a rearrangement of the atomic bonds and an increase of the distance between non-bonded fragments when this is set too close in the starting geometry. If we further inspect the statistics of Figure 4.6, we can observe that, for the first cell optimization, in only one case there was an increase in the energy: this is related to COF 12011N2 and results from a false step of the BFGS minimizer, which did not have the time to relax back in the limit of 20 steps. For as many as 153 COFs (49.68%) the cell optimization converged within the first 20 steps (i.e., at the “preliminary” cell optimization stage), but only the 19.61% of these had a negligible energy drop (<1 meV) in the final cell optimization, meaning that still for the most of the cases the AIMD helped to find a lower minimum. More detailed statistics are reported in Table 4.1, distinguishing between two- and three-dimensional COFs.

Table 4.1 – Statistics on the optimization process of the 308 COFs that succeeded DFT optimization.

	2D COFs	**3D COFs**	**all COFs**
Number of optimized structures	261	47	308
Converged in Preliminary Cell Opt.	50.96%	42.55%	49.68%
Converged, but ΔE Final Cell Opt. > -1.0 meV/atom	12.03%	70.00%	19.61%
Avg. steps in Preliminary Cell Opt	16.3	16.5	16.3
Avg. steps in Final Cell Opt.	150.3	113.3	144.7

The difference in the structures before and after the geometry and cell optimization can be appreciated from the change of the geometric properties, as plotted in Figure 4.7.

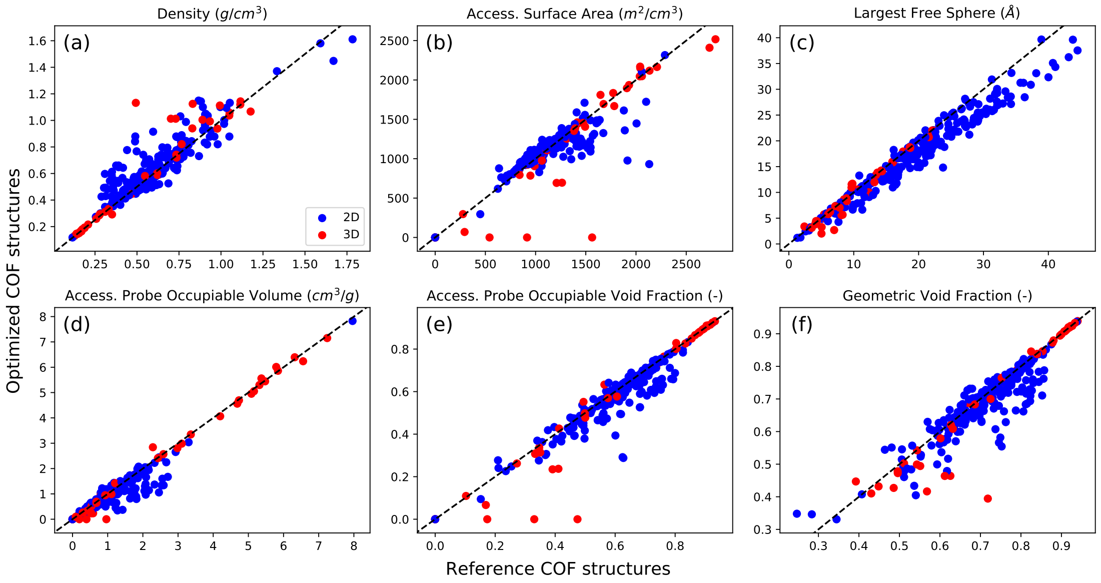


Figure 4.7 – Parity plots that compare the geometric properties of COFs before and after the optimization: (a) density, (b) accessible surface area, (c) largest free sphere, (d) accessible probe occupiable volume, (e) accessible probe occupiable void fraction and (f) geometric void fraction. To determine the accessibility of the pore volume and surface, a spherical probe with radius 1.86 Å (i.e., the conventional kinetic radius of nitrogen) was considered. The “geometric” void fraction is considered as all the portion of the pore volume that does not overlap with the atoms.[195] The colors distinguish between 2- (blue) and 3-dimensional structures.

We can see that in the many cases the DFT optimization leads to lower density, corresponding to a shrinking of the unit cell of the COF. The most evident case is for COF 18120N3 (Figure 4.8), where the density increases from 0.49 g/cm^3 to 1.13 g/cm^3, with a correspondent decrease of the geometric void fraction from 0.72 to 0.39 After the optimization this structure becomes non-porous (i.e., null N_2-accessible pore volume and surface). By inspecting the change in the atomic positions and the unit cell lengths we can notice that the structure shrunk to reach a more favourable configuration that optimizes the conformation of the ligands and their van der Waals interactions. For this COF, the measured pore volume from the nitrogen uptake is 0.36 cm^3/g,[226] i.e., well below the value computed from the reported structure (accessible N_2 probe occupiable volume equal to 0.96 cm^3/g). We can therefore expect that this structure can shrink after desolvation. For this COF the rigid crystal assumption would not hold to model the adsorption: the reported structure would lead to artificial results.

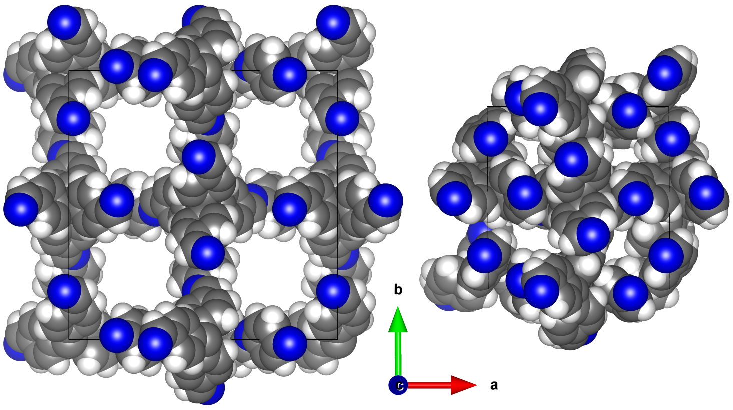


Figure 4.8 – Structure of COF 18120N3 before (left) and after (right) the cell optimization.

For the two-dimensional COFs, we observe in most of the cases a shift in the layers (i.e., the "inclined" configuration) that maximizes the non-covalent interactions. This results in a general reduction of the accessible surfaces area, pore volume, and largest free sphere, as shown in Figure 4.7. We observe that the energy stabilization due to this shift is very low. Considering the 133 layered structures where the geometry converged at the first cell optimization stage, the energy stabilization at the end of the final cell optimization, i.e., due to the tilting, is on average -13.1 meV/atom. This is a relatively low energy, and one may therefore expect that, at finite temperature, there is no unique stacking of these materials. While our protocol successfully identifies a close-by local minimum for both 2D and 3D structures, for 2D materials, a complete screening of all possible stacking configurations and the averaging over those accessible at finite temperature needs to be further explored.

4.4.3 CO_2 separation performances

The systematically optimized structures with high quality charges, that have been obtained for 308 COFs (CURATED DFT-optimized set), allows us to model the interaction with polar molecules, e.g., considering Coulombic attractions, and evaluate the performance of these materials for gas adsorption. As an example, we investigate the use of COFs for the removal of carbon dioxide from coal's post-combustion flue gasses.

To highlight the importance of high quality input structures (i.e., DFT optimized and with DDEC charges) for the evaluation of CO_2 adsorption, we computed also the empirical Qeq charges for the non-optimized structures: this is a cheaper protocol that is usually exploited for high-throughput screenings. In Figure 4.9, we compare the CO_2 Henry coefficient and adsorption energy at infinite dilution (i.e., computed from the Widom insertion method) for the two protocols, where the absolute change in the void fraction is shown by the color of the markers. It is striking to see the impact of both the optimization and the use of accurate partial charges. One can observe at best a weak correlation between the two protocols. The implication of these results is that using the COF-structures as reported combined with a simple charge assignment scheme will potentially result in many false positives and neglecting good-performing materials. This observation is one of the main motivation of this work, and illustrates the need to extend our CURATED approach to other databases. More details of the contributions that charges and geometry have to the CO_2 interactions are included in the Supporting Information.

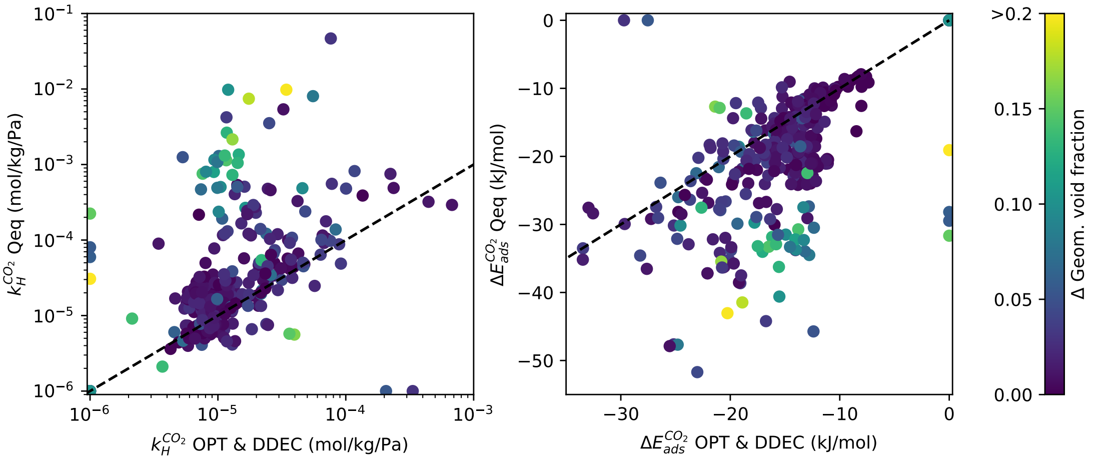


Figure 4.9 – Comparison of the CO_2 Henry coefficient and the adsorption energy at infinite dilution for non-optimized COFs with Qeq partial charges and DFT-optimized COFs with DDEC charges. Dashed lines show the condition of equal values for both axis. The color of the markers indicates the absolute change of the geometric void fraction after the optimization. Points that are laying close to the axes (i.e., null Henry Coefficient and adsorption energy) indicate non-permeable structures, where all the pores are inaccessible. Blocking spheres have been used in all the calculations to exclude inaccessible pores, as described in the Methods section.

With our DFT-optimized set of CURATED COFs with DDEC charges, we are now in a position to evaluate their performance for CO_2 capture and geological storage from a coal-fired power plant. We used the same protocol to evaluate different classes of microporous materials as previously employed by our group.[233, 228] In this protocol we assume a pressure-temperature swing adsorption process to separate CO_2 from a CO_2/N_2 mixture and subsequent compression of purified CO_2 to 150 bar, the requirement for underground storage. The key parameter to assess the performance of this process is the minimal parasitic energy, i.e., the energy that is required to separate and compress one kg of CO_2. However, other key parameters are also important for the evaluation: one should aim for high purity of the final high-CO_2 concentration gas, and high working capacity of the materials, to achieve the same productivity with less material or with less adsorption/desorption cycles.

Of all the COFs we considered, 12 COFs are non-porous for CO_2 or N_2 (according to the atomic radii definition from UFF, i.e., half of the Lennard-Jones sigma) and need to be excluded from the calculation. In addition, eight structures have inaccessible pores, which need to be blocked to prevent GCMC from growing particles in these narrow pockets. In the present study, we assumed that the COFs are rigid and therefore maintain their stacking upon adsorption. One can envision changes in the stacking upon adsorption. It would be prohibitively expensive to include framework flexibility in our screening structures, but it would be important to study these effects if one further considers top performing structures.

In Figure 4.10, we compared the simulated CO_2 isotherms for the first seven entries of our CURATED DFT-optimized COF database (i.e., the oldest synthesized) with experimental data.[234]

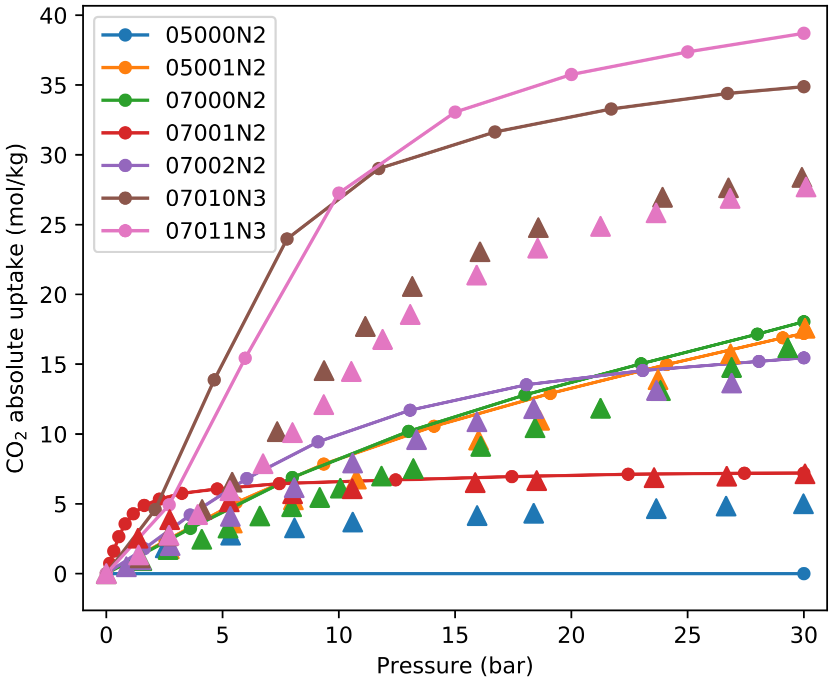


Figure 4.10 – Comparison of simulated (lines) CO_2 isotherms with experimental ones (triangles).[234] The isotherms were obtained at 300K and 298K, respectively. Note that the simulated uptake of COF 05000N2 is set to zero because from geometric analysis, no accessible volume was found for CO_2 nor N_2. The conventional names of these COFs are, in the same order as in the legend: COF-1, COF-5, COF-10, COF-6, COF-8, COF-102, and COF-103.

A notable difference is found for COF 05000N2 (COF-1), which we found to be non-porous under the rigid framework assumption: the small uptake that was measured is possibly the consequences of the motion of the layers that creates interstices where the CO_2 molecule can percolate and occupy the pores. Considering the uptake in the low pressure range, we note from the isotherms a systematic overestimation of the CO_2-framework interactions, that reflects in higher Henry coefficients. Also, in the 3D COFs 07010N3 and 07011N3 the apparently higher saturation found in simulations, may be related with the imperfect cristallinity of the sample or partial desolvation. A more detailled comparison is provided in the Supporting Information. Considering that we carry out a high-throughput approach, we conclude that the agreement with experiments is acceptable.

To compute the minimal parasitic energy of a material, we need as input the pure component isotherms and the heats of adsorption for CO_2 and N_2. The adsorbing bed is assumed to have a void fraction of 0.3, due to the pelletization and packing of the crystal. The volumetric working capacity measures the amount of CO_2 that one cubic meter of bed can evacuate between adsorption and desorption.

In Figure 4.11 the minimal parasitic energy is plotted as a function of the Henry coefficient, the volumetric working capacity, and the final molar purity. In all three comparisons we see a strong correlation. However, for materials with low parasitic energy, i.e., below 1 MJ/kg where the process is potentially more energy efficient than ammine-based technologies, the correlation becomes less evident, and one can for example select among the materials with low parasitic energies (that is an index for operative costs) the ones with higher working capacity, that would correspond to the need of less adsorbent (and therefore lower capital costs).

The envelope for the minimum parasitic energy versus the Henry coefficient that one can draw in Figure 4.11 can also be compared with the results obtained for Lin et al. for a database of more than 300,000 hypothetical zeolites and zeolitic imidazolate frameworks (ZIFs),[233] and later by Huck et al., for MOFs and porous polymer networks (PPNs).[228] In the present work, the lowest parasitic energies are obtained in the range of the Henry coefficient between 10^{-4} to 10^{-3} mol/kg/Pa. These results are coherent with the earlier findings on other materials. Hower, the Henry coefficient for COFs rarely exceeds 10^{-4} mol/kg/Pa, and none was found to exceed 10^{-3} mol/kg/Pa. Our effort consistently extends the “materials genome” for carbon capture and sequestration, now including also COFs. When compared with the best material found in the work of Huck et al.,[228] we observe that the performance of these COFs are still below the best-performer MOF-74, for which we get 0.705 MJ per kg of CO_2, with a purity of 0.943 and a volumetric working capacity of 64.87 kg of CO_2 per cubic meter of bed. In total, 14 materials over 60 were indicated by Huck et al. to have a low parasitic energy for coal flue gas (in the range of 0.7-0.85 MJ/kg), including 6 cation exchanged zeolites, 5 MOFs and 3 porous polymers. However, the consideration that COFs materials contains only light atoms, being a potentially cheap, can be an important criterion of choice, making this class of materials appealing for CO_2 separation.

Since we used AiiDA for the data-tracing of the full workflow, from the reported structure to the performances evaluation, we can now easily inspect every intermediate stage, backtracking the provenance of the final results. We illustrate this for three COFs: 18041N3, having the

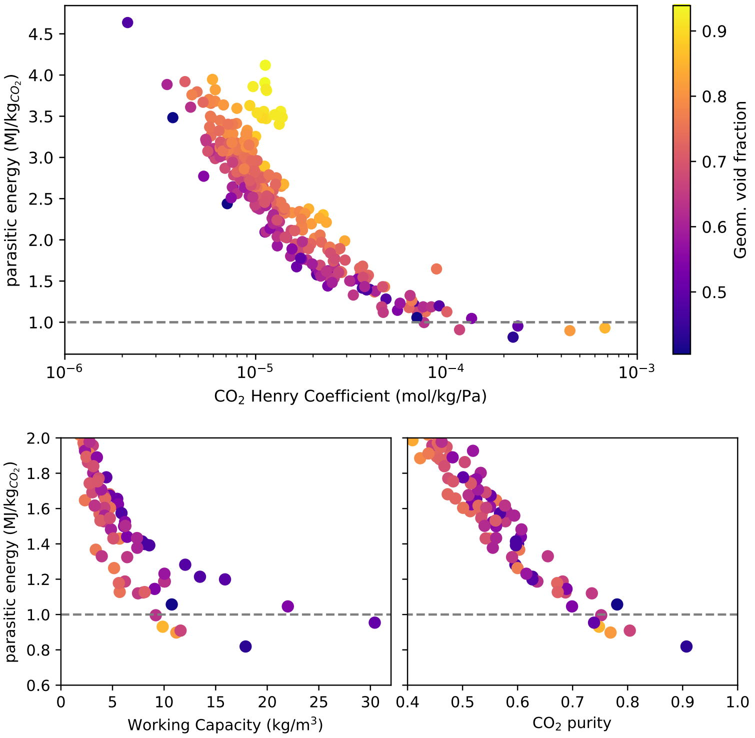


Figure 4.11 – For a set of 296 porous and DFT-optimized COFs the CO_2 parasitic energy is plotted versus the Henry coefficient for CO_2. The dotted line gives the comparable parasitic energy of the amine-based capture process. In the lower plots the parasitic energy is plotted versus the other two main outputs from the process modelling: the CO_2 working capacity per cubic meter of adsorbent bed and the final CO_2 molar purity of the mixture.

lowest parasitic energy (0.819 MJ/kg), 13180N3 having the maximum working capacity (30.42 kg/m^3) and the test case 05001N2 (COF-5), which has a relatively high parasitic energy of 2.681 MJ/kg, and therefore is not promising for this particular application. For these materials we can inspect in one glance (Figure 4.12) the energy profile during the optimization, the changes in the structures, the uptake sampling and the details of the process simulation. The reader can access a similar visualization from the Materials Cloud, in the Discovery section.

It is interesting to compare our results with the original articles, that reports the synthesis details for these COFs. For example, it is important to confirm that the experimental structure can be desolvated. This is the case for 18041N3[235] but not for 13180N3,[236] where the COF is reported to loose its crystallinity and porosity after solvent removal. Therefore, to use this material for gas adsorption, the fact that our optimization routine converges to a stable structure suggests that this material can be stable without a solvent, this indicates that it would be worthwhile to investigate different synthesis protocols that may allow for milder conditions for the activation and possibly retain the crystallinity of the sample.

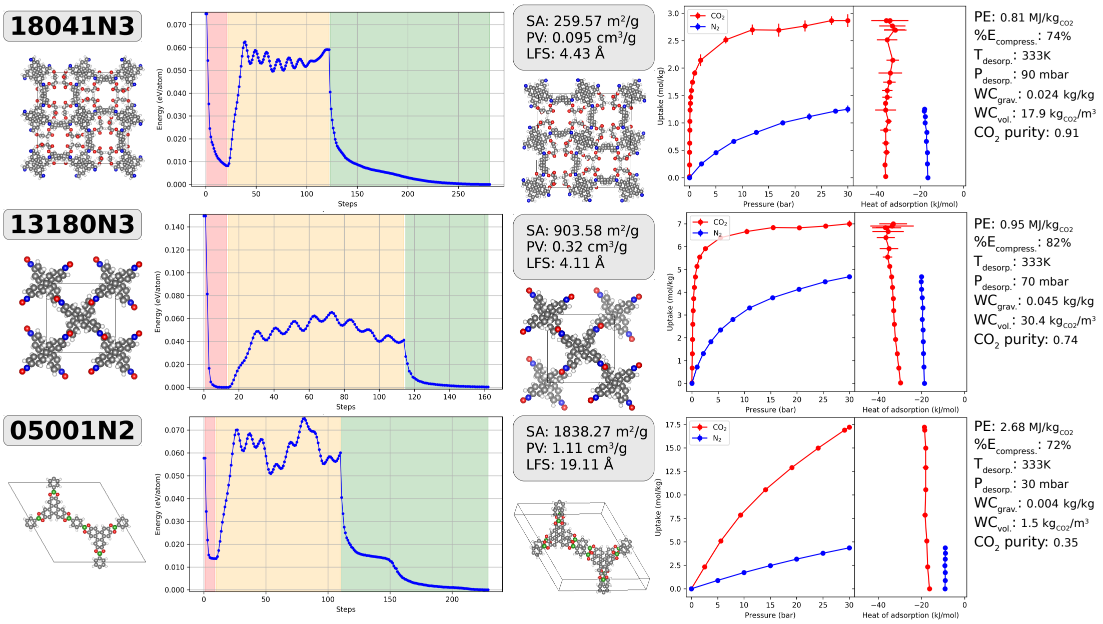


Figure 4.12 – The results obtained from the full workflow are sketched for three COFs: 18041N3, 13180N3 and 05001N2. From the left to the right we can find the initial geometry, the energy profile during the three-stages optimization, the final geometry and its main geometric properties, the CO_2 and N_2 isotherms and heat of adsorption, and finally the results from the process modelling. "SA", "PV" and "LFS" labels indicate respectively the nitrogen accessible surface area, pore volume, and the largest free sphere's diameter. "PE" and "WC" stand for the minimal parasitic energy and the CO_2 working capacity in the adsorbent.

4.5 Conclusions

We presented a systematic way to optimize the structures of Covalent Organic Frameworks (COFs) using density functional theory. In addition, we have assessed their performance for CO_2 capture as estimated from classical simulations. The optimization revealed some substantial changes, mainly in the layers rearrangement of these materials, and finally provided a set of framework that have been obtained from a consistent and reproducible protocol. The Automated Interactive Infrastructure and Database for Computational Science (AiiDA) has been used to achieve the result, and this set of 308 CURATED DFT-optimized COFs is made available, with optimized structure and high quality partial charges, for further computational investigations. The acronym stands for *Clean, Uniform and Refined with Automatic Tracking from Experimental Database*. We plan to extend our CURATED set of COFs periodically with the most recently reported frameworks. It can serve as a reference for consulting and for molecular simulations.

We show that the computed adsorption isotherms for CO_2 are in fair agreement with experiments, allowing for a reliable ranking of these materials for carbon capture. The modular design of our workchain allows for future testing of different force field and settings, in order to improve the overall match of simulation with experiments: we believe that this will be possible with the parallel effort of building a consistent repository to collect experimental adsorption measurement.

We plan to extend the same concept to other classes of materials, e.g., Metal Organic Framework (MOFs) and zeolites, aiming for an extensive database for adsorption properties, useful not only for comparison but also for training machine learning models that can later be used to pre-screen new materials from quickly computed geometric properties. We will need to face more complex challenges, e.g., dealing with the DFT modelling of transition metals, or tuning the standard force fields to better describe the interaction with open metal sites in MOF. In this perspective, AiiDA can serve as a common language for the whole scientific community to collaborate, systematically improving and re-combining workchains that need to model more and more complex systems.

4.6 Methods

4.6.1 DFT calculations

DFT calculations were performed using the PBE exchange-correlation functional [178] with DFT-D3(BJ) dispersion corrections [237]. The Quickstep code of the CP2K package was used,[238] employing GTH pseudopotentials [239], DZVP-MOLOPT-SR contracted Gaussian basis sets and an auxiliary plane wave basis set. The plane waves cutoff is set to 600 Ry cutoff, and these are mapped on a 4-levels multi-grid, with relative cutoff of 50 Ry and progression factor of 3. The orbital transformation (OT) method was used, and if a band gap < 0.1 eV is found, the calculation was re-run from scratch using Broyden diagonalization and Thomas-Fermi smearing at 300K.

For geometry optimization, we considered the atomic positions to be converged once the maximum force on the atoms dropped below 1.0 mHartree/bohr (as well as a root-mean squared value below 0.7). The threshold for the pressure is set to 100 bars for the cell optimization. The first stage (preliminary cell optimization) the BFGS minimizer is used. For the second stage (AIMD) an NPT_F[240] simulation is performed at 400K and 1 bar, using the CSVR thermostat[241] and the barostat from Martyna et al.[242] As for the third stage (final cell optimization), the Limited-memory BFGS (L-BFGS) minimizer is employed. Further details on the choice of the convergence threshold and minimizer can be found in the Supporting Information.

The workchain is kept efficient by always restarting the wave functions from the calculation of the previous stage, and using the Always Correct Predictor-Corrector (ASPC) [243] to have a more accurate first guess of coefficients of the wave function at every cell optimization or MD step.

4.6.2 Partial charges

Density derived electrostatic charges (DDEC)[136] are evaluated, using the software Chargemol,[244] and feeding the electron density of the optimized structures as computed from CP2K. The version 6 of the protocol, i.e., DDEC-6,[245] is used.

Qeq charges are computed with the same protocol we tested in our previous work:[246, 247] the periodic-Qeq (PQeq) calculator egulp[248] was used with GMP parameters. Out of 310 neutral COFs, only one did not converge the Qeq calculation (17131N2).

4.6.3 Geometry-based descriptors

Geometric properties are evaluated using the software Zeo++ (version 0.3).[249] The software's default definition of the atomic radii was used. The accessibility of the internal pore volume and surface was assessed using a spherical probe of 1.86 Å, i.e., the conventionally used kinetic radius for nitrogen.

When computing the probe-occupiable accessible pore volume[195] and the blocking spheres[250] to be used for the molecular simulation, we adopted a different set of radii, consistently with the force field of the simulation. The frameworks' atomic diameter were set equal to Lennard-Jones sigma in UFF, and we considered spherical probes with diameters of 3.05 Å and 3.31 Å for CO_2 (oxygen's sigma in TraPPE) and N_2 (nitrogen's sigma in TraPPE), respectively.

4.6.4 Parasitic energy evaluation

CO_2 and N_2 isotherm are computed using the Raspa package,[189] at 300K, within the range of 0.001-30 bars. The TraPPE force field is employed to model the gasses[190] and the dispersion interaction with the framework are computed using Lorentz-Berthelot mixing rules, employing the UFF parametrization[127] for the atoms of the COFs. The pressure points for the sampling are selected using a novel protocol that we describe in the Supporting Information. The uptakes are computed from the lowest pressure, running GCMC for 1,000 cycles for initialization and 10,000 for production, and restarting from the final configuration, for the next pressure calculation. The heat of adsorption is computed during GCMC using the particle fluctuation method.[251] Blocking spheres are considered, to prevent the insertion of gas molecules in non-accessible pores. COFs with null probe-occupiable accessible pore volume are considered non-porous and for these materials the isotherms were not computed.

Finally, we used the in-house code from the work of Huck et al.,[228, 252] to compute the optimal parameters for the temperature-pressure swing process, the parasitic energy, the working capacity and the final purity of the CO_2-rich mixture.

Conclusions and future work

Considering from a broader perspective the work presented in this thesis, some general conclusions can be drafted. In the field of molecular adsorption, two key properties are essential: the number of adsorbate molecules that the framework can host in its pores (i.e., the saturation loading) and the framework-adsorbate interactions. As one may expect for the first property, the number of sites is very much dependent on the pore volume of the material, which needs to be accurately assessed and compared with the experimental characterization that uses nitrogen or argon. The method we proposed, the calculation of the "accessible probe-occupiable pore volume", needs to be compared to experiments as the first check that allows us to verify if the experimental crystal was correctly desolvated and retained its porosity. In this case, the unit cell model is a representative model of the system and one can expect a good match with experiments at saturation, for different type of molecules, and proceed with further simulations. The probe-occupiable pore volume calculation is a cheaper approximation of a Grand Canonical Monte Carlo (GCMC) simulation, that can provide a similar information on the saturation loading and the extra information of the accessibility, at a fraction of the computational time.

The second property, the molecule-framework interaction, is much more influential is particularly important for assessing loading at lower pressures and the selectivity. These interactions are shown to be a more critical aspect to model: in the case of metal organic frameworks (MOFs) the main problem is represented by the interaction with the "open metal sites" (also known as "unsaturated metal centers"). In the first chapter we showed that these are not sufficiently described when using force field methods (i.e., van der Waals and Coulomb contributions only) but they can also suffer from poor modelling when using density functional theory (DFT) methods. The final tuning of the force field was shown to be mathematically easy and effective, but still the complexity of the problem makes it difficult to define a universal tuning protocol for the different MOF-molecule combinations.

Moreover, even when neglecting the problem of open metal sites, and considering only the van der Waals and Coulomb contributes for a first evaluation, we highlighted the key role of partial charges in the modelling of the interaction with polar molecules. For high-throughput calculations, in order to have a preliminary ranking of the adsorption strength at low pressures, one might be tempted to use cheaper empirical methods to derive these partial charge. These run faster as they do not require a DFT calculation and a partitioning of the energy density and/or a fitting of the produced electrostatic potential. However, only by comparing these cheaper "Qeq" methods for a large number of structures, one can have an idea of the accuracy that can be achieved. Care should also be used to select the proper parameters and to report them diligently in order to ensure the reproducibility of each calculation.

A general conclusion on the second and the third chapters concerns the importance of comparing different approximated protocols, for thousands of materials of the same class (MOFs in this case) to understand their validity and the artifacts that they can lead to. Ideally, one would like to compare directly the results from a calculation with the experimental value, but this is rarely possible except for small subsets of materials that have been measured and reported consistently. Therefore, we showed that the protocol of comparison used for the pore volume and the partial charge provide lots of additional insights on the different methods that are benchmarked against each other.

Finally, the work described in the first three chapters informed us on the best protocols to effectively exploit this synergy of classical and quantum methods, searching for promising microporous materials for carbon capture and storage. The selection of covalent organic frameworks (COFs) was motivated by the relatively small number of reported structures and the absence of open metal sites that we are still unable to systematically model with a sufficient accuracy. The delicate balance of accuracy/cost that we chose was shown to be in fair agreement with experimental isotherms and provided a final ranking of these materials for a carbon capture application. Moreover, we could get insights on other properties of these materials, e.g., their stacking, and test the use of DFT methods for systematic geometry optimization and partial charges evaluation. The lesson we learned from building a full workflow was on the importance of the initial input parameters to feed at the different stages of the workflow, and on the importance of forecasting all the possible sources of failure. The final gain that we got from this effort was to have a final protocol that can serve as a reference in the field for further improvements in the directions of both accuracy and computational efficiency. Moreover, the use of AiiDA (Automated Interactive Infrastructure and Database for Computational Science) allowed us to track the path between these calculations and provide a common language for other researchers to understand our routines, that can now be re-used, improved, or branched with new calculations.

Future work

The first direction we are proceeding in aims to extend our workflows to new classes of materials and applications. We are now looking into MOFs that pose more severe challenges in DFT modeling due to the complexity of their chemistry. Other classes that will be interesting to investigate are zeolites and carbon structures (e.g. schwarzites and polymers). As for the the plethora of gas-applications that can be investigated and branched using our workflows, several recent works showed the power of high-throughput simulations to identify best candidates for, e.g., the separation of noble gases or warfare agents, or the storage of gasses like methane, hydrogen, or oxygen. Still, what we are aiming for is a systematic and well documented protocol for these simulations that would allow for a direct comparison of any new material with the ones already present in the database; this would lead to reliable identification of the more promising application for a given new crystal.

The next step after building such a database of materials and properties is the application of machine learning techniques to evaluate these properties in a fraction of the time needed for the full simulations. This is possible only after the construction of a consistent and extensible database. Our current investigations show that the crucial aspects are not only the training, but also the transferability of a trained model from one set of structures to another. Indeed, the necessity of a large number of structures for training these models requires extending the training to hypothetical frameworks, i.e., the structures that can be generated *in silico* by "taking inspiration" from previously-synthesized structures. The number of published hypothetical structures is now much larger than the number of the experimental materials. Their use can nonetheless easily induce some bias in the training, and a seemingly good match with the validation test is not necessarily related to a good generalization on the structures that can be effectively synthesized. This is a problem that definitively needs more understanding in order to reliably exploit the power of machine learning.

Another challenging follow-up is the systematic investigation of saturated phases inside the micropores. These systems are particularly relevant for the separation of liquid-mixtures and for heterogeneous catalysis. However, their thermodynamics and kinetics are particularly complex, and advanced tailor-made methods need to be developed to efficiently compute the relevant properties for an application-based ranking. This makes it even harder to envision a systematic comparison of thousands of materials instead of the investigation of a few ones. The challenge is again to find, at the actual state of computational power, a possible compromise between the speed of the calculation and the need to capture the essential physics to adequately describe the performance of these materials.

Bibliography

[1] Hiroyasu Furukawa, Kyle E Cordova, Michael O'Keeffe, and Omar M Yaghi. The chemistry and applications of metal-organic frameworks. *Science*, 341(6149):1230444, 2013.

[2] Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.*, 4(2):83–89, 2012.

[3] Russell E Morris and Paul S Wheatley. Gas storage in nanoporous materials. *Angew. Chem., Int. Ed.*, 47(27):4966–4981, 2008.

[4] Jian-Rong Li, Ryan J Kuppler, and Hong-Cai Zhou. Selective gas adsorption and separation in metal–organic frameworks. *Chem. Soc. Rev.*, 38(5):1477–1504, 2009.

[5] Teng Zhang and Wenbin Lin. Metal–organic frameworks for artificial photosynthesis and photocatalysis. *Chem. Soc. Rev.*, 43(16):5982–5993, 2014.

[6] Lauren E Kreno, Kirsty Leong, Omar K Farha, Mark Allendorf, Richard P Van Duyne, and Joseph T Hupp. Metal–organic framework materials as chemical sensors. *Chem. Rev.*, 112(2):1105–1125, 2011.

[7] Jiewei Liu, Lianfen Chen, Hao Cui, Jianyong Zhang, Li Zhang, and Cheng-Yong Su. Applications of metal–organic frameworks in heterogeneous supramolecular catalysis. *Chem. Soc. Rev.*, 43(16):6011–6061, 2014.

[8] Roberta Poloni, Kyuho Lee, Robert F Berger, Berend Smit, and Jeffrey B Neaton. Understanding trends in CO_2 adsorption in metal–organic frameworks with open-metal sites. *J. Phys. Chem. Lett.*, 5(5):861–865, 2014.

[9] Jack D Evans, Guillaume Fraux, Romain Gaillac, Daniela Kohen, Fabien Trousselet, Jean-Mathieu Vanson, and Francois-Xavier Coudert. Computational chemistry methods for nanoporous materials. *Chem. Mater.*, 29(1):199–212, 2016.

[10] Samuel O Odoh, Christopher J Cramer, Donald G Truhlar, and Laura Gagliardi. Quantum-chemical characterization of the properties and reactivities of metal–organic frameworks. *Chem. Rev.*, 115(12):6051–6111, 2015.

[11] Stefan Grimme, Andreas Hansen, Jan Gerit Brandenburg, and Christoph Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.*, 116(9):5105–5154, 2016.

[12] Stephen S-Y Chui, Samuel M-F Lo, Jonathan PH Charmant, A Guy Orpen, and Ian D Williams. A chemically functionalizable nanoporous material [cu3 (tma) 2 (h2o) 3] n. *Science*, 283(5405):1148–1150, 1999.

[13] Jarad A Mason, Mike Veenstra, and Jeffrey R Long. Evaluating metal–organic frameworks for natural gas storage. *Chem. Sci.*, 5(1):32–51, 2014.

[14] Qing Min Wang, Dongmin Shen, Martin Bülow, Miu Ling Lau, Shuguang Deng, Frank R Fitch, Norberto O Lemcoff, and Jessica Semanscin. Metallo-organic molecular sieve for gas separation and purification. *Microporous Mesoporous Mater.*, 55(2):217–230, 2002.

[15] Lomig Hamon, Elsa Jolimaître, and Gerhard D Pirngruber. CO_2 and CH_4 separation by adsorption using cu-btc metal- organic framework. *Ind. Eng. Chem. Res.*, 49(16):7497–7503, 2010.

[16] Lei Ge, Wei Zhou, Victor Rudolph, and Zhonghua Zhu. Mixed matrix membranes incorporated with size-reduced cu-btc for improved gas separation. *Journal of Materials Chemistry A*, 1(21):6350–6358, 2013.

[17] Lik H Wee, Nikki Janssens, Sneha R Bajpe, Christine EA Kirschhock, and Johan A Martens. Heteropolyacid encapsulated in cu 3 (btc) 2 nanocrystals: an effective esterification catalyst. *Catal. Today*, 171(1):275–280, 2011.

[18] Elena Pérez-Mayoral and Jiří Čejka. [cu3 (btc) 2]: a metal–organic framework catalyst for the friedländer reaction. *ChemCatChem*, 3(1):157–159, 2011.

[19] Jing-Yun Ye and Chang-Jun Liu. Cu 3 (btc) 2: Co oxidation over mof based catalysts. *Chem. Comm.*, 47(7):2167–2169, 2011.

[20] Hui Wu, Jason M Simmons, Gadipelli Srinivas, Wei Zhou, and Taner Yildirim. Adsorption sites and binding nature of CO_2 in prototypical metal- organic frameworks: A combined neutron diffraction and first-principles study. *J. Phys. Chem. Lett.*, 1(13):1946–1951, 2010.

[21] Lukáš Grajciar, Petr Nachtigall, Ota Bludský, and Miroslav Rubeš. Accurate ab initio description of adsorption on coordinatively unsaturated cu2+ and fe3+ sites in mofs. *J. Chem. Theory Comput.*, 11(1):230–238, 2014.

[22] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.

[23] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys*, 132(15):154104, 2010.

[24] Max Dion, Henrik Rydberg, Elsebeth Schröder, David C Langreth, and Bengt I Lundqvist. Van der waals density functional for general geometries. *Phys. Rev. Lett.*, 92(24):246401, 2004.

[25] Kyuho Lee, Éamonn D Murray, Lingzhu Kong, Bengt I Lundqvist, and David C Langreth. Higher-accuracy van der waals density functional. *Phys. Rev. B*, 82(8):081101, 2010.

[26] Kyuho Lee, Joshua D Howe, Li-Chiang Lin, Berend Smit, and Jeffrey B Neaton. Small-molecule adsorption in open-site metal–organic frameworks: a systematic density functional theory study for rational design. *Chem. Mater.*, 27(3):668–678, 2015.

[27] Roberta Poloni, Berend Smit, and Jeffrey B Neaton. CO_2 capture by metal–organic frameworks with van der waals density functionals. *J. Phys. Chem. A*, 116(20):4957–4964, 2012.

[28] Roberta Poloni, Berend Smit, and Jeffrey B Neaton. Ligand-assisted enhancement of CO_2 capture in metal–organic frameworks. *J. Am. Chem. Soc.*, 134(15):6714–6719, 2012.

[29] Li-Chiang Lin, Kyuho Lee, Laura Gagliardi, Jeffrey B Neaton, and Berend Smit. Force-field development from electronic structure calculations with periodic boundary conditions: applications to gaseous adsorption and transport in metal–organic frameworks. *J. Chem. Theory Comput.*, 10(4):1477–1488, 2014.

[30] Rocio Mercado, Bess Vlaisavljevich, Li-Chiang Lin, Kyuho Lee, Yongjin Lee, Jarad A Mason, Dianne J Xiao, Miguel I Gonzalez, Matthew T Kapelewski, Jeffrey B Neaton, et al. Force field development from periodic density functional theory calculations for gas separation applications using metal–organic frameworks. *J. Phys. Chem. C*, 120(23):12590–12604, 2016.

[31] Lukas Grajciar, Andrew D Wiersum, Philip L Llewellyn, Jong-San Chang, and Petr Nachtigall. Understanding CO_2 adsorption in cubtc mof: comparing combined dft–ab initio calculations with microcalorimetry experiments. *J. Phys. Chem. C*, 115(36):17925–17933, 2011.

[32] Davide Tiana, Christopher H Hendon, and Aron Walsh. Ligand design for long-range magnetic order in metal–organic frameworks. *Chem. Comm.*, 50(90):13990–13993, 2014.

[33] Lukas Grajciar, Ota Bludsky, and Petr Nachtigall. Water adsorption on coordinatively unsaturated sites in cubtc mof. *J. Phys. Chem. Lett.*, 1(23):3354–3359, 2010.

[34] Konstantinos D Vogiatzis, Andreas Mavrandonakis, Wim Klopper, and George E Froudakis. Ab initio study of the interactions between CO_2 and n-containing organic heterocycles. *ChemPhysChem*, 10(2):374–383, 2009.

[35] Konstantinos D Vogiatzis, Wim Klopper, Andreas Mavrandonakis, and Karin Fink. Magnetic properties of paddlewheels and trinuclear clusters with exposed metal sites. *ChemPhysChem*, 12(17):3307–3319, 2011.

[36] Peter G Boyd, Seyed Mohamad Moosavi, Matthew Witman, and Berend Smit. On the force field prediction of materials properties in metal organic frameworks. *J. Phys. Chem. Lett.*, 8(2):357–363, 2017.

[37] Colin R Groom, Ian J Bruno, Matthew P Lightfoot, and Suzanna C Ward. The cambridge structural database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 72(2):171–179, 2016.

[38] Yongchul G Chung, Jeffrey Camp, Maciej Haranczyk, Benjamin J Sikora, Wojciech Bury, Vaiva Krungleviciute, Taner Yildirim, Omar K Farha, David S Sholl, and Randall Q Snurr. Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.*, 26(21):6185–6192, 2014.

[39] John P Perdew, Adrienn Ruzsinszky, Gábor I Csonka, Oleg A Vydrov, Gustavo E Scuseria, Lucian A Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100(13):136406, 2008.

[40] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, 2009.

[41] Georg Kresse and D Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758, 1999.

[42] Peter E Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953, 1994.

[43] Yan Zhao and Donald G Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120(1-3):215–241, 2008.

[44] Peter J Knowles, Jamie S Andrews, Roger D Amos, Nicholas C Handy, and John A Pople. Restricted møller-plesset theory for open-shell molecules. *Chem. Phys. Lett.*, 186(2-3):130–136, 1991.

[45] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, GA Petersson, et al. Gaussian 09, revision d. 01, 2009.

[46] Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys*, 90(2):1007–1023, 1989.

[47] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys*, 96(9):6796–6806, 1992.

[48] Ernest R Davidson. Comment on dunning's correlation-consistent basis sets. *Chem. Phys. Lett.*, 260(3-4):514–518, 1996.

[49] Nikolai B Balabanov and Kirk A Peterson. Systematically convergent basis sets for transition metals. i. all-electron correlation consistent basis sets for the 3 d elements sc–zn. *J. Chem. Phys.*, 123(6):064107, 2005.

[50] S F Boys and Fiorenza De Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Mol. Phys.*, 19(4):553–566, 1970.

[51] Björn O Roos, Peter R Taylor, and Per EM Si. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chem. Phys.*, 48(2):157–173, 1980.

[52] Kerstin Andersson, Per Aake Malmqvist, Björn O Roos, Andrzej J Sadlej, and Krzysztof Wolinski. Second-order perturbation theory with a casscf reference function. *Journal of Physical Chemistry*, 94(14):5483–5488, 1990.

[53] Francesco Aquilante, Jochen Autschbach, Rebecca K Carlson, Liviu F Chibotaru, Mickaël G Delcey, Luca De Vico, Nicolas Ferré, Luis Manuel Frutos, Laura Gagliardi, Marco Garavelli, et al. Molcas 8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.*, 37(5):506–541, 2016.

[54] Per-Olof Widmark, Per-Åke Malmqvist, and Björn O Roos. Density matrix averaged atomic natural orbital (ano) basis sets for correlated molecular wave functions. *Theor. Chim. Acta*, 77(5):291–306, 1990.

[55] Markus Reiher and Alexander Wolf. Exact decoupling of the dirac hamiltonian. ii. the generalized douglas–kroll–hess transformation up to arbitrary order. *J. Chem. Phys*, 121(22):10945–10956, 2004.

[56] Francesco Aquilante, Laura Gagliardi, Thomas Bondo Pedersen, and Roland Lindh. Atomic cholesky decompositions: a route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency. *J. Chem. Phys*, 130(15):154107, 2009.

[57] David Dubbeldam, Sofía Calero, Donald E Ellis, and Randall Q Snurr. Raspa: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.*, 42(2):81–101, 2016.

[58] Jeffrey J Potoff and J Ilja Siepmann. Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.*, 47(7):1676–1682, 2001.

[59] Anthony K Rappé, Carla J Casewit, KS Colwell, WA Goddard Iii, and WM Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114(25):10024–10035, 1992.

[60] Stephen L Mayo, Barry D Olafson, and William A Goddard. Dreiding: a generic force field for mol. simul.s. *Journal of Physical Chemistry*, 94(26):8897–8909, 1990.

[61] Carlos Campaná, Bastien Mussard, and Tom K Woo. Electrostatic potential derived atomic charges for periodic systems using a modified error functional. *J. Chem. Theory Comput.*, 5(10):2866–2878, 2009.

[62] Alberto Otero-de-la Roza, Erin R Johnson, and Víctor Luaña. Critic2: A program for real-space analysis of quantum chemical interactions in solids. *Comp. Phys. Comm.*, 185(3):1007–1018, 2014.

[63] Michael Fischer, José RB Gomes, and Miguel Jorge. Computational approaches to study adsorption in mofs with unsaturated metal sites. *Mol. Simul.*, 40(7-9):537–556, 2014.

[64] AI Liechtenstein, VI Anisimov, and J Zaanen. Density-functional theory and strong interactions: Orbital ordering in mott-hubbard insulators. *Phys. Rev. B*, 52(8):R5467, 1995.

[65] Gregory W Mann, Kyuho Lee, Matteo Cococcioni, Berend Smit, and Jeffrey B Neaton. First-principles hubbard u approach for small molecule binding in metal-organic frameworks. *J. Chem. Phys*, 144(17):174104, 2016.

[66] Lei Wang, Thomas Maxisch, and Gerbrand Ceder. Oxidation energies of transition metal oxides within the gga+ u framework. *Phys. Rev. B*, 73(19):195107, 2006.

[67] T Thonhauser, S Zuluaga, CA Arter, K Berland, E Schröder, and P Hyldgaard. Spin signature of nonlocal correlation binding in metal-organic frameworks. *Phys. Rev. Lett.*, 115(13):136402, 2015.

[68] Ikutaro Hamada. van der waals density functional made accurate. *Phys. Rev. B*, 89(12):121103, 2014.

[69] Jiří Klimeš and Angelos Michaelides. Perspective: Advances and challenges in treating van der waals dispersion forces in density functional theory. *J. Chem. Phys*, 137(12):120901, 2012.

[70] Jan Řezáč and Pavel Hobza. Benchmark calculations of interaction energies in noncovalent complexes and their applications. *Chem. Rev.*, 116(9):5038–5071, 2016.

[71] Banglin Chen, M Eddaoudi, TM Reineke, JW Kampf, M O'keeffe, and OM Yaghi. Cu2 (atc) 6h2o: Design of open metal sites in porous metal- organic crystals (atc: 1, 3, 5, 7-adamantane tetracarboxylate). *J. Am. Chem. Soc.*, 122(46):11559–11560, 2000.

[72] Rémi Maurice, Kanthen Sivalingam, Dmitry Ganyushin, Nathalie Guihéry, Coen De Graaf, and Frank Neese. Theoretical determination of the zero-field splitting in copper acetate monohydrate. *Inorg. Chem*, 50(13):6229–6236, 2011.

[73] Jianbo Hu, Jing Liu, Yang Liu, and Xiao Yang. Improving carbon dioxide storage capacity of metal organic frameworks by lithium alkoxide functionalization: a mol. simul. study. *J. Phys. Chem. C*, 120(19):10311–10319, 2016.

[74] Paolo Aprea, Domenico Caputo, Nicola Gargiulo, Fabio Iucolano, and Francesco Pepe. Modeling carbon dioxide adsorption on microporous substrates: Comparison between cu-btc metal- organic framework and 13x zeolitic molecular sieve. *J. Chem. Eng. Data*, 55(9):3655–3661, 2010.

[75] Pradip Chowdhury, Samuel Mekala, Frieder Dreisbach, and Sasidhar Gumma. Adsorption of co, CO_2 and CH_4 on cu-btc and mil-101 metal organic frameworks: Effect of open metal sites and adsorbate polarity. *Microporous Mesoporous Mater.*, 152:246–252, 2012.

[76] Rajamani Krishna. Adsorptive separation of co 2/ch 4/co gas mixtures at high pressures. *Microporous and Mesoporous Mater.*, 156:217–223, 2012.

[77] Baiyan Li, Zhijuan Zhang, Yi Li, Kexin Yao, Yihan Zhu, Zhiyong Deng, Fen Yang, Xiaojing Zhou, Guanghua Li, Haohan Wu, et al. Enhanced binding affinity, remarkable selectivity, and high capacity of CO_2 by dual functionalization of a rht-type metal–organic framework. *Angew. Chem., Int. Ed.*, 51(6):1412–1415, 2012.

[78] I Spanopoulos, I Bratsos, C Tampaxis, D Vourloumis, E Klontzas, GE Froudakis, G Charalambopoulou, TA Steriotis, and PN Trikalitis. Exceptional gravimetric and volumetric CO_2 uptake in a palladated nbo-type mof utilizing cooperative acidic and basic, metal–CO_2 interactions. *Chem. Comm.*, 52(69):10559–10562, 2016.

[79] Christopher E Wilmer, Ki Chul Kim, and Randall Q Snurr. An extended charge equilibration method. *J. Phys. Chem. Lett.*, 3(17):2506–2511, 2012.

[80] Avelino Corma. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.*, 97(6):2373–2420, 1997.

[81] Patricia Horcajada, Tamim Chalati, Christian Serre, Brigitte Gillet, Catherine Sebrie, Tarek Baati, Jarrod F Eubank, Daniela Heurtaux, Pascal Clayette, Christine Kreuz, et al. Porous metal-organic-framework nanoscale carriers as a potential platform for drug delivery and imaging. *Nat. Mat.*, 9(2):172–178, 2010.

[82] Jeffrey R Long and Omar M Yaghi. The pervasive chemistry of metal–organic frameworks. *Chem. Soc. Rev.*, 38(5):1213–1214, 2009.

[83] Xiao Feng, Xuesong Ding, and Donglin Jiang. Covalent organic frameworks. *Chem. Soc. Rev.*, 41(18):6010–6022, 2012.

[84] Kyo Sung Park, Zheng Ni, Adrien P Côté, Jae Yong Choi, Rudan Huang, Fernando J Uribe-Romo, Hee K Chae, Michael O'Keeffe, and Omar M Yaghi. Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. *Proc. Natl. Acad. Sci. U. S. A.*, 103(27):10186–10191, 2006.

[85] Weigang Lu, Daqiang Yuan, Dan Zhao, Christine Inge Schilling, Oliver Plietzsch, Thierry Muller, Stefan Brase, Johannes Guenther, Janet Blumel, Rajamani Krishna, et al. Porous

polymer networks: synthesis, porosity, and applications in gas storage/separation. *Chem. Mater.*, 22(21):5964–5972, 2010.

[86] Lev Sarkisov and Alex Harrison. Computational structure characterisation tools in application to ordered and disordered porous materials. *Mol. Simul.*, 37(15):1248–1257, 2011.

[87] Thomas F Willems, Chris H Rycroft, Michaeel Kazi, Juan C Meza, and Maciej Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.*, 149(1):134–141, 2012.

[88] Kenneth Sing. The use of nitrogen adsorption for the characterisation of porous materials. *Colloids Surf., A*, 187:3–9, 2001.

[89] AL Spek. Single-crystal structure validation with the program platon. *J. Appl. Crystallogr.*, 36(1):7–13, 2003.

[90] AL Myers and PA Monson. Adsorption in porous materials at high pressure: theory and experiment. *Langmuir*, 18(26):10261–10273, 2002.

[91] Michael L Connolly. Computation of molecular volume. *J. Am. Chem. Soc.*, 107(5):1118–1124, 1985.

[92] L Gurvich. *J. Phys. Chem. Soc. Russ.*, 47:49–56, 1915.

[93] Youn-Sang Bae, A Ozgur Yazaydın, and Randall Q Snurr. Evaluation of the bet method for determining surface areas of mofs and zeolites that contain ultra-micropores. *Langmuir*, 26(8):5475–5483, 2010.

[94] SC Van der Marck. Network approach to void percolation in a pack of unequal spheres. *Phys. Rev. Lett.*, 77(9):1785, 1996.

[95] Joseph Hoshen and Raoul Kopelman. Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm. *Phys. Rev. B*, 14(8):3438, 1976.

[96] Richard L Martin, Prabhat, David D Donofrio, James A Sethian, and Maciej Haranczyk. Accelerating analysis of void space in porous materials on multicore and gpu platforms. *Int. J. High Perform. Comput. Appl.*, 26(4):347–357, 2012.

[97] Jihan Kim, Richard L Martin, Oliver Rübel, Maciej Haranczyk, and Berend Smit. High-throughput characterization of porous materials using graphics processing units. *J. Chem. Theory Comput.*, 8(5):1684–1693, 2012.

[98] AW Thornton, KE Jelfs, K Konstas, CM Doherty, AJ Hill, AK Cheetham, and TD Bennett. Porosity in metal–organic framework glasses. *Chem. Comm.*, 52(19):3750–3753, 2016.

[99] T Chokbunpiam, S Fritzsche, C Chmelik, J Caro, W Janke, and S Hannongbua. Gate opening, diffusion, and adsorption of co2 and n2 mixtures in zif-8. *J. Phys. Chem. C*, 120(41):23458–23468, 2016.

[100] D Dubbeldam, S Calero, D Ellis, and R Snurr. Raspa 2.0: Molecular software package for adsorption and diffusion in (flexible) nanoporous materials. *Mol. Simul.*, 42(2):81–101, 2016.

[101] Joseph Hirschfelder, R Byron Bird, and Charles F Curtiss. Molecular theory of gases and liquids. 1964.

[102] Simone Cavenati, Carlos A Grande, Alirio E Rodrigues, Christoph Kiener, and Ulrich Müller. Metal organic framework adsorbent for biogas upgrading. *Ind. Eng. Chem. Res.*, 47(16):6333–6335, 2008.

[103] Andrew D Wiersum, Jong-San Chang, Christian Serre, and Philip L Llewellyn. An adsorbent performance indicator as a first step evaluation of novel sorbents for gas separations: application to metal–organic frameworks. *Langmuir*, 29(10):3301–3309, 2013.

[104] J Moellmer, A Moeller, F Dreisbach, R Glaeser, and R Staudt. High pressure adsorption of hydrogen, nitrogen, carbon dioxide and methane on the metal–organic framework hkust-1. *Microporous Mesoporous Mater.*, 138(1):140–148, 2011.

[105] Irena Senkovska and Stefan Kaskel. High pressure methane adsorption in the metal-organic frameworks cu 3 (btc) 2, zn 2 (bdc) 2 dabco, and cr 3 f (h 2 o) 2 o (bdc) 3. *Microporous Mesoporous Mater.*, 112(1):108–115, 2008.

[106] Lev Sarkisov and Jihan Kim. Computational structure characterization tools for the era of material informatics. *Chem. Eng. Sci.*, 121:322–330, 2015.

[107] Orhan Talu and Alan L Myers. Molecular simulation of adsorption: Gibbs dividing surface and comparison with experiment. *AIChE J.*, 47(5):1160–1168, 2001.

[108] Paul A Wright, Martin J Maple, Alexandra MZ Slawin, Véronique Patinec, R Alan Aitken, Simon Welsh, and Paul A Cox. Cation-directed syntheses of novel zeolite-like metalloaluminophosphates sta-6 and sta-7 in the presence of azamacrocycle templates. *Dalton Trans.*, (8):1243–1248, 2000.

[109] Johanna Heine, Małgorzata Hołyńska, Marco Reuter, Benedikt Haas, Sangam Chatterjee, Martin Koch, Katharina I Gries, Kerstin Volz, and Stefanie Dehnen. In (sar) 3 as a building block for 3d and helical coordination polymers. *Crystal Growth & Design*, 13(3):1252–1259, 2013.

[110] Thazhe Kootteri Prasad, Dae Ho Hong, and Myunghyun Paik Suh. High gas sorption and metal-ion exchange of microporous metal–organic frameworks with incorporated imide groups. *Chem. Eur. J.*, 16(47):14043–14050, 2010.

[111] Yabing He, Hiroyasu Furukawa, Chuande Wu, Michael O'Keeffe, Rajamani Krishna, and Banglin Chen. Low-energy regeneration and high productivity in a lanthanide–hexacarboxylate framework for high-pressure co 2–ch 4–h 2 separation. *Chem. Comm.*, 49(60):6773–6775, 2013.

[112] Ronny Grünker, Irena Senkovska, Ralf Biedermann, Nicole Klein, Martin R Lohe, Philipp Müller, and Stefan Kaskel. A highly porous flexible metal–organic framework with corundum topology. *Chem. Comm.*, 47(1):490–492, 2011.

[113] Yabing He, Zhangjing Zhang, Shengchang Xiang, Hui Wu, Frank R Fronczek, Wei Zhou, Rajamani Krishna, Michael O'Keeffe, and Banglin Chen. High separation capacity and selectivity of c2 hydrocarbons over methane within a microporous metal–organic framework at room temperature. *Chem. Eur. J.*, 18(7):1901–1904, 2012.

[114] Dan Zhao, Daqiang Yuan, Andrey Yakovenko, and Hong-Cai Zhou. A nbo-type metal–organic framework derived from a polyyne-coupled di-isophthalate linker formed in situ. *Chem. Comm.*, 46(23):4196–4198, 2010.

[115] Zhiyong Guo, Hui Wu, Gadipelli Srinivas, Yaming Zhou, Shengchang Xiang, Zhenxia Chen, Yongtai Yang, Wei Zhou, Michael O'Keeffe, and Banglin Chen. A metal–organic framework with optimized open metal sites and pore spaces for high methane storage at room temperature. *Angew. Chem., Int. Ed.*, 50(14):3178–3181, 2011.

[116] Christopher E Wilmer, Omar K Farha, Taner Yildirim, Ibrahim Eryazici, Vaiva Krungleviciute, Amy A Sarjeant, Randall Q Snurr, and Joseph T Hupp. Gram-scale, high-yield synthesis of a robust metal–organic framework for storing methane and other gases. *Energy Environ. Sci.*, 6(4):1158–1163, 2013.

[117] Hye Jeong Park, Young Eun Cheon, and Myunghyun Paik Suh. Post-synthetic reversible incorporation of organic linkers into porous metal–organic frameworks through single-crystal-to-single-crystal transformations and modification of gas-sorption properties. *Chem. Eur. J.*, 16(38):11662–11669, 2010.

[118] Daqiang Yuan, Dan Zhao, Daofeng Sun, and Hong-Cai Zhou. An isoreticular series of metal–organic frameworks with dendritic hexacarboxylate ligands and exceptionally high gas-uptake capacity. *Angew. Chem., Int. Ed.*, 49(31):5357–5361, 2010.

[119] Berend Smit and Theo L. M. Maesen. Towards a molecular understanding of shape selectivity. *Nature*, 451(7179):671–678, 2008.

[120] JeongYong Lee, Omar K. Farha, John Roberts, Karl A. Scheidt, SonBinh T. Nguyen, and Joseph T. Hupp. Metal organic framework materials as catalysts. *Chem. Soc. Rev.*, 38(5):1450, 2009.

[121] Jian-Rong Li, Ryan J. Kuppler, and Hong-Cai Zhou. Selective gas adsorption and separation in metal organic frameworks. *Chem. Soc. Rev.*, 38(5):1477, 2009.

[122] Shengqian Ma and Hong-Cai Zhou. Gas storage in porous metal organic frameworks for clean energy applications. *Chem. Commun.*, 46(1):44–53, 2010.

[123] Cory M. Simon, Jihan Kim, Diego A. Gomez-Gualdron, Jeffrey S. Camp, Yongchul G. Chung, Richard L. Martin, Rocio Mercado, Michael W. Deem, Dan Gunter, Maciej Haranczyk, David S. Sholl, Randall Q. Snurr, and Berend Smit. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.*, 8(4):1190–1199, 2015.

[124] Peyman Z. Moghadam, Aurelia Li, Seth B. Wiggin, Andi Tao, Andrew G. P. Maloney, Peter A. Wood, Suzanna C. Ward, and David Fairen-Jimenez. Development of a Cambridge Structural Database Subset: A Collection of Metal Organic Frameworks for Past, Present, and Future. *Chem. Mater.*, 29(7):2618–2625, apr 2017.

[125] Peter G. Boyd, Yongjin Lee, and Berend Smit. Computational development of the nanoporous materials genome. *Nat. Rev. Mater.*, 2:17037, 2017.

[126] Christopher E. Wilmer, Omar K. Farha, Youn-Sang Bae, Joseph T. Hupp, and Randall Q. Snurr. Structure property relationships of porous materials for carbon dioxide separation and capture. *Energy Environ. Sci.*, 5(12):9849, 2012.

[127] A. K. Rappe and William a. Goddard. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.*, 114:10024–10035, 1992.

[128] Stephen L. Mayo, Barry D. Olafson, and William a. Goddard. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.*, 94(26):8897–8909, dec 1990.

[129] R. S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.*, 23(10):1833–1840, 1955.

[130] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta*, 44(2):129–138, 1977.

[131] Patrick Bultinck, Christian Van Alsenoy, Paul W. Ayers, and Ramon Carbó-Dorca. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.*, 126(14), 2007.

[132] Graeme Henkelman, Andri Arnaldsson, and Hannes Jónsson. A fast and robust algorithm for Bader decomposition of charge density. *Comput. Mater. Sci.*, 36(3):354–360, 2006.

[133] Christopher I. Bayly, Piotr Cieplak, Wendy D. Cornell, and Peter A. Kollman. A well behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[134] Curt M. Breneman and Kenneth B. Wiberg. Determining atom centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11(3):361–373, 1990.

[135] Carlos Campana, Bastien Mussard, and Tom K. Woo. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.*, 5(10):2866–2878, oct 2009.

[136] Thomas A. Manz and David S. Sholl. Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *J. Chem. Theory Comput.*, 6(8):2455–2468, 2010.

[137] Thomas A. Manz and David S. Sholl. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *J. Chem. Theory Comput.*, 8(8):2844–2867, 2012.

[138] Nidia Gabaldon Limas and Thomas A. Manz. Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv.*, 6(51):45727–45747, 2016.

[139] Anthony K. Rappe and William A. Goddard. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, 95(8):3358–3363, 1991.

[140] Emmanuel Haldoupis, Sankar Nair, and David S. Sholl. Finding MOFs for Highly Selective CO_2 /N_2 Adsorption Using Materials Screening Based on Efficient Assignment of Atomic Point Charges. *J. Am. Chem. Soc.*, 134(9):4313–4323, mar 2012.

[141] Yongchul G. Chung, Jeffrey Camp, Maciej Haranczyk, Benjamin J. Sikora, Wojciech Bury, Vaiva Krungleviciute, Taner Yildirim, Omar K. Farha, David S. Sholl, and Randall Q. Snurr. Computation-Ready, Experimental Metal Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.*, 26(21):6185–6192, nov 2014.

[142] Dalar Nazarian, Jeffrey S. Camp, and David S. Sholl. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal Organic Frameworks. *Chem. Mater.*, 28(3):785–793, feb 2016.

[143] Wilfried J. Mortier, Swapan K. Ghosh, and S. Shankar. Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.*, 108(15):4315–4320, 1986.

[144] Sunder Ramachandran, T G Lenz, W M Skiff, and A K Rappé. Toward an Understanding of Zeolite Y as a Cracking Catalyst with the Use of Periodic Charge Equilibration. *J. Phys. Chem.*, 100(14):5898–5907, 1996.

[145] Akifumi Oda and Ohgi Takahashi. Parameter determination for the charge equilibration method including third- and fourth-order terms applied to non-metallic compounds. *Chem. Phys. Lett.*, 495(1-3):155–159, 2010.

[146] Christopher E. Wilmer, Ki Chul Kim, and Randall Q. Snurr. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.*, 3(17):2506–2511, sep 2012.

[147] Brad A. Wells, Caspar De Bruin-Dickason, and Alan L. Chaffee. Charge Equilibration Based on Atomic Ionization in Metal Organic Frameworks. *J. Phys. Chem. C*, 119(1):456–466, jan 2015.

[148] Eugene S. Kadantsev, Peter G. Boyd, Thomas D. Daff, and Tom K. Woo. Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption. *J. Phys. Chem. Lett.*, 4(18):3056–3061, sep 2013.

[149] Geoffrey C. Martin-Noble, David Reilley, Luis M. Rivas, Matthew D. Smith, and Joshua Schrier. EQeq+C: An Empirical Bond-Order-Corrected Extended Charge Equilibration Method. *J. Chem. Theory Comput.*, 11(7):3364–3374, 2015.

[150] Razvan A. Nistor, Jeliazko G. Polihronov, Martin H. Müser, and Nicholas J. Mosey. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.*, 125(9):094108, sep 2006.

[151] Sean P. Collins and Tom K. Woo. Split-Charge Equilibration Parameters for Generating Rapid Partial Atomic Charges in Metal Organic Frameworks and Porous Polymer Networks for High-Throughput Screening. *J. Phys. Chem. C*, 121(1):903–910, jan 2017.

[152] R T Sanderson. An Interpretation of Bond Lengths and a Classification of Bonds. *Science*, 114(2973):670–672, dec 1951.

[153] R. T. Sanderson. *Chemical bonds and bond energy*. Academic Press: New York, 1976.

[154] Raymond P. Iczkowski and John L. Margrave. Electronegativity. *J. Am. Chem. Soc.*, 83(17):3547–3551, 1961.

[155] Robert G. Parr and Ralph G. Pearson. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.*, 105(26):7512–7516, 1983.

[156] Saber Naserifar, Daniel J. Brooks, William A. Goddard, and Vaclav Cvicek. Polarizable charge equilibration model for predicting accurate electrostatic interactions in molecules and solids. *J. Chem. Phys.*, 146(12), 2017.

[157] N. Rosen. Calculation of Interaction between Atoms with s-Electrons. *Phys. Rev.*, 38(2):255–276, jul 1931.

[158] Wilfried J. Mortier, Karin Van Genechten, and Johann Gasteiger. Electronegativity Equalization: Application and Parametrization. *J. Am. Chem. Soc.*, 107(4):829–835, 1985.

[159] Patrick Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J. P. Tollenaere. The electronegativity equalization method I: Parametrization and validation for atomic charge calculations. *J. Phys. Chem. A*, 106(34):7887–7894, 2002.

[160] Patrick Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, C. Van Alsenoy, and J. P. Tollenaere. The electronegativity equalization method II: Applicability of different atomic charge schemes. *J. Phys. Chem. A*, 106(34):7895–7901, 2002.

[161] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369(3):253–287, 1921.

[162] Akifumi Oda and Shuichi Hirono. Geometry-dependent atomic charge calculations using charge equilibration method with empirical two-center Coulombic terms. *J. Mol. Struct.: THEOCHEM*, 634(1-3):159–170, sep 2003.

[163] Archana DasGupta and Sigeru Huzinaga. New developments in CNDO molecular orbital theory. *Theor. Chim. Acta*, 35(4):329–340, 1974.

[164] Min Zhang and Rene Fournier. Self-Consistent Charge Equilibration Method and Its Application to Au 13 Na n (n = 1,10) Clusters. *J. Phys. Chem. A*, 113(13):3162–3170, apr 2009.

[165] Christopher E. Wilmer and Randall Q. Snurr. Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: Calculation of framework charges via charge equilibration. *Chem. Eng. J.*, 171(3):775–781, 2011.

[166] Charlotte E Moore. Ionization potentials and ionization limits derived from the analyses of optical spectra. Technical report, NATIONAL STANDARD REFERENCE DATA SYSTEM, 1970.

[167] T. Andersen, H. K. Haugen, and H. Hotop. Binding Energies in Atomic Negative Ions: III. *J. Phys. Chem. Ref. Data*, 28(6):1511–1533, 1999.

[168] Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.

[169] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371, 1993.

[170] T. Hashimoto, K. Hirao, and H. Tatewaki. Comment on Dunning's correlation-consistent basis sets. *Chem. Phys. Lett.*, 243(1-2):190–192, 1995.

[171] Riccardo Chelli, Piero Procacci, Roberto Righini, and Salvatore Califano. Electrical response in chemical potential equalization schemes. *J. Chem. Phys.*, 111(18):8569–8575, nov 1999.

[172] Ljiljana Dosen-Micovic, Dragoslav Jeremic, and Norman L. Allinger. Treatment of electrostatic effects within the molecular-mechanics method. 1. *J. Am. Chem. Soc.*, 105(7):1716–1722, apr 1983.

[173] G. Lee Warren, Joseph E. Davis, and Sandeep Patel. Origin and control of superlinear polarizability scaling in chemical potential equalization methods. *J. Chem. Phys.*, 128(14), 2008.

[174] T. Verstraelen, P. Bultinck, V. Van Speybroeck, P. W. Ayers, D. Van Neck, and M. Waroquier. The significance of parameters in charge equilibration models. *J. Chem. Theory Comput.*, 7(6):1750–1764, 2011.

[175] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.*, 138(7), 2013.

[176] Wei Li, Zizhen Rao, Yongchul G. Chung, and Song Li. The Role of Partial Atomic Charge Assignment Methods on the Computational Screening of Metal-Organic Frameworks for CO 2 Capture under Humid Conditions. *ChemistrySelect*, 2(29):9458–9465, oct 2017.

[177] Julian D. Gale. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc. Faraday Trans.*, 93(4):629–637, 1997.

[178] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.

[179] Zhiwei Qiao, Kang Zhang, and Jianwen Jiang. In silico screening of 4764 computation-ready, experimental metal organic frameworks for CO_2 separation. *J. Mater. Chem. A*, 4(6):2105–2114, 2016.

[180] Aleksandr V. Marenich, Steven V. Jerome, Christopher J. Cramer, and Donald G. Truhlar. Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.*, 8(2):527–541, 2012.

[181] Qing Xu and Chongli Zhong. A General Approach for Estimating Framework Charges in Metal - Organic Frameworks. *J. Phys. Chem. C*, pages 5035–5042, 2010.

[182] Edwin Argueta, Jeena Shaji, Arun Gopalan, Peilin Liao, Randall Q. Snurr, and Diego A. Gómez-Gualdrón. Molecular Building Block-Based Electronic Charges for High-Throughput Screening of Metal-Organic Frameworks for Adsorption Applications. *J. Chem. Theory Comput.*, 14(1):365–376, 2018.

[183] http://titan.chem.uottawa.ca/chm8309i/faps/html/quickstart.htmlcharge-equilibration.

[184] M. J. Frisch, G.W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, and D. J Sonnenber. Gaussian 09, 2009.

[185] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297, 2005.

[186] Dmitrij Rappoport and Filipp Furche. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.*, 133(13):134105, oct 2010.

[187] Noel M. O'Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, and Geoffrey R. Hutchison. Open Babel: An Open chemical toolbox. *J. Cheminform.*, 3(10), 2011.

[188] Rappe, a. k., private communication, july 2018.

[189] David Dubbeldam, Sofía Calero, Donald E. Ellis, and Randall Q. Snurr. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.*, 42(2):81–101, jan 2016.

[190] Jeffrey J Potoff and J Ilja Siepmann. Vapor liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.*, 47(7):1676–1682, jul 2001.

[191] Mansi S. Shah, Michael Tsapatsis, and J. Ilja Siepmann. Development of the Transferable Potentials for Phase Equilibria Model for Hydrogen Sulfide. *J. Phys. Chem. B*, 119(23):7041–7052, 2015.

[192] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369(3):253–287, 1921.

[193] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*. Elsevier, 2002.

[194] Ding-Yu Peng and Donald B. Robinson. A New Two-Constant Equation of State. *Ind. Eng. Chem. Fundam.*, 15(1):59–64, feb 1976.

[195] Daniele Ongari, Peter G Boyd, Senja Barthel, Matthew Witman, Maciej Haranczyk, and Berend Smit. Accurate Characterization of the Pore Volume in Microporous Crystalline Materials. *Langmuir*, 33(51):14529–14538, dec 2017.

[196] Qingyuan Yang, Dahuan Liu, Chongli Zhong, and Jian-Rong Li. Development of Computational Methodologies for Metal–Organic Frameworks and Their Application in Gas Separations. *Chem. Rev.*, 113(10):8261–8323, oct 2013.

[197] Allison L. Dzubak, Li-Chiang Lin, Jihan Kim, Joseph A. Swisher, Roberta Poloni, Sergey N. Maximoff, Berend Smit, and Laura Gagliardi. Ab initio carbon capture in open-site metal organic frameworks. *Nat. Chem.*, 4(10):810–816, aug 2012.

[198] Rocio Mercado, Bess Vlaisavljevich, Li-Chiang Lin, Kyuho Lee, Yongjin Lee, Jarad A. Mason, Dianne J. Xiao, Miguel I. Gonzalez, Matthew T. Kapelewski, Jeffrey B. Neaton, and Berend Smit. Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal Organic Frameworks. *J. Phys. Chem. C*, 120(23):12590–12604, 2016.

[199] Daniele Ongari, Davide Tiana, Samuel J. Stoneburner, Laura Gagliardi, and Berend Smit. Origin of the Strong Interaction between Polar Molecules and Copper(II) Paddle-Wheels in Metal Organic Frameworks. *J. Phys. Chem. C*, 121(28):15135–15144, jul 2017.

[200] N. N. Greenwood and A. Earnshaw. Butterworth-Heinemann, 1997.

[201] Toon Verstraelen, Veronique Van Speybroeck, and Michel Waroquier. The electronegativity equalization method and the split charge equilibration applied to organic systems: Parametrization, validation, and comparison. *J. Chem. Phys.*, 131(4), 2009.

[202] Christian S. Diercks and Omar M. Yaghi. The atom, the molecule, and the covalent organic framework. *Science*, 355(6328):eaal1585, mar 2017.

[203] Adrien P. Coté, Annabelle I. Benin, Nathan W. Ockwig, Michael O'Keeffe, Adam J. Matzger, and Omar M. Yaghi. Porous, crystalline, covalent organic frameworks. *Science*, 310(5751):1166–1170, nov 2005.

[204] H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortes, A. P. Cote, R. E. Taylor, M. O'Keeffe, and O. M. Yaghi. Designed synthesis of 3D covalent organic frameworks. *Science*, 316(5822):268–272, apr 2007.

[205] Xiao Feng, Xuesong Ding, and Donglin Jiang. Covalent organic frameworks. *Chem. Soc. Rev.*, 41(18):6010–6022, 2012.

[206] Maria S. Lohse and Thomas Bein. Covalent organic frameworks: structures, synthesis, and applications. *Adv. Funct. Mater.*, 28(33):1705553, aug 2018.

[207] Abass A. Olajire. Recent advances in the synthesis of covalent organic frameworks for CO2 capture. *J. CO2 Util.*, 17:137–161, 2017.

[208] Song Lin, Christian S. Diercks, Yue-Biao Zhang, Nikolay Kornienko, Eva M. Nichols, Yingbo Zhao, Aubrey R. Paris, Dohyung Kim, Peidong Yang, Omar M Yaghi, and Christopher J Chang. Covalent organic frameworks comprising cobalt porphyrins for catalytic CO2 reduction in water. *Science*, 349(6253):1208–1213, sep 2015.

[209] Linus Stegbauer, Katharina Schwinghammer, and Bettina V. Lotsch. A hydrazone-based covalent organic framework for photocatalytic hydrogen production. *Chem. Sci.*, 5(7):2789–2793, 2014.

[210] Wei Zhao, Lieyin Xia, and Xikui Liu. Covalent organic frameworks (COFs): perspectives of industrialization. *CrystEngComm*, 20(12):1613–1634, 2018.

[211] Richard L. Martin, Cory M. Simon, Berend Smit, and Maciej Haranczyk. In silico design of porous polymer networks: high-throughput screening for methane storage materials. *J. Am. Chem. Soc.*, 136(13):5006–5022, apr 2014.

[212] Richard L Martin, Cory M Simon, Bharat Medasani, David K Britt, Berend Smit, and Maciej Haranczyk. In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species. *J. Phys. Chem. C*, 118(41):23790–23802, oct 2014.

[213] Rocío Mercado, Rueih-Sheng Fu, Aliaksandr V. Yakutovich, Leopold Talirz, Maciej Haranczyk, and Berend Smit. In silico design of 2D and 3D covalent organic frameworks for methane storage applications. *Chem. Mater.*, 30(15):5069–5086, aug 2018.

[214] Youshi Lan, Xianghao Han, Minman Tong, Hongliang Huang, Qingyuan Yang, Dahuan Liu, Xin Zhao, and Chongli Zhong. Materials genomics methods for high-throughput construction of COFs and targeted synthesis. *Nat. Commun.*, 9(1):1–10, 2018.

[215] Cory M. Simon, Jihan Kim, Diego a. Gomez-Gualdron, Jeffrey S. Camp, Yongchul G. Chung, Richard L. Martin, Rocio Mercado, Michael W. Deem, Dan Gunter, Maciej Haranczyk, David S. Sholl, Randall Q. Snurr, and Berend Smit. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.*, 8(4):1190–1199, 2015.

[216] Minman Tong, Youshi Lan, Zhenglong Qin, and Chongli Zhong. Computation-ready, experimental covalent organic framework for methane delivery: screening and material design. *J. Phys. Chem. C*, 122(24):13009–13016, jun 2018.

[217] Frank H. Allen. The Cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B Struct. Sci.*, 58(3):380–388, jun 2002.

[218] ccdc.cam.ac.uk/Community/depositastructure/, accessed on april 2019.

[219] Tianqiong Ma, Eugene A. Kapustin, Shawn X. Yin, Lin Liang, Zhengyang Zhou, Jing Niu, Li-Hua Li, Yingying Wang, Jie Su, Jian Li, Xiaoge Wang, Wei David Wang, Wei Wang, Junliang Sun, and Omar M. Yaghi. Single-crystal x-ray diffraction structures of covalent organic frameworks. *Science*, 361(6397):48–52, jul 2018.

[220] Giovanni Pizzi, Andrea Cepellotti, Riccardo Sabatini, Nicola Marzari, and Boris Kozinsky. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.*, 111:218–230, 2016.

[221] github.com/core-cof/CoRE-COF-Database, accessed on november 2018.

[222] Minman Tong, Youshi Lan, Qingyuan Yang, and Chongli Zhong. Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. *Chem. Eng. Sci.*, 168:456–464, aug 2017.

[223] Tongan Yan, Youshi Lan, Minman Tong, and Chongli Zhong. Screening and design of covalent organic framework membranes for CO2/CH4 separation. *ACS Sustainable Chemistry & Engineering*, page acssuschemeng.8b04858, nov 2018.

[224] github.com/danieleongari/egulp.

[225] Chang Wang, Yu Wang, Rile Ge, Xuedan Song, Xueqing Xing, Qike Jiang, Hui Lu, Ce Hao, Xinwen Guo, Yanan Gao, and Donglin Jiang. A 3D covalent organic framework with exceptionally high iodine capture capability. *Chem. - A Eur. J.*, 24(3):585–589, 2018.

[226] Xinyu Guan, Yunchao Ma, Hui Li, Yusran Yusran, Ming Xue, Qianrong Fang, Yushan Yan, Valentin Valtchev, and Shilun Qiu. Fast, ambient temperature and pressure ionothermal synthesis of three-dimensional covalent organic frameworks. *J. Am. Chem. Soc.*, 140(13):4494–4498, 2018.

[227] Shun-qi Xu, Rong-ran Liang, Tian-guang Zhan, Qiao-yan Qi, and Xin Zhao. Construction of 2D covalent organic frameworks by taking advantage of the variable orientation of imine bonds. *Chem. Commun.*, 53(16):2431–2434, 2017.

[228] Johanna M. Huck, Li-Chiang Lin, Adam H. Berger, Mahdi Niknam Shahrak, Richard L. Martin, Abhoyjit S. Bhown, MacIej Haranczyk, Karsten Reuter, and Berend Smit. Evaluating different classes of porous materials for carbon capture. *Energy Environ. Sci.*, 7(12):4132–4146, 2014.

[229] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications.* Elsevier, 2002.

[230] Adrien P. Côté, Hani M. El-Kaderi, Hiroyasu Furukawa, Joseph R. Hunt, and Omar M. Yaghi. Reticular synthesis of microporous and mesoporous 2D covalent organic frameworks. *Journal of the American Chemical Society*, 129(43):12914–12915, oct 2007.

[231] Binit Lukose, Agnieszka Kuc, and Thomas Heine. The structure of layered covalent-organic frameworks. *Chem. - A Eur. J.*, 17(8):2388–2392, feb 2011.

[232] Dalar Nazarian, Jeffrey S. Camp, Yongchul G. Chung, Randall Q. Snurr, and David S. Sholl. Large-scale refinement of metal-organic framework structures using density Functional Theory. *Chem. Mater.*, 29(6):2521–2528, mar 2017.

[233] Li-Chiang Lin, Adam H. Berger, Richard L. Martin, Jihan Kim, Joseph a. Swisher, Kuldeep Jariwala, Chris H. Rycroft, Abhoyjit S. Bhown, Michael W. Deem, Maciej Haranczyk, and Berend Smit. In silico screening of carbon-capture materials. *Nat. Mater.*, 11(7):633–641, jul 2012.

[234] Hiroyasu Furukawa and Omar M Yaghi. Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications. *J. Am. Chem. Soc.*, 131(25):8875–8883, 2009.

[235] Qiuyu Lu, Yunchao Ma, Hui Li, Xinyu Guan, Yusran Yusran, Ming Xue, Qianrong Fang, Yushan Yan, Shilun Qiu, and Valentin Valtchev. Postsynthetic functionalization of three-dimensional covalent organic frameworks for selective extraction of lanthanide ions. *Angew. Chemie Int. Ed.*, 57(21):6042–6048, may 2018.

[236] Daniel Beaudoin, Thierry Maris, and James D. Wuest. Constructing monocrystalline covalent organic networks by polymerization. *Nat. Chem.*, 5(10):830–834, oct 2013.

[237] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7):1456–1465, may 2011.

[238] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103–128, apr 2005.

[239] S. Goedecker and M. Teter. Separable dual-space Gaussian pseudopotentials. *Physical Review B - Condensed Matter and Materials Physics*, 54(3):1703–1710, 1996.

[240] manual.cp2k.org/cp2k-5_1-branch/CP2K_INPUT/MOTION/MD.html, accessed on april 2019.

[241] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[242] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant-pressure molecular-dynamics algorithms. *Journal of Chemical Physics*, 101:4177–4189, 1994.

[243] Jiří Kolafa. Time-reversible always stable predictor-corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.*, 25(3):335–342, 2004.

[244] Chargemol, release 09_26_2017, available at https://sourceforge.net/projects/ddec/.

[245] Thomas A. Manz and Nidia Gabaldon Limas. Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.*, 6(53):47771–47801, 2016.

[246] Anthony K. Rappe and William A. Goddard. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, 95(8):3358–3363, apr 1991.

[247] Daniele Ongari, Peter G Boyd, Ozge Kadioglu, Amber K. Mace, Seda Keskin, and Berend Smit. Evaluating charge equilibration methods to generate electrostatic fields in nanoporous materials. *J. Chem. Theory Comput.*, 15(1):382–401, jan 2019.

[248] github.com/danieleongari/egulp.

[249] Thomas F. Willems, Chris H. Rycroft, Michaeel Kazi, Juan C. Meza, and MacIej Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.*, 149(1):134–141, 2012.

[250] MacIej Haranczyk and James A. Sethian. Automatic structure analysis in high-throughput characterization of porous materials. *J. Chem. Theory Comput.*, 6(11):3472–3480, nov 2010.

[251] David Dubbeldam, Ariana Torres-Knoop, and Krista S. Walton. On the inner workings of monte carlo codes. *Mol. Simul.*, 39(14-15):1253–1292, 2013.

[252] github.com/danieleongari/calc_pe.

List of publications

- D. Ongari, D. Tiana, S. J. Stoneburner, L. Gagliardi, and B. Smit, Origin of the strong interaction between polar molecules and copper (II) paddle-wheels in metal–organic frameworks, The Journal of Physical Chemistry C 2017, 121, 28, 15135-15144, **10.1021/acs.jpcc.7b02302**.

- D. Ongari, P. G. Boyd, S. Barthel, M. Witman, M. Haranczyk, and B. Smit, Accurate characterization of the pore volume in microporous crystalline materials, Langmuir 2017, 33, 51, 14529-14538, **10.1021/acs.langmuir.7b01682**.

- S.K. Elsadi, D. Ongari W. Xu, M. H. Mohamed, M. Haranczyk, and P. Thallapally, Xenon recover at room temperature using metal–organic frameworks, Chemistry - A European Journal 2017, 23, 45, 10758-10762, **10.1002/chem.201702668**.

- G. Bauer, D. Ongari, X. Xu, D. Tiana, B. Smit, and M. Ranocchiari, Metal–organic frameworks invert molecular reactivity: Lewis acid phosphonium zwitterion catalyze the Aldol-Tishchenko reaction, Journal of the American Chemical Society 2017, 139, 50, 18166-18169, **10.1021/jacs.7b10928**.

- S. Kampouri, T. N. Nguyen, C. P. Ireland, B. Valizadeh, F. M. Ebrahim, G. Capano, D. Ongari, A. Mace, N. Guijarro, K. Sivula, A. Sienkiewicz, L. Forró, B. Smit, and K. C. Stylianou, Photocatalytic hydrogen generation from a visible-light responsive metal–organic frameworks system: the impact of nickel phosphide nanoparticles, Journal of Materials Chemistry A 2018, 6, 2476-2481, **10.1039/c7ta10225a**.

- T.N. Nguyen, S. Kampouri, B. Valizadeh, W. Luo, D. Ongari, O. Marie Planes, A. Züttel, B. Smit, and K. C. Stylianou, Photocatalytic hydrogen generation from a visiblelight responsive MOF system: stability versus activity of molybdenum sulfide cocatalysts, ACS Applied Materials & Interfaces 2018, 10, 36, 30035-30039, **10.1021/acsami.8b10010**.

- A. Gladysiak, K. S. Deeg, I. Dovgaliuk, A. Chidambaram, K. Ordiz, P. G. Boyd, S. M. Moosavi, D. Ongari, J. A. R. Navarro, B. Smit, and K. C. Stylianou, Biporous metal–organic framework with tunable CO2/CH4 separation performance facilitated by intrinsic flexibility, ACS Applied Materials & Interfaces 2018, 10, 42, 36144-36156, **10.1021/acsami.8b13362**.

- D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin and B. Smit, Evaluating charge equilibration methods to generate electrostatic fields in nanoporous materials, Journal of Chemical Theory and Computation 2018, 15, 1, 382-401, **10.1021/acs.jctc.8b00669**.

- K. M. Jablonka, D. Ongari, and B. Smit, Applicability of tail-corrections in the molecular simulations of porous materials, Journal of Chemical Theory and Computation 2019, in press, **10.1021/acs.jctc.9b00586**.

- D. Ongari*, Y. Liu*, and B. Smit, Can metal–organic frameworks be used for cannabis breathalyzers?, ACS Applied Materials & Interfaces 2019, 11, 38, 34777-34786, **10.1021/acsami.9b13357**. **Contributed equally.*

- D. Ongari, A. V. Yakutovich, L. Talirz, and B. Smit, Building a consistent and reproducible database for adsorption evaluation in COFs, ACS Central Science 2019, in press, **10.1021/acscentsci.9b00619**.

PERSONAL INFORMATION

Daniele Ongari

Avenue de Tourbillon 72, Sion (VS) 1950, Switzerland

+41 787994472

daniele.ongari@gmail.com

github.com/danieleongari

Skype: danieleongari

Date of birth 07/11/1990 | Nationality Italian

JOB APPLIED FOR

Chemical Engineer / Data Scientist

WORK EXPERIENCE

September 2015 – October 2019

PhD

École Polytechnique Fédérale de Lausanne, Sion, Switzerland

- "Synergy of classical and quantum computational methods to investigate the adsorption properties of microporous materials", under the direction of Prof. Berend Smit

February 2015 – August 2015

Internship as Plant Process Engineer

Solvay Solutions Italia, Ospiate (MI)

- "SolWatt" project: data collection and analysis of utility flows to evaluate and reduce energy consumptions of the plant
- Checking and updating of existing plant P&ID and documentation

EDUCATION AND TRAINING

October 2012 – October 2014

Master of Science in Chemical Engineering

Politecnico di Milano

Final score: 110 cum laude /110

Master thesis: computational investigation of the "backbiting" reaction in polystyrene, by evaluating molecular mobility via molecular mechanics methods and the reactive final event via quantum mechanics analysis.

October 2009 – July 2012

Bachelor of Science in Chemical Engineering

Politecnico di Milano

Final score: 105/110

Final project: computational model of a biomass gasifier on a travelling grate
(A. Frassoldati, T. Maffei, D. Ongari, A. Cuoci, T. Faravelli, S. Pierucci, E. Ranzi, *"A Chemical Approach to Biomass Gasification"*, European Biomass Conference and Exhibition)

September 2004 – July 2009

High School

Liceo scientifico statale "G. Falcone e P. Borsellino", Arese (MI)

Final score: 100/100

Final thesis about the "limits" of Science: Gödel's Incompleteness Theorems, Heisenberg's Uncertainty Principle, "Palomar" by Italo Calvino

Foreign languages: English, French

© European Union, 2002-2013 | http://europass.cedefop.europa.eu

Ce document a été imprimé au Centre d’impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.