PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Chimie ParisTech

Développements méthodologiques pour la simulation de l'adsorption des gaz simples dans les matériaux nanoporeux

Methodology developments for the simulation of simple gas adsorption in nanoporous materials

Soutenue par
Lionel ZOUBRITZKY
Le

Composition du jury :

François-Xavier COUDERT
Directeur de Recherche CNRS, Chimie ParisTech — *Directeur de thèse*

École doctorale n°388
Chimie Physique et Chimie Analytique de Paris Centre

Spécialité
Chimie Physique

Chimie Paris ParisTech | PSL

REMERCIEMENTS

TABLE DES MATIÈRES

Notations and abbreviations

$\boldsymbol{r}$	triplet of coordinates in real space.
$\boldsymbol{k}$	triplet of coordinates in Fourier space.
$\boldsymbol{p}$	configuration of a system.
$\mathcal{X}$	system or sub-system.
$\|\boldsymbol{x}\|$	Euclidean norm of vector $\boldsymbol{x}$.
$\|\boldsymbol{x}\|_\mathrm{p}$	periodic distance, *i.e.* the smallest Euclidean norm of a periodic image of vector $\boldsymbol{x}$.
pcu	topology from the RCSR.
FAU	topology from IZA-SC.
MC	Monte-Carlo

1

CATION PLACEMENT IN ZEOLITES

Among the different families of adsorbents, zeolites are among the most used in industry. Small gases tend to adsorb preferentially on cations when they are present, which makes the study of cationic zeolites particularly relevant to the understanding of industrial adsorption processes. This chapter presents the chemical nature of cationic zeolites and explain the challenges and methods used to obtain representative models of them for subsequent adsorption simulations. The central part of the chapter focuses on the newly developed "shooting star" methodology, which significantly reduces the amount of resource needed to compute the repartition of cations in zeolites.

1.1 Cationic zeolites

In this section, we detail the general structure of cationic zeolites and highlight the challenges associated with the simulation of such systems.

1.1.1 Structure and properties

Zeolites are a family of aluminosilicates, first discovered by Axel Fredrik Cronstedt in 1756. Similar to glass, their atomic structure is made of tetrahedra where the central atom, called a T-atom, is usually silicon, bridged together by oxygens. They are crystalline however, with typical unit cell lengths varying between 10 Å and 100 Å, and have nanopores.

The different kinds of zeolites are separated by topology: each known zeolite topology is given a capital three-letter code by the International Zeolite Association Structure Committee (IZA-SC). As explained in **??**, the topology designates the information extracted from the network of chemical bonds: in the case of zeolites, this network can be simplified by taking one vertex per T-atom, and abstracting each oxygen bridge as a simple edge between the two corresponding T-atoms. This simple protocol makes it clear that the net of a zeolite is very close to the structure it represents – in other words, simplifying the crystal into its net does not lose much information on the initial structure – which explains why a classification by topology is the primary way to distinguish different zeolite structures. It is noteworthy that such a classification is vastly insufficient in the case of MOFs, where the chemistry of metal nodes and the nature of the ligands is thus used first; while for other kinds of materials such as glasses, it is downright impossible to establish because of the absence of cell periodicity.

While millions of hypothetical zeolite topologies have been identified by numerical methods [10.1039/C0CP02255A], only 256 have been experimentally observed[1] as of 2024. This discrepancy may stem from the metastability of zeolites, whose porosity make them thermodynamically less favourable than glass [10.1021/acs.cgd.3c00893]. On the other hand, new experimental zeolite topologies have been discovered each year in the last twenty years, which indicate that accessing new structures is also limited by the current synthetic methods used. Moreover, new synthetic routes can allow new chemistry on existing topologies [10.1039/D2SC06010H, REF].

Indeed, each given topology may give rise to several zeolites that differ by their exact chemical composition. Isoelectronic metal substitution consists in the replacement of silicon by germanium or, in some cases, titanium, zinc or tin [REF]. A much more common kind of substitution consists in replacing some T-atoms by aluminium, or, more anecdotally, boron, gallium [REF]: each such substitution introduces one extra electron in the system, hence these zeolites contain cations to maintain electroneutrality. The nature of the cation itself leads to some more variability: sodium cations are usually the ones used during synthesis [REF], and can then be substituted by other cations [REF]. For simplicity, all zeolites discussed will be considered having only either Si or Al as T-atoms.

Ion exchange is actually the most prevalent industrial use case for cationic zeolites, especially as water softener for laundry agents [REF]. Their considerable internal surface and the accessibility of active sites such as the cations or the substituted metals make them also useful for catalysis [REF]. Other applications include nuclear waste storage, building material additives, water absorbents, soil treatment [REF], oxygen and contrast agent carrier for cancer treatment [10.1039/D3QI00169E] and others. Finally, they can be used as adsorbents for gas separation or storage, which is also of capital industrial interest.

[1] plus 9 intergrowth families, which are not strictly crystalline

Zeolite structures are mostly considered rigid structures in simulations, although some notable exceptions exist [10.1021/acs.chemmater.5b02103, 10.1126/science.abn7667], and most actually posess a flexibility window [10.1039/C003977B] which is exploited, especially during adsorption [10.1016/j.micromeso.2016.10.005]. In addition, T-atom substitution tends to modify only slightly the T-O bond length (Si-O: REF while Al-O: REF) and does not affect the O-T-O angles [REF]. As a consequence, the structure of a substituted zeolite can usually be well approximated by taking the reference structure of the topology (provided as a pure silicate by the IZA-SC) and replacing the relevant silicon atoms. However, not all placements of the substituents may correspond to existing zeolites

1.1.2 Aluminium placement

Cationic zeolites have some aluminiums as part of their structure, however their repartition among the T-atoms raises many questions.

Rules

The first rule, which is obeyed by almost all known zeolites, was discovered by Löwenstein [REF]. It states that no two aluminiums may be neighbours, in the topological sense of being bridged by an oxygen. This hard rule limits the maximum number of aluminiums to half the number of T-atoms in general, and less for topologies that contain odd cycles: for instance, if there is a cycle of 5 T-atoms, only two among them can be Al.

The amount of aluminiums in a zeolite is usually discussed in terms of Si/Al ratio. Löwenstein's rule thus limits the minimal Si/Al ratio to 1 in general, which corresponds to the maximum number of Al and thus of cations. Zeolites with high Si/Al ratios, usually called high-silica zeolites, contain few cations and tend to be hydrophobic [REF]; in general, they also tend to adsorb fewer molecules since cations are privileged adsorption sites.

While Löwenstein's rule is widely accepted as an axiom, and is mainly sustained by the absence of experimental synthesis of any zeolite with Si/Al ratio lower than 1, it is not currently proven. This comes from the cost of the few spectroscopic methods that allow identifying the precise locations of Al atoms: ^{27}Al and ^{29}Si magic-angle spinning NMR can be used to provide some limited information, while [10.1039/B203966B, 10.1039/B301634J] used UV-Visible-NIR spectroscopy on Co^{2+}-exchanged ZSM-5 with high Si/Al ratio to retrieve their Al distribution. But this method cannot be generalized to low-silica zeolites. DFT computation confirms the rule in general, but also show potential violations for some zeolite topologies such as HEU [10.1016/j.mtcomm.2021.102028].

The second rule, proposed by [REF Dempsey et al, 10.1021/j100722a020] on the basis of FAU zeolites with varying Si/Al ratios, states that when possible, the number of next-neighbours Al atoms should be minimized to decrease the number of unfavourable interactions between close aluminiums. This rule is actually known to be violated in many real zeolites, including dealuminated FAU. Speculatively, this may come from the fact that the [REF Dempsey et al] made their observations on zeolites whose synthesis lead them to reach their global energy minimum, while many other synthetic routes can create structures trapped in a local energy minimum with respect to their Si/Al ordering.

Simulation methodologies

Overall, the lack of experimental evidence attesting to the precise position of aluminiums in zeolites make the theoretical prediction of such placement delicate, since it lacks verifiable comparison points. In particular, a crucial question is whether Si/Al ordering is truly guided by thermodynamical consideration or not [10.1007/s11814-021-0796-2]. The latter case would appear for zeolites whose structure is in a metastable state, which is likely to be the case for many real zeolites as evidenced by [10.1039/B301634J] in one of the few experimental studies tackling that very issue.

Even if the aluminium placement can be obtained as that which minimizes the global energy of the structure, this does not necessarily mean that the placement itself is periodic, nor that it follows the same periodicity as the unit cell. Yet, as it is difficult to provide a concrete result on the localisation of aluminiums without making such hypotheses, the studies in the literature only focus on the study of the placement of aluminiums in a single unit cell, supposing that the rest is periodic.

In these conditions, the last remaining question is the simulation methodology to use to find the global energy minimum of the structure. The most basic method consists in generating all possible configurations, computing their energy, and choosing that which minimizes the energy. This approach has actually been widely followed to find the preferred location of aluminiums when there are only one or two aluminiums per unit cell, using Density Functional Theory (DFT) to compute the energy [10.1007/s11814-021-0796-2, REF]. The involved computational cost makes it impractical for lower Si/Al ratios however, which are of major interest for adsorption purposes.

[REF Marie Jeffroy] proposed a more refined Monte-Carlo (MC) protocol for this problem, based on classical force fields for energy computation. Starting from a configuration with the target Si/Al ratio under Löwenstein's rule, each MC step consists in exchanging the positions of a Si and an Al while maintaining Löwenstein's rule, and accepting or refusing the step according to the Metropolis-Hastings criterion (explained in Section 1.2.1). For low Si/Al ratios, this algorithm requires adding some "restart" option that entirely repopulates the aluminiums at random under Löwenstein's rule, because that very rule can lead to blocked situation where no single Si/Al exchange is possible, as illustrated on Figure 1.1.

[REF Findley et al., 10.1021/acs.jpcc.8b03475] proposed an innovative method to assess the validity of theoretically-obtained aluminium placements, by studying CO_2 adsorption isotherms. They found that different Si/Al orderings lead to different isotherms, which, conversely, means that obtaining a precise experimental isotherm can allow retrieving the degree of Al ordering. In accordance to other results of Marie Jeffroy [REF], they conclude that some topologies (REF) are more likely to have their aluminiums placed at random compared to others (REF).

Hence, given the uncertainty that any aluminium placements obtained by thermodynamical simulation actually reflects those observed in real zeolites, we choose to consider that the aluminium placement in our zeolites is taken at random. In practise, each zeolite with a given Si/Al ratio will be modelled as the average of six zeolites of that topology with the same Si/Al ratio, all following Löwenstein's rule but having different Al repartition (taking symmetry into account as well). The results will be computed on each of these six models, and averaged to obtain the result for the zeolite.

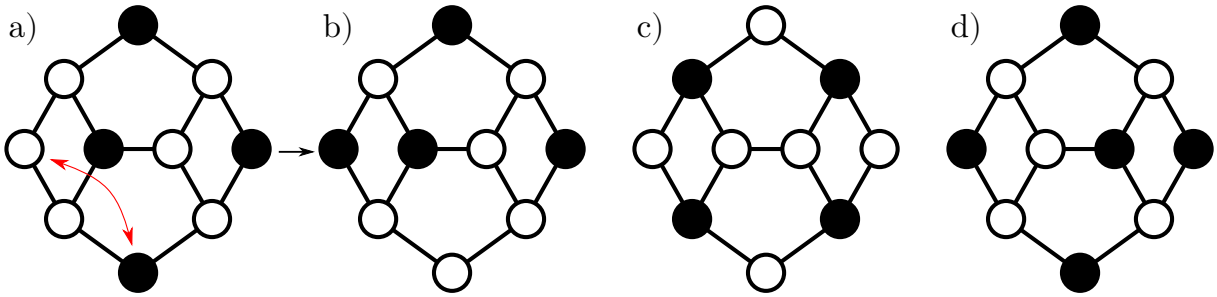


Figure 1.1: Schematic representation of aluminium placement in a zeolite fragment. Empty circles represent silicium, filled circles represent aluminium. Black edges represent -O- bridges. Löwenstein's rule forbids two aluminiums from being neighbors.
a) An exchange move between a silicium and an aluminium.
b) The result of the exchange.
c) A blocked configuration where no move is possible.
d) A configuration with one more aluminium, which cannot be reached from c).

To obtain the zeolite structures with random aluminium placements obeying Löwenstein's rule and the target Si /Al ratio, we use the following algorithm:

- The starting point is either the zeolite filled with Si if the required Si /Al ratio is above 3, else a zeolite filled with as many Al as possible using a greedy algorithm. That greedy technique consists in doing a graph traversal of the T-atoms, transforming each atom into an Al if it is not the neighbour of an Al, or making it a Si otherwise: if all cycles in the graph are even, this is guaranteed to provide an aluminium placement with the minimal Si /Al of 1.

- The main algorithm then alternates between adjusting the number of Al and changing their positions, for a few steps each, back and forth.

  - The position update follows the principle of Marie Jeffroy's MC scheme but disregarding the energy change: one Si and one Al are exchanged at random if the resulting structure still follows Löwenstein's rule.

  - The number of Al atoms is modified if the current Si /Al ratio does not correspond to the target: if it is too high, an Al is converted into a Si, and otherwise a Si that has no Al neighbour is converted into an Al.

- Since the situation can be blocked (see [FIGURE]), the algorithm sometimes restart from scratch, or sometimes does a partial restart by converting multiple Al atoms into Si.

- Configurations where the correct Si /Al ratio is reached are kept in a set. Each configuration is assigned a hash, which is simply a sequence of 0 if the T-atom is a Si or 1 otherwise, in a fixed order of the T-atom. To avoid considering symmetry-equivalent configurations as different, the signature of a configuration is computed as the lexicographically minimum of all the hashes corresponding to the application of a symmetry operation on the configuration. For each unique signature, a corresponding configuration is then retained, and the six models are taken at random among these configurations.

One last consideration is the choice of the supercell. It appears that, for a large number of zeolites topologies, the smallest unit cell cannot be filled with aluminiums so that it both obeys Löwenstein's rule and reaches the minimal Si/Al ratio for that topology. In the case of LTA for instance, the smallest compatible cell is a $2 \times 2 \times 2$ supercell of the smallest unit cell. As a consequence, the greedy algorithm used before is actually attempted on several supercells, and the one leading to the smallest Si/Al ratio is retained.

The full list of minimum Si/Al ratio found for each zeolite topology (except interrupted ones) and the respective supercell is presented in Table 1.1. The supercell is considered with respect to the idealized framework available for each topology on the IZA-SC database[2]. This table presents the lowest Si/Al ratio found, but apart for the cases where it is 1, the values here are not proven to be minimal, since the algorithm relies on random attempts. Similarly, the supercell used could be not minimal.

It is worth mentioning that a previous screening study of adsorption in zeolite by [10.1021/acsami.0c20892] claimed to use the lowest possible Si/Al ratios to screen through zeolites, but failed to reach the ones presented here, systematically in the case where a supercell needed to be used. Another [10.1021/acs.jpcc.3c06115, 10.1021/acs.jpcc.4c00066] alluded to that issue in the case of LTA, yet failed to identify some other topologies (ATT, JBY, LEV, LTL, MRT, OFF) which could reach an Si/Al ratio of 1. The choice of the supercell during aluminium placement is thus an easily overlooked, yet critical, point to consider.

1.1.3 Cations

The electroneutrality of the framework requires that each negative introduced by an Al substitution be compensated by the introduction of a cation in the framework. In many cases, Na^+ or H^+ are used during synthesis; they are sometimes replaced by other cations in later steps.

These cations have a primordial role for adsorption processes, because they constitute privileged adsorption sites for many small gases. The precise mechanism by which they contribute to the overall adsorption capacity of a material depends on their nature, their location, and the nature of the guest molecule however. Moreover, in addition to simple one-to-one adsorption cases of a small molecule on a cation, X-ray spectroscopy evidences the presence of adsorption complexes involving several cations and guest molecules at the same time [10.1039/C2CP23237E]. Finally, the very nature of the cations present in the structure can be quite diverse, ranging from monovalent alkali to polyvalent metal ions, and even molecular ions such as many of the organic structure directing agents (OSDAs) used during zeolite synthesis.

In order to perform numerical simulation of the adsorption process, it is thus necessary to have a model of the zeolite that places its cations in the correct positions. Fortunately, and opposite to the previous case of aluminiums, X-ray diffraction is a convenient technique which can be used to obtain the crystallographic sites in which the cations are located, as well as the respective occupancy of each such site.

[2] accessible at https://europe.iza-structure.org/IZA-SC/ftc_table.php

Topology	Si/Al	Supercell
ABW	1	1×1×1
ACO	1	1×1×1
AEI	1	1×1×1
AEL	1	1×1×1
AEN	1	1×1×1
AET	1	1×1×1
AFG	1	1×1×1
AFI	1	1×1×1
AFN	1	1×1×1
AFO	1	1×1×1
AFR	1	2×2×4
AFS	1	1×1×1
AFT	1	1×1×1
AFV	1	3×3×2
AFX	1	1×1×1
AFY	1	1×1×1
AHT	1	1×1×1
ANA	1	1×1×1
ANO	1	1×1×1
APC	1	1×1×1
APD	1	1×1×1
AST	1	1×1×1
ASV	1	1×1×1
ATN	1	1×1×1
ATO	1	1×1×1
ATS	1	1×1×1
ATT	1	3×4×3
ATV	1	1×1×1
AVE	1	1×1×1
AVL	1	3×3×2
AWO	1	1×1×1
AWW	1	2×2×4
BCT	1	1×1×1
BEC	1.29	2×2×2
BIK	1.4	4×2×5
BOF	1.4	2×1×1
BOG	1.4	1×1×1
BOZ	1.88	1×1×1
BPH	1	1×1×1
BRE	1.29	4×2×4
BSV	1	1×1×1
CAN	1	1×1×1
CAS	1.4	1×1×1
CDO	1.62	4×2×2
CFI	1.46	2×5×1
CGF	1	2×2×4
CGS	1	1×1×1
CHA	1	1×1×1
CON	1.33	1×1×1
CSV	1.22	1×2×1
CZP	1	1×1×1
DAC	1.57	1×3×2
DDR	1.73	1×2×1
DFO	1	2×2×2
DFT	1	4×4×3
DOH	1.62	1×2×1
DON	1.29	1×1×1
EAB	1	1×1×1
EDI	1	4×4×4
EEI	1.44	1×1×1
EMT	1	1×1×1
EON	1.4	2×1×1
EOS	1.18	1×2×2
EPI	1.74	4×2×3
ERI	1	1×1×1
ESV	1.4	1×1×1
ETL	1.48	2×1×1
ETR	1	1×1×1
ETV	1.33	1×1×1
EUO	1.49	1×1×1
EWF	1.49	2×1×1
EWO	1.4	2×2×5
EWS	1.53	1×1×1
EZT	1	1×1×1
FAR	1	1×1×1
FAU	1	1×1×1
FER	1.57	1×1×3
FRA	1	1×1×1
GIS	1	1×1×1
GIU	1	1×1×1
GME	1	1×1×1
GON	1.29	1×1×1
GOO	1	1×1×1
HEU	1.25	1×1×2
IFO	1	1×1×1
IFR	1.29	1×1×2
IFW	1.29	1×1×1
IFY	1	1×1×2
IHW	1.55	1×1×1
IMF	1.55	1×1×1
IRN	1.19	1×1×1
IRR	1.6	2×2×2
ISV	1.31	2×2×1
ITE	1.56	1×2×1
ITG	1.33	1×1×2
ITH	1.33	1×2×1
ITR	1.33	1×1×1
ITT	1.56	2×2×3
ITW	1.4	1×1×1
IWR	1.33	2×2×2
IWS	1.27	1×1×2
IWV	1.62	1×1×2
IWW	1.33	1×1×1
JBW	1	5×4×3
JNT	1	1×1×1
JOZ	1.5	2×1×1
JRY	1	1×1×1
JSN	1	3×4×2
JSR	1.67	1×1×1
JST	2	1×1×1
JSW	1	1×1×1
JSY	1	1×1×1
JZO	1.37	1×1×1
JZT	1.35	1×1×1
KFI	1	1×1×1
LAU	1	2×2×4
LEV	1	3×3×2
LIO	1	1×1×1
LOS	1	1×1×1
LOV	1.25	2×2×1
LTA	1	2×2×2
LTF	1.27	1×1×4
LTJ	1	1×1×1
LTL	1	2×2×4
LTN	1	1×1×1
MAR	1	1×1×1
MAZ	1.4	1×1×2
MEI	1.43	1×1×1
MEL	1.67	1×1×1
MEP	1.71	1×1×1
MER	1	1×1×1
MFI	1.67	1×1×1
MFS	1.45	3×1×1
MON	1.67	2×2×1
MOR	1.67	1×1×2
MOZ	1	1×1×4
MRT	1	2×4×2
MSE	1.55	1×1×1
MSO	1	2×2×2
MTF	1.44	1×1×2
MTN	1.72	1×1×1
MTT	1.4	1×1×1
MTW	1.33	1×5×3
MVY	1	1×1×1
MWF	1	1×1×1
MWW	1.53	2×2×1
NAB	1.5	2×2×1
NAT	1.22	2×2×4
NES	1.52	1×1×1
NON	1.44	1×1×1
NPO	2	4×4×5
NPT	2	2×2×2
NSI	1.4	2×5×3
OBW	1.92	1×1×1
OFF	1	3×3×4
OKO	1.62	1×1×1
OSI	1	1×1×1
OSO	2	1×1×1
OWE	1	2×4×3
PAU	1	1×1×1
PCR	1.5	1×1×1
PHI	1	1×1×1
PON	1	1×1×1
POR	1	1×1×1
POS	1.29	1×1×1
PSI	1	1×1×1
PTF	1.5	1×1×1
PTO	1.4	1×1×1
PTT	1	1×1×1
PTY	1.5	1×1×1
PUN	1.25	1×1×1
PWN	1	1×1×1
PWO	1.5	1×1×1
PWW	1.5	1×1×1
RFE	1.4	1×1×1
RHO	1	1×1×1
RRO	1.25	2×1×1
RSN	1.59	4×1×4
RTE	1.4	1×1×2
RTH	1.53	2×1×3
RUT	1.57	2×2×3
RWR	1.6	4×4×1
RWY	3	1×1×1
SAF	1	1×1×1
SAO	1	1×1×1
SAS	1	1×1×1
SAT	1	1×1×1
SAV	1	1×1×1
SBE	1	1×1×1
SBN	1.5	1×2×1
SBS	1	1×1×1
SBT	1	1×1×1
SEW	1.36	1×2×1
SFE	1.33	1×1×1
SFF	1.46	2×1×2
SFG	1.47	1×2×2
SFH	1.29	5×1×2
SFN	1.29	1×5×2
SFO	1	2×2×4
SFS	1.67	1×1×1
SFW	1	1×1×1
SGT	1.49	2×2×1
SIV	1	1×1×1
SOD	1	1×1×1
SOF	1.5	1×1×1
SOR	1.18	1×1×2
SOS	1.4	2×4×3
SOV	1.29	1×1×1
SSF	1.25	2×2×2
SSY	1.33	1×1×1
STF	1.46	1×1×2
STI	1.25	1×1×1
STT	1.46	1×1×1
STW	1.5	1×1×1
SVV	1.33	1×1×1
SWY	1	1×1×1
SZR	1.25	1×1×2
TER	1.35	1×1×1
THO	1	2×4×4
TOL	1	1×1×1
TON	1.4	2×2×5
TSC	1	1×1×1
TUN	1.53	1×1×1
UEI	1	1×1×1
UFI	1.29	2×2×1
UOS	1.18	2×4×3
UOV	1.38	1×1×1
UOZ	1	1×1×1
USI	1	1×1×1
UTL	1.62	1×2×2
UWY	1.37	1×2×3
VET	1.43	1×1×1
VFI	1	1×1×1
VNI	1.73	3×3×1
VSV	1.78	4×4×1
WEI	1.5	1×1×1
YFI	1.4	1×1×1
YUG	1.56	3×2×4
ZON	1	4×2×2

Table 1.1: Minimal Si/Al and corresponding supercell obtained for all 239 non-interrupted topologies

1.2 Prediction of cation placement

While experimental cation maps are available on some zeolite structures, two seminal problems remain. First, when screening through a large amount of zeolite topologies and Si/Al ratio, only a small fraction of structures have their experimental cation map available. Second, even with the crystallographic sites, not all cation repartitions that obey the given occupancies have the same energy. For example, the site I and I' of FAU being very close to one another, the occupation of a site I prevents either of the neighbouring sites I' from being occupied at the same time. Deciding on the exact location of cations in a zeolite model is thus no trivial matter.

1.2.1 Methodologies

To obtain equilibrium data, such as average positions, on molecular systems, there are two general simulation frameworks which can be used: molecular dynamics and Monte-Carlo (MC) simulation. We only used MC and its variations, because they are the only strategy available to study a variable number of molecules, which is necessary for the simulation of adsorption as we will explain in more details in Section 2.2. The possible use of molecular dynamics and its variants will be discussed later in Section 1.4.

Canonical Monte-Carlo

The Monte-Carlo (MC) method designates a large variety of algorithms which all converge towards their result by performing multiple random samplings. A wide subset of these form the Markov Chain Monte-Carlo (MCMC) schemes, which are a class of algorithms that aim at sampling a target probability distribution. In the context of statistical physics, such methods provide practical ways of obtaining macroscopic observable values from the simulation of the microscopic behaviour of particles. Indeed, an observable $\mathcal{O}$ can be computed as the ensemble average of its microscopic counterpart $\langle o \rangle$, which can be computed as an integral over the configuration space. For example, in the canonical ensemble where the number of particles, the external temperature and the volume of the system is fixed,

$$\mathcal{O} = \langle o \rangle = \frac{1}{Z} \int \mathrm{d}\boldsymbol{p}\, o(\boldsymbol{p}) \exp\left(-\beta U(\boldsymbol{p})\right) \tag{1.1}$$

where $U(\boldsymbol{p})$ is the energy of configuration $\boldsymbol{p}$ and Z is the partition function. In other words, $\mathcal{O}$ can be computed as an integral of values of o on a probability distribution of $\boldsymbol{p}$ given by $p \mapsto \frac{1}{Z} \exp\left(-\beta U(\boldsymbol{p})\right)$.

The direct evaluation of this probability distribution is impossible in general, because of the partition function Z whose value cannot be computed. Instead, the MCMC method gives a way to directly evaluate the observable $\mathcal{O}$ by providing a sequence of configurations $\boldsymbol{p}$ that follows the same probability distribution as the target. The integral can then be computed as simply the average value of o on this specific set of configurations $\boldsymbol{p}$.

The Metropolis-Hastings algorithm is the most common MCMC scheme used to provide such a sequence of configurations. It consists in starting from an initial state, *i.e.* an initial position for all the mobile particles of the system, and making it evolve through a sequence of steps. At each step, a trial move, *i.e.*, a modification of the state, is attempted and may, or may not, be

accepted with a certain probability. In the case of a simulation in the canonical ensemble, the difference of energy ΔE between after and before the move gives the probability

$$P_{\text{accept}}(\Delta E) = \min\left(1, \exp\left(-\beta \Delta E\right)\right) \tag{1.2}$$

that the move be accepted. Intuitively, this means that any move that decreases the energy of the system is accepted, while the others are accepted with a probability that decreases as the energy difference increases, sharply at low temperature and more slowly at high temperature. In the limit of infinite temperature, all the moves are accepted; at zero temperature, only the moves that decrease the energy are. The details on how to compute energy are deferred to Section 1.3.

To avoid drift, the algorithm obeys detailed balance which, conceptually, represents the microreversibility of the simulated physical processes. In practise, this means that, for each MC move that brings the system from configuration $\boldsymbol{p}_{\mathcal{A}}$ to $\boldsymbol{p}_{\mathcal{B}}$, there must exist a "converse" MC moves that can bring the system from $\mathcal{B}$ to $\mathcal{A}$ such that $\pi_{\mathcal{A}} P_{\mathcal{A}\to\mathcal{B}} = \pi_{\mathcal{B}} P_{\mathcal{B}\to\mathcal{A}}$ where $\pi_{\mathcal{X}}$ is the stationary probability of state $\boldsymbol{p}_{\mathcal{X}}$ and $P_{\mathcal{X}\to\mathcal{Y}}$ is the probability of the move from $\boldsymbol{p}_{\mathcal{X}}$ to $\boldsymbol{p}_{\mathcal{Y}}$. In the canonical ensemble, $\pi_{\mathcal{X}}$ is equal to to $\frac{1}{Z}\exp\left(-\beta U\left(\boldsymbol{p}_{\mathcal{X}}\right)\right)$, while the probability of transition is $\min\left(1, \exp\left(-\beta \Delta E\right)\right)$. Hence, detailed balance simply translates to the algorithmic condition that the probability of choosing the trial move $\mathcal{A} \to \mathcal{B}$ must be equal to that of choosing the trial move $\mathcal{B} \to \mathcal{A}$, irrespective of the probability of accepting these moves.

The MC moves which are used for simulations in the canonical ensemble are usually the following:

- translation: displace each atom of the particle by the same vector, taken uniformly at random in a sphere.
- rotation: rotate the particle around one of its atoms (called the "bead") along three uniformly random Euler angles, possibly constrained to some values only, most often all values under a maximum angle.
- reinsertion: remove and reintroduce the particle uniformly at random in the system.

Overall, the algorithm thus consists in a sequence of steps, each of which is an attempt to displace a cation either locally (translation) or not (reinsertion) and is accepted or not depending on the difference of energy the move generates. Configurations corresponding to neighbour steps are thus either identical if the move was refused, or only separate by one atom move, so they are strongly correlated. To reduce computational cost, the computation of the microscopic state function o is thus only realized once every cycle, each cycle corresponding to a number of steps. We fix the number of steps per cycle to 100.

To satisfy detailed balance, the sphere in which the translation vector is taken must have a fixed radius, the bead must be fixed and the angle constraints must be symmetric. However, the best maximum translation length $d_{\max}$ and maximum angle $\theta_{\max}$ are those that yield a MC move acceptance rate of 1/2. Since these optimal values depend on the precise simulation, it is customary to actually make these two constraints evolve along the simulation until reaching an

acceptance rate of 1/2. To do so, at the end of each cycle, they are updated with the following formulas:

$$\begin{aligned} d_{\max} &\leftarrow \text{clamp}\left(d_{\max} \times \left(1 + (\mathrm{T}_{\text{ratio}} - 1/2) \times \sqrt{\frac{10}{99+i}}\right),\ 0.1\,\text{Å},\ 3\,\text{Å}\right) \\ \theta_{\max} &\leftarrow \frac{i-1}{i} \times \theta_{\max} + \frac{\mathrm{R}_{\text{ratio}}}{i} \times 120^\circ \end{aligned} \tag{1.3}$$

where i is the number of the current cycle (starting at 1), $\mathrm{T}_{\text{ratio}}$ (respectively $\mathrm{R}_{\text{ratio}}$) is the ratio of accepted over attempted translations (respectively rotations) in the entire simulation, and clamp(x, m, M) is x if $m \leq x \leq \mathrm{M}$, otherwise m if $x \leq m$, otherwise M (if $x \geq \mathrm{M}$). Having $d_{\max}$ and $\theta_{\max}$ change across the simulation breaks detailed balance, which is theoretically forbidden; yet it does not actually pose a problem in because the formulas used ensure that the two values evolve smoothly towards the optimum, and become almost constants for long enough simulations.

At the beginning of each step, an MC move is chosen at random. In our simulations, all three moves (translation, rotation and reinsertion) are chosen equiprobably. For the particular case of the movement of mono-atomic species, such as most cations in zeolites, the rotation MC move is useless and thus not used, while the reinsertion move collapses to a simple translation with infinite $d_{\max}$, both moves being chosen equiprobably.

The Metropolis-Hastings algorithm thus provides a way to walk through the configuration space. By privileging moves that decrease energy, the system spends more steps in low energy regions; by yet allowing opposite moves sometimes, the scheme ensures that all the configuration space is eventually visited. Hence, the algorithm guarantees ergodicity: in an infinitely long simulation, the fraction of steps spent in a part of the configuration space is proportional to its volume.

Combined aluminium and cation placement

Specifically in the case cationic zeolites, [Marie Jeffroy REF] proposes a simulation method to obtain both the placement of cations in the structure and the location of aluminium atoms across the T-sites. This consists in running an MC simulation where the cations are allowed all previously mentioned MC steps, while the aluminiums are only allowed exchanging positions with siliciums, or a restart step that randomly positions aluminiums and siliciums, both while obeying Löwenstein's rule as explained in **??**.

The statistical ensemble simulated in this setting includes both aluminium and cations, which means that the distribution sampled through such a simulation method represents the equilibrium for both species. From a chemical point of view, this situation is representative of a framework whose synthesis was done with that very cation, and was thermodynamically-driven. While this does correspond to reality in some cases, there is also evidence that the synthesis is kinetically-driven in many other occurences: for such situation, obtaining the aluminium positions from statistical equilibrium is not representative of the actual framework. And in yet other situations, the cation is exchanged post-synthesis, which decorrelates the aluminium placement in the framework from the cation used for adsorption purposes.

Nonetheless, even in those cases, there is no less value in the aluminium placement obtained from such a hybrid strategy than from random sampling of the T-atoms. In fact, by virtue of

the added dimensions of the free energy landscape, a kinetically trapped cation may move more freely following the rearrangement of the aluminiums in the structure, which can make the entire simulation converge more efficiently. If not for the hope of obtaining more relevant aluminium placements, this combined strategy remains valuable for its diminished computationally cost compared to parallel tempering, explained later in Section 1.2.2.

Unfortunately, the fact that the framework is not strictly constant as before prevents the use of precomputed energy grid, later discussed in Section 1.3.4. This was not a problem in the context of [Marie Jeffroy]'s study because the force field they used was such that a silicium-aluminium swap would not change the van der Waals interactions, but this is not the case of most force fields, including the one we use in our study [BoulfelfelSholl2021]. Being unable to precompute the short-term energy contribution of the framework would thus make the simulations significantly slower, tipping the balance against the use of this method.

Site-hopping

Cations are free to move in a canonical MC simulation with translation moves, but some reconfigurations can be problematic. In particular, the cations may not have enough energy to overcome the barriers inherent to the structure of the zeolitic framework in any reasonable simulation time. For example, a cation trapped in the site I of a FAU zeolite will never manage escape into a site III in a room-temperature simulation of reasonable length without a reinsertion move: this simply comes from the high activation barrier it needs to overcome while physically moving out of the sodalite cage. Reinsertion moves are thus useful to allow large configuration movements to occur without having to pass through physical energy barriers.

However, reinsertion consists in attempting to displace a cation to a random position in the framework, which, in the vast majority of cases, ends up rejected because it is too energetically unfavorable. A better move would consist in displacing the cation towards a favorable position, to increase the likelihood of the move being accepted: this is called biasing, which requires unbiasing the acceptation rate to maintain micro-reversibility.

Taking this idea to the extreme, all cations can be forced to sit on one of a fixed set of possible positions, called a site, and to only move by hopping onto another of these positions, irrespective of the distance with the previous position. One hop is then accepted or rejected based on the Metropolis-Hastings criterion, like any other MC simulation. This "site-hopping" approach entirely bypasses the issue of spatial energy barrier since the cations directly hop onto a site, without having to move through an unfavorable region to access it. Hence, it makes this kind of simulation particularly well adapted to sharp energy landscapes, like that of cations in zeolites.

A major constraint of this simulation method however, is that it requires knowing the cationic sites, which is precisely the purpose of the current development. One could weaken this constraint by allowing many pseudo-sites, for instance taking all the points of a regular grid, but this basically amounts to reverting to reinsertion moves without sites, which defeats the purpose of this protocol. Site hopping is thus useful in the later stages of cation placement, once the positions of the sites have already been identified, in order to evaluate their population.

1.2.2 Meta-algorithms

The previous algorithms allow sampling the configuration space at a fixed temperature. However, even with reinsertion or site-hopping, only one particle can move at a time, which is sometimes not enough. For instance, in FAU zeolites, filling both sites I' when starting from an initial configuration where only the central site I is occupied requires that both the central cation move to the edge of the cage and that an external cation be set in the opposite window. Since both movements require passing through an activation barrier, the probability of it happening naturally is low, and thus requires particularly long simulation times to be observed. More generally, collective movements can remain difficult.

To circumvent this issue, the temperature of the simulation can be raised: this mechanically makes all MC steps more likely to be accepted, so higher-energy regions become more reachable. At the same time, a simulation at a hot temperature is much less likely to explore the details of the energy landscape enough to fall into the energy minima of interest, since the accessible configuration space became so much larger. Several meta-algorithms rely on the previously explained algorithms but use different temperatures in different settings, to attempt finding the optimal compromise between wide and detailed explorations.

Simulated annealing

The simplest idea consists in running the simulation while smoothly varying the temperature between cold and hot, alternating between increasing and decreasing phases, possibly interspersed with plateaus. This is called simulated annealing, because of the similarity with the recrystallization method used in metallurgy.

Parallel tempering

Simulated annealing explores the configuration space by alternating between cold phases, where the exploration is detailed but localized, and hot phases, where the exploration is fast but superficial. Parallel tempering, also called "replica exchange" proposes doing both phases at the same time.

In more detail, a parallel tempering simulation consists in running several MC simulations, the "replicas", in parallel, each at a different temperature. The temperatures are taken between room temperature and a hot temperature. An extra MC step is introduced, called the "exchange" step, which consists in exchanging the configurations of two replicas with neighbour temperatures. As with any MC step, the Metropolis-Hastings algorithm ensures that the exchange is accepted only if the difference of energy resulting from this exchange is not too high: as a consequence, for the exchange step to be accepted regularly, the temperatures must be taken so that the typical energies of the system between neighbour temperatures overlap, as shown on [FIGURE].

In this strategy, the coldest simulation is constantly exploring a local free energy minimum, while the hottest simulation easily accepts MC moves and thus walks through the entire configuration space. By exchanging configurations, new configurations regularly trickle down from the hot simulations to the colder ones, while the general Metropolis-Hastings strategy ensures convergence towards a global minimum. It is also more efficient than simulated annealing since the exploration of the configuration space is not constrained by the small duration of the hot phase. Parallel tempering is therefore the method of choice used to obtain convergence when facing difficult energy landscapes such as the one encountered for cation placement.

Naturally, its main downside is its computational cost, proportional to the number of simulations and thus effectively related to the difference between the highest and the lowest temperatures. The choice of intermediate temperatures must be guided by an initial discovery of the typical energies encountered by system at each temperature, which requires a few dedicated simulations prior to the parallel tempering. Finding the optimal number of simulations requires striking a balance with the proportion of accepted MC exchange steps, whose impact on the overall simulation convergence can be difficult to measure. Additionally, simulations with neighbour temperatures must regularly synchronize to attempt an exchange, which strongly limits the amount of parallelization usable in the implementation of parallel tempering.

Shooting star methodology

Previously exposed methods such as simple canonical MC, simulated annealing or parallel tempering focus on making the system (or at least, one of its replicas) reach equilibrium after a certain number of initial steps, so that the rest of the simulation occurs at equilibrium. This is crucial for the problem of sampling the equilibrium distribution of states in general. For the case of cation placement in zeolites however, the observed crystallographic positioning of the cations indicates that this distribution is sharply peaked: only a discrete number of sites can be occupied, and the question is to find their positions and occupancy. As a consequence, we can use another, simpler, strategy, which is potentially less efficient for the actual sampling of the distribution, but can uncover relevant local minima faster.

This new methodology takes the same simulation elements as the ones previously detailed, but arranges them in a different fashion. We call it "shooting star" because it can be decomposed into one hot simulation, like the bright head of the meteor, and a series of cold simulations that stem from it, like the cooling incandescent debris left in its wake.

The first element is the exploration of the phase space. In order to fully explore it, a naive approach consists in doing steps, each of which randomly reinserts a particle in the system. Without the possibility for step rejection, this leads to sampling extremely unlikely states almost all the time however. We propose to perform global exploration by simply running a canonical MC simulation at a very high temperature, for instance 2000 K. Such an approach makes most kinetic obstacles irrelevant, but still prevents absurd configurations like when two atoms collide. This MC simulation is simply called the hot simulation.

The second element is the local minima exploration, which is grafted onto the previous one: every Δ_{spawn} cycles of the hot simulation, a snapshot of its current configuration is taken and used as the starting point for a cold simulation, *i.e.* another canonical MC simulation that runs at room temperature. Δ_{spawn} is chosen so that two consecutive starting points should be uncorrelated. The higher the temperature of the hot simulation, the less correlated its consecutive steps are, so Δ_{spawn} can actually be very low. Each cold simulation runs for M_{init} initial steps, that allow the system to cool down, and M_{prod} production steps at the target room temperature.

Similar to the previous accelerated methodologies, only a fraction of the explored states can actually be used to perform statistics. In the case of simulated annealing, it corresponded to the cold regions of the simulation; for parallel tempering, only the coldest replica is useful to obtain statistics at that temperature; for the current methodology, the relevant steps are

the production of the cold simulation. The main advantage of the shooting star resides in the efficient exploration of the configuration space, rooted in the hot simulation that is independent from the cold ones. Conversely, it lacks the physical meaning of simulated annealing – which is the numerical counterpart to the annealing process used in metallurgy – or the time-reversibility of parallel tempering and bare canonical MC.

In terms of parameters, the shooting star method requires choosing T_{hot} the hot temperature, T_{cold} the cold one, N_{cold} the total number of cold simulations to launch, the previously mentioned M_{init}, M_{prod} and Δ_{spawn}, as well as the number N_{init} of initial steps of the hot simulation before launching the first cold simulation. Of all these parameters, T_{cold}, N_{init} and $N_{prod} = M_{prod} \times N_{cold}$ must also be chosen for a single canonical MC simulation and Δ_{spawn} can always be taken around 100 (its value has little effect on overall performance and quality as long as it is not too low nor too high). The extra parameters T_{hot} is equivalent to the hottest temperature used in simulated annealing and parallel tempering. Finally, M_{init} must be evaluated by an initial run, similar to how the intermediate temperatures used in parallel tempering are obtained or the slope and plateau lengths of simulated annealing are usually chosen. The complexity of the overall setup is thus similar to that of simulated annealing.

1.2.3 Extraction of sites and population from simulation

Cation site locations from density

Assuming we do not know the position of the cation sites, a canonical MC simulation (possibly with shooting star or another meta-algorithm) yields a sequence of cation configurations sampled at room temperature. This sequence can be flattened into a map of the zeolite framework, with one additional point at each position occupied by a cation in one of the steps. Looking at this map allows qualitatively identifying the crystallographic sites as the zones where the density of points is the highest. In order to get quantitative information on the site localization and population, we devise a clustering algorithm that extract the sites from the previous density map.

The first step consists in dividing the space by a regular grid, whose voxels have the same angles as the unit cell but possibly different lengths, taken to be closest to a target value around 0.15 Å. To each voxel is then attributed the number of cations encountered within across the recorded simulation steps. This first step thus bins the density onto a regular grid.

In the second step, the bins are sorted by decreasing order of their value. Each voxel is then taken in order: if it is closer than a set distance to a previously recorded site, then the location of the site is updated as the average of its current position and that of the voxel, weighted by the respective density of both; the density of the voxel is then aded to that of the site. Otherwise, if it is far enough from all existing sites, it is added to the (initially empty) list of recorded sites, along with its density. This yields a series of sites and their respective occupancy. To be more efficient, the main loop early stops when the density of the inquired voxels go below a threshold value, taken as the maximum density of all bins divided by 100: as a consequence, the occupancies are corrected by a multiplicative factor so that the sum of the occupancies of all sites correspond to the number of cations. We note that this should not be required in theory, but it is the necessary fix to a double discretization artifact, once inherent to the MC scheme, and one due to our binning.

We observe slightly better results by sorting the bins according the decreasing order their smoothed counterparts, where smoothing is done by convolving the density map by a small Gaussian, of standard deviation the size of the diagonal of a voxel. Indeed, this erases some small statistical fluctuations that may change the precise order of the voxels.

Direct site population

Population combined with site hopping

1.3 Energy computation

All of the previously exposed methods rely on Monte-Carlo steps which may, or may not, be accepted, depending on the difference of energy between the two configurations. By far, the greatest computational cost of any of these methods lies in these energy computations; it is therefore crucial to be able to compute it efficiently.

The RASPA2 software

RASPA2 is an open-source software developed by [REF] for classical simulations of molecular systems. Although general purpose, it is mainly used with nanoporous materials and in particular for the study of adsorption and diffusion within. RASPA2 implements Monte-Carlo schemes with the Metropolis-Hastings algorithm to perform canonical simulations, but also grand canonical for adsorption (see Chapter 2), as well as parallel tempering. Furthermore, it includes Gibbs ensemble simulation, flexible guest species and frameworks, and many other functionalities which were not needed here.

RASPA2 is written in C, a low-level programming language which allows reaching optimal performance, but is tedious to program in. It cannot be used on GPU and is single-threaded. In order to develop new simulation strategies in an easier setting, I re-implemented the algorithms used for both canonical and grand-canonical simulation of rigid species in rigid frameworks using the Julia programming language (see **??**).

The energy computation algorithms detailed herein are those which were implemented in this new code, but most of them follow that implemented in RASPA2. Moreover, RASPA2 was used to validate the new code, by checking that the simulations yielded the correct physical values (energies, adsorption capacities, and such). By contrast, the shooting star methodology presented before does not exist in RASPA2.

1.3.1 Force fields

The energy of a molecular structure results from the interaction of its constituent atoms, which are intrinsically quantum. At the molecular level, a popular method used is the quantum Density Functional Theory (DFT), which includes a variety of particular methods, each presenting its own compromise between precision and performance. Some alternatives include methods derived from molecular orbital theory. The number of energy computations for a single molecular simulation can be of the order of 10 000 however, which would be much too costly with such methods.

What is commonly used instead are classical force fields, which approximate quantum interactions as a sum of classical interactions between each pair of atoms. These pairwise interactions are defined by one formula per pair of nature of atoms; in the case of non-bonded atoms, each formula is simply a function of the distance between the two atoms. Zeolites are mostly rigid

materials, which means that their deformation due to their vibration modes, or subsequent to the organization of cations or adsorbed gas inside, is negligible for the purpose of the molecular simulation. As a consequence, the interactions between framework atoms can be skipped, the only ones to consider for cation placement are those between the cation and the different atoms of the zeolites, which are not bonded.

The pairwise interaction formula is usually divided between the energy contribution of the charges of the atoms, and a short-range van der Waals term. Designing a force fields thus customarily consists in choosing the atomic charges and the formula for the van der Waals term. Several options have been used in the literature for the latter: a very common option is the Lennard-Jones potential

$$r \mapsto 4\varepsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right) \tag{1.4}$$

where r is the distance between atoms. One alternative is the Buckingham potential

$$r \mapsto \mathrm{A} \exp\left(-\mathrm{B} \times r\right) - \frac{\mathrm{C}}{r^6} \tag{1.5}$$

The potential parameter ε, σ, A, B or C are part of the design of the force field.

1.3.2 Short-term interactions

Since the van der Waals term decays quickly towards zero, it is computationnally useful to set it to exactly zero beyond a cutoff distance. To do so while avoiding the introduction of unphysical forces, one possibility consists in shifting the formula by a constant, so that the function remains continuous at the cutoff. If no force is required, such as in the MC methods, it is possible to leave the potential unshifted and discontinuous at the cutoff; a tail correction is then added to the total energy to account for the missing interactions. In any case, knowing that the interaction becomes zero after the cutoff allows some computational optimizations, such as using cell lists in order to iterate over the relevant atom pairs only, instead of all atom pairs.

For simplicity of the implementation, the dimensions of the simulated unit cell must be taken so that for any atoms A and B, there is at most one periodic image of B which is lower than the cutoff distance from atom A. This translates to the constraint that the orthogonal lengths of the unit cell must be as long as at least twice the cutoff. We take this cutoff as 12 Å, as is common in the literature [REF] and is the default in RASPA2.

Given this constraint, the total energy of a configuration due to the van der Waals interaction can be computed by summing the contribution of each the pair of atoms which are closer than the cutoff distance. This can done naively with a double loop over all atoms, computing the distance of each pair and rejecting that above the threshold. Here, the distance considered is the periodic distance, *i.e.* the smallest distance between periodic images of the two atoms, respective to the periodicity of the crystal. Computing it is up to 6 times more costly than without periodicity, and can become a computational bottleneck for simple enough force fields.

To avoid unecessary distance computations, a possibility consists in binning the space of the unit cell, and skipping the pairs of atoms that belong to bins further than the cutoff distance. This idea underlies the principle of neighbour list which can be used to obtain the list of relevant

pairs of atoms efficiently. We use the CellListMap.jl [10.1016/j.cpc.2022.108452] julia package for this purpose, although we only enable it when the cell is large enough. Indeed, the naive approach can actually be more efficient in many cases where the orthogonal lengths of the cell are close to twice the cutoff distance.

1.3.3 Ewald summation

Energy decomposition

In opposition to the van der Waals term, charge interactions do not have a cutoff, because Coulomb's law does not decay fast enough towards zero. The interaction between atoms A and B, of respective charge q_A and q_B, is:

$$\mathrm{E}_{\mathrm{A,B}}^{\mathrm{Coulomb}} = \frac{q_\mathrm{A} q_\mathrm{B}}{4\pi\varepsilon_0 \times \left\|\boldsymbol{r}_\mathrm{B} - \boldsymbol{r}_\mathrm{A}\right\|} = \frac{q_n q_m}{4\pi\varepsilon_0 \times \left\|\boldsymbol{r}_m - \boldsymbol{r}_n + \boldsymbol{l}\right\|} \tag{1.6}$$

where n and m are the unique atom numbers in a reference unit cell of respectively A and B, and $\boldsymbol{l}$ is the lattice vector between the origin of the unit cell of A and that of B. Therefore the total contribution of charge interactions in a crystal with N atoms per unit cell is:

$$\mathrm{E}^{\mathrm{Coulomb}} = \frac{1}{2}\sum_{\mathrm{A}\neq\mathrm{B}} \mathrm{E}_{\mathrm{A,B}}^{\mathrm{Coulomb}} = \frac{1}{8\pi\varepsilon_0}\sum_{\boldsymbol{l}} \sum_{\substack{n=1 \\ n\neq m \text{ if } \boldsymbol{l}=0}}^{\mathrm{N}} \sum_{m=1}^{\mathrm{N}} \frac{q_n q_m}{\left\|\boldsymbol{r}_n - \boldsymbol{r}_m + \boldsymbol{l}\right\|} \tag{1.7}$$

where $\boldsymbol{l}$ iterates over the lattice vectors.

The semi-convergence of the sum over $\boldsymbol{l}$ makes it impossible to compute directly. Hence, several methods have been developped to obtain the result such as particle mesh Ewald (PME) and particle-particle-particle-mesh ($\mathrm{P^3M}$). We focus on the initial Ewald summation technique underlying them, which is used in RASPA2 and which we also implemented.

Since the crystal is a periodic system, any function of the system that depends on a position in space can be decomposed into a simple sum by Fourier transform. This is in particular true of the electric potential, hence the problematic sum above could be computed in Fourier space. However, the charges are located on the atoms, and are thus discrete, which causes singularities that prevent the computation. The idea behind Ewald summation consists in fictitiously smoothing the point charges into Gaussians: these smooth charges yield a potential computable in Fourier space, while the potential resulting from the real point charge minus their Gaussian counterpart is computed in real space. Electric potential is additive, so the total electric potential $\Phi(\boldsymbol{r})$ is then obtained by adding the real space term $\Phi^{\mathrm{direct}}(\boldsymbol{r})$ and the Fourier term $\Phi^{\mathrm{reciprocal}}(\boldsymbol{r})$. Finally, the resulting energy is simply the action of the point charges on the potential:

$$\mathrm{E}^{\mathrm{Coulomb}} = \frac{1}{2}\sum_{n=1}^{\mathrm{N}} q_n \left(\Phi^{\mathrm{direct}}(\boldsymbol{r}_n) + \Phi^{\mathrm{reciprocal}}(\boldsymbol{r}_n)\right) - \mathrm{E}^{\mathrm{self}} \tag{1.8}$$

$\mathrm{E}^{\mathrm{self}}$ designates a spurious self-interaction energy term which must be substracted.

Direct term

To each atom A of the crystal is associated a Gaussian charge density

$$\boldsymbol{r} \mapsto q_\mathrm{A}\left(\frac{\alpha}{\sqrt{\pi}}\right)^3 \exp\left(-\alpha^2\left\|\boldsymbol{r}_\mathrm{A} - \boldsymbol{r}\right\|^2\right) \tag{1.9}$$

Thus, the direct term is the potential resulting from the point charge minus this Gaussian. This can be obtained by solving Poisson's law:

$$-\nabla\Phi(\boldsymbol{r}) = 4\pi\rho(\boldsymbol{r}) \tag{1.10}$$

which yields, after some simplification steps not detailed here:

$$\Phi_{\mathrm{A}}^{\mathrm{direct}}(r) = q_{\mathrm{A}}\frac{\operatorname{erfc}(\alpha r)}{r} \tag{1.11}$$

where r is the distance to atom A and erfc is the complementary error function erfc = 1 − erf, with erf : $x \mapsto \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t$. Since erfc quickly decays towards 0, the total direct potential at a given point can be computed by only considering the contribution of the closest charges around, so

$$\Phi^{\mathrm{direct}}(\boldsymbol{r}) \approx \sum_{n=1}^{\mathrm{N}} \frac{q_n}{\left\|\boldsymbol{r}-\boldsymbol{r}_n\right\|_{\mathrm{p}}}\left(\operatorname{erfc}\left(\alpha\left\|\boldsymbol{r}-\boldsymbol{r}_n\right\|_{\mathrm{p}}\right)\right) \tag{1.12}$$

where $\left\|\boldsymbol{x}\right\|_{\mathrm{p}}$ designates the distance to the origin of the closest periodic image of $\boldsymbol{x}$.

Reciprocal term

For the reciprocal term, we start by writing the total charge density coming from the Gaussians given in Equation (1.9):

$$\rho^{\mathrm{reciprocal}}(\boldsymbol{r}) = \sum_{\boldsymbol{l}}\sum_{n=1}^{\mathrm{N}} \frac{q_n}{4\pi\varepsilon_0}\left(\frac{\alpha}{\sqrt{\pi}}\right)^3 \exp\left(-\alpha^2\left\|\boldsymbol{r}_n+\boldsymbol{l}-\boldsymbol{r}\right\|^2\right) \tag{1.13}$$

Its Fourier decomposition over the periodic unit cell of volume V is:

$$\begin{aligned}
\hat{\rho}^{\mathrm{reciprocal}}(\boldsymbol{k}) &= \frac{1}{\mathrm{V}}\int_{\mathrm{V}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\rho^{\mathrm{reciprocal}}(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \\
&= \frac{1}{\mathrm{V}}\int_{\Omega} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{n=1}^{\mathrm{N}}\frac{q_n}{4\pi\varepsilon_0}\left(\frac{\alpha}{\sqrt{\pi}}\right)^3\exp\left(-\alpha^2\left\|\boldsymbol{r}_n+\boldsymbol{r}\right\|^2\right)\mathrm{d}\boldsymbol{r} \\
&= \frac{2\pi\alpha^3}{\mathrm{V}\pi^{3/2}}\sum_{n=1}^{\mathrm{N}}\frac{q_n}{4\pi\varepsilon_0}e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_n}\int_0^{\infty}\left\|\boldsymbol{r}'\right\|^2\exp\left(-\alpha^2\left\|\boldsymbol{r}'\right\|^2\right)\int_0^{\pi}e^{-i\boldsymbol{k}\cdot\boldsymbol{r}'}\sin(\theta)\mathrm{d}\theta\mathrm{d}\boldsymbol{r}' \\
&= \frac{1}{4\pi\varepsilon_0\mathrm{V}}\exp\left(\frac{-k^2}{4\alpha^2}\right)\sum_{n=1}^{\mathrm{N}}q_n e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_n}
\end{aligned} \tag{1.14}$$

after several simplification steps not detailed here. Using Poisson's law (Equation (1.10)), the electric potential due to the Gaussians can then finally be expressed as:

$$\Phi^{\mathrm{reciprocal}}(\boldsymbol{r}) = \frac{1}{\varepsilon_0\mathrm{V}}\sum_{\boldsymbol{k}\neq\boldsymbol{0}}\sum_{n=1}^{\mathrm{N}}\frac{q_n}{k^2}e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}_n)}\exp\left(\frac{-k^2}{4\alpha^2}\right) \tag{1.15}$$

The term $\kappa_{\boldsymbol{k}} = \frac{1}{k^2}\exp\left(-k^2/(4\alpha^2)\right)$ decays fast, so the sum over $\boldsymbol{k} \neq \boldsymbol{0}$ is actually computed as a sum over a finite number of values of $\boldsymbol{k}$ close to $\boldsymbol{0}$. These values are chosen so that the remaining sum of $\kappa_{\boldsymbol{k}}$ for any $\boldsymbol{k}$ not retained is below the target precision of the computation.

Self-interaction term

The last remaining term is the self-interaction, which must be removed. Its existence comes from the way the reciprocal term is computed: $\Phi^{\text{reciprocal}}(\boldsymbol{r})$ includes the contribution of all charges on point $\boldsymbol{r}$, but there should not be any contribution of atom A at position $\boldsymbol{r}_{\text{A}}$.

Moreover, in the case of rigid polyatomic molecules, like the small gases we will discuss in Chapter 2, there should not be any contribution of atom A at the positions $\boldsymbol{r}_{\text{B}}$ of other atoms B part of the same molecule.

Each atom $1 \leq n \leq \text{N}$ thus contributes a spurious potential term:

$$\Phi_n^{\text{self}} = \frac{2\alpha q_n}{\sqrt{\pi}} + \frac{1}{2} \sum_{\substack{m \\ m \smile n,\ m \neq n}} \frac{q_m \operatorname{erf}\left(\left\|\boldsymbol{r}_n - \boldsymbol{r}_m\right\|_{\text{p}}\right)}{4\pi\varepsilon_0 \left\|\boldsymbol{r}_n - \boldsymbol{r}_m\right\|_{\text{p}}} \tag{1.16}$$

where $m \smile n$ is true when both atoms m and n belong to the same molecule.

Total charge energy

Combining the last three terms, we obtain the overall Coulomb energy:

$$\begin{aligned} \text{E}^{\text{Coulomb}} = &\sum_{\boldsymbol{l}} \sum_{n=1}^{\text{N}} \sum_{\substack{m=1 \\ m \not\smile n}}^{n} \frac{q_n q_m}{8\pi\varepsilon_0} \frac{\operatorname{erfc}\left(\alpha \left\|\boldsymbol{r}_n - \boldsymbol{r}_m + \boldsymbol{l}\right\|\right)}{\left\|\boldsymbol{r}_n - \boldsymbol{r}_m + \boldsymbol{l}\right\|} \\ &+ \frac{1}{2\varepsilon_0 \text{V}} \sum_{\boldsymbol{k} \neq \boldsymbol{0}} \sum_{n=1}^{\text{N}} \sum_{m=1}^{\text{N}} \frac{q_n q_m}{k^2} e^{i\boldsymbol{k}\cdot(\boldsymbol{r}_m - \boldsymbol{r}_n)} \exp\left(\frac{-k^2}{4\alpha^2}\right) \\ &- \cancel{\sum_{n=1}^{\text{N}} \frac{2\alpha q_n}{\sqrt{\pi}}} - \frac{1}{2} \sum_{n=1}^{\text{N}} \sum_{\substack{m=1 \\ m \smile n,\ m \neq n}}^{\text{N}} \frac{q_n q_m}{4\pi\varepsilon_0} \frac{\operatorname{erf}\left(\left\|\boldsymbol{r}_n - \boldsymbol{r}_m\right\|_{\text{p}}\right)}{\left\|\boldsymbol{r}_n - \boldsymbol{r}_m\right\|_{\text{p}}} \end{aligned} \tag{1.17}$$

The term summing the $2\alpha q_n/\sqrt{\pi}$ cancels out because the overall system needs to be electrically neutral to carry a physical meaning, so the sum of all q_n is 0. This invariant is checked in the code to avoid any error.

1.3.4 Precomputation

Different algorithms that solve a particular problem generally differ in the specific space-time tradeoff they choose. Obtaining the energy of a configuration presents the same possibility: while entirely computing the energy from scratch requires a set amount of computation, it is possible to make subsequent computations faster at the cost of a greater memory footprint of the algorithm, by storing some intermediate results which can be reused afterwards.

Energy grid

In the case of simulations involving a rigid framework, the most obvious precomputation that can be done is that of the short-term interactions of the guest species with the framework atoms. Indeed, such interactions only depend on the nature and the position of the guest atom, so it is possible to store them on an energy grid.

More precisely, fore each possible guest atom, the unit cell of the framework is divided into a regular grid, with a set spacing between consecutive points. We use 0.15 Å, which is fine

enough to be accurate but not too fine to avoid using too much memory. For each grid point, the sum of the interactions of all framework atoms with a guest atom in that position is stored at that point. The grid is then written to disk.

During the simulation, in order to compute the short-term interactions between the framework atoms and a guest atom at a given position, the value is interpolated from the surrounding grid points. Using a simple linear interpolation of the values does not yield very satisfactory results however, since derivative discontinuity would push that atoms towards the grid points, which would result in simulation artifacts. To prevent this, RASPA2 stores not only the value of the energy, but also the three first derivatives, three values of the second derivatives and one of the third, following a scheme by [REF]. Moveover, the grid is forced to be orthorombic: this means that it may not be aligned with the unit cell of the framework, but it does not matter as long as each point in the unit cell can be surrounded by recorded grid points. I reimplemented the same grid algorithm, which ensures a smooth interpolation everywhere in space.

Ewald precomputation

Short-term interactions that can be stored in grids include the short-term direct part of the Ewald summation. The reciprocal term cannot however be simply stored as such.

Yet, it is still possible to precompute a large part of the reciprocal interaction which is due to the framework. To take the potential given by Equation (1.15) back to the energy, it is useful to introduce the sums:

$$\mathrm{S}_{\boldsymbol{k}} = \sum_{n=1}^{\mathrm{N}} q_n e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_n} \tag{1.18}$$

for each $\boldsymbol{k}$. With this notation, the contribution of the reciprocal term to the energy is

$$\begin{aligned}
\mathrm{E}^{\text{reciprocal}} &= \frac{1}{2}\sum_{m=1}^{\mathrm{N}} q_m \Phi^{\text{reciprocal}}(\boldsymbol{r}_m) \\
&= \frac{1}{2\varepsilon_0 \mathrm{V}} \sum_{m=1}^{\mathrm{N}} q_m \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}_m} \mathrm{S}_{\boldsymbol{k}} \quad \text{then using } \overline{\mathrm{S}_{\boldsymbol{k}}} = \sum_{m=1}^{\mathrm{N}} q_m e^{i\boldsymbol{k}\cdot\boldsymbol{r}_m} \\
&= \frac{1}{2\varepsilon_0 \mathrm{V}} \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} \overline{\mathrm{S}_{\boldsymbol{k}}} \times \mathrm{S}_{\boldsymbol{k}} = \frac{1}{\varepsilon_0 \mathrm{V}} \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} \left\|\mathrm{S}_{\boldsymbol{k}}\right\|^2
\end{aligned} \tag{1.19}$$

Hence, to avoid redundant computations, the list of relevant $\boldsymbol{k}$, associated to their factor $\kappa_{\boldsymbol{k}} = \exp\left(-k^2/\left(4\alpha^2\right)\right)/k^2$, is computed once and for all. For each atom n of the framework, the factors $q_n \exp\left(i\boldsymbol{k}\cdot\boldsymbol{r}_n\right)$ are also computed once. Distinguishing between the atoms of the framework and the guests allow separating $\mathrm{S}_{\boldsymbol{k}}$ into $\mathrm{S}_{\boldsymbol{k}}^{\text{framework}} + \mathrm{S}_{\boldsymbol{k}}^{\text{guest}}$, so that finally, the energy can be computed as

$$\mathrm{E}^{\text{reciprocal}} = \underbrace{\frac{1}{2\varepsilon_0 \mathrm{V}} \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} \left\|\mathrm{S}_{\boldsymbol{k}}^{\text{framework}}\right\|^2 + \frac{1}{2\varepsilon_0 \mathrm{V}} \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} \left(\overline{\mathrm{S}_{\boldsymbol{k}}^{\text{framework}}} + \overline{\mathrm{S}_{\boldsymbol{k}}^{\text{guest}}}\right) \times \mathrm{S}_{\boldsymbol{k}}^{\text{guest}}}_{\text{precomputed part}} \tag{1.20}$$

Finally, the entire self-interaction term can be computed once for all, since it does not depend on the configuration. This is only the case here because the simulated moleculses are rigid bodies; for deformable structures, the self-interaction term depends on the actual conformation of the guests, which may depend on the state.

Optimizations for computing the difference of energies

A simulation that follows the Metropolis-Hastings algorithm walks across configurations, each pair of contiguous states being separated by a single MC move. In this context, computing the energy of each trial configuration to compare it to the previous one is wasteful: the only part of the energy impacted by the MC move is that which stems from the moved species. Instead of computing the energy of the trial configuration, it is thus better to directly compute the energy difference due to the move.

For the short-term interactions, this is straightforward: the energy difference is simply the sum of the interactions of the species at the new position minus that at its previous position. If there are N species, moving one thus costs O(N) computations for these terms, instead of $O(N^2)$ required to compute this energy the first time. The direct part of the Ewald summation follows the same principle.

For the reciprocal part, the difference in energy is derived from Equation (1.20) by removing the first constant term and computing:

$$\Delta E^{\text{reciprocal}} = \frac{1}{2\varepsilon_0 V} \sum_{\boldsymbol{k}} \kappa_{\boldsymbol{k}} \left(\overline{S_{\boldsymbol{k}}^{\text{framework}}} \times \Delta S_{\boldsymbol{k}}^{\text{guest}} + \Delta \left\| S_{\boldsymbol{k}}^{\text{guest}} \right\|^2 \right) \tag{1.21}$$

Since both values of $S_{\boldsymbol{k}}^{\text{guest}}2$ need to be computed before and after the move to take the difference of their norm, it is actually equally efficient to directly compute both values of $E^{\text{reciprocal}}$ before and after the move, directly following Equation (1.20), and taking the difference, instead of the previous formulation.

In order to allow computing this efficiently, each term $q_n e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_n}$ is thus stored in memory (instead of recomputing it each time). The new values $q_n e^{-i\boldsymbol{k}\cdot\boldsymbol{r}'_n}$ computed for the trial move are also temporarily kept in memory, so they can be used to replace the previous ones if the trial move is accepted. Finally, the value of each $S_{\boldsymbol{k}}^{\text{guest}}$ is also kept in memory, so that it can be updated by reusing the previous computation.

RASPA2 does not implement this optimization, and simply recomputes the $e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_n}$ terms each time, which is slower but requires less memory.

Blocking pockets

Some MC trial moves attempt to displace an atom to a location that would lead to unphysically high energies – typically, too close to a framework atom. Such positions can be determined once and for all, and the relevant MC moves automatically refused without the need for any energy computation.

To do so, the simplest strategy consists in making a boolean grid, aligned with the energy grids used for short-term interactions, which stores a 1 when the voxel is blocked or 0 if it is allowed. The first step of any MC trial move then consists in checking whether each atom is in an allowed voxel, or refusing the move otherwise. In practise, the simplest implementation consists in making one such boolean grid per atom kind (since the energy threshold depends on the atom), and for each atom, checking the closest grid point of its kind.

In addition to its use as computational optimization, such boolean grids can also be used to mark some parts of the framework as inaccessible, irrespective of the corresponding energy.

This is useful when small window size prevents a species – for instance a gas molecule – from entering a pore – for instance the sodalite cages in LTA zeolites [10.1021/ja953856q]. Indeed, without explicitly blocking the pocket, an MC move could make a species appear in the pore, whereas it should be kinetically impossible to enter. This strategy will thus be used when studying the adsorption of gases in zeolites (see Chapter 2), but for the small cations under study here, there is no need to additionally block any part of the frameworks.

RASPA2 only implements blocking for the latter case – *i.e.* explicit pocket removal. It does not use a boolean grid: instead, the blocked space is described as a combination of blocking spheres. For each trial move, the distance between the atoms of the moved guest and the center of each sphere is computed: if one is below its radius, the move is rejected. This approach requires less precomputation and thus less memory than storing the boolean grids, but its cost is much greater, and increases with the number of blocking spheres, preventing it from being used as an optimization like before. It is thus less efficient than the approach I implemented.

Site-hopping precomputations

1.4 Possible alternatives

Molecular dynamics (MD) could also be used to place cations in zeolites. Generally speaking, the principle of MD consists in running a simulation where the particles follow Newton's laws of motion, with forces based on force fields like for MC schemes. The simulation is divided in time steps, between which the forces are computed on each particle to derive their position at the next time step. Hence, the trajectories of the particles in MD follow a physical path.

Temperature can be included in different ways in MD, all of which use an element of randomness that eventually impacts the velocities of particles, in order to ensure that the average values extracted from the simulation correspond to those which would be obtained at that temperature. This means that the precedent discussion on the inaccessibility of high-energy regions to the system at room temperature still applies in MD.A simple MD simulation is thus doomed to only explore local energy minima, failing to reach ergodicity like with the simple canonical MC scheme.

To circumvent this issue, it is necessary to bias the simulation in order to force the system to explore regions of high energy. One such strategy consists in doing metadynamics [REF], which is a variation of MD in which the potential energy landscapes is modified during the simulation. The effective free energy of a point in configuration space is computed as the sum of the real free energy of that point without bias, added to the sum of some Gaussian functions centred on previously explored points. Every few steps, another Gaussian function centred on the current configuration point is added to the bias, so that the effective free energy of explored regions increases through the simulation, which allows the system to eventually escape local minima and explore other regions.

A Gaussian is a function of the form $f : \boldsymbol{x} \mapsto \mathrm{A}e^{-(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma(\boldsymbol{x}-\boldsymbol{\mu})}$. In practise, the variable $\boldsymbol{x}$ of the Gaussians used in metadynamics are not directly the 3N coordinates of the N particles, but a small-dimensional vector whose coordinates are called "collective variables". These collective variables are themselves function of the 3N initial coordinates. The parameter A and Σ of the Gaussians being usually fixed, each Gaussian is stored by its centre $\boldsymbol{\mu}$, whose coordinates are thus the collective variables of the current configuration point. Therefore, compared to regular

MD, running a metadynamics simulation incurs the additional cost of computing the collective variables $\boldsymbol{x}$ of each visited configuration point, storing the Gaussians' centres, and computing their value at each step.

For cation placement the main issue is to find a small set of relevant collective variables, since the cost of metadynamics grows with their number. We did not explore this venue because we expected it to be similar in difficulty to the MC alternatives, with no clear reason why it should be more efficient. MD in general is also known to converge more slowly than MC, because large configuration change made possible by MC steps are prohibited by the physics-driven approach of MD, and the same argument could be made of metadynamics compared to other accelerated MC schemes. Exploring metadynamics or its variants could nonetheless be of interest in the search for more alternative, and potentially more efficient, algorithms.

2

MOLECULAR SIMULATION OF ADSORPTION

Adsorption is the physical phenomenon by which a molecule, the adsorbate, attaches to a surface, the adsorbent. One particular application of adsorption in the industry is for gas separation, by using an adsorbent that specifically binds one constituent of the target gas mixture more than the other.

In this chapter, we detail the nature of this phenomenon in the case of small gases in zeolites. We then explain the usual techniques which are used for its numerical simulation, as well as a possible alternative.

2.1 Molecular description

2.1.1 Physisorption and chemisorption

The atomic nature of adsorption depends on the kind of chemical interaction the supports it. On the one hand, if the adsorbate forms a strong (*i.e.* covalent, sometimes ionic) bond with the adsorbent, then the adsorption is called a chemisorption, because it involves a chemical reaction. In that case, the surface is modified. On the other hand, if the adsorbate and adsorbent

are only bond by a weak interaction, the adsorption is called a physisorption and the surface is left intact.

Chemisorption involves a stronger bond that physisorption, and is thus less reversible. Since it changes the surface of the adsorbent, it cannot be used industrially to separate large amounts of gas, as the adsorbent would need to be provided in equivalently large amount. It can however be used in the industry as an intermediate step of a catalytic cycle that ends up regenerating the surface. In our case however, we will implicitly refer to physisorption when mentioning adsorption.

2.1.2 Adsorption sites in crystalline materials

An adsorption site designates a particular space where the adsorbate is likely to remain trapped on the surface of the adsorbent. In other words, the interaction between the adsorbate and the adsorbent is maximally attractive when the adsorbate is located in one of the adsorption sites.

In the case of crystalline materials, adsorption sites can be identified through spectroscopy

The position of the adsorption site depends on the adsorbent, but also the adsorbate. At very low density for example, a small polar molecule like H_2O will preferentially adsorb in locations where it can orient itself to maximize the stabilizing Coulomb interaction of the framework. A large apolar molecule like a xylene on the other hand, will adsorb in locations that maximize its Van der Waals interactions with the adsorbent. Such sites are not necessarily related.

Moreover, the location of the adsorption sites becomes less relevant when the density of the adsorbate becomes high. Indeed, the adsorbate-adsorbate interactions can play a key role in the global adsorption capacity of a material, and the stronger these interactions, the more the internal structure of the adsorbate will reorganize to accommodate these interactions, which may displace some molecules from the ideal adsorption site.

Overall, the situation is quite different from the cationic sites, discussed in Section 1.2.1

2.1.3 Small gases in zeolites

Small gases such as CO_2, N_2, O_2, CH_4 or Ar, among others, are typical adsorbates that need to be separated for industrial purposes. In general, molecular sieves like zeolites can be used for this purpose

Given their rather similar sizes

2.2 Grand Canonical Monte-Carlo

2.2.1 Principle

The adsorption capacity of a gas in a material at a fixed temperature and pressure is the number of molecules of gas that are retained by the material when in contact with a gas in these conditions, at equilibrium. Since it is a macroscopic observable $\mathcal{O}$, which can be experimentally measured, it can also be expressed as the statistical average of its microscopic counterpart o in a well-chosen statistical ensemble. For the current case, the adsorption capacity results from a chemical equilibrium between the gas out of the material, at its given temperature T and

pressure P, and the gas in the pores of the material, at the same temperature T. The pressure of the gas in the material is an ill-defined concept, because it depends on the density of the gas, which itself is not uniform at any mesoscopic scale in a porous material. Instead, the chemical equilibrium translates to the equality of the chemical potentials μ of the gas in both conditions, hence μ is the relevant statistical variable to describe the system in the material, which accounts for the fixed gas pressure P out of the material. Finally, we will focus on rigid crystalline materials, hence the volume V of a unit cell can be used as an extensive variable to limit the size of the system.

The statistical ensemble resulting from the constraints of fixing the chemical potential μ, the temperature T, and the volume V is called the grand canonical ensemble. Similar to Equation (1.1) used for the canonical ensemble (where the number of particles N is fixed instead of their chemical potential μ), $\mathcal{O}$ can be expressed using an integral over the microstates $(\mathrm{N}, \boldsymbol{p}_\mathrm{N})$:

$$\mathcal{O} = \langle o \rangle = \frac{1}{\Xi} \sum_\mathrm{N} \int \mathrm{d}\boldsymbol{p}_\mathrm{N}\, o\left(\boldsymbol{p}_\mathrm{N}\right) \exp\left(-\beta\left(\mathrm{U}\left(\boldsymbol{p}\right) - \mu\mathrm{N}\right)\right) \tag{2.1}$$

where N is the number of gas particles, $\boldsymbol{p}_\mathrm{N}$ is their configuration, and Ξ is the grand canonical partition function.

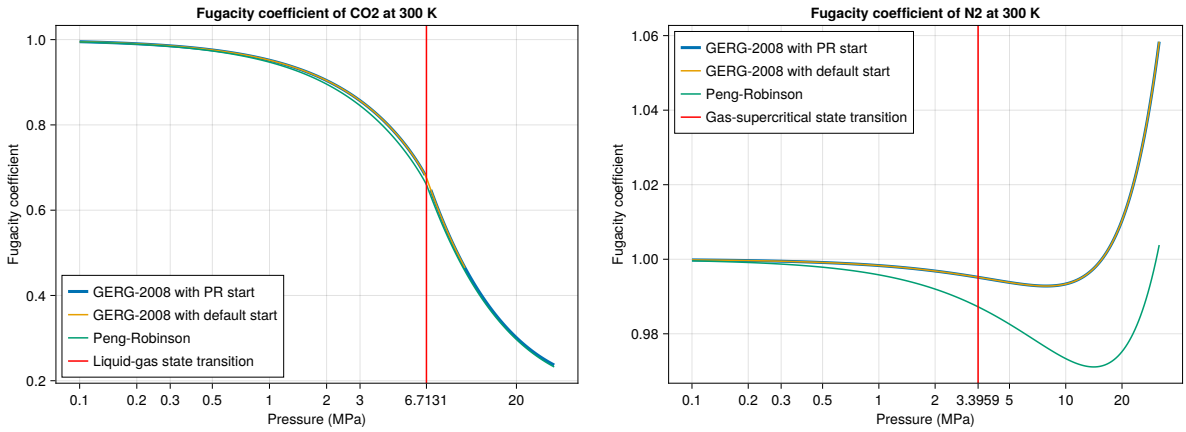
Likewise, this integral can be numerically evaluated by relying on a Monte-Carlo Markov Chain to provide a sequence of configurations $(\mathrm{N}, \boldsymbol{p}_\mathrm{N})$ such that the statistical average $\langle o \rangle$ can be retrieved by computing the arithmetic average of $o\left(\boldsymbol{p}_\mathrm{N}\right)$ over these configurations: this particular kind of simulation is naturally called a Grand Canonical Monte-Carlo (GCMC) simulation. In the specific case of the adsorption capacity, the microscopic function is simply $o\left(\boldsymbol{p}_\mathrm{N}\right) = \mathrm{N}$, thus it is computed by running a GCMC simulation and simply averaging the number of particles in the system across the simulation.

2.2.2 Implementation

In practice, GCMC is implemented like a canonical MC simulation with one additional element: a swap MC move that can either insert or delete a particle. The insertion move simply consists in adding one additional gas particle to the system, while the deletion move removes one. If there are no particle in the systems anymore, the deletion move systematically fails; yet, it is crucial the probability of trying an insertion or a deletion remains equal independently of the number of particles in the system, to ensure micro-reversibility. For this reason, only the probability of trying a swap move is specified as a simulation parameter, while the probability of trying either an insertion or a deletion cannot be individually chosen and are necessarily equal to half the probability of a swap move.

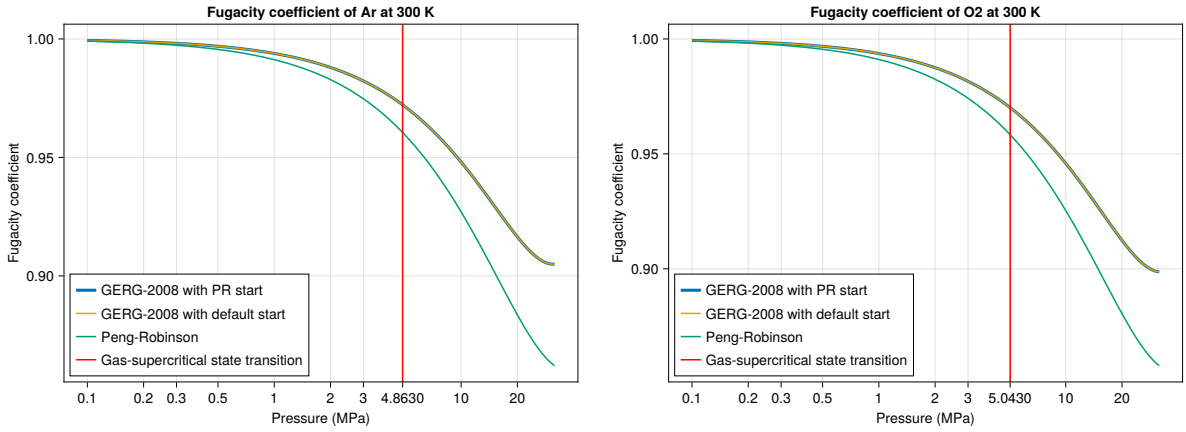
The acceptance probability of a swap move is not decided purely on the basis of the difference of energy in the system, as is the case in Equation (1.2) for the canonical ensemble, because it also depends on the chemical potential μ of the system. With N, the number of particles in the system before the trial swap move, the acceptance probabilities for insertion and deletion are:

$$\begin{aligned} \mathrm{P}_\mathrm{accept}(\Delta\mathrm{E}, \mathrm{N} \to \mathrm{N}+1) &= \min\left(1,\ \frac{\beta\varphi\mathrm{PV}}{\mathrm{N}+1} \exp\left(-\beta\Delta\mathrm{E}\right)\right) \\ \mathrm{P}_\mathrm{accept}(\Delta\mathrm{E}, \mathrm{N} \to \mathrm{N}-1) &= \min\left(1,\ \frac{\mathrm{N}}{\beta\varphi\mathrm{PV}} \exp\left(-\beta\Delta\mathrm{E}\right)\right) \end{aligned} \tag{2.2}$$



(a) CO_2 at 300 K

(b) N_2 at 300 K



(c) Ar at 300 K

(d) O_2 at 300 K

Figure 2.1: Fugacity coefficients of small gases computed with different `Clapeyron.jl` *models*

The dependency to the chemical potential μ is conveyed by the pressure P of the gas exterior to the zeolite. This value is corrected by the fugacity coefficient φ, which accounts for the difference between the real gas and the ideal gas model.

While the pressure P and volume V of the systems are simulation inputs, the fugacity coefficient φ of the gas needs to be computed prior to the simulation. To do so, RASPA uses the Peng-Robinson equation of state, which is adequate for many common fluids. Since our range of gas applications is more restricted than the general-purpose RASPA software, we use the GERG-2008 [10.1021/je300655b] equation of state, which is the ISO standard for natural gases due to its accuracy compared to experimental measurements, with an extended validity range up to 700 K and 70 MPa, in the scope of our study.

In terms of programming, the fugacity coefficient can directly be obtained from the `Clapeyron.jl` [10.1021/acs.iecr.2c00326] Julia package, which provides a convenient unified access to many useful thermodynamic tools. Due to the way it is implemented, Clapeyron.jl used to fail and return `NaN` instead of the valid fugacity coefficient for CO_2 at high pressures: to solve this issue, we do an initial volume computation for the gas in the target conditions using the Peng-Robinson model, and we use the result as the starting point to compute the coefficient with the GERG-2008 model. Figure 2.1a illustrates the differences between these approaches:

the latter allows to compute fugacities at any pressure except those very close to the critical point (304.13 K and 7.3773 MPa for CO_2), while the former simply fails at high pressures (above 1 MPa). The other gases illustrated on Figure 2.1 show that the Peng-Robinson fugacity is usually close but slightly lower than that obtained with GERG-2008, which would result in an equally slight underestimation of the adsorption capacity. The difference between the two equations of state remains in the uncertainty margin of the GCMC simulation however, so it should not lead to a visible difference in the measured adsorption capacity.

2.2.3 Computational aspects

Compared to canonical MC, the efficient implementation of GCMC poses the additional challenge of changing the number of particles across the simulation. This requires almost no accommodation for the computations that operate in the real space, *i.e.* framework precomputed grids, and pairwise interactions (including the direct part of Ewald summation). For these, the only adaptation consists in making the neighbour lists resizable, which is already implemented in `CellListMap.jl`, and otherwise keeping the list of atoms in the system up to date, to correctly check each pair of atoms. The direct part of Ewald summation falls in this category.

The reciprocal-space part of Ewald summation requires adapting the optimization described in Section 1.3.4 to account for the possible introduction or removal of a species. For such a move, both values of $S_{\boldsymbol{k}} = \sum_n q_n e^{i\boldsymbol{k}\cdot\boldsymbol{r}_n}$ before and after the swap need to be computed, exactly like in Equation (1.21). Likewise, the value before the swap is already stored in memory, but the computation of the new value is slightly different: instead of computing the difference as $\sum_n q_n \left(e^{i\boldsymbol{k}\cdot\boldsymbol{r}'_n} - e^{i\boldsymbol{k}\cdot\boldsymbol{r}_n} \right)$, the difference is now $\pm \sum_{n \text{ swapped}} q_n e^{i\boldsymbol{k}\cdot\boldsymbol{r}_n}$. In the case of a deletion, each element of the latter is already stored in memory, whereas in the case of an insertion, this new computation is stored in a temporary place, like for any other trial move, ready to be stored permanently if the move is accepted.

The self-interaction term is no longer constant: each new molecule contributes a term equal to $\sum_m \frac{q_n q_m}{4\pi\varepsilon_0} \operatorname{erf}\left(\|\boldsymbol{r}_n - \boldsymbol{r}_m\|_{\mathrm{p}} \right) / \|\boldsymbol{r}_n - \boldsymbol{r}_m\|_{\mathrm{p}}$ where m iterates over the atoms of the molecule. This term is computed once for each kind of molecules and stored, so that the contribution of the self-interaction term can be updated with a single addition or subtraction when a swap move occurs. Of course, any molecule whose number can vary must be electrically neutral to ensure the the overall system remains electrically neutral; this is also enforced in the code.

Another difference is the tail correction, which is not constant anymore either.

Overall, having a variable number of species requires extra bookkeeping, to make sure that all atoms are accounted for. To maximize performance in current computers, the data structures must have good cache locality: this means that, when memory needs to be read, it should be done in contiguous accesses as much as possible. For this purpose, the positions of the atoms, the $\kappa_{\boldsymbol{k}}$ factors, the precomputed $q_n e^{i\boldsymbol{k}\cdot\boldsymbol{r}_n}$ and $S_{\boldsymbol{k}}$, the self-interaction terms, etc., are all stored in tight arrays, with the appropriate indexing for each. When an atom is removed, there are two choices: either leave a gap in the array, and keep track separately of which parts of the array actually refer to a valid data if need be, or move the contents of the array to fill the gap.

Both algorithmic strategies have their merit, and warrant a trial to check which performs best. We did not take the time to implement both however, and instead chose the latter systematically, because it is slightly simpler to reason about. In practice, such an approach consists in replacing the content of the gap in the array by the data at the end of the array, before reducing the size of the array, and thus removing the temporarily duplicated data at the end. The indices relative to the content of the array also need to be updated. The complexity of each array update is thus proportional to the number of elements to remove, which is generally optimal. Insertions are operated at the end of the arrays, with an optimal complexity.

2.3 GCMC on a grid

Even with an efficient implementation, regular GCMC simulation remain resource-consuming, especially in the context of a large scale material screening. Improving this performance is thus valuable, especially if the precision cost can be controlled.

2.3.1 Principle

Similar to the methods exposed in Section 1.3.4, a typical strategy for reducing the computational cost consists in precomputing as much as possible, to decrease the workload of the main loop of the algorithm. In the case of GCMC, the main computation occurring at each move trial is that of the interactions between species, both in direct and reciprocal spaces. Since the configuration space for the molecules in the unit cell of the framework is continuous, and thus infinite, it is impossible to precompute all the possible pair interactions.

However, that space can be discretized: to do so, the mobile species are forced to occupy one position each among a fixed set of possibilities. The conceptually simplest set is a regular grid: each molecule must reside atop one grid point. For polyatomic species, like all gases of interest with the exception of rare gases, this definition is incomplete however. There are thus three options:

1. the grid can include additional dimensions to account for the orientation of the molecule. In that case there must be a fixed convention dictating which atom of the molecule is placed on the grid point. That atom, called the “bead”, is typically that around which the rotation MC move operates.

2. each grid point can come with an associated fixed orientation of the molecule sitting on it. In that case, the lowest energy orientation with respect to the framework is probably the most reasonable choice, but it may grossly misrepresent the actual orientation of the molecules if those are mostly driven by the interaction with other molecules.

3. the molecule can be represented as a single point sitting on the grid, with no orientation. In that case, the interaction energy of the molecule with the framework can be taken as the average of the energies corresponding to the different orientations and, likewise, the molecule-molecule interactions can be computed by taking averages over the orientation of both molecules. These averages could be either arithmetic or Boltzmann averages, depending on the interpretation: a Boltzmann average implicitly assumes that the molecules are at equilibrium, which is not very well defined for a single molecule sit-

ting on a grid point, while an arithmetic average implies has an orientation independent of the corresponding energy, which is not exact either.

Our study is restricted to small gases of the air which are all linear and centrosymmetric: CO_2, N_2, O_2, etc., so the bead is conventionally chosen as the center of the molecule. Choosing an orientation thus amounts to choosing a direction for a rod (the molecule) across possible angles. To accurately cover this configuration space while limiting the number of used angles, we use a Lebedev grid, which is a set of N_l points (θ_i, φ_i) on the surface of the sphere such that the integral of a function f can be well approximated by the sum

$$I[f] = \int_0^{\pi} sin(\theta) d\theta \int_0^{2\pi} d\varphi f(\theta, \varphi) \approx 4\pi \sum_{i=1}^{N_l} w_i f(\theta_i, \varphi_i) \tag{2.3}$$

with the appropriate weights w_i. The values of w_i, θ_i and φ_i have been precomputed once and for all for different values of N_l. Because of orthogonality constraints on the Lebedev points, not all values of N_l are possible: the ones which are usable in my code are N_l = 6, 14, 26, 38, 50, 74, 86, 110, 146, 170, 194, 230, 266, 302, 350, 434, 590, 770, 974, 1202, 1454, 1730, 2030, 2354, 2702, 3074, 3470, 3890, 4334, 4802, 5294 and 5810, allowing for any fine-tuning of the precision level. Only half of these points are actually used since the both the target gases and the Lebedev grids are centrosymmetric.

The number of possible configurations of a molecule is $\alpha \times a \times b \times c$ where (a, b, c) is the size of the spatial grid and α is either N_l in the first of the three options above, or 1 for the second and third option. Since we use a Van der Waals cutoff of 12 Å, the framework with the smallest grid will necessarily have its perpendicular lengths greater or equal to 2×12 Å = 24 Å. Thus, with a distance of 0.2 Å between two grid points, a, b and c are greater or equal to than 120 so the size of the configuration space is at least $\alpha \times 120^3$.

The goal of the grid MC approach is to reduce the amount of computations done at each step, to accelerate the GCMC simulations, by precomputing the pair interactions between grid points. The number of pair interactions to precompute is thus, in theory, at least $\left(\alpha \times 120^3\right)^2/2 = \alpha^2 \times 120^6/2$. It can actually be reduced further with translation symmetry considerations: the interaction between grid points $(x_1, y_1, z_1, \theta_1)$ and $(x_2, y_2, z_2, \theta_2)$ is identical to that between grid points $(x_1 + \Delta, y_1 + \Delta, z_1 + \Delta, \theta_1)$ and $(x_2 + \Delta, y_2 + \Delta, z_2 + \Delta, \theta_2)$ for any Δ. It is thus sufficient to precompute the interaction between $(0, 0, 0, \theta)$ and all other grid points, so the minimum number of precomputations is actually $\alpha^2 \times 120^3 \approx \alpha^2 \times 2 \cdot 10^6$. Each precomputation yields an energy, represented by a floating-point number, which takes 8 B of space, and the typical RAM of a computer is on the order of 10 GB, so the maximum value of α is approximately $\sqrt{10^{10}/\left(8 \times 2 \cdot 10^6\right)} = 25$ for the smallest possible spatial grid.

In conclusion, using explicit angles is theoretically doable only for small enough materials and by using the smallest numbers of angles, otherwise the precomputation results cannot be stored in memory. Moreover, we expect the numerical cost of the precomputations to be considerable itself, since it requires $\alpha^2 \times 2 \cdot 10^6$ pair interaction computations. This is similar to the cost of simulating a system of $N = \alpha^2$ particles in interaction across one million MC steps, which is a typical GCMC simulation length, since each MC trial move requires computing 2N pair interactions (N for the particle before the move, N for the particle after).

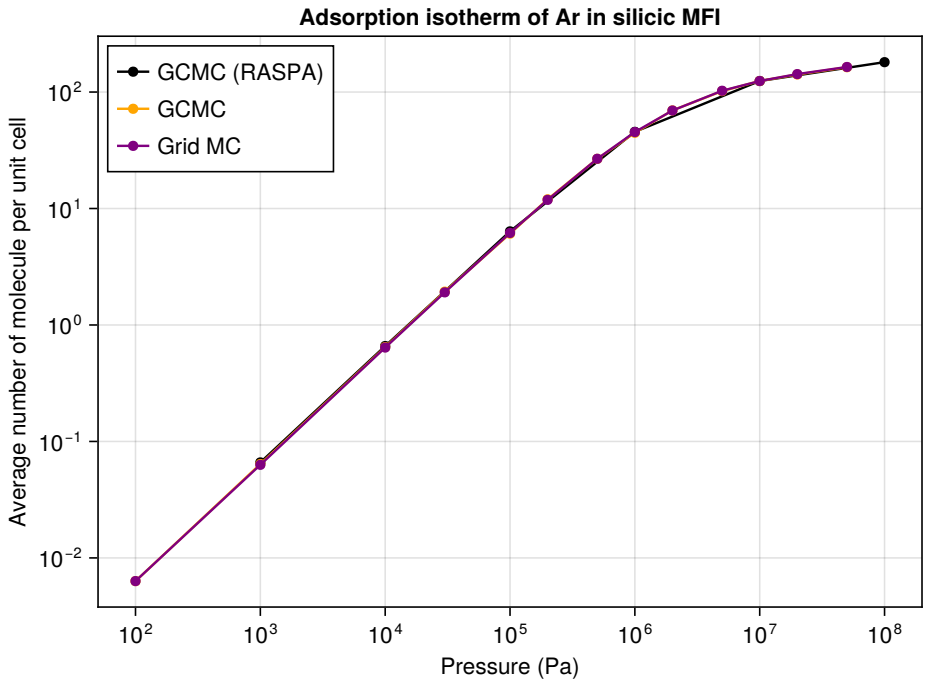


Figure 2.2: Adsorption isotherm of Ar on silicic MFI obtained with different simulation strategies. The black line is the RASPA implementation of GCMC while the orange line is our implementation.

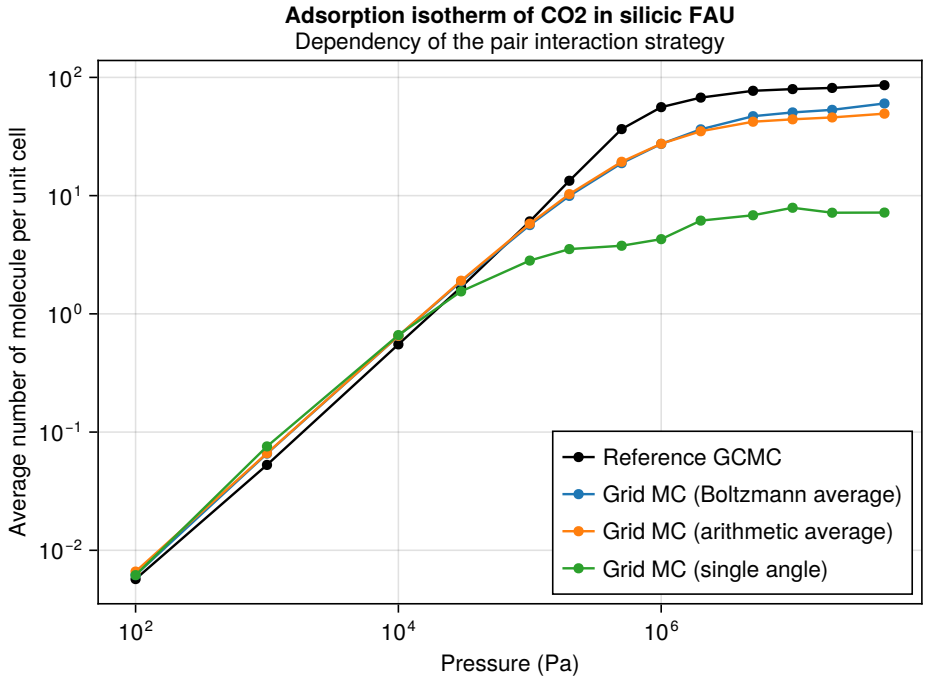
2.3.2 Validation

To check the validity of this simulation strategy, we start by comparing the loading of Ar in a zeolite obtained by regular GCMC and through this grid MC approach. The result for zeolite MFI is presented on Figure 2.2, using a grid step size of 0.2 Å. The perfect superposition of the three isotherms indicate that the grid MC simulation method is apt to reproduce the adsorption property of noble gas in zeolites with great accuracy, even at high pressure.

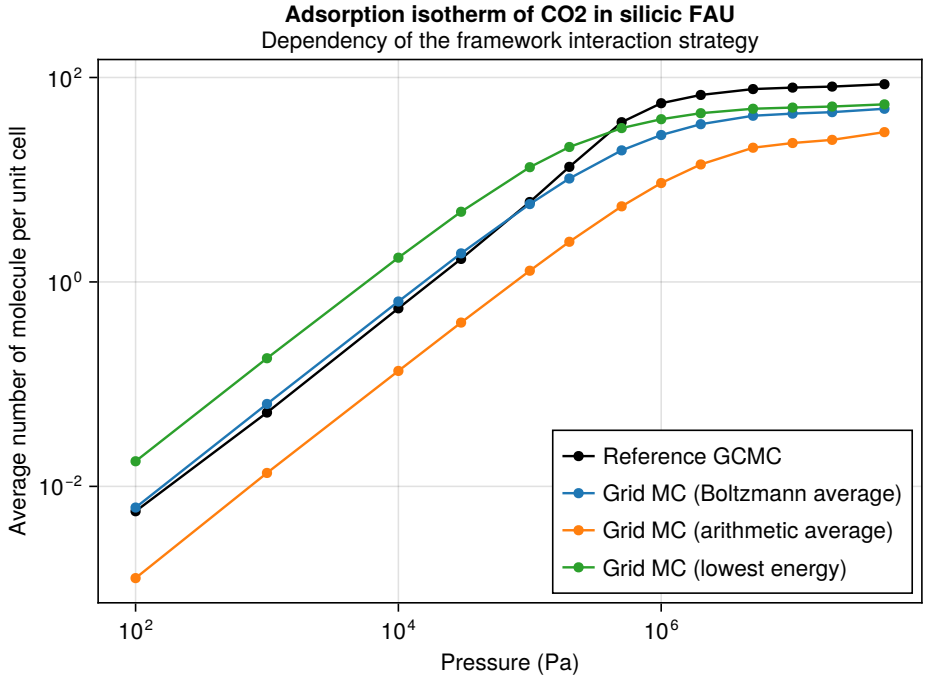
With more complex molecules, angles have to be taken into account, using either of the two last options among the three presented before. Their comparison is illustrated with the adsorption of CO_2 in FAU in Figure 2.3a, shows that averaging orientations (option 3) provide more qualitative results compared to selecting one single orientation per grid point (option 2). Yet, none manage to qualitatively reproduce the adsorption behavior at intermediate to high pressures.

As evoked previously, the nature of the angular averaging, either arithmetic or Boltzmannian, of energies stems from a hypothesis on the distribution of microstates, neither of which is actually valid. The influence of this choice on the simulated loading of CO_2 in FAU is presented in Figure 2.3.

In Figure 2.3a, it appears that the choice of averaging for the molecule-molecule interaction has little influence on the result until the high pressures. Unsurprisingly, the Boltzmann average, which favors attractive interaction, increases the loading compared the arithmetic average, but the general trend is very similar and neither appear to be particularly more right than the other.



(a) Influence of the molecule-molecule interaction strategy. "single angle" refers to option 2, while for option 3, the fixed framework-molecule interaction strategy is the Boltzmann average.



(b) Influence of the framework-molecule interaction strategy with option 3. The fixed molecule-molecule interaction strategy is the arithmetic average. "lowest energy" refers to taking the minimum instead of taking an average.

Figure 2.3: Adsorption isotherm of CO_2 on silicic FAU obtained with different grid MC strategies.

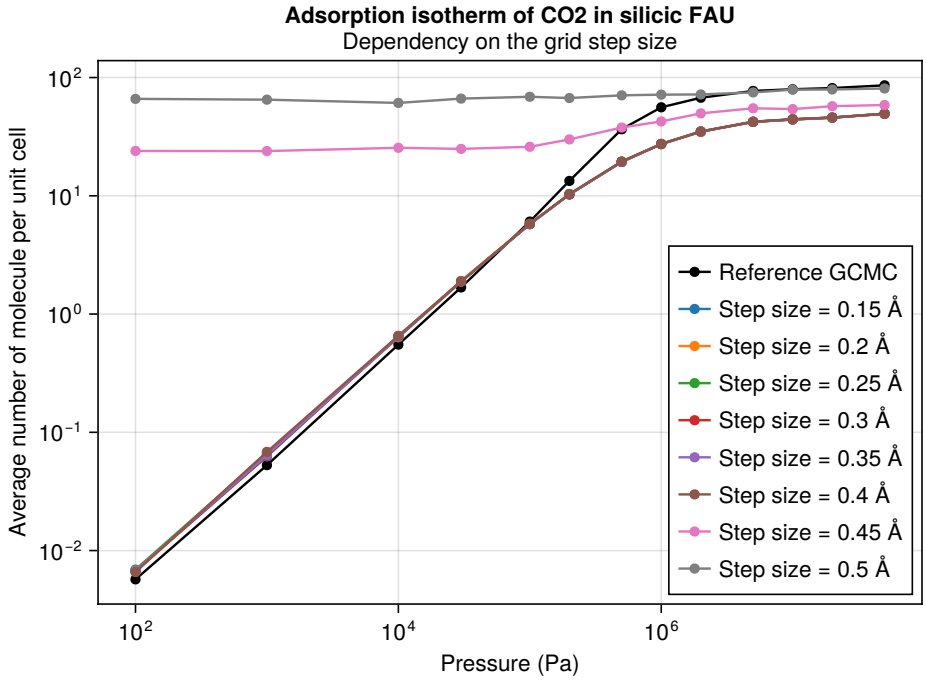


Figure 2.4: Adsorption isotherm of CO_2 on silicic FAU obtained with different grid MC step sizes

On the other hand, Figure 2.3b illustrates the drastic difference in the choice of averaging for the framework-molecule interaction. Only the Boltzmann average yields an isotherm closely following that of the reference GCMC simulation at low pressures: the arithmetic averaging of angles yields a large underestimation of the loading, while taking only the lowest energy across angles yields a large overestimation.

From these observations, it appears that the molecules of CO_2 preferentially orient themselves according to the interactions with the framework at low pressure, which should be obvious since low pressures correspond to a regime where molecule-molecule interactions are negligible. More importantly, irrespective of the pressure regime, the entropic contribution is paramount: the molecules do not simply orient themselves according to the lowest energy configuration, but freely rotate in accordance with Boltzmann distribution.

Regarding the numerical method itself, the main limit is the behaviour at intermediate and high pressure, which significantly parts from expectation. While this necessarily stems from the use of a grid, it does not depend on the grid step size: Figure 2.4 shows the evolution of the isotherm using option 3 with Boltzmann averaging for framework-molecule interactions and arithmetic averaging for molecule-molecule interactions. The isotherm does not evolve much when the grid step size is small enough, which points to the fact the issue lies in the nature of the simulation method itself.

2.3.3 Numerical aspects

3

DATABASE APPROACH

3.1 Screening studies

Research on the adsorption properties of zeolites in particular, and nanoporous materials in general, aims at understanding the nature and respective importance of the different interactions that drive the adsorption phenomenon. This understanding can afterwards be turned into principles, which are then applied to predict the adsorption capacities of similar materials on similar gases, where these principles remain valid.

The natural starting point for this research is a study of a particular material with a particular gas, which leads to particular observations that may then be generalized. Many such studies have been performed and published on zeolites, each involving a considerable amount of work, which provide very precise results on a given topology, with up to a few different Si/Al ratios. Yet, the conclusions can seldom be carried over to provide predictions on other topologies.

The converse strategy consists in doing screening studies, in which a vast array of structures is tackled at the same time. Given the cost of doing so experimentally, these studies are mostly computational, thus providing less precise results than actual experiments, but on a broader scale.

Most zeolite screening studies focus on all-silica zeolites, an approach that sidesteps the issue of placing cations. However, most zeolite adsorption applications require cationic zeolites, since cations critically increase the adsorption capacity of the materials.

Even among the few studies which tackle cationic zeolites across topologies and Si/Al ratios, not all properly place cations in the framework, although it is known that this can lead to incorrect subsequent adsorption measurements.[1] For example, [2] targets 57 different topologies with 5 different possible cations, but fails to mention where these are placed before the start of the Grand Canonical Monte-Carlo (GCMC) simulations, likely resulting in them being trapped in a non-optimal placement. [3] also screens cationic zeolites, but places cations at the so-called "minimum energy positions" instead of using a robust method like parallel tempering; they also use a force field made for FAU zeolites only, which is inadequate to screen other zeolites. We are only aware of the works of [1], which tackles many zeolite with various Si/Al ratios for screening purposes, that properly places cations through parallel tempering.

In this chapter, we explain how we constitute a database of cationic zeolite structures and how it can be used for isotherm prediction.

3.2 Adsorption isotherm

3.2.1 Definition and classification

The behaviour of a material with respect to the adsorption of a gas is typically assessed by measuring the evolution of its adsorption capacity with pressure, at a fixed temperature. The resulting curve is called an isotherm, and it is usually classified according to its general shape into one of several possible categories, represented on Figure 3.1.

The shape of the isotherm depends on the surface of the materials, the strength of its interaction of the adsorbate and of the adsorbate-adsorbate interactions. For example, type I isotherms are often observed on materials with very small external surface and monodisperse pores of small size, where adsorption quickly reaches saturation. On the other hand, type III isotherms occur when the adsorbent-adsorbate interaction is weak and adsorption is not limited by the pore size (either because adsorption occurs on the external surface or because the pores are very large). Some of these types exhibit hysteresis loops: for example, type IV(a) isotherms, typical of mesoporous adsorbents, have the adsorption branch lower than the desorption branch (when the gas is withdrawn), which comes from the capillary condensation of the adsorbate at high pressure. Real isotherms are often combinations or slight variations of these idealized types.[4]

At very low pressure, all isotherms are expected to be linear: this is Henry's law, and Henry's adsorption constant K is proportional to the slope of the isotherm. In that limit, all adsorbates behave like ideal gases since, by definition, there is no adsorbate-adsorbate interaction at low enough pressure.

The other end of the pressure spectrum is more complex. When adsorption occurs on the external surface of the adsorbent, for instance with clays, and if adsorbent-adsorbent interactions are strong enough to allow multi-layer adsorption, then the number of adsorbed species never stops growing with increasing pressures. However, with zeolites and more generally any

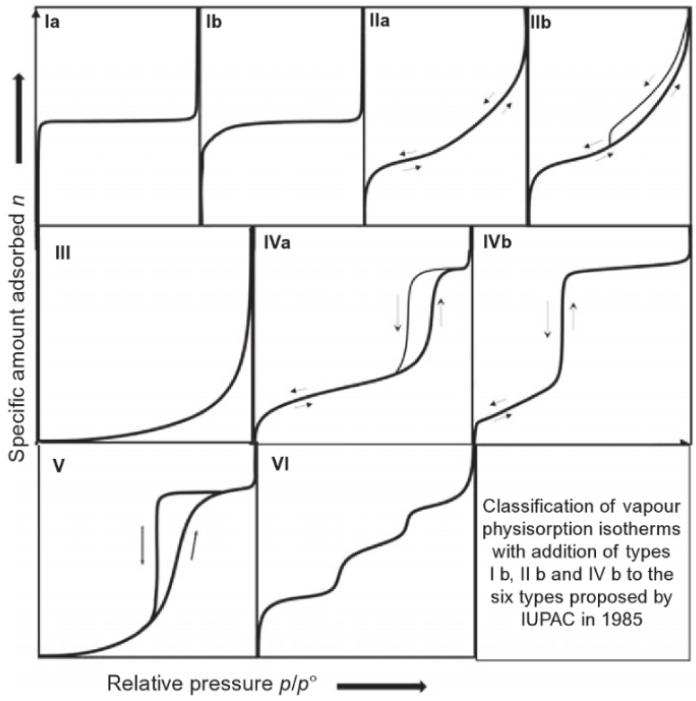


Figure 3.1: Classification of gas isotherm by IUPAC, taken from [5], adapted from [6]. p^0 designates the saturation pressure of the adsorbate.

adsorbent where adsorption occurs in finite pores, the adsorption capacity plateaus at high pressure when all the pores are filled with the adsorbate. It should be mentioned though that the adsorption capacity can still increase beyond that limit, because of the compressibility of the adsorbate in liquid or supercritical state. These regimes of extremely high pressure (beyond 100 MPa) are of lesser industrial relevance because reaching these conditions is costly and hazardous. They are, however, visible in simulations.

3.2.2 Experimental observation and numerical prediction

Experimentally, adsorption isotherms are measured by progressively increasing the input pressure of the adsorbate in the material at the fixed temperature, and measuring the amount adsorbed. The desorption isotherm follows up by progressively decreasing the input pressure, and similarly measuring the amount of adsorbate.

The measurement itself usually consists in either of the three following methods:[4]

- volumetry: a fixed amount of gas at the target pressure is put in contact with the adsorbent, and what is measured is the difference of pressure before and after adsorption.
- breakthrough: the gas starts flowing through the adsorbent at a given time, and the outlet concentration and volumetric flowrates are measured.
- gravimetry: the difference in mass of the sample containing the adsorbent before and after adsorption is measured by balancing its weight against buoyancy.

By reproducing the experiment with a reference gas (usually N_2 or Ar), these techniques give access to the net adsorption, which is the absolute amount of adsorbed species minus that which would be in a fluid occupying the same space as the adsorbent at the same pressure and

temperature. Knowledge of the volume of adsorbent allows retrieving the absolute adsorption capacity, which is directly accessible in simulations.

Some more anecdotic isotherm measurement techniques including impedance spectroscopy as well as NMR allow obtaining directly the absolute adsorption,[4] but these methods are much too costly to be used routinely.

The adsorption capacity can be predicted numerically through GCMC simulations, as explained in Section 2.2. Beyond the limits due to the precision of the energy computation, either from a force field or from DFT, and convergence of the simulations, those are also dependent on a model of the adsorbent, which cannot take into account the complexity of real materials. In the case of zeolites, the location of the aluminium and of the cations has already been discussed through Chapter 1, but more generally, all kinds of defects and small crystal size effects can lead to experimental isotherms differing from the simulated counterparts. This limits the accuracy of individual numerical simulation compared to precise experimental adsorbents. They remain useful however to extract tendencies and direct the search for new adsorbents, which will be experimentally tested eventually.

3.3 Database constitution

Finding general tendencies in the adsorption behaviour across zeolites requires doing the analysis of many different structures. To do so, the first step consists in establishing a database of models for the materials themselves. The general workflow for creating these structures, starting from the idealized zeolite structure defined by its topology, down to the cation placement, has already been described in Chapter 1. To summarize, the two steps are:

1. Aluminium placement. Start by finding the minimum Si/Al ratio by replacing as many silicium with aluminium atoms as possible, with random exchanges and restarts. Check multiple supercell sizes. Then, starting from these filled structures, replace some aluminium by silicium atoms at random and do more exchange steps to generate 6 aluminium placements for each target Si/Al ratio.

2. Cation placement: using either parallel tempering or shooting star simulations. The only cation used in the current study is sodium.

For the entire database, we used the same force field which is that developed by [7] for adsorption of small gases on cationic zeolites with mobile cations.

3.3.1 Aluminium placement

The current database contains 6 aluminium placements for all 239 known and not interrupted zeolite topologies, with Si/Al ratios of 1, 1.23, 1.4, 1.7, 2, 3, 5 and 10. The actual Si/Al ratios of each structures are those closest to the previous references, which may sometimes differ. When the minimum Si/Al ratio (presented in Table 1.1) is not among the previous references, it is added separately, and the lower Si/Al ratios cannot be generated. For example, the list of actual Si/Al ratios for topology YUG is 1.56, $\frac{121}{71} \approx 1.704$, 2, 3, 5 and $\frac{349}{35} \approx 9.971$.

The algorithm sometimes fails at finding 6 distinct aluminium placements: in that case, the maximum number of distinct placements is kept. To check whether a new placement is distinct

from previous ones, the following algorithm is used. First, for each topology, the list of T-sites is mapped to the list of integers between 1 and N, the number of T-sites of the topology. Each placement is represented by a sequence of N bits, where a bit of 1 at position i indicates an aluminium in T-site i and 0 stands for silicium. Then, all M symmetries of the topology are applied to the placement: this results in M N-bit sequences. The (lexicographically) smallest among those serves as a unique signature that identifies the aluminium placement. When a new placement is found, is signature is computed as previously and compared to those already encountered: if it is different from all of them, the new placement is stored.

In the case where Si/Al = 1 for example, there can be up to only 2 distinct aluminium placements that obeys Löwenstein's rule: taking an arbitrary T-atom as a reference, one placement corresponds to that reference set to Al, and the other to that reference set to Si. For many topologies, these two placements are actually symmetry-equivalent, so they share the same signature and only one of them is actually stored. This is not necessarily the case however: for example, topology JSW has two distinct aluminium placements for Si/Al = 1.

Overall, 8983 different structures with aluminium placed were generated.

3.3.2 Cation placement

For all 239 known and non-interrupted topologies, and for each of the up to 6 different aluminium placements corresponding to the minimum Si/Al, a shooting star simulation was launched with the hot run at 2000 K for 20 000 cycles (and 2000 initialization cycles), spawning one cold simulation every 100 cycles. Each of the resulting 200 cold simulations ran at 300 K for 10 000 cycles. This served as an experiment to assess the convergence of the simulations: out of the 239 topologies, 57 converged in less than 5000 cycles (averaged across aluminium placements). For these topologies, shooting star simulations were run for all previously created aluminium placements across the different Si/Al ratios, using the same parameters (except for the number of cold cycles raised to 15 000). For each shooting star simulation, the 200 cold simulations are divided into six groups, and for each of these groups, the structure with the lowest corresponding energy is kept. Hence, this methodology yields 6 cation placement per structure.

Overall, 14 268 different structures with both aluminium and cations placed were obtained.

3.4 Prediction

Using the structures from the database, it is possible to study the adsorption behaviour of zeolites at scale, across topologies and Si/Al ratios. To do so, one simply needs to run a GCMC simulation for each combination of zeolite structure (including Si/Al ratio, aluminium placement, nature and placement of the cations), temperature, pressure, and gas.

In order to investigate the relationship between the topology of zeolites and their adsorption properties, we attempted to build a predictive model that could output a plausible isotherm for each combination of three parameters, which are 1) the zeolite topology, 2) the Si/Al ratio and 3) the temperature, using Na^+ as cation and either CO_2 or N_2 as gas.

3.4.1 Why isotherms?

Adsorption isotherms are often used for the characterization of experimental structures. The technique consists in measuring the isotherm using one of the methods presented in Section 3.2.2, then fitting the obtained curve against a set of pre-computed isotherms, called the kernels. Each kernel is computed based on a pore geometry (*i.e.* its shape, like a slit or a sphere, and its size) and the thermodynamical properties of the gas, often modeled using simple fluid theory. Fitting the isotherm then corresponds to finding coefficient on each kernel such that the sum of the kernels weighted by the coefficients yields the best approximation of the isotherm. The best fit thus provides the repartition of pore geometries based on their corresponding coefficients.

Another typical context of use for isotherm fitting is for the identification of a material, or the evaluation of its purity or degree of crystallization. To do so, a simple method consists in measuring one or more isotherms and comparing them to those obtained on a reference structure, to check if they match. In general data science, two curves can be compared by computing their root-mean-square deviation (RMSD), but this does not provide much physical meaning to the comparison, and requires that both isotherms have points measured at the same pressures. More often, the reference isotherm has been fitted against one model, hence the isotherms of the experimental structure can be fitted against the same model, so that the coefficients of the model, which carry some physical meaning, can be compared.

In our setting, the goal is quite different since we aim at predicting the adsorption capacity from a theoretical structure, instead of an experimental one. The concept of isotherm is actually not strictly necessary to this goal: one could imagine making a black-box model that directly outputs the adsorption capacity given the three parameters (topology, Si/Al ratio, temperature) as well as the gas pressure. This approach, however, leads to a model that cannot be physically interpreted, and that offers no guarantee on the general shape of the isotherms, with the risk of them being unphysical. On the contrary, making a model that directly predicts the coefficients of a given isotherm model from the three parameters yields a result that already embeds the physics of adsorption through the isotherm mode. Of course, it also offers the additional guarantee of a smooth and plausible interpolation of the simulated data points across pressures.

To make such a model, the first step thus consists in fitting the simulated isotherm against a given adsorption model, in order to obtain the coefficients associated with each triplet of parameters. Retrieving the relation between the parameters and the coefficient can then be cast as a data science problem.

3.4.2 Adsorption models

The literature is rife with adsorption models, either grounded in theory or empirical ones. The precise choice of a model for an isotherm stems from nature of the adsorption when known, otherwise the shape of the isotherm, the nature of the material or even the scientific community. Since we use isotherm models simply as a mean for extracting a common mathematical descriptor of the isotherms across numerous settings, our only criterion is that the model should be rich enough to accurately represent the isotherms. Beyond that, the simpler the model, *i.e.* the smaller the number of coefficients necessary to represent an isotherm, the better it will be for prediction.

Next are the isotherm models which were investigated. This list is not exhaustive, but it covers a sizeable proportion of the most commonly used models for small gas adsorption in zeolites.[8]

Langmuir

The simplest adsorption model corresponds to a physical situation where the adsorbate behaves as an ideal gas, and binds with the adsorbent in a reversible reaction of equilibrium constant β. In that setting, the Langmuir adsorption model states that the fraction of occupied adsorption sites is $\beta P/(1 + \beta P)$ where P is the pressure, hence the total adsorption capacity is:

$$n_{\text{Langmuir}}(P) = \alpha \frac{\beta P}{1 + \beta P} \tag{3.1}$$

where α is the maximum population of the site.

This very simple model is not accurate for zeolites however. One of the reasons is that zeolites contain multiple adsorption sites, both spatially – there are multiple adsorption reactions happening simultaneously – and energetically – they have different values of β.

N-site Langmuir

One simple way to circumvent the previous issues consists in simply summing multiple Langmuir adsorption models. This physically assumes some kind of independence between the different adsorption sites, which is debatable, and the general formula is

$$n_{\text{N-site Langmuir}}(P) = \sum_{i=1}^{N} \alpha_i \frac{\beta_i P}{1 + \beta_i P} \tag{3.2}$$

Such a general model can be made to fit any isotherm, simply by increasing the number of sites N until reaching a high enough complexity of the model. In practice, N = 3 is good enough to fit many isotherms, simply because the fit can be optimized across six coefficients, although the underlying physical model may not be accurate at all.

Freundlich

This simple empirical model is commonly used when the adsorption is known to happen on a heterogeneous surface, and depends on two parameters A and γ:

$$n_{\text{Freundlich}}(P) = A \times P^{\gamma} \tag{3.3}$$

Beyond its simplicity, the main limit of the model is its behaviour at high pressure: the model diverges to $+\infty$ whereas, in practice, adsorption reaches saturation (before the highly condensed phase).

Sips

The Sips model combines both Langmuir and Freundlich isotherms, with the following formula:

$$n_{\text{Sips}}(P) = \alpha \frac{(\beta P)^{\gamma}}{1 + (\beta P)^{\gamma}} \tag{3.4}$$

This model has two limit regimes: at low pressure, it is equivalent to the Freundlich model while at high pressure, it reaches a plateau similarly to the Langmuir model. It also reduces to the Langmuir model in the particular case of $\gamma = 1$.

Toth

The Toth model is another empirical variation of the Langmuir model:

$$n_{\text{Toth}}(\text{P}) = \alpha \frac{\beta \text{P}}{\left(1 + (\beta \text{P})^{\gamma}\right)^{1/\gamma}} \tag{3.5}$$

Like the Sips model, it reduces to the Langmuir model for $\gamma = 1$, and reaches a plateau at high pressure. A slight difference with the Sips model is the adsorption capacity at low pressure, which becomes proportional to P and thus obeys Henry's law, whereas both Freundlich and Sips models have a dependency in P^{γ}.

Redlich-Peterson

Like the Sips model, the Redlich-Peterson model combines both Langmuir and Freundlich models:

$$n_{\text{Redlich-Peterson}} = \alpha \frac{\text{P}}{1 + (\beta \text{P})^{\gamma}} \tag{3.6}$$

It also follows Henry's law by reducing to a linear function of P at low pressure, but its behaviour at high pressure is similar to Freundlich's, making it only adapted to pressure regions below the saturation point.

Jensen-Seaton

By adding a fourth parameter, Jensen and Seaton propose an adsorption model that present both Henry's law linear behaviour at low pressure, as well as a non-constant affine behaviour at high pressure that models the compressibility region.

$$n_{\text{Jensen-Seaton}} = \alpha \beta \text{P} \times \left(1 + \left(\frac{\beta \text{P}}{1 + \delta \text{P}}\right)^{\gamma}\right)^{-1/\gamma} \tag{3.7}$$

This model reduces to the Toth model when $\delta = 0$ and to the Langmuir model when $\gamma = 1$ (with second parameter $\beta' = \beta + \delta$).

Sips-Toth

A custom-made adsorption model, absent from the literature, consists in using a fourth parameter ζ to act as a continuous switch between the Toth model ($\zeta = 0$) and the Sips model ($\zeta \to +\infty$):

$$n_{\text{Sips-Toth}} = \alpha \times \left(\frac{(\beta \text{P})^{\gamma}}{1 + (\beta \text{P})^{\gamma}}\right)^{\frac{1+\gamma\zeta\text{P}}{\gamma(1+\zeta\text{P})}} \tag{3.8}$$

Like its two building blocks, this model collapses to a simple Langmuir isotherm when $\gamma = 1$, and reaches a plateau at high pressure. It also obeys Henry's law at low pressure.

Summary

Table 3.1 summarizes the key properties of the previously presented isotherm models. The number of coefficients characterizes the complexity of the model. Its behaviour at low pressure should obey Henry's law, while it should be affine at very high pressure to correctly model the compressibility phase, or at least behave as a plateau otherwise to account for the stationary phase at high pressures.

Name	Coefficients	Henry's law at low P	Behaviour at high P
Linear	1	No	Affine
N-site Langmuir	2N	Yes	Plateau
Freundlich	2	No	Polynomial
Sips	3	No	Plateau
Toth	3	Yes	Plateau
Redlich-Peterson	3	Yes	Polynomial
Jensen-Seaton	4	Yes	Affine
Sips-Toth	4	Yes	Plateau

Table 3.1: Main characteristics of different adsorption models

Several of these models can be combined by simply adding them together. For this purpose, one extra dummy model was added, called "Linear", which simply consists in doing a linear regression of the isotherm:

$$n_{\text{Linear}} = \eta \times \text{P} \tag{3.9}$$

Of course, this model does not fit any actual isotherm by itself, but it can be added to any model that reaches a plateau at high pressure to fix its behaviour for the compressibility phase, by making it affine as expected.

The result of isotherm fitting according to all of these models, with and without linear addition, is represented in Figure 3.2.

The quality of the fit generally increases with the number of parameters, simply because of the increased flexibility of the model. Similarly, adding one extra linear parameter improves the quality of the overall fit, in particular at high pressure, since it corresponds to the linear compressibility regime. The case of the Jensen-Seaton adsorption model complements this point: there is no improvement in the fit due to the linear parameter in that case, because Jensen-Seaton isotherms already behave linearly at high pressure.

However, not all models perform equivalently. For example, even with three sites and thus six parameters, Langmuir fits very poorly compared to the Sips-Toth model plus linear term, which is the overall best and only require five parameters. With only four parameters, Toth plus linear is equivalent to Jensen-Seaton. Both Freundlich and Redlich-Peterson fail to fit because of the polynomial behaviour at high pressure.

3.4.3 Isotherm fitting

In practice, fitting an isotherm means finding the parameters α, β, γ, ... of the model such that the theoretical isotherm resulting from the model is as close as possible to the input one. This closeness can be properly defined through the RMSD for instance. Finding the best parameters can thus be formulated as an optimization problem, *i.e.* a problem of the form

$$\text{minimize } f(\alpha, \beta, \gamma) \quad \text{such that } \alpha > 0, \beta > 0, 0 < \gamma < 5 \tag{3.10}$$

for example, where f represents the deviation between the model with the given parameters and the actual isotherm.

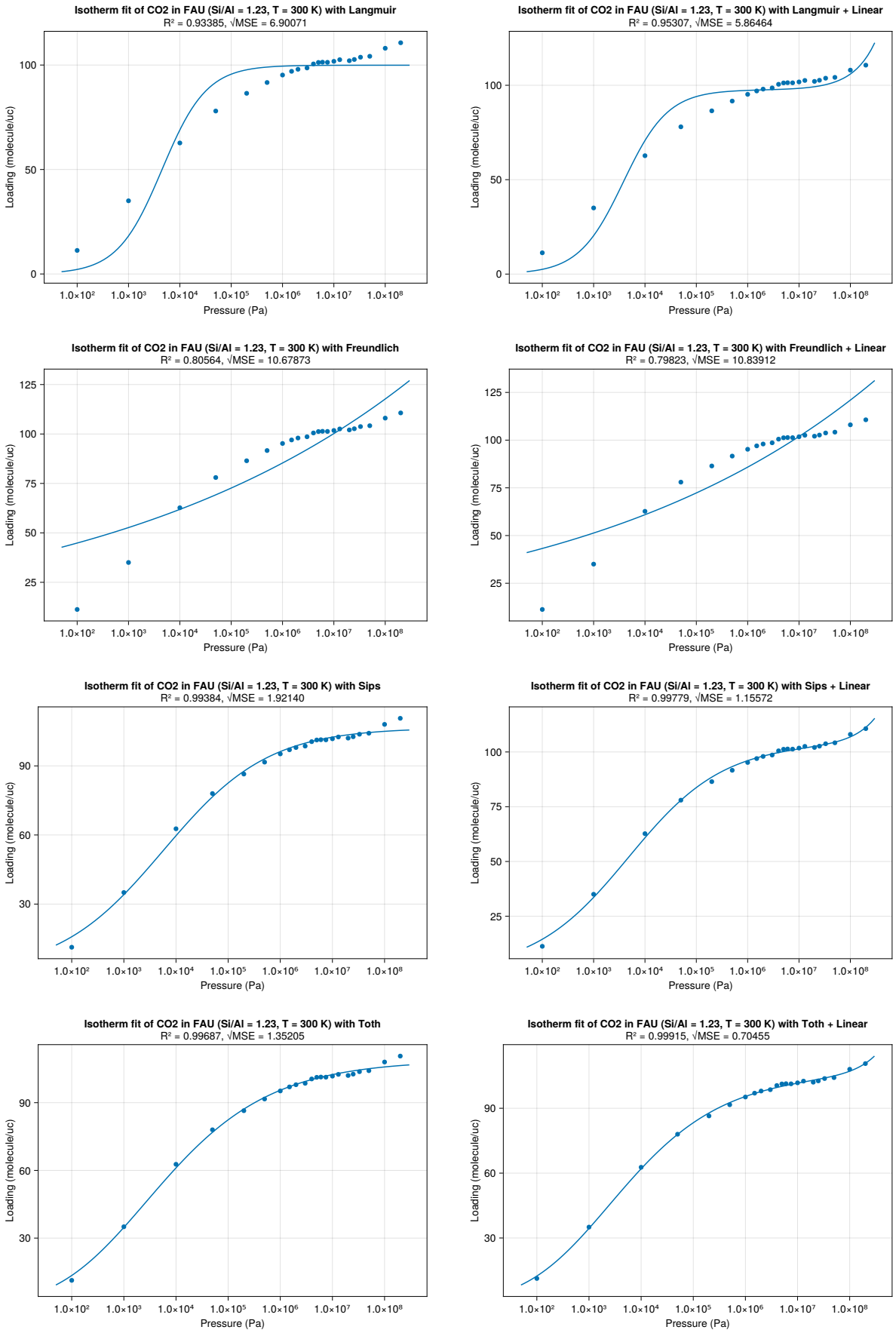


Figure 3.2: Simulated adsorption isotherm fitted with different models

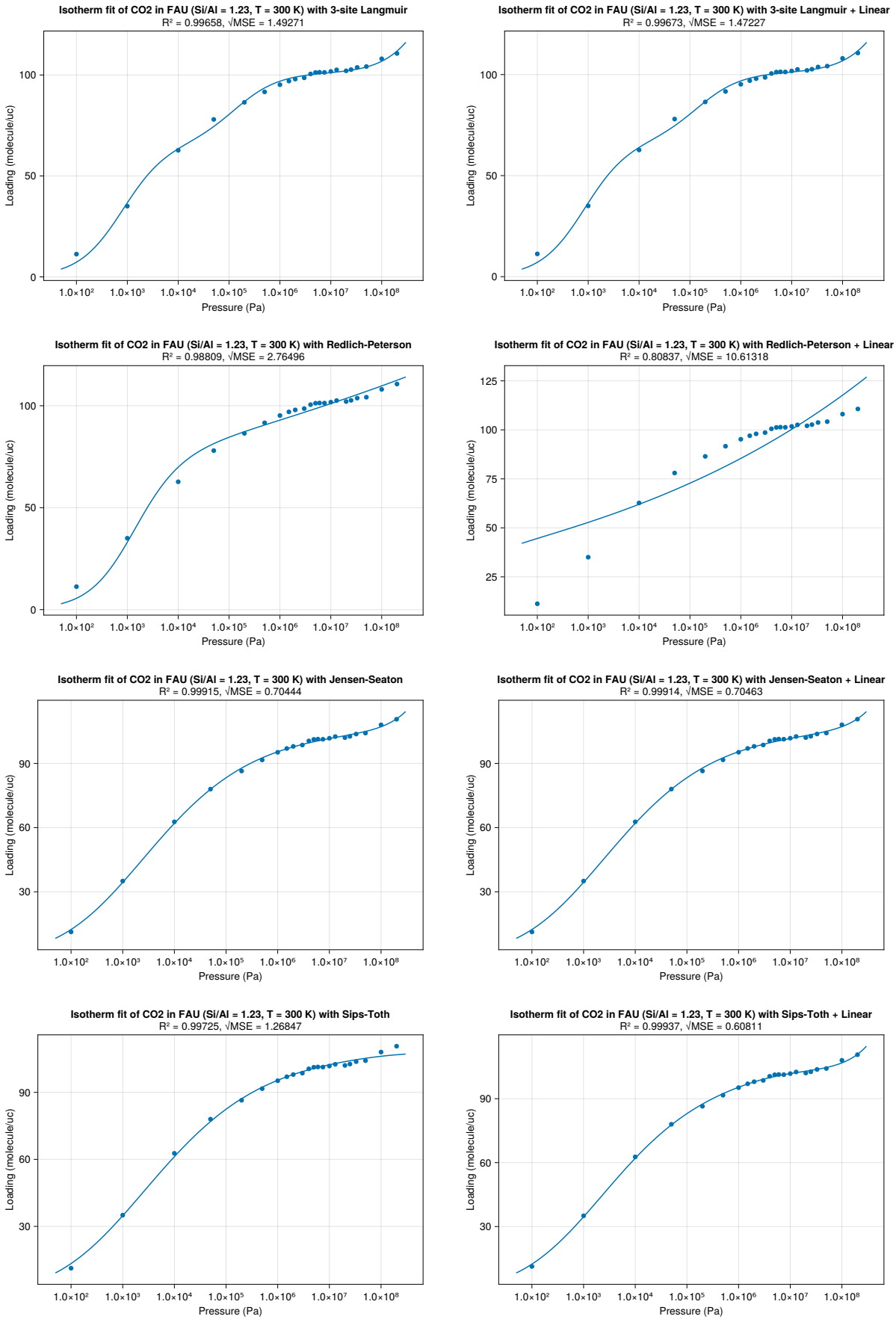


Figure 3.2: Simulated adsorption isotherm fitted with different models (continued)

Unfortunately, the expressions of the different adsorption models make this problem non-convex, which is a category of optimization problems generally considered difficult. To solve it, we use the `LsqFit.jl` and `BlackBoxOptim.jl` Julia packages to do global optimization. Crucially, these two packages do not rely on f being differentiable, since that would require finding the derivative of the deviation with respect to the parameters for all investigated adsorption models and their combination.

Both packages use meta-heuristic and stochastic algorithms to perform the actual optimization, but they need a starting point, *i.e.* an initial value for the parameters, to work. The result often depends on the quality of this initial point: the closer it is to the optimum, the better the result. Hence, we tried to use `LsqFit.jl`, with an initial set of parameters found from linearizations of the model:

- For Langmuir, the linear regression $n^{-1} = f(\mathrm{P}^{-1})$ yields $n^{-1}_{\text{Langmuir}} = \alpha'^{-1}_0 \left(1 + \beta'^{-1}_0 \mathrm{P}^{-1}\right)$. We then test three different initial points: $\alpha_0 = \alpha'_0 \times \theta$ and $\beta_0 = \beta'_0/\theta$ with $\theta = 0.833$, 1.44 and 0.579.
- For Freundlich, the linear regression $\log n = f(\log \mathrm{P})$ directly gives $\log n_{\text{Freundlich}} = \log \mathrm{A}_0 + \gamma_0 \log \mathrm{P}$.
- For Sips, a first linear regression $\log n = f(\log \mathrm{P})$ gives an approximate value γ_0, then a second linear regression $n^{-1} = f(\mathrm{P}^{\gamma_0})$ yields $n^{-1}_{\text{Sips}} = \alpha_0 \left(1 + \beta^{-\gamma_0} \mathrm{P}^{-\gamma_0}\right)$.
- For Toth, we take the three initial points α_0 and β_0 from the Langmuir isotherm and test $\gamma_0 = 0.667$, 2.25 and 0.296 respectively.
- For Redlich-Peterson, a first linear regression $\log(\mathrm{P}/n) = f(\log \mathrm{P})$ gives $\log(\mathrm{P}/n_{\text{Redlich-Peterson}}) = \gamma_0 \log \mathrm{P} + \mathrm{C}$, then a second linear regression $\mathrm{P}/n = f(\mathrm{P}^{\gamma_0})$ yields $\mathrm{P}/n_{\text{Redlich-Peterson}} = \alpha_0^{-1} \left(1 + \beta_0^{\gamma_0} \mathrm{P}^{\gamma_0}\right)$.
- For Jensen-Seaton, we try the three sets of initial parameters α_0, β_0 and γ_0 taken with the method for Toth, and fix $\delta_0 = 8 \times 10^{-10}$.
- For Sips-Toth, we try both Sips and Toth parameters, with the switching parameter ζ_0 equal to either 1, 10^{-5} or 10^{-10}.

When adding a linear term to a model, the linear slope is taken from a linear regression of the four last point in the isotherm, then the rest of the initial point is taken from the isotherm minus this initial slope.

Unfortunately, this rational but simplistic method does not give good results even for simple models, because the initial points are too far from the optimum. A more systematic approach consists in checking a larger set of initial parameters with the `BlackBoxOptim.jl` package. We then use the best set of parameters found as the initial point for an `LsqFit.jl` optimization, and keep whichever set of parameters yields the lowest RMSD.

The only requirement for `BlackBoxOptim.jl` is to delimit the minimum and maximum values for each parameter. We use $0\,\mathrm{Pa}^{-1} < \beta < 1\,\mathrm{Pa}^{-1}$, $0 < \gamma < 5$, $0\,\mathrm{Pa}^{-1} < \delta < 1\,\mathrm{Pa}^{-1}$, $0\,\mathrm{Pa}^{-1} < \zeta < 0.5\,\mathrm{Pa}^{-1}$, and $2\mathrm{M}/3 < \alpha < 4\mathrm{M}/3$ where M is the maximum value of the isotherm. Overall, we

find that this approach, starting with `BlackBoxOptim.jl` to have an initial point followed by `LsqFit.jl` for fine-grained optimization, gives the best result.

External to the Julia ecosystem, it is also possible to perform isotherm fitting using proprietary softwares like OriginLab, or other free open-source alternatives like the PyGAPS package in Python.[9] The latter provides isotherm fitting capability for (up to 3-site) Langmuir, Freundlich, Toth, Jensen-Seaton and a few other models, but not Sips, Redlich-Peterson, nor custom functions like the Sips-Toth model presented before. The PyGAPS framework is flexible enough to allow adding custom models however, so it could be interesting to compare the quality of our fits with theirs.

3.4.4 Simple models

Once a sufficient number of fitted isotherms are available, the next goal is to somehow interpolate across them across Si/Al ratio and temperatures for a given topology. Through fitting, this problem is reduced to finding function $\alpha(r, \mathrm{T})$, $\beta(r, \mathrm{T})$, $\gamma(r, \mathrm{T})$, … *i.e.* to expressing each parameter of the adsorption model as a function of the Si/Al ratio r and the temperature T.

For each parameter p, we tried to express $p(r, \mathrm{T})$ through simple models, using `GLM.jl`, a Julia package for fitting linear models. In detail, we tried to fit expressions of the form

$$f_p^{(p)}(p) = \mathrm{A}^{(p)} \times f_r^{(p)}(r) + \mathrm{B}^{(p)} \times f_\mathrm{T}^{(p)}(\mathrm{T}) + \mathrm{C}^{(p)} \tag{3.11}$$

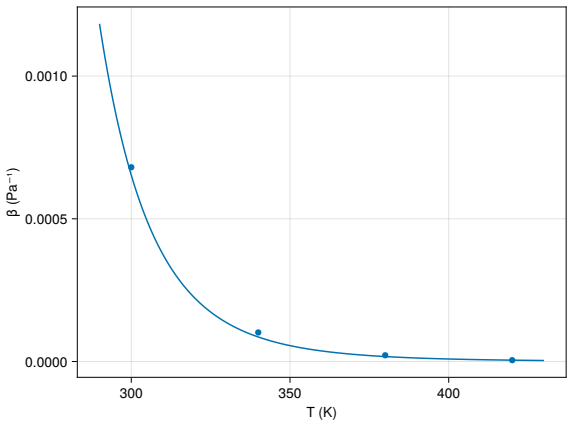
where each function $f^{(p)}$ was either the identity, the inverse, or the log. Such a fit consists in finding the optimal parameters $\mathrm{A}^{(p)}$, $\mathrm{B}^{(p)}$ and $\mathrm{C}^{(p)}$ to minimize the RMSD between both sides of the expected equality. There are 27 such expressions to test for each parameter p since each of the three function $f^{(p)}$ only has three choices, and the expression which yields the lowest RMSD is kept as the model, along with the corresponding parameters.

For example, using the Jensen-Seaton adsorption model for CO_2 adsorption in FAU zeolites, the following expression is found for the β parameter:

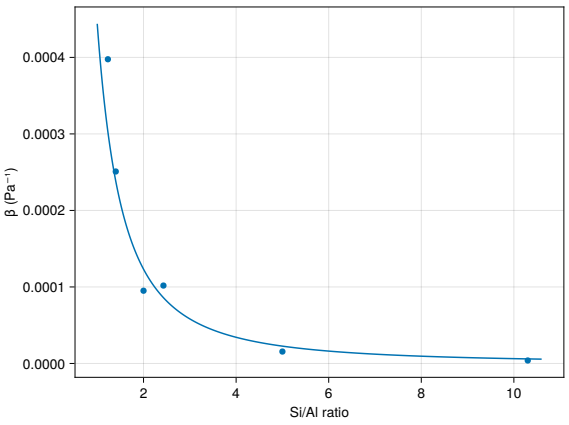
$$\log \beta(r, \mathrm{T}) = -22.927688538843174 - 1.8489495470102486 \log r + 5170.651401050577/\mathrm{T} \tag{3.12}$$

Figure 3.3 represents the comparison between the interpolated value of $\beta(r, \mathrm{T})$ and the values collected from the isotherms fitting in the previous setting. In this case, the fit works well, but the simple model above does not always provide such a good level of agreement.

Overall, this approach provides a simple model for the evolution of the parameters which works reasonably well to interpolate an isotherm. For example, Figure 3.4 shows the difference in quality between the direct Jensen-Seaton fit of an isotherm, and the one obtained from the simple model. In this example, the simple model was built using Si/Al ratios of 1, 1.23, 1.4, 2, 5 and 10 and temperatures of 300 K, 340 K and 420 K, thus excluding the investigated point with Si/Al ratio of 2.4 at $\mathrm{T} = 380$ K. While not as quantitative as the actual fit, the simple model allows retrieving an isotherm which is close to the reference without needing any additional computation.

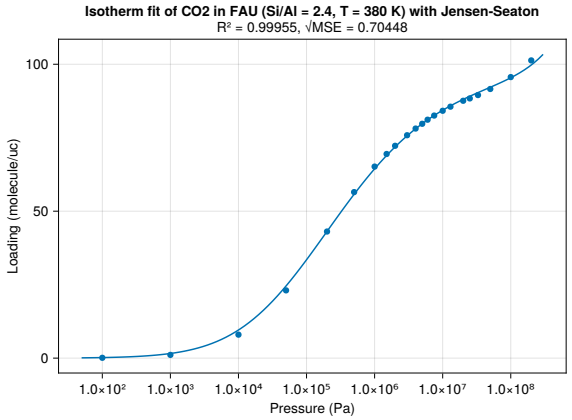


(a) Evolution with temperature for Si/Al = 2.43

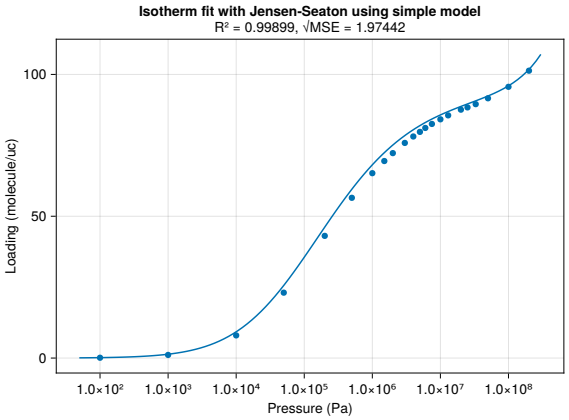


(b) Evolution with Si/Al ratio for T = 340 K

Figure 3.3: Evolution of the β parameter in the Jensen-Seaton fit of CO_2 adsorption isotherms on FAU. The full line is the simple model obtained in Equation (3.12).



(a) Direct Jensen-Seaton fit



(b) Interpolated fit using simple model

Figure 3.4: Fitted adsorption isotherms of CO_2 in FAU with Si/Al = 2.4 at T = 380 K

3.5 Perspectives

At the time of writing, we have only computed gas adsorption isotherm on three topologies: CHA, LTA and FAU. The previous methodology using simple models to interpolate isotherms constitute a convenient method to obtain adsorption capacities given enough data for the target topology, but does not allow generalizing to other topologies. To reach this target, more isotherm data is required on various topologies, from which point it would become possible to create yet another simple model whose target are the coefficients A, B and C of the previous simple model, depending only on the topology.

Another approach consists in relying on the more general tools of machine learning, instead of the custom approach detailed before. Deep learning could be used as a black-box method to directly learn the $p(r, T, \tau)$ functions, where τ designates the framework topology, numerically represented using ring statistics or coordination sequences for instance. Yet, once again, this requires much more isotherm data than what is currently available.

Fortunately, our cationic zeolite structure database now contains 64 topologies with representative structures across Si/Al ratios. Once the adsorption isotherms are computed on all of these structures, it should provide enough data to start creating a more robust predictive model.

LIST OF PUBLICATIONS

Peer-reviewed papers

1. Lionel Zoubritzky and François-Xavier Coudert. “CrystalNets.jl: Identification of Crystal Topologies”. In: *SciPost Chem.* 1 (2022), p. 005. DOI: 10.21468/SciPostChem.1.2.005.

Preprint

Publication from a previous project

BIBLIOGRAPHY

[1] H. Fang, A. Kulkarni, P. Kamakoti, R. Awati, P. I. Ravikovitch, and D. S. Sholl. "Identification of high-CO2-capacity cationic zeolites by accurate computational screening". In: *Chemistry of Materials* 28.11 (2016), pp. 3887–3896.

[2] S. H. Mousavi, K. Chen, J. Yao, A. Zavabeti, J. Z. Liu, and G. K. Li. "Screening of alkali metal-exchanged zeolites for nitrogen/methane separation". In: *Langmuir* 39.3 (2023), pp. 1277–1287.

[3] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, et al. "In silico screening of carbon-capture materials". In: *Nature materials* 11.7 (2012), pp. 633–641.

[4] S. Brandani, E. Mangano, and L. Sarkisov. "Net, excess and absolute adsorption and adsorption of helium". In: *Adsorption* 22 (2016), pp. 261–276.

[5] J. Rouquerol, F. Rouquerol, P. Llewellyn, G. Maurin, and K. Sing. *Adsorption by powders and porous solids: principles, methodology and applications*. Academic press, 2013.

[6] K. Sing, D. Everett, R. Haul, L. Moscou, R. Pierotti, J. Rouquerol, and T. Siemieniewska. "International union of pure and applied chemistry physical chemistry division commission on colloid and surface chemistry including catalysis. vol. 57, issue 4". In: *Pure Appl Chem* (1985), pp. 603–619.

[7] S. E. Boulfelfel, J. M. Findley, H. Fang, A. S. Daou, P. I. Ravikovitch, and D. S. Sholl. "A Transferable Force Field for Predicting Adsorption and Diffusion of Small Molecules in Alkali Metal Exchanged Zeolites with Coupled Cluster Accuracy". In: *The Journal of Physical Chemistry C* 125.48 (2021), pp. 26832–26846.

[8] N. Ayawei, A. N. Ebelegi, and D. Wankasi. "Modelling and interpretation of adsorption isotherms". In: *Journal of chemistry* 2017.1 (2017), p. 3039817.

[9] P. Iacomi and P. L. Llewellyn. "pyGAPS: a Python-based framework for adsorption isotherm processing and material characterisation". In: *Adsorption* 25.8 (2019), pp. 1533–1542.

RÉSUMÉ EN FRANÇAIS

Introduction

RÉSUMÉ

MOTS CLÉS

simulation moléculaire, adsorption, matériaux nanoporeux, topologie, zéolites

ABSTRACT

KEYWORDS

molecular simulation, adsorption, nanoporous materials, topology, zeolites