Faculty of Sciences
Department of Physics and Astronomy
Chairman: Prof. Dr. D. RYCKBOSCH

Explicit sampling of the unit cell shape for a more accurate estimation of the free energy profile for flexible crystals

by

Senne CAROES

Supervisors: Prof. Dr. ir. A. GHYSELS, Prof. Dr. ir. V. VAN SPEYBROECK
Counsellors: Dr. ir. L. VANDUYFHUYS, ir. R. DEMUYNCK, ir. S. ROGGE

Master's dissertation submitted in order to obtain the academic degree of
MASTER OF SCIENCE IN PHYSICS AND ASTRONOMY

Academic year 2015–2016

Center for
Molecular Modeling

This research was conducted at the Center for Molecular Modeling.

Preface

1 juni 2016, aan het einde van de laatste etappe. Nu de finishlijn in zicht komt, denk ik met weemoed terug aan het lange, mooie parcours dat de Fysica en de Sterrenkunde voor me uitgestippeld had. Het is een hele eer om in het spoor van de groten der aarde de Wetenschap te ontdekken.

De laatste etappe was lastig, met berg, dal, bloed, zweet en hard werk bovenal; precies daarom de meest belonende. Zonder mijn geweldige gedreven promotoren en begeleiders, had ik de eindstreep echter niet gehaald. Daarom een speciale bedanking voor mijn twee promotoren: Professor Ghysels voor het vertrouwen en de vrijheid die ik kreeg, en voor de nauwe samenwerking; de talloze uren besprekingen, richtlijnen en discussies. Professor Van Speybroeck, voor de kritische voortgangsbesprekingen, waar steeds nieuwe inzichten en toekomstige ideeën uit voortkwamen. Aan mijn begeleiders: Sven, Ruben, Louis. Bedankt om steeds klaar te staan met raad als ik een doodlopend straatje ingereden was. Jullie gedrevenheid heeft een blijvende indruk op me nagelaten.

Toen ik aankwam op het Centrum voor Moleculaire Modellering, voelde ik meteen dat dit een plaats was waar ik me thuis kon voelen. Er volgden dan ook vele mooie momenten; eeuwige irrelevante discussies, koffie- en theepauzes, in het bijzonder gedeeld met mijn collega studenten Titus, Pieter, Klaas, Michiel en Yentl, of tijdens de bezoekjes van Sam, lachend dat Valentino Rossi dit weekend geen goede motorrace had gereden.

Bedankt aan mijn ouders, die me alle kansen hebben gegeven om te studeren, in het bijzonder om me een kot ter beschikking te stellen ver weg van het bruisende Gent, waar ik in afzondering een nieuw soort leven kon opbouwen gevuld met regelmaat en ambitie. Bedankt aan mijn grootouders, voor de sporadische bezoekjes in het weekend, en de vele telefoontjes.

Senne Caroes, juni 2016

Copyright agreement

Senne Caroes
June 1st, 2016

Expliciete sampling van de eenheidscel om het vrije-energieoppervlak van flexibele kristallen nauwkeuriger op te stellen

door

Senne Caroes

Scriptie ingediend tot het behalen van de academische graad van
Master of Science in de Fysica en de Sterrenkunde

Academiejaar 2015–2016

Promotoren: Prof. Dr. ir. A. Ghysels, Prof. Dr. ir. V. Van Speybroeck
Scriptiebegeleiders: Dr. ir. L. Vanduyfhuys, ir. R. Demuynck, ir. S. Rogge
Universiteit Gent
Faculteit Wetenschappen
Vakgroep Fysica en Sterrenkunde
Voorzitter: Prof. Dr. D. Ryckbosch

Samenvatting

Dit werk beschrijft hoe de expliciete sampling van de eenheidscelvorm het vrije-energieprofiel van het metaal-organisch rooster MIL-53(Al) beïnvloedt. De verschillende metastabiele configuraties van het systeem worden verkend, en een bistabiliteit in de eenheidscelvorm wordt gevonden in een bepaald volumebereik. Verschillende samplingmethodes en methodes om het vrije-energieprofiel op te stellen (thermodynamische integratie, histogrammethode, metadynamica) worden onderzocht. Er wordt geconcludeerd dat tweedimensionale metadynamica de enige onderzochte methode is die de bistabiliteit in de eenheidscelvorm correct kan samplen. Voor vrije-energiemethodes die deze bistabiliteit niet samplen, wordt een correctieterm voorgesteld om de onderschatting van de entropie te compenseren. De koppeling van de collectieve bewegingen van de atoomkernen met deze van de eenheidscel wordt onderzocht via hoofdcomponentenanalyse. Methodes ter kwantificatie van de verschillende collectieve bewegingen worden ontwikkeld.

Trefwoorden: *Metaal-organische roosters, ademend gedrag, flexibiliteit, moleculaire dynamica, hoofdcomponentenanalyse, vrije energie, eenheidscel, sampling*

Explicit sampling of the unit cell shape for a more accurate estimation of the free energy profile for flexible crystals

SENNE CAROES
Supervisors: Prof. Dr. Ir. A. Ghysels, Prof. Dr. Ir. V. Van Speybroeck
Counsellors: Dr. Ir. L. Vanduyfhuys, Ir. R. Demuynck, Ir. S. Rogge

Abstract—**This work explains how the explicit sampling of the unit cell shape influences the free energy profile of a study system, the metal-organic framework MIL-53(Al). The multiple stable configurations of the system are explored, and a bistability in the shape space is found at low volumes. Different methods of sampling and constructing the free energy profile (thermodynamic integration, histogram method and metadynamics) are investigated, and it is concluded that from this set of methods only the 2D metadynamics correctly samples the bistability in the shape space. For free energy profiles obtained by methods failing to sample the bistability in shape space at low volumes, an additive correction factor is proposed to account for the underestimation of entropy. The coupling of the collective motion of the nuclei with the unit cell shape is investigated using principal component analysis. Methods of quantifying the description of the different collective motions are developed.**

Keywords—**Metal-organic frameworks, breathing, flexibility, molecular dynamics, principal component analysis, free energy, unit cell, sampling**

I. Introduction

Molecular simulations are a very helpful tool for scientists to obtain information on the properties of materials. The great complexity of the governing laws of molecular systems renders the application of mathematical tools to find exact, analytical solutions insufficient. Also, the computational power to numerically approximate these exact solutions of these laws within a given accuracy is orders of magnitude too weak, resulting in a quest to find more accurate and efficient approximative methods. To obtain information about the study system, the different configurations of interest are sampled. In most works, the unit cell shape fluctuations are sampled implicitly, but here it is investigated how this can be carried out explicitly. The implications of sampling the unit cell shape are not known *a priori*. In this work, it is demonstrated that the sampling of the unit cell shape is necessary for obtaining the correct free energy profile as a function of the system's volume. Next, the different stable shape configurations are explored, using a variety of methods. The most appropriate set of simulation techniques for the correct sampling of the unit cell shape space is proposed. Finally, it is investigated how the collective motion of the nuclei couples with the unit cell motion using principal component analysis.

The system on which these calculations are performed is MIL-53(Al) [1], a metal-organic framework (MOF). MOFs form a class of crystalline materials exhibiting a framework structure with metal ions or clusters at nodal points and organic linkers connecting those points. Because of this structure, most MOFs have a large inner surface, which makes them suitable for adsorption, among other applications.

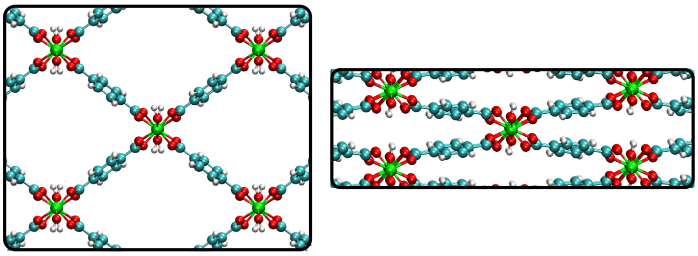


Fig. 1. MIL-53(Al) in the large-pore (left) and the narrow-pore phase (right)

MIL-53(Al) exhibits a framework consisting of 1,4-benzenedicarboxylate (BDC) organic linkers connecting onedimensional Al 3^+(OH) chains (see Figure 1). The system exhibits breathing behavior [2]: a transition between two metastable states of the material, corresponding to expansion and contraction induced by external stimuli. In this work, the empty MIL-53(Al) state at low volumes is labelled the narrow-pore (np) phase, while the empty MIL-53(Al) state at high volumes is labelled the large-pore (lp) phase. The transition from the lp to the np phase in MIL-53(Al) corresponds to a wineracack motion [3]: In the lp phase, the angle between neighboring linker planes in the z-direction is around 70°–90°, and in the np phase this is around 30°–50°, see Figure 1. This transition has a large amplitude in volume, as well as a large shift in unit cell shape.

II. Methodology

A. The unit cell

The unit cell is determined by its unit cell vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$, constituting the rows in the unit cell matrix $\mathbf{h}$:

$$\mathbf{h} = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix} \quad (1)$$

In this work, the unit cell is always rotated in such a way that the unit cell matrix is symmetric (corresponding to six degrees of freedom). The relation between the original $\mathbf{h}$ and the symmetric variant $\mathbf{h}_s$ is found by right multiplying the latter with the rotation matrix $\mathbf{R}$:

$$\mathbf{h} = \mathbf{h}_s\mathbf{R} \quad (2)$$

For the remainder of the discussion, the unit cell matrix $\mathbf{h}$ is assumed symmetric, possibly after the application of the inverse transformation in Equation 2. The volume V and shape $\mathbf{h}_0$ of the unit cell are defined as:

$$V = \det(\mathbf{h}) \quad \text{and} \quad \mathbf{h}_0 = \frac{\mathbf{h}}{V^{1/3}} \tag{3}$$

B. Simulations

The system is simulated by molecular dynamics (MD), using the in-house developed software package Yaff [4]. The integration timestep of the simulations is 0.5 fs, which are performed on a double unit cell containing 152 nuclei. To evaluate the potential energy of the system, a force field for MIL-53(Al) [5] is used. A force field is an analytical model of the potential energy surface as a function of only the nuclear coordinates, parametrized to describe the system around a stable equilibrium state (in this case, the lp state).

All simulations have a fixed number of particles N and happen at a temperature $T = 300$ K, making use of the Nosé-Hoover chains thermostat with three beads [6]. Applying isotropic pressures or fixing the unit cell shape $\mathbf{h}_0$ is done by an extension of the MTTK barostat [7], in such a way that the motion of the unit cell tensor $\mathbf{h}$ can be split up in the motion of the volume V and the motion of the normalized unit cell tensor $\mathbf{h}_0$ [8].

C. Free energy calculations

The Helmholtz free energy F, for a system with energy E and entropy S, in thermal equilibrium with a heat bath at a temperature T, is defined as

$$F = E - TS \tag{4}$$

The following relation holds:

$$\left(\frac{\partial F}{\partial V}\right)_T = -P \tag{5}$$

The free energy $F(q)$ for a macrostate q is defined as

$$F(q) = -k_B T \ln\left(p(q)/c\right) \tag{6}$$

where the constant c is used to make the expression $p(q)/c$ dimensionless. Every free energy profile can be arbitrarily shifted by a constant, and for further reference the constant is tacitly assumed unity, while carrying the dimensions of p.

C.1 Histogram method

The histogram method estimates the probability of visiting a state q, by constructing a histogram $h(q)$ with bin width W for N simulated data points.

$$\begin{aligned} p(q) &= \frac{h(q)}{NW} && (7) \\ F(q) &= -k_B T \ln[p(q)] && (8) \end{aligned}$$

C.2 Thermodynamic integration

In thermodynamic integration, the relation between the derivative of the free energy profile with respect to the volume, yielding the negative of the pressure (see Equation 5) is used to construct free energy profiles as a function of the volume [8]:

$$F(V) - F(V_{ref}) = -\int_{V_{ref}}^{V} \langle P \rangle_{V'} dV' \tag{9}$$

where $\langle P \rangle_{V'}$ is the ensemble average of the pressure in the system fixed at a volume V'.

D. Enhanced sampling

To improve sampling in regions with low probability (regions with a high free energy, see Equation 6), enhanced sampling methods are employed. These methods add a bias potential $U_b(q)$ as a function of the collective variable q to the free energy profile, forcing the system towards states which are not sampled sufficiently in regular MD simulations.

D.1 Umbrella sampling

In an umbrella simulation, the bias potential is added to the system *a priori* and remains unaltered during the MD simulation. Performing the histogram method results in a biased free energy profile $F_b(q)$, from which the original free energy profile $F(q)$ is found by correcting the profile for the bias $U_b(q)$:

$$\begin{aligned} F(q) &= F_b(q) - U_b(q) && (10) \\ &= -k_B T \ln[p_b(q)] - U_b(q) && (11) \end{aligned}$$

D.2 Metadynamics

Metadynamics [9] is a sampling technique that adds a history-dependent bias potential $U_b(t, q)$ to the system that changes during the MD simulation. After sufficient simulation time, the histogram of the system as a function of the collective variable becomes flat when the bias potential is kept fixed: $p(q) \equiv c$. It follows

$$F_b(q) = -k_B T \ln p_b(q) = c' \tag{12}$$

Combining this with the expression for the biased free energy:

$$F(q) = -U_b(q) + c' \tag{13}$$

Hence the free energy of the system as a function of q is the negative of the applied bias potential, once the histogram of the simulation has become flat.

A metadynamics step consists of a simulation with a length on the order of picoseconds, after which (at time t_i) a gaussian hill $he^{\frac{-(q-q_i)^2}{2w^2}}$ is added to the bias potential, where q_i, h and w are respectively the mean, height and width of the gaussian. The bias potential at time t during the simulation becomes:

$$U_b(t, q) = \sum_{i|t_i<t} he^{\frac{-(q-q_i)^2}{2w^2}} \tag{14}$$

In this work, the collective variables for metadynamics are components of the symmetric unit cell and/or the volume V. The derivative of the bias potential $U_b(q)$ with respect to the transposed unit cell matrix $\mathbf{h}^T$ (not yet assumed to be symmetric) is given by:

$$\frac{\partial U_b(q)}{\partial \mathbf{h}^T} = \mathbf{\Xi}_b^T \cdot \mathbf{h}^{-1} \tag{15}$$

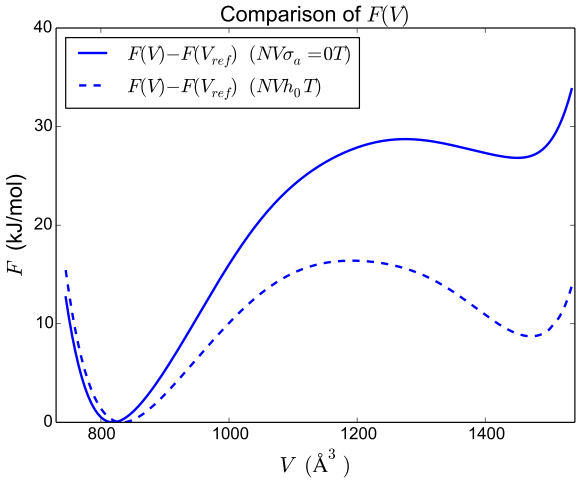


Fig. 2. Comparison of the free energy profiles $F(V)$ as a function of the unit cell volume V obtained by performing thermodynamic integration of the average pressure. The average pressure is determined in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble (solid line) on the one hand and $NV\mathbf{h}_0T$ ensemble (dotted line) on the other hand.

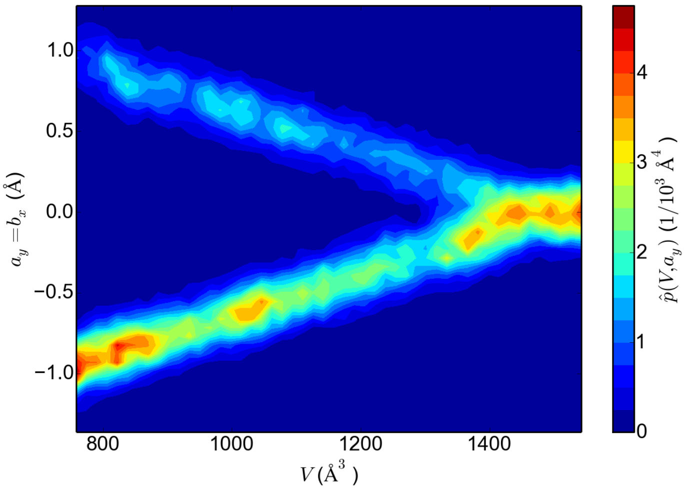


Fig. 3. The estimated probability density $\hat{p}(V, a_y) = h(V, a_y)/NS$, with $h(V, a_y)$ the 2D histogram of the extended umbrella simulation set as a function of the unit cell volume V and the $a_y = b_x$ component, N the total number of observations and S the surface of the histogram bin (here in Å^4).

with Ξ_b the virial tensor due to the bias potential U_b and $\mathbf{h}^{-1}$ the inverse of the unit cell matrix. The previous equation can be transformed into

$$\Xi_b = \mathbf{h}^T \cdot \left(\frac{\partial U_b(q)}{\partial q} \frac{\partial q}{\partial \mathbf{h}_s} \cdot \mathbf{R}^{-1} \right) \tag{16}$$

Where $\mathbf{h}_s$ is the symmetrized unit cell matrix and $\mathbf{R}$ the rotation matrix from Equation 2. The bias potential can finally be applied to the system by altering the virial tensor:

$$\Xi \rightarrow \Xi + \Xi_b \tag{17}$$

E. Principal component analysis

In this work, principal component analysis (PCA) is used as a method to get an accurate description of the important coordinate fluctuations observed within an MD simulation [10]. PCA constructs a covariance matrix $\mathbf{C}$ of the fluctuations in a coordinate set $\vec{q}$ during the simulation:

$$C_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \tag{18}$$

The diagonalization of this covariance matrix results in eigenvectors $\vec{v}^{(l)}$ that specify mutually orthogonal directions of collective motions. The corresponding eigenvalues $\lambda^{(l)}$ are the variances associated with the motion in the directions $\vec{v}^{(l)}$. After diagonalizing, $\mathbf{C}$ is written as:

$$C_{ij} = \sum_l \lambda^{(l)} v_i^{(l)} v_j^{(l)} \tag{19}$$

The eigenmodes are ordered in a nonincreasing fashion: $\lambda^{(1)} \geq \lambda^{(2)} \geq \cdots \geq \lambda^{(6+3N)}$. The position of a nucleus α can be represented by its Cartesian coordinates $(x_\alpha, y_\alpha, z_\alpha)$ or by its fractional coordinates $(s_{\alpha a}, s_{\alpha b}, s_{\alpha c})$ along the lattice vectors. The relation between these two representations is given by

$$\vec{r}_\alpha = \vec{s}_\alpha \mathbf{h} \tag{20}$$

Since the unit cell matrix is rotated such that it is symmetric (see Equation 2), the nuclei need to be rotated appropriately to a set of positions $\vec{r}_{R\alpha}$. Their fractional coordinates $\vec{s}_\alpha$ remain unaltered after rotation of the unit cell such that:

$$\vec{r}_{R\alpha} = \vec{s}_\alpha \mathbf{h}_s = \vec{s}_\alpha \mathbf{h} \mathbf{R}^{-1} = \vec{r}_\alpha \mathbf{R}^{-1} \tag{21}$$

The coordinate set used for the analysis here is the merged set of symmetric unit cell components and rotated nucleic positions.

III. Results

A. Implicit unit cell shape sampling

Free energy profiles as a function of the volume are constructed using thermodynamic integration. To determine the average pressure $\langle P \rangle_{V'}$ at certain well-chosen volumes V', two different ensembles are used. One is the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, where unit cell shape fluctuations occur under isotropic stress, the other the $NV\mathbf{h}_0T$ ensemble where the unit cell shape is fixed to $\mathbf{h}_0$. Hence in the latter, the unit cell shape fluctuations remain unsampled.

The two free energy profiles are compared in Figure 2. It is shown that not sampling the unit cell shape fluctuations has a large influence on the free energy profile: the free energy difference between the lp and np phase lowers from 26.8 to 8.7 kJ/mol. The lp-to-np transition has a higher barrier, increasing from 1.9 to 7.7 kJ/mol with respect to the lp minimum. It is concluded that there is a great need for (at least implicit) sampling of the unit cell shape fluctuations in MD simulations of MIL-53(Al).

B. Bistability in shape space

To further explore the stable configurations of MIL-53(Al), an umbrella simulation of length $\approx$ 1 ns is performed in the collective variable V in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble at zero pressure. The bias potential $U_b(V)$ applied to the system is the negative free energy profile $-F(V)$ obtained in the previous subsection, using the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble. Application of this bias potential results in sampling of the complete volume region between the np and lp phase, and this umbrella simulation is hence labelled "the extended umbrella". A bistability of

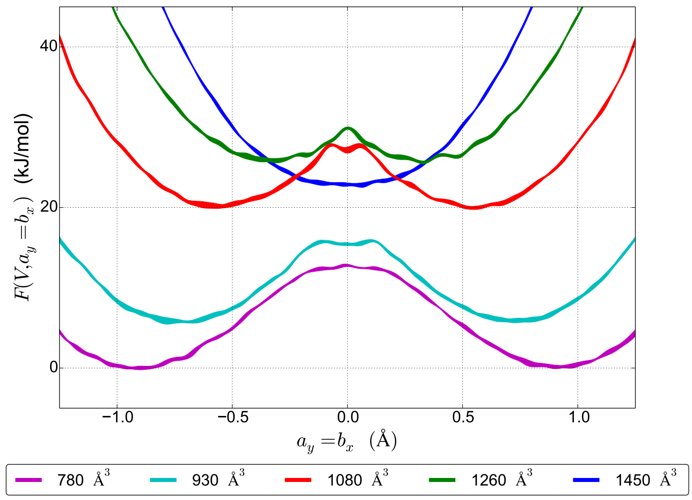


Fig. 4. Estimate of the free energy curve $F(V, a_y)$ as a function of the cell component $a_y = b_x$ for five different unit cell volumes 780, 930, 1080, 1260 and 1450 Å^3. The curves are shifted with respect to each other by a constant such that $p(V, a_y) \propto e^{-\beta F(V, a_y)}$. The linewidth denotes the region $F(V, a_y) \pm 2\ \mathrm{SE}_{F(V, a_y)}$, with SE the estimation of the standard error on the free energy profiles.

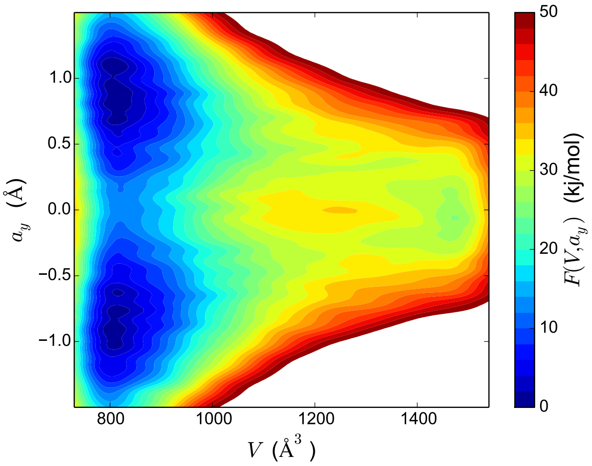


Fig. 5. 2D free energy profile $F(V, a_y = b_x)$ obtained by performing metadynamics in the components $(V, a_y = b_x)$.

the unit cell component $a_y = b_x$ is detected in the np phase, and the branched structure of the complete 2D histogram $h(V, a_y)$ is shown in Figure 3. The two unit cell configurations at volumes away from the lp phase are labelled the positive and negative branch, corresponding to positive and negative $a_y = b_x$ respectively. Although there is a large difference in estimated probability between the two branches, these differences result in minor free energy differences. It is not known if this observation can be attributed to a statistical fluctuation. An optimization and space group calculation can not distinguish structures from both branches. Hence, additional, detailed research using metadynamics is required.

C. Metadynamics

The first set of metadynamics simulations has the cell component $a_y = b_x$ at fixed volume V as collective variable ($NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble). It renders detailed information on the relative stability of both branches at a discrete set of volumes, with an estimation of the standard error on the profile. At each unit cell volume in the set [780, 930, 1080, 1260, 1450] Å^3, ten independent fixed volume metadynamics simulations are performed in three stages, each of length 1 ns. The first stage contains 2000 hills of width 0.05 Å and of height 0.3 kJ/mol, placed each 0.5 ps. The second stage contains 2000 narrower and lower hills (a width of 0.025 Å and a height of 0.1 kJ/mol), while the last stage performed a metadynamics step every 2 ps, hence 500 hills are added in this stage (with the same height and width as in the previous stage). The results are plotted in Figure 4. It is concluded that within the error margin, there is no reason to assume both branches are inequivalent in terms of free energy.

The next set of metadynamics simulations has a two dimensional collective variable, containing the volume V and variable cell component $a_y = b_x$, in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble at zero pressure. The obtained 2D free energy profile is a result of summing over 7200 Gaussian hills of height 1 kJ/mol placed every 0.6 ps, 0.05 Å wide in the $a_y = b_x$ direction, and 100 Å^3 in the V direction. The profile is shown in Figure 5. It clearly shows the structure of two stable minima in the np phase and one in the lp.

D. Free energy profile as a function of the unit cell volume

The 2D free energy profile $F(V, a_y = b_x)$ obtained by performing 2D metadynamics (see Figure 5) is integrated into a onedimensional profile $F(V)$:

$$F(V) = -k_B T \ln \left[\int_{-|q|}^{|q|} e^{-\beta F(V, a_y = b_x)} da_y \right] \qquad (22)$$

with $\pm|q| = \pm 2$ Å the complete range of $a_y = b_x$.

In addition, a set of 1D metadynamics simulations is performed with the volume as collective variable, simulated in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble at zero pressure. Ten independent simulations are performed to be able to determine an estimation of the standard error on the profile. The simulation consists of two stages, one having a simulation time of 1.2 ns, containing 2000 hills of width 50 Å^3 and height 0.3 kJ/mol, the other stage having a simulation time of 1 ns, with 2000 hills of width 25 Å^3 and height 0.1 kJ/mol. The calculated profile shows perfect correspondence with the profile obtained in Subsection III-A.

The obtained profile is compared to the other $F(V)$ profiles obtained via thermodynamic integration and 2D metadynamics, see Figure 6. It is shown that the top of the np-to-lp free energy barrier and the lp minimum in the $F(V)$ of the 2D metadynamics simulation have increased.

It is concluded that the 2D metadynamics simulation renders the correct $F(V)$ because it samples both branches at the same time, at all volumes. The 1D methods (thermodynamic integration and 1D metadynamics in the volume), however, sample both branches in the lp phase (low free energy barrier between the + and − branch), but only one branch in the np phase. The lack of sampling explains the difference in free energy, because adding the degeneracy to the np phase lowers its minimum by $-k_B T \ln 2 = -1.73$ kJ/mol.

This behavior is explained based on the following explanation. Assume that both branches are completely equivalent, such that the law of total probability renders for the probability of visiting a volume V in branch $\pm$:

$$p(V|+) = p(V|-) = \frac{1}{2}p(V) \tag{23}$$

Suppose a simulation is performed that at V_{np} only samples the positive branch, but at V_{lp} samples both branches. For this simulation, a probability $\hat{p}(V)$ of visiting unit cell volume V and a free energy $\hat{F}(V)$ is obtained:

$$\begin{aligned} \hat{p}(V_{\text{np}}) &= p(V_{\text{np}}|+) = \frac{1}{2}p(V_{\text{np}}) && (24) \\ \hat{p}(V_{\text{lp}}) &= p(V_{\text{lp}}|+) + p(V_{\text{lp}}|-) = p(V_{\text{lp}}) && (25) \end{aligned}$$

Now, from Equation 8, for the true free energy difference between the lp and np phase $\Delta F_{\text{np-lp}}$:

$$\Delta F_{\text{lp-np}} = -k_B T \ln\Big(p(V_{\text{lp}})\Big) + k_B T \ln\Big(p(V_{\text{np}})\Big) \tag{26}$$

For the free energy difference between the lp and np phase $\Delta\hat{F}_{\text{np-lp}}$, found by incomplete sampling, we instead obtain:

$$\begin{aligned} \Delta\hat{F}_{\text{lp-np}} &= -k_B T \ln\Big(\hat{p}(V_{\text{lp}})\Big) + k_B T \ln\Big(\hat{p}(V_{\text{np}})\Big) && (27) \\ &= -k_B T \ln\Big(p(V_{\text{lp}})\Big) + k_B T \ln\Big(\frac{1}{2}p(V_{\text{np}})\Big) && (28) \\ &= \Delta F_{\text{lp-np}} - k_B T \ln\Big(2\Big) && (29) \end{aligned}$$

Hence it is seen that the correct $\Delta\hat{F}_{\text{lp-np}}$ can be obtained by accounting for the missed degeneracy in the np phase, and so lowering the free energy in the np phase by $k_B T \ln(2)$. Note that if in both the np as in the lp phase only one branch is visited, both correction terms cancel in the calculation of $\Delta\hat{F}_{\text{lp-np}}$.

The discussion here holds for a system in two volume states, the np and lp phase, with an a twofold degeneracy in each volume state. In one volume state this degeneracy is sampled, and it the other state it is not. It is seen that lowering the np phase minimum of the 1D free energy simulations by 1.73 kJ/mol makes the np and lp phase minimum, as well as the top of the np-to-lp transition barriers correspond, see Figure 7. However, expanding this reasoning to the complete free energy profile as a function of the volume is not straightforward.

E. Principal component analysis

The collective fluctuations found in the system are calculated for three simulations, one is the extended umbrella simulation from Subsection III-B, the other two are umbrella simulations with a parabolic bias, one in the np phase and the other in the lp phase. The important fluctuations are characterized by their contribution to the unit cell volume ($(c^{(l)}_{\text{vol}})^2$), the unit cell shape ($(c^{(l)}_{\text{shape}})^2$), the nuclear ($(c^{(l)}_{\text{nuc}})^2$) and the internal nuclear motion ($(c^{(l)}_{\text{nuc,int}})^2$). These components are found by projection of the eigenvectors $\vec{v}^{(l)}$ on the direction of volume change $\vec{e}_{\text{vol}}$, the unit cell shape space $(\vec{e}_{\text{shape}})_k$ with $k \in [0,5]$, and the nuclear coordinate space $(\vec{e}_{\text{nuc}})_m$ with $m \in [0,3N]$:

$$(c^{(l)}_{\text{vol}})^2 = (\vec{v}^{(l)} \cdot \vec{e}_{\text{vol}})^2 \tag{30}$$

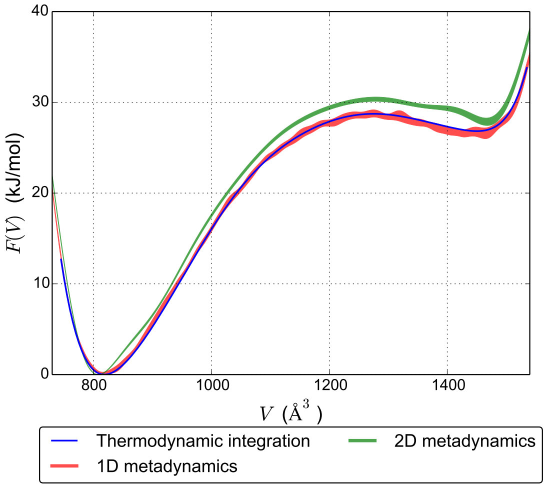


Fig. 6. The free energy profiles $F(V)$ as a function of the unit cell volume V obtained by the different sampling methods. The np phase minimum is set to zero for all curves. The profile of the 1D metadynamics simulation denotes the region $F(V) \pm 2\text{SE}_{F(V)}$.

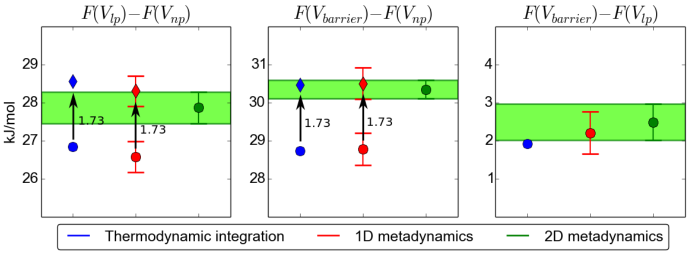


Fig. 7. Values of the free energy difference between the lp phase and np phase (left), the height of the free energy barrier from the np phase (middle), and the height of the free energy barrier with respect to the lp phase (right). The original values are denoted by dots, the values for which $F(V_{np})$ is corrected by a constant $-k_B T \ln 2 = -1.73$ kJ/mol are denoted by diamonds. Error bars denote 2 standard errors deviation.

$$(c^{(l)}_{\text{shape}})^2 = \sum_{k=1}^{5} (\vec{v}^{(l)} \cdot (\vec{e}_{\text{shape}})_k)^2 \tag{31}$$

$$(c^{(l)}_{\text{nuc}})^2 = \sum_{m=1}^{3N} (\vec{v}^{(l)} \cdot (\vec{e}_{\text{nuc}})_m)^2 \tag{32}$$

The contribution of the eigenmodes to the internal nuclear motion ($(c^{(l)}_{\text{nuc,int}})^2$) is the contribution due to a change in fractional coordinates. In other words, it is the motion of the nuclei that happens independently of the unit cell.

The results of the calculation are shown in Figure 8. It is concluded that a large set of collective nuclear motion is associated with a change in the unit cell shape. While some unit cell shape changes contain no internal nuclear motion, i.e. the fractional coordinates of nuclei remain fixed during the motion, most unit cell shape changes correlate with nuclear motion. Hence, to correctly sample the collective fluctuations of the nuclei, there is a great need to include the cell fluctuations in MD simulations. Again, the $NV\mathbf{h}_0T$ ensemble, in which the unit cell is kept fixed, is insufficient to simulate the collective nuclear fluctuations present in the system.

One of the important results in the discussion of the modes is

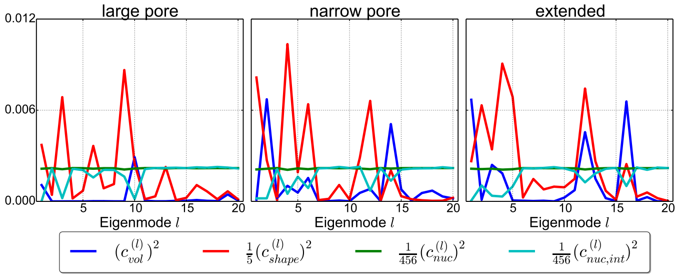


Fig. 8. The component sizes $(c^{(l)}_{\text{vol}})^2$, $(c^{(l)}_{\text{shape}})^2$, $(c^{(l)}_{\text{nuc}})^2$ and $(c^{(l)}_{\text{nuc,int}})^2$ for modes $l = 1, .., 20$ in the set of biased simulations (lp, np and extended bias). Terms are scaled by their dimension (1 for volume, 5 for shape, and 456 for the nuclei) for visualization.

that the np-to-lp transition does not contain any kneecap motion in the lp phase, in contrast to the np phase. This is concluded because the winerack motion does not contain any internal nuclear motion, while the kneecap does. Also, linker rotation around the linker axis in the np state is very suppressed in comparison to the lp state.

IV. Conclusion

This work is a thorough investigation of the extended and efficient sampling of the unit cell shape space. It describes methods for the construction of an accurate and precise free energy profile, describing the breathing behavior of the flexible crystal MIL-53(Al). Explicit sampling yields a better understanding of the influence of the unit cell shape on the calculated free energy profile.

To construct a correct free energy profile as a function of the unit cell volume for MIL-53(Al), it is necessary to sample the unit cell shape. At low volumes, a pair of stable and equivalent configurations of the unit cell, separated by a free energy barrier, is detected. This bistability in shape remains unsampled by methods that only implicitly sample the unit cell shape space, such as thermodynamic integration and 1D metadynamics in the volume direction. Of the methods described in this work, 2D metadynamics is the only method that can explicitly sample the bistability, and hence yield the correct free energy profile as a function of the unit cell volume. The free energy profiles obtained by methods that fail to sample both branches at np phase volumes (thermodynamic integration, 1D metadynamics) need to be corrected at these volumes around this np phase by a term $-k_B T \ln 2$. Correcting the free energy profile at intermediate states between the np and lp is not straightforward, and additional analysis is required.

Principal component analysis is performed on the np phase, lp phase and complete volume range between the two phases of MIL-53(Al) at 300 K. This analysis shows that the lp-to-np transition contains a winerack motion in both phases, while it contains no kneecap motion in the lp phase, in contrast to the np phase. The unit cell component fluctuations and collective nuclear motions are heavily coupled, hence sampling of the unit cell fluctuations is vital for the accurate sampling of the collective nuclear motions.

Acknowledgements

This work was performed at the Center for Molecular Modeling (CMM), under supervision of Prof. Dr. Ir. A. Ghysels, Prof. Dr. Ir. V. Van Speybroeck, Dr. Ir. L. Vanduyfhuys, Ir. R. Demuynck and Ir. S. Rogge. I wish to thank them for their support and guidance throughout the realization of this thesis.

REFERENCES

[1] T. Loiseau et al. “A Rationale for the Large Breathing of the Porous Aluminum Terephthalate (MIL-53) Upon Hydration”. In: *Chemistry–A European Journal* 10.6 (2004), pp. 1373–1382.

[2] Y. Liu et al. “Reversible structural transition in MIL-53 with large temperature hysteresis”. In: *Journal of the American Chemical Society* 130.35 (2008), pp. 11813–11818.

[3] P. Serra-Crespo et al. “Experimental evidence of negative linear compressibility in the MIL-53 metal–organic framework family”. In: *CrystEngComm* 17.2 (2015), pp. 276–280.

[4] T. Verstraelen et al. *Yaff, yet another force field*. 2013.

[5] L. Vanduyfhuys et al. “Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al)”. In: *Journal of Chemical Theory and Computation* 8.9 (2012), pp. 3217–3231. ISSN: 15499618.

[6] H. W. G. “Canonical Dynamics: Equilibrium Phase-Space Distributions”. In: *Phys. Rev. A: At., Mol., Opt. Phys.* 31.3 (1985), p. 1695.

[7] G. J. Martyna, D. J. Tobias, and M. L. Klein. “Constant pressure molecular dynamics algorithms”. In: *The Journal of Chemical Physics* 101.September (1994), p. 4177. ISSN: 00219606.

[8] S. M. J. Rogge et al. “A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks”. In: *Journal of Chemical Theory and Computation* 11.12 (2015), pp. 5583–5597. ISSN: 15499626.

[9] R. Martoňák, A. Laio, and M. Parrinello. “Predicting crystal structures: the Parrinello-Rahman method revisited”. In: *Physical review letters* 90.7 (2003), p. 075503.

[10] M. A. Balsera et al. “Principal component analysis and long time protein dynamics”. In: *The Journal of Physical Chemistry* 100.7 (1996), pp. 2567–2572.

Contents

List of abbreviations

BDC	benzenedicarboxylate
DFT	density functional theory
FEP	free energy profile
lp	large-pore/large pore
MC	Monte Carlo
MD	molecular dynamics
MIL	Matériaux de l'Institut Lavoisier
MOF	metal-organic framework
np	narrow-pore/narrow pore
PCA	principal component analysis
PES	potential energy surface
SE	standard error
SVD	singular value decomposition
WHAM	weighted histogram analysis method

Chapter 1

Introduction

> *"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."*
>
> — Paul A.M. Dirac, Proceedings of the Royal Society of London, April 6th 1929

The prophetic words by one of the brightest physicists in history marks the birth of the field of computational physics, decennia before the age of digital revolution. However, even the great Paul Dirac could not have foreseen the vast advances of mankind in theoretical physics, from the prediction of giant black holes, to the Standard Model of tiny subatomic particles, crescendoing together with findings in experimental physics, from the Hubble Telescope above the Thermosphere, to the Large Hadron Collider under the Alps.

The quote from Dirac particularly holds true for quantum and computational chemistry, a growing research field strongly intertwined with computational physics, describing molecular systems by means of computater simulations. The governing rules for the interaction between particles in these molecular systems are described by the Schrödinger equation from Quantum Mechanics. The mathematical tools for the solution to this equation are insufficient and the computational power to numerically find solutions is orders and orders of magnitude too weak.

Nonetheless, the scientific world is closing the gap on three frontiers. Firstly, the computers are slowly getting more powerful, performing more complex calculations in less time. Secondly, the methods to numerically solve the Schrödinger equation are becoming more accurate. This goes hand in hand with recent developments in accurate classical approximations of molecular systems (in particular force fields, see Section 2.4). Lastly, the simulation and analysis methods become increasingly efficient, leading to less simulation time required to get accurate information about the systems (see Section 2.5 - 2.7).

This work classifies as a contribution to this last type of advances in the field of computational physics, in which the unit cell properties of one specific molecular system, the metal-organic framework MIL-53(Al), are investigated. A good understanding of these properties yields valuable information of how the sampling of the unit cell shape results in an accurate determination of the free energy profile for flexible crystals.

1.1 Metal-organic frameworks

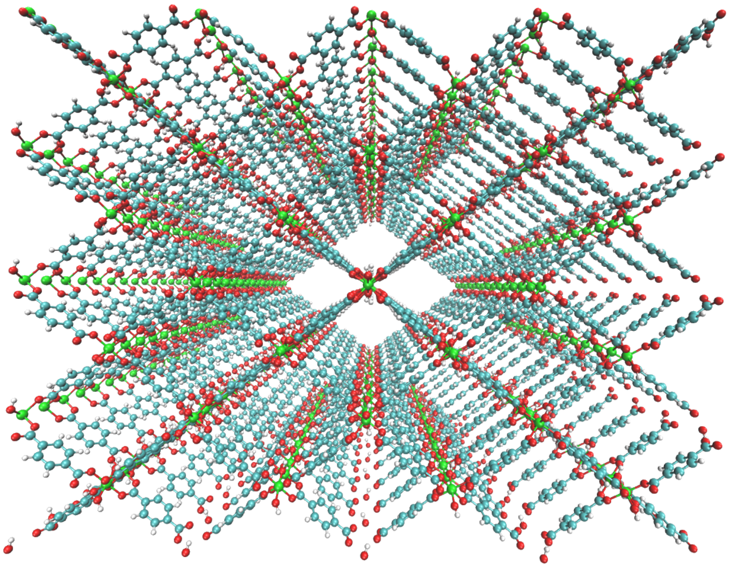


Figure 1.1: Visualization of the 3D structure of a metal-organic framework, MIL-53(Al).

Metal-organic frameworks (MOFs) form a new class of materials rapidly growing in size (the doubling time of the number of known 3D MOFs is estimated to be 3.9 years [1]). They are crystalline, periodic materials forming a framework structure with metal ions or clusters at nodal points and organic linkers connecting those points. Because of this framework structure, most MOFs have a large internal surface [2], which makes these materials suitable for adsorption and other applications.

The first stable MOFs were synthesized at the turn of the millenium: syntheses of new porous frameworks were described by Kitagawa et al. in 1998 [3]. The construction of a whole range of different MOFs rapidly followed: MOF-5 was synthesized by Yaghi and co-workers in 1999 [4], HKUST-1 by Charmant et al. in 1999 [5], and MIL-47 by Férey et al. in 2002 [6]. The crystallinity of the MOFs was confirmed by X-ray powder diffraction, their porosity by performing gas adsorption isotherm measurements [7]. The syntheses were performed by joining different building blocks, which makes them in general easily synthesizeable. These building blocks, metal-containing units and organic linkers, are joined together by strong bonds. Yaghi is seen as the founding father of this method to create the crystalline frameworks, called the reticular synthesis [8], patented in 1997 [9].

To date, an impressive number of other MOFs have been synthesized, and many more stable MOF structures have been computationally predicted. They can be divided into different classes, such as for example the MIL-53 family [10], the ZIFs [11], and the IRMOF series [2]. Because of the building block structure of MOFs, there is an almost endless number of hypothetical combinations. Rather than synthesizing this vast and growing number of combinations, the challenge lies in simulating and computationally characterizing these materials. By means of high-throughput screening procedures, MOFs suited for particular applications can be exposed [12, 13, 14].

1.1.1 Breathing

A large number of MOFs are stimuli-responsive: they can undergo large transformations in response to guest adsorption or external stimuli such as pressure, temperature or light [15]. A good example of such a transformation is breathing, which is of particular interest in this work. Breathing is a transition, induced by external stimuli, from a closed or narrow pore structure to an open or large pore structure of a material, corresponding to expansion and contraction respectively. A review of the breathing behavior in porous hybrid matter has been written by Férey and Serre in 2009 [16]. As a response to temperature, some MOFs show single-crystal to single-crystal phase transitions (MIL-53(Cr) [17]), and even negative thermal expansion (IRMOF-1 [18]).

Adsorption

The large inner surface of most MOFs makes them suitable for adsorption of molecules, by maximizing the attractive guest-host interactions [2, 19]. The framework structure may undergo large volume changes under guest adsorption. This is a consequence of the fact that the intramolecular interactions in the MOF (i.e. the flexibility of the linker-oxide connections and linker-linker dispersion) and intermolecular interactions between the MOF host and the guest molecules (the van der Waals force and electrostatic polarisation) are comparable in strength. Some of these MOFs show “sponge-like” behavior, i.e. they continuously swell upon solvent uptake (e.g. MIL-100 [20]). Others exhibit gate opening transformations, these are mostly MOFs with layered or interdigitated frameworks [21]. Multistep adsorption with multiple well-defined free energy minima has been observed in Co(BDP) [22]. Under the evacuation of solvents, MOFs can undergo reversible single-crystal to single-crystal or single-crystal to amorphous transitions [23].

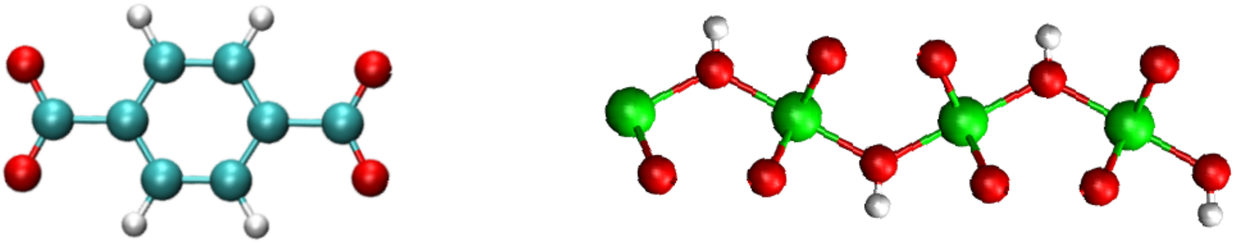
Mechanical stress

MOFs exhibit a wide variety of interesting mechanical properties. Under pressure, a high number of MOFs show anisotropic behavior [24]. Negative linear compressibility, i.e. expansion in certain directions under application of pressure, has recently been observed for the MOFs Ag(mim) [25], $NH_4Zn(HCOO)_3$ [26] and the MIL-53 family [27]. This is recognized in MOFs with a “winerack” framework motif. Auxeticity, the possession of directions with negative Poisson ratios, has been predicted to be present for some MOFs [24, 28], but as of the beginning of 2016 have not yet been verified experimentally.

A large number of MOFs have very low elastic moduli, which means that they are easily compressed. Experimentally, this behavior was first measured for ZIF-8, with a remarkably low shear modulus of 1 GPa [29]. The elastic moduli of MOFs can also show high anisotropy: they are much more easily compressed in some directions than in others, dependent on the topology of the framework. This last property is seen as a key signature for the breathing phenomenon mentioned earlier [24].

1.1.2 Applications

Industrial applications of MOFs include gas storage (CO_2 and CH_4 [30]), gas purification and separation applications (xenon-krypton separation [31]) as well as catalysis [32, 33]. The structure of MOFs allows for a design in which pore size and functionality could be varied systematically, for example to construct MOFs with a high capacity for methane storage [2]. Tuning MOFs for selective inclusion of large molecules and proteins and the exploitation of the pores as reaction vessels is possible [1]. Extensive research on MOFs is also performed in the fields of biomedical imaging [34] and proton conduction [35].



(a) BDC linker.

(b) Aluminum oxide chain.

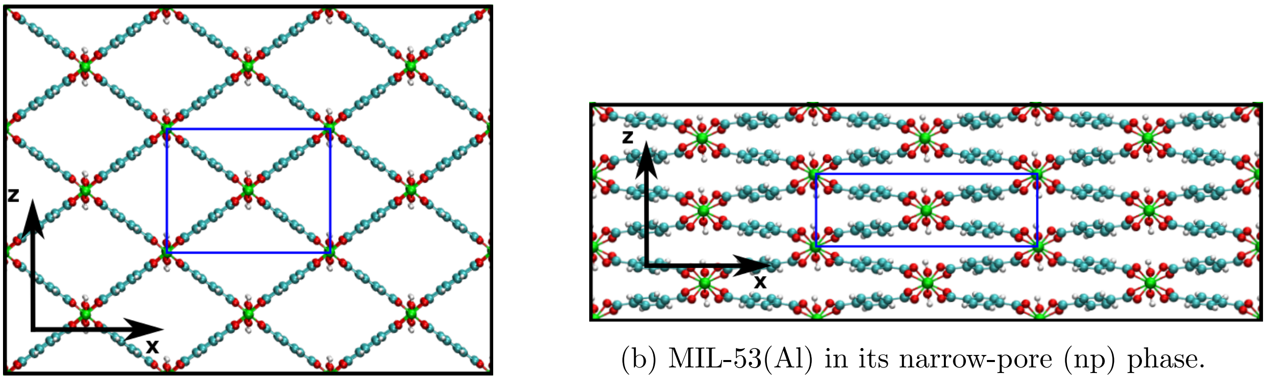
Figure 1.2: Building blocks of MIL-53(Al).

Some recent efforts are made for the computational screening of large sets of MOFs to find ones with appropriate properties for methane storage (a possible alternative fuel for cars) [12, 13, 14]. The porous properties of MOFs makes them also suitable to work as potential carriers for drug delivery [20]. The adsorption of guest molecules can be used to influence the mechanical behavior of MOFs. A joint experimental-modelling study has demonstrated a large flexibility of the MIL-47(V^{IV}) upon mechanical pressure which strongly deviates from its rigid behaviour in presence of guest molecules [36].

1.2 A more extensive look at MIL-53

The MIL-53 family (MIL = Matériaux de l'Institut Lavoisier) is characterized by a framework that consists of 1,4-benzenedicarboxylate (BDC) organic linkers (Figure 1.2a) connecting one dimensional M^{3+}(OH) chains (Figure 1.2b). These one dimensional chains result in the existence of 1D-channels, an important feature of MIL-53. The metal atom M^{3+} determines the subtype of MIL-53. The first type of MIL-53, named MIL-53(Cr), was first synthesized in 2002 by Serre et al [37]. Other known types include MIL-53(V) [6], MIL-53(Al) [10], MIL-53(Fe) [38], Mil-53(Ga) [39], MIL-53(Sc) [40] and finally a variant with indium, MIL-53(In) [41].

The MIL-53 family shows breathing behavior with a large magnitude in volume. This transition happens without breaking of the bonds between the building blocks, hence, this behavior is due to the structural flexibility of the framework [10, 37, 42]. The area of structural flexibility lies at the connection between the inorganic chain and the organic linker. In this work, the state of empty MIL-53(Al) at low volumes is labelled the narrow-pore (np) phase, the state of empty MIL-53(Al) at high volumes is labelled the large-pore (lp) phase. The transition from the lp to np phase is mostly due to a winerack



(a) MIL-53(Al) in its large-pore (lp) phase.

(b) MIL-53(Al) in its narrow-pore (np) phase.

Figure 1.3: The framework structure of MIL-53(Al) perpendicular to the 1D channels, aluminum oxide chains (in the channel direction) connected by BDC organic linkers. The blue box denotes the unit cell used in the simulations of this work. Nuclei are colored according to their type: green-Al, red-O, blue-C, white-H.

motion [27] (see Figure 1.4a): In the lp phase , the angle between neighboring linker planes in the z-direction is around $70° - 90°$, and in the np phase this is $30° - 50°$, see Figure 1.3. Another motion that can contribute to the lp-to-np transition is the kneecap motion [43] (see Figure 1.4b). This motion is due to a combination of nondirectional ionic AlO bonds with the rigidity of the triangular OCO carboxylate group in the BDC linker. A rotation of the linkers happens around the O-O axis between the linkers and the aluminum oxide chains resulting in a shrinkage of the volume. In the lp phase, the C-C-O-Al dihedral angle is distributed around 180°, while in the np phase these angles are distributed between ±120° and ±160° (using the IUPAC convention [44]). This behavior is verified both experimentally [45] and computationally [46].

1.2.1 Experiment

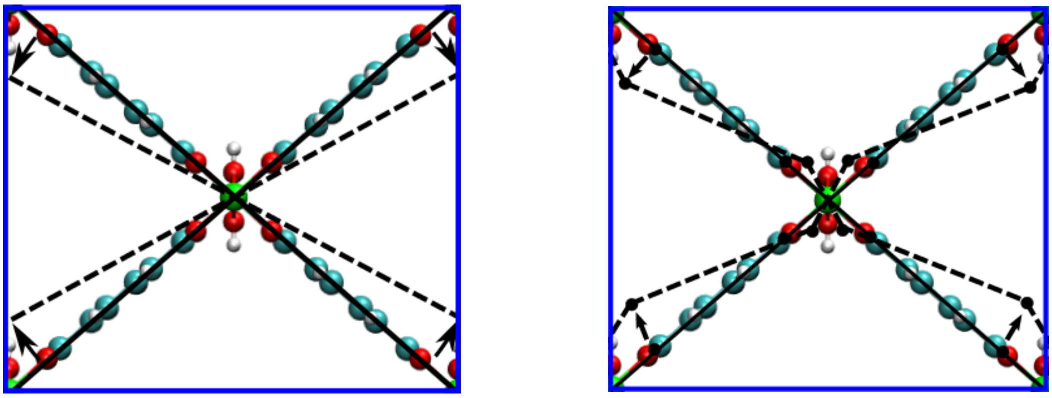
The synthesis and breathing behavior of MIL-53(Cr) is described by Serre et al. [37]. It was found that at room temperature, the MOF absorbs atmospheric water, and undergoes a fully reversible transition from its “ht” (high temperature) into its “lt” (low temperature) form. Both structures have been determined by X-ray diffraction, the thermal behavior investigated using thermogravimetric analysis and X-ray thermodiffractometry.

The first experiments on degassed MIL-53(Al) were performed by Liu et al. [42]. They found that MIL-53 exhibits a reversible structural transition between an open-pored and a closed-pored structure (the lp and np phase respectively) as a function of temperature without the presence of any guest molecules, using both neutron powder diffraction and inelastic neutron scattering techniques. In conclusion, at high temperatures and low pressures, empty MIL-53(Al) exists in the lp phase with a large unit cell volume (see Figure 1.3a). At high pressure, the unit cell transforms to the np phase with a smaller unit cell volume (see Figure 1.3b).

1.2.2 Simulation

The first ab initio simulation on MIL-53(Al) was performed by Walker et al. [47]. It was concluded that the breathing of the metalorganic framework MIL-53(Al) is controlled by a competition between short- and long-range interactions (dispersion) and entropic factors. The first force field simulation that could reproduce the breathing behavior in the presence of CO_2 for MIL-53(Cr) states was performed by Salles et al. [48]. Vanduyfhuys et al. constructed a force field for MIL-53(Al), parametrized by ab initio data [49]. The free energy profile of empty MIL-53(Al) was qualitatively reproduced and the breathing behavior was observed [46]. However, this force field predicts the np to be the most stable at room temperature and atmospheric pressure, while experiments show the lp to be more stable [42]. Another force field specifically designed for MOFs, is the UFF4MOF force field developed by Addicoat et al. [50].

The mechanism that induces the transition from the lp to np phase and backwards, due to adsorption is not fully understood. Most theoretical models on the thermodynamics of fluid adsorption use rigid framework hosts, but these models are insufficient for the description of breathing for these flexible materials. Coudert et al. use the grand potential in the osmotic ensemble for both the lp and np states to describe the breathing [51]. Ghysels et al. use a generic parametrized free energy model to describe the transition [52]. A semi-analytical mean-field model for MIL-53(Al) and MIL-53(Cr) was also tested and applied (Vanduyfhuys et al. [53]).



(a) The winerack motion.

(b) The kneecap motion.

Figure 1.4: Movement associated with the breathing of MIL-53(Al).

1.3 Objective

In this work computational research on the different unit cell configurations of the empty MIL-53(Al) is performed. The findings for this system are expected to be applicable to other systems, in particular metal-organic frameworks with similar topologies. First, the need for the explicit sampling of the unit cell shape space is demonstrated. This is done by comparing the results of two sets of simulations: one in which the shape space is sampled, and another in which it is not. Secondly, the different configurations of the study system are explored. It is already known that the empty MIL-53(Al) material exhibits a narrow-pore and a large-pore phase. This behavior is simulated, and the existence of an additional range of unit cell shapes of the study system is investigated.

The (relative) stability of the different configurations is determined by constructing free energy profiles. A variety of advanced simulation techniques are tested, and their accuracy is compared. The most appropriate set of simulation techniques is proposed, and the free energy differences (as well as the barriers) between configurations of the system are derived as accurately as possible. Finally, using principal component analysis, it is investigated how the motion of the nuclei couples to the motion of the unit cell. It was already recognized that the lp-to-np transition of MIL-53(Al) concurred together with a winerack and kneecap motion of the nuclei. The tools for analyzing the collective fluctuations in a molecular dynamics simulation are applied and extended. Additional nuclear motions present in the various configurations of the system under investigation, coupled with the unit cell, are studied by means of eigenvalue decomposition techniques.

Chapter 2

Methodology

All research in this work is performed computationally, by simulating a double unit cell of MIL-53(Al) containing 152 nuclei. With these simulations, the evolution of the microscopical system through time is studied. This chapter describes the underlying theory of the simulation techniques used in this work. Firstly, a definition is given of the unit cell of the system, together with the unit cell volume and shape. Next, the basics of statistical physics are explained, which will couple the microscopic configurations of the system to observable macroscopic properties. The third section elucidates simulation techniques that return sets of microscopic states in agreement with macroscopic constants, as efficiently as possible. A force field is the underlying functional form utilized to calculate energies and forces on particles in this simulation. Further, a set of physical ensembles is described, which simulate various experimental conditions. Among the ensembles of interest in this thesis, the free energy for different macrostates is calculated. The basic routines are explained, together with some more advanced techniques (umbrella sampling and metadynamics) to get a maximal resolution of the free energy at a minimal simulation effort.

2.1 The unit cell

MOFs form periodic crystals. Because of this periodicity, the properties of the MOF can be modelled by its properties inside a small periodic box of the complete crystal, the so-called unit cell. A unit cell is a periodic box (a representative subset of a material) that contains all unique atoms and interactions present in the complete crystal, while neglecting surface effects. The main properties of a sizeable MOF are determined by its bulk structure behavior, whereas surface effects are mostly negligible. The choice of the unit cell is not

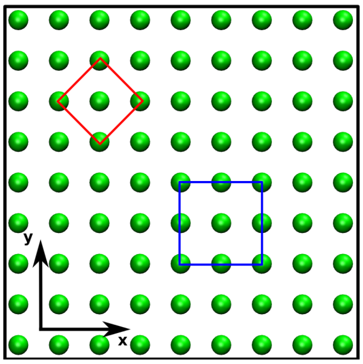


Figure 2.1: Visualization of a primitive (red) and an alternative (blue) unit cell of a crystalline 3D material (in this case, fcc aluminum).

unique. A special role is reserved for the primitive unit cell, i.e. the smallest cell that satisfies the previous definition. The primitive and an alternative unit cell for fcc Al are depicted in Figure 2.1. Unit cells can take different shapes and sizes, and are characterized by a 3×3 unit cell matrix $\mathbf{h}$, which contains three unit cell vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$, denoting the directions and lengths of periodicity.

$$\mathbf{h} = \begin{pmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} & \mathrm{h}_{13} \\ \mathrm{h}_{21} & \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{31} & \mathrm{h}_{32} & \mathrm{h}_{33} \end{pmatrix} = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix} \tag{2.1}$$

The shape and volume of the cell is completely determined by the relative lengths a, b and c of these vectors, and the angles between these vectors α (between $\vec{b}$ and $\vec{c}$), β (between $\vec{a}$ and $\vec{c}$) and γ (between $\vec{a}$ and $\vec{b}$). The complete unit cell matrix however, contains nine degrees of freedom. Six denote the volume and shape of the unit cell, while the remaining three describe the rotation of the coordinate system of the unit cell with respect to the "physical" coordinate system. The most simple unit cell shape is cubic ($a = b = c$, $\alpha = \beta = \gamma = 90°$), the most irregular triclinic ($a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90°$). The volume of the unit cell is defined as

$$V = \det(\mathbf{h}) \tag{2.2}$$

The unit cell shape can be described by normalizing the cell matrix $\mathbf{h}$:

$$\mathbf{h}_0 = \frac{\mathbf{h}}{V^{1/3}} \tag{2.3}$$

The unit cell shape and volume can change during the simulation. The system under investigation in this work is a double unit cell of MIL-53(Al), expanded in the $\vec{b}$ direction. Except for the volume V and unless explicitly mentioned, all unit cell coordinates are given for the doubled system.

2.2 Statistical physics

In thermodynamics, observable macroscopic properties for the studied system are defined, as well as empirically derived relations between them. These properties, e.g. volume or temperature, are a result of what happens to the system on a microscopic level. An observed macroscopic property can be the result of many different microscopic configurations of the system. In other words, a macrostate (a state of the system which is described by macroscopic quantities) can be the result of different microstates (microscopical configurations of the system). The space in which all degrees of freedom of the microstate are described is called the phase space. For molecular systems and within the Born-Oppenheimer approximation, the phase space usually consists of all degrees of freedom associated with the positions and momenta of the nuclei, sometimes augmented with the magnetic spins of these nuclei. The Born-Oppenheimer approximation is the assumption that the electronic motion can be separated from the nuclear motion, because the mass of the electrons is negligible in comparison to the mass of the nuclei. Hence, the electrons move much faster than the nuclei such that the equilibration time of the electronic motion is much smaller than the equilibration time of the nuclear motion. As a result, the potential energy $V(\vec{r}^N, \vec{p}^N)$ of the system depends parametrically on the nuclear configuration in phase space.

Statistical physics uses the information about the microstates to build a probabilistic model and to predict the properties of the macrostates. The underlying idea behind the methods in statistical physics is the ergodic hypothesis, first proposed by Boltzmann in his work "Vorlesungen über Gastheorie" in 1898. In its original form, the ergodic hypothesis states that all accessible microstates of the system are equiprobable over long periods of time. Here, the time spent in a certain macrostate relates to the number of microstates available in the macrostate. In its more modern form, the ergodic hypothesis states that the average of a process parameter over time equals the average of that parameter over the statistical ensemble (statistical ensembles are sets of macrostates, their definition is given in Subsection 2.5). By defining a partition function Z, an appropriate mathematical description of the statistical properties of the complete set of microstates present in an ensemble is obtained. This partition function is a function of the thermodynamic macrostate variables, such as temperature and volume.

2.2.1 Discrete systems

Consider a system characterized by a discrete set of energy levels E_j and degeneracies g_j (with $j = 1, ..., M$). If this system is in thermal equilibrium with a heat bath at temperature T, each state E_j is visited with a probability p_j proportional to the Boltzmann factor $g_j e^{-E_j/k_B T}$:

$$p_j = \frac{g_j e^{-E_j/k_B T}}{Z} \equiv \frac{g_j e^{-\beta E_j}}{Z} \tag{2.4}$$

where we defined $\beta = 1/k_b T$, and Z is a constant normalization factor. Since the probability of finding the system in one of the observable states j is one, the sum over the probabilities of all states j satisfies:

$$\sum_j p_j = \sum_j \frac{g_j e^{-\beta E_j}}{Z} = 1 \tag{2.5}$$

The normalization factor can be determined from the previous equation, it is named the discrete partition function Z:

$$Z = \sum_j g_j e^{-\beta E_j} \tag{2.6}$$

2.2.2 Continuous systems

In a continuous system with N (non identical) particles having a mass m, each having a position $\vec{r}$ and a momentum $\vec{p}$, the phase space is defined as the $6N$ dimensional space of all positions and momenta $(\vec{r}^N, \vec{p}^N)$. The probabilities p_j now become probability densities $p(\vec{r}^N, \vec{p}^N)$, having a dimension of $1/[\mathrm{Js}]^{3N}$. The energy in $(\vec{r}^N, \vec{p}^N)$ is described by a Hamiltonian $H(\vec{r}^N, \vec{p}^N)$. The probability $p(\vec{r}^N, \vec{p}^N) d\vec{r}^N d\vec{p}^N$ for the system being in a state around $(\vec{r}^N, \vec{p}^N)$ becomes:

$$p(\vec{r}^N, \vec{p}^N) d\vec{r}^N d\vec{p}^N = \frac{1}{h^{3N}} \frac{e^{-\beta H(\vec{r}^N, \vec{p}^N)}}{Q_{(r,p)}} d\vec{r}^N d\vec{p}^N \tag{2.7}$$

$Q_{(r,p)}$ is the total partition function which integrates the Boltzmann factor over the complete phase space $(\vec{r}^N, \vec{p}^N)$, in the case of $H(\vec{r}^N, \vec{p}^N) = (\vec{p}\ ^N)^2/2m + U(\vec{r}\ ^N)$:

$$\begin{aligned} Q_{(r,p)} &= \frac{1}{h^{3N}} \int e^{-\beta H(\vec{r}^N, \vec{p}^N)} d\vec{r}^N d\vec{p}^N && (2.8) \\ &= \frac{1}{\Lambda^{3N}} \int e^{-\beta U(\vec{r}^N)} d\vec{r}^N && (2.9) \\ &= \frac{1}{\Lambda^{3N}} Q_{(r)} && (2.10) \end{aligned}$$

and Λ, the thermal de Broglie wavelength, a normalisation factor, resulting from the integration over $\vec{p}^N$ in Equation 2.8:

$$\frac{1}{\Lambda^{3N}} = \frac{1}{h^{3N}} \int e^{-\beta \frac{p^2}{2m}} d\vec{p}^N \tag{2.11}$$

$$= \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2} \tag{2.12}$$

In this formula, h is Planck's constant and m the mass of the particles in the system. If the particles have different masses m_i $(i = 1, .., N)$, the factor $m^{3N/2}$ becomes $\Pi_{i=1}^{N} m_i^{3/2}$. Equation 2.5 becomes, with integration over the complete range $(\vec{r}^N, \vec{p}^N)$:

$$\int p(\vec{r}^N, \vec{p}^N) d\vec{r}^N d\vec{p}^N = \frac{1}{h^{3N}} \frac{1}{Q_{(r,p)}} \int e^{-\beta H(\vec{r}^N, \vec{p}^N)} d\vec{r}^N d\vec{p}^N = 1 \tag{2.13}$$

As can be seen from the last equation, not $p(\vec{r}^N, \vec{p}^N)$ but $p(\vec{r}^N, \vec{p}^N) d\vec{r}^N d\vec{p}^N$ is dimensionless. $Q_{(r)}$ is the partition function for the positions of the particles $\vec{r}^N$:

$$Q_{(r)} = \int e^{-\beta U(\vec{r}^N)} d\vec{r}^N \tag{2.14}$$

Note that $Q_{(r)}$ is not a dimensionless quantity, but has the dimensions $[\mathrm{m}^{3N}]$.

2.3 Molecular dynamics

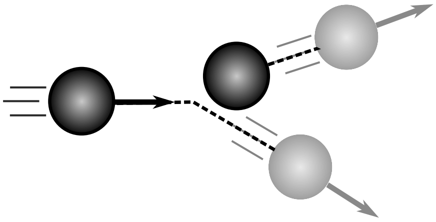


Figure 2.2: The first application of molecular dynamics: Elastic collisions between hard spheres.

The two main methods of simulation are Monte Carlo (MC) and molecular dynamics (MD). The form of MC as used today, Markov Chain Monte Carlo, was first developed in the 1940s by Ulam [54], and programmed by von Neumann on the ENIAC computer. In the 1950s, the technique was used at Los Alamos for early work relating to the development of the hydrogen bomb, and later became popularized in the fields of physics, physical chemistry, and operations research. A MC simulation uses random numbers to accept or reject random trial states/configurations of the material. A MC simulation step consists of generating a new configuration in which there is a certain random component present. Different algorithms exist for applying this scheme, of which the Metropolis algorithm is one of the more widely applied [55]. The Metropolis algorithm varies the system in one or more degrees of freedom (in a MOF this can for example be a nuclear position), and calculates the difference in potential energy between the trial and the initial configuration. If the energy of the trial configuration is lower, the trial state is accepted as the new initial state for the next simulation step. If the trial state results in a rise in energy, it can still be accepted, but with a reduced probability proportional to an exponentially decaying function of the energy difference. If it is not accepted, the next simulation step is performed with the old initial configuration.

In this work, however, molecular dynamics will be used. First proposed in 1959 by Alder and Wainwright [56], the technique was programmed on the IBM 704 computer to simulate elastic collisions between hard spheres, see Figure 2.2. With this technique, the equations of motion are integrated numerically to simulate the evolution of the system in time. A MD simulation step takes into account the positions and velocities of all particles in the system, and calculates the forces on these particles. Once these forces are known, the positions and velocities of the particles after a time Δt, the timestep, are calculated using an appropriate discretization of simple Newtonian mechanics. The negative gradient with respect to the nuclear position $\vec{r_i}$ of the potential energy $V(\vec{r}^N, \vec{p}^N)$ returns the force $\vec{F_i}$ on the i^{th} nucleus, such that the equations of motion can be derived:

$$\vec{F}_i(\vec{r}^N, \vec{p}^N) = -\frac{\partial V(\vec{r}^N, \vec{p}^N)}{\partial \vec{r}_i} \tag{2.15}$$

The gradients of the potential energy surface, i.e. the forces, can be calculated quantummechanically by constructing the electronic wavefunction or electronic density. These techniques, which in principle do not require any prior assumptions on the properties of the studied system (hence their pseudonym “ab initio techniques”), are computationally very demanding. Also, most low level ab initio techniques fail to describe long range dispersion interactions adequately. These interactions are key for the description of the MIL-53(Al) framework [47, 49].

2.4 Force fields

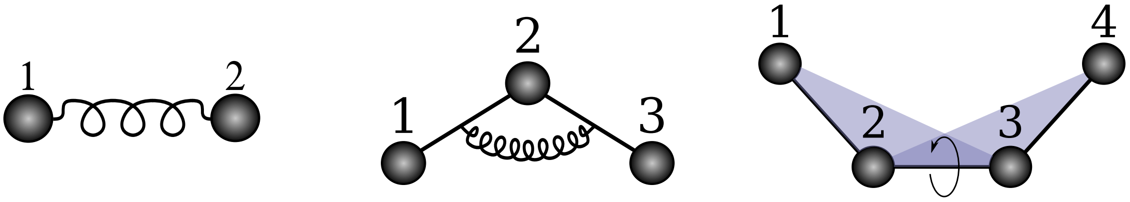


Figure 2.3: Covalent terms of the force field used for MIL-53(Al), respectively a bond, a bend and a dihedral angle. Figures courtesy of Louis Vanduyfhuys [57].

It is computationally more feasible to use a force field to obtain the energy and the forces in the system, instead of considering the electronic wavefunction or density. The research of the properties of MIL-53(Al) requires simulation times on the order of nanoseconds, the number of particles in a unit cell on the order of $100-1000$. Long range dispersion interactions are implemented by introducing a van der Waals potential between the nuclei [58], and the parameters of this potential can be manually implemented or tuned, if needed.

A force field is an approximate description of the potential energy surface, which parametrizes the potential energy surface (PES) V as an analytical function $V^{ff}(\vec{r}^N)$ only of the set of nuclear coordinates $\vec{r}^N$. An analytical expression of the gradients of this energy surface can easily be calculated. However, an accurate mathematical model is needed to approximate the potential energy surface. The mathematical model used here divides the potential energy into two parts, the covalent part, which contains all energy terms due to the bonding of the nuclei (bonds, bends and dihedral angles), and a noncovalent part, which contains the electrostatic and van der Waals interactions between nonbonded atoms.

The total potential energy V^{ff} is the sum of the different terms:

$$V^{ff} = V_{\text{bond}} + V_{\text{bend}} + V_{\text{dih}} + V_{\text{EI}} + V_{\text{vdW}} \tag{2.16}$$

$$V_{\text{bond}} = \sum_{i=1}^{N_r} \frac{1}{2} K_{r,i}(r_i - r_{i,0})^2 \tag{2.17}$$

$$V_{\text{bend}} = \sum_{i=1}^{N_\theta} \frac{1}{2} K_{\theta,i}(\theta_i - \theta_{i,0})^2 \tag{2.18}$$

$$V_{\text{dih}} = \sum_{i=1}^{N_\phi} \frac{1}{2} K_{\phi,i}(1 - \cos(m_i \phi_i)) \tag{2.19}$$

$$V_{EI} = \frac{1}{2} \sum_{i \neq j}^{N_{at}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.20}$$

$$V_{\text{vdW}} = \frac{1}{2} \sum_{i \neq j}^{N_{at}} \epsilon_{ij} \left[1.84 \times 10^5 \exp\left(-12 \frac{r_{ij}}{\sigma_{ij}} \right) - 2.25 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \tag{2.21}$$

$$\tag{2.22}$$

Various methods to obtain a parametrization of the force field terms exist:

- obtain parameters from a highly transferable, though less accurate, force field, e.g. UFF [59] or DREIDING [60];
- fit the model to reproduce experimental data;
- fit the model to ab initio calculations.

This work uses the ab initio parametrized force field developed by Vanduyfhuys et al. [49].

2.4.1 Covalent terms

In the work of Vanduyfhuys et al. [49], a set of ab initio density functional theory (DFT) calculations is performed on two nonperiodic clusters to optimize their geometry. Once this geometry is optimized, the gradients and Hessian matrix can be determined for the nuclei in the core region of the cluster, which are used to determine the force field parameters of the covalent energy terms.

2.4.2 Noncovalent terms

The noncovalent part consists of the electrostatic and van der Waals interaction. For the electrostatic interaction, nuclear charges are fitted as point charges on the nuclear positions. A variety of methods or partitioning schemes exists. The valence Hirschfeld-I scheme [61] is chosen to be the best candidate. It is a recent extension of the Hirshfeld-I partitioning scheme [62], which partitions the electron density. The MIL-53(Al) charges are derived from the cluster calculations. To get the covalent interactions, gradients of the electrostatic interaction have to be subtracted from the gradients in the DFT cluster calculation. The van der Waals parameters are taken from the MM3 force field of Allinger et al. [58].

2.4.3 Yaff

The force field simulations are performed using Yaff (Yet Another Force Field [63]). This is an in-house developed software to perform MD simulations. The code is open-source and Python based. The advantages of using in-house software are clear: the problems / error handlers are known, the developers are reachable, and very recent implementations which are not yet published are available. The code is a collaborative work, started and coordinated by Prof. T. Verstraelen, and the number of implemented methods is rapidly growing. The specific applications which make this work possible, are the barostats implemented by S. M. J. Rogge [46], and the metadynamics methods implemented by R. Demuynck.

2.5 Ensembles

In experiments, systems are subject to certain constraints: they can be kept at constant temperature, pressure, chemical potential, etc. Choices are made between conjugate pairs about what observables are controlled: the number of particles or the chemical potential, the energy or the temperature, the unit cell tensor or the stress tensor. If an observable is kept constant during an experiment, we only want to obtain the microstates that correspond to this observable. A statistical ensemble of a system is a collection of different microstates of that system, corresponding to that observable. Ensembles are named after the constant observable properties, e.g. the NVE ensemble contains all microstates with a given number of particles N, volume V and at a certain energy E.

First a definition of temperature, volume, pressure, unit cell shape and anisotropic pressure has to be given. In a 3D microscopic system with N particles, the instantaneous temperature T_i is defined as a measure of the kinetic energy of the particles:

$$\frac{3}{2}Nk_BT_i = \frac{1}{2}\sum_{j=1}^{N} m_j v_j^2 \tag{2.23}$$

During simulations, this instantaneous temperature can fluctuate, depending on what algorithm is used for the thermostat (see subsection 2.5.1). The temperature T of the ensemble is defined as the average over all simulation steps of the instantaneous temperature $T = \langle T_i \rangle$. The volume of the system in a simulation, i.e. the volume V of the unit cell, is described by Equation 2.2. The normalized unit cell shape $\mathbf{h}_0$ is defined in Equation 2.3. The stress on a system is defined by the stress tensor $\boldsymbol{\sigma}$. It can be divided in an isotropic contribution $P\mathbf{1}$ ($\mathbf{1}$ is the unit tensor) and an anisotropic contribution $\boldsymbol{\sigma}_\mathrm{a}$ (also referred to as the deviatoric stress):

$$\boldsymbol{\sigma} = P\mathbf{1} + \boldsymbol{\sigma}_\mathrm{a} \tag{2.24}$$

Now, the pressure is defined as the magnitude of the isotropic part P, and the deviatoric stress corresponds to $\boldsymbol{\sigma}_\mathrm{a}$. The stress tensor $\boldsymbol{\sigma}$ is defined as the ensemble average of the instantaneous stress tensor $\langle \boldsymbol{\sigma}_i \rangle$, the same holds for the pressure P and the anisotropic contribution $\boldsymbol{\sigma}_\mathrm{a}$.

The number of particles in an MD simulation remains fixed. The appropriate set of microstates at a given temperature T is found by running an MD-simulation and applying a thermostat, which makes the instantaneous temperature T_i fluctuate around a predefined T. The same holds for the pressure, for which a barostat is applied. Several techniques and algorithms exist for applying these thermostats and barostats, and some of them are discussed in the next subsections.

2.5.1 Energy or temperature?

Since Newton's equations of motion conserve the total energy, basic MD simulations (NVE simulations) automatically conserve the energy. To control the temperature instead, a variety of methods exist:

- Velocity rescaling [64][65]: after some number of simulation steps, the difference between the instantaneous and the imposed temperature is calculated and all velocities are scaled so that this difference is zero. This method is deterministic but very artificial since it suppresses the fluctuations in the instantaneous temperature.
- Andersen thermostat [66]: stochastic forces are applied on a randomly chosen set of particles. These particles are assigned a new velocity that is drawn from a Maxwell-Boltzmann distribution at the imposed temperature T. An important parameter to consider is the collision frequency: a high application frequency of the stochastic forces leads to a faster thermal equilibruim, but strongly disturbs the dynamic behavior.

- Nosé-, Nosé-Hoover and Nosé-Hoover chains thermostat [67]: The Lagrangian equations of motion are adapted, and extra variables are introduced. In the Nosé formulation, time and distance are scaled, while in the Nosé-Hoover formulation, two thermodynamic friction coefficients are implemented. They describe the kinematics of a heat bath. It can be shown that the Lagrangian of the Nosé-Hoover chains thermostat leads to a correct temperature distribution.

2.5.2 Unit cell tensor or stress tensor?

In basic MD simulations, the unit cell tensor is fixed, and periodic boundary conditions are applied to the system. However, most experiments are performed under constant isotropic pressure. To mimic these conditions during a simulation, a barostat technique is applied. Applying a barostat can be done using the method proposed by Berendsen et al. [68]. This results in an exponential dampening of the difference between the imposed and the required stress. Because of this dampening, strong fluctuations in kinetic and potential energy are suppressed, which results in unphysical behavior [69].

A better barostat for the purposes of this work is the MTTK barostat, named after Martyna, Tuckerman, Tobias and Klein [70]. It applies the same approach to extend the Lagrangian as in the Nosé-Hoover thermostat. The main advantage of the MTTK barostat is that the equations of motion can be rewritten in such a way that the motion of the unit cell matrix $\mathbf{h}$ can be split up in the motion of the volume V and the motion of the normalized unit cell matrix $\mathbf{h}_0$ [46]. The isotropic pressure P is related to the volume, while the deviatoric stress $\boldsymbol{\sigma}_a$ is related to the normalized unit cell tensor. With this method, different combinations of ensembles are possible, and the most commonly used ensembles here will be the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$, $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ and $NV\mathbf{h}_0T$ ensemble.

A last barostat technique is the Langevin piston method [71]. A damping force is added to the equations of motion, aside from the existing stochastic force similar to the Langevin thermostat. Both the MTTK barostat and Langevin piston are verified to reproduce the physically correct ensembles, although they converge more slowly than the Berendsen barostat method [46].

2.6 Free energy methods

As a result of the second law of thermodynamics, for a closed system with energy E, volume V and number of particles N, the entropy S is maximum at equilibrium. In contrast, for a system in thermal equilibrium connected to a heat bath at a temperature T, the Helmholtz free energy F is at a minimum. The free energy F is defined as:

$$F = E - TS \tag{2.25}$$

It can be proven that

$$\left(\frac{\partial F}{\partial V}\right)_T = -P \tag{2.26}$$

For a system in thermal equilibrium with a heat bath and kept at a constant isotropic pressure, i.e. the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, the Gibbs free energy G is at a minimum.

$$G = F + PV \tag{2.27}$$

If the system exhibits two or more equilibria, the stable equilibrium is the state with the lowest free energy. For further reference, the free energy is denoted as F and the ensemble is explicitly mentioned. The free energy for a global state of a molecular system in the $NV\mathbf{h}_0T$ ensemble can be written as a function of the partition function $Q_{(r,p)}$ (see Equation 2.8):

$$F = -k_BT\ln(Q_{(r,p)}) = -k_BT\ln\left(\frac{1}{h^{3N}}\int e^{-\beta H(\vec{r}^N,\vec{p}^N)}d\vec{r}^N d\vec{p}^N\right) \tag{2.28}$$

Suppose a coordinate $C(\vec{r}^N,\vec{p}^N)$ is defined for all points in phase space. All microstates with $C(\vec{r}^N,\vec{p}^N) = q$ form an ensemble. When discussing the sampling of states, this macrostate q is more commonly referred to as the collective variable. With each of these ensembles corresponds a partition function $Q_{(r,p)}(q)$:

$$Q_{(r,p)}(q) = \frac{1}{h^{3N}}\int e^{-\beta H(\vec{r}^N,\vec{p}^N)}\delta\Big(C(\vec{r}^N,\vec{p}^N) - q\Big)d\vec{r}^N d\vec{p}^N \tag{2.29}$$

Note that $\int Q_{(r,p)}(q)dq = Q_{(r,p)}$. The probability $p(q)dq$ for the system to visit the macrostate $C \in [q, q+dq]$ is the integral of the Boltzmann factor over that part of phase space that satisfies $C(\vec{r}^N,\vec{p}^N) = q$. The probability density $p(q)$ can be calculated as follows:

$$\begin{aligned}
p(q) &= \int p(\vec{r}^N,\vec{p}^N)\delta\Big(C(\vec{r}^N,\vec{p}^N) - q\Big)d\vec{r}^N d\vec{p}^N && (2.30)\\
&= \frac{1}{h^{3N}}\frac{1}{Q_{(r,p)}}\int e^{-\beta H(\vec{r}^N,\vec{p}^N)}\delta\Big(C(\vec{r}^N,\vec{p}^N) - q\Big)d\vec{r}^N d\vec{p}^N && (2.31)\\
&= \frac{Q_{(r,p)}(q)}{Q_{(r,p)}} && (2.32)
\end{aligned}$$

The free energy $F(q)$ for a collective variable q is defined as

$$F(q) = -k_B T \ln\left(p(q)/c\right) = -k_B T \ln\left(p(q)\right) + c' \tag{2.33}$$

where the constant c is introduced, making the expression $(p(q)/c)$ dimensionless. Due to the fact that changing the factor c only introduces a non relevant shift of the entire free energy profile, we tacitly assume this factor to be unity for the remainder of this thesis.

The previous discussion here holds for systems simulated in the $NV\mathbf{h}_0T$ ensemble, and collective variables q which are only a function of $(\vec{r}^N, \vec{p}^N)$. A generalisation of the partition function is needed for systems in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ and the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, for collective variables q as a function of the unit cell matrix $\mathbf{h}$. The previous definition of the partition functions $Q_{(r,p)} \equiv Q_{(r,p)}(\mathbf{h})$ (Equation 2.8) and $Q_{(r,p)}(q)$ (Equation 2.29) is extended. An expression for the partition function $Q_{(r,p,\mathbf{h})}$ in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble is proposed by Martyna, Tobias and Klein [72]:

$$Q_{(r,p,\mathbf{h})} = \int \exp\big[-\beta P_{\text{ext}} \det(\mathbf{h})\big] Q_{(r,p)}(\mathbf{h}) \det(\mathbf{h})^{1-d} d\mathbf{h} \tag{2.34}$$

Here d is the dimension of the system (in this case $d = 3$), and $\mathbf{h}$ is assumed symmetric, so $d\mathbf{h}$ is an integration of six components, it has the dimension Å^6. The expression for the partition function as a function of the collective variable q becomes:

$$Q_{(r,p,\mathbf{h})}(q) = \int \exp\big[-\beta P_{\text{ext}} \det(\mathbf{h})\big] Q_{(r,p)}(\mathbf{h}) \det(\mathbf{h})^{1-d} \delta\Big(C(\vec{r}^N, \vec{p}^N, \mathbf{h}) - q\Big) d\mathbf{h} \tag{2.35}$$

Note that equivalently $\int Q_{(r,p,\mathbf{h})}(q) dq = Q_{(r,p,\mathbf{h})}$. The previous definitions of $p(q)$ (Equation 2.32) and $F(q)$ (Equation 2.33) are adapted:

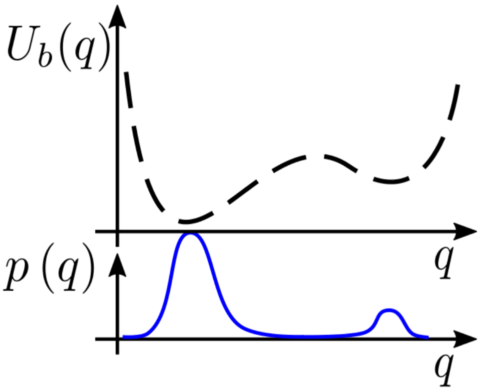
$$p(q) = \frac{Q_{(r,p,\mathbf{h})}(q)}{Q_{(r,p,\mathbf{h})}} \tag{2.36}$$

The free energy $F(q)$ for a collective variable q is defined as

$$F(q) = -k_B T \ln\left(p(q)/c\right) = -k_B T \ln\left(p(q)\right) + c' \tag{2.37}$$

where again the constant c is introduced, making the expression $(p(q)/c)$ dimensionless and this factor is assumed to be unity for the remainder of this thesis.

2.6.1 Histogram method



An estimate of the probability density $p(q)$ for each collective variable q can be obtained from an MD simulation with N simulation steps, by constructing a histogram $h(q)$ with binwidth W:

$$p(q) \;=\; \frac{h(q)}{NW} \tag{2.38}$$

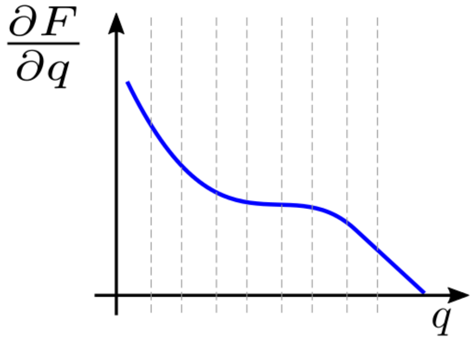
Equation 2.37 can then be used to obtain an estimate of the free energy in q:

$$F(q) \;=\; -k_B T \ln[p(q)] \tag{2.39}$$

One of the problems with the histogram method is that each region of q has to be sufficiently sampled. In other words, the set of microstates $(\vec{r}^N, \vec{p}^N)$ should be dense in the accessible phase space corresponding to a macrostate q. In a normal MD simulation, the macrostates which have a high probability are sufficiently sampled. This is in contrast to those with low probability, of which a very low number of microstates have been visited. In these last regions, the histogram $h(q)$ contains a lot of noise, resulting in a bad estimate of $p(q)$. Also, with simulations of finite time and low temperature, the system can get stuck in free energy minima (regions with high probability can not be reached if the path towards them has a low probability) and never sample the complete range of accessible microstates corresponding to q. This contributes to the fact that a lot of noise can be expected from a free energy profile obtained by the histogram method.

A solution to this problem is given by the application of a bias potential to the system, forcing it to visit a complete range of global coordinates $[q_1, q_2]$. Afterwards, the obtained free energy profile must be corrected for this bias potential. These methods are discussed in Section 2.7.

2.6.2 Thermodynamic integration



In the $NV\mathbf{h}_0T$ ensemble, if the Hamiltonian $H(\vec{r}^N, \vec{p}^N; q)$ depends on the collective variable q, the free energy difference between two states can be calculated by taking the derivative of Equation 2.28:

$$\frac{\partial F(q)}{\partial q} = -k_B T \frac{\int -\beta \frac{\partial H}{\partial q} e^{-\beta H(\vec{r}^N, \vec{p}^N)} d\vec{r}^N d\vec{p}^N}{\int e^{-\beta H(\vec{r}^N, \vec{p}^N)} d\vec{r}^N d\vec{p}^N} = \langle \frac{\partial H}{\partial q} \rangle \tag{2.40}$$

$$F(q) - F(q_0) = \int_{q_0}^{q} \frac{\partial F(q')}{\partial q'} dq' = \int_{q_0}^{q} \langle \frac{\partial H}{\partial q'} \rangle_{q'} dq' \tag{2.41}$$

In the case that the collective variable q is the volume V, the extension of the partition function in Equation 2.34, and the relation between P and $\partial F/\partial V$ from Equation 2.26 can be used. Thermodynamic integration of $\partial F/\partial V$ finally leads to the following identity:

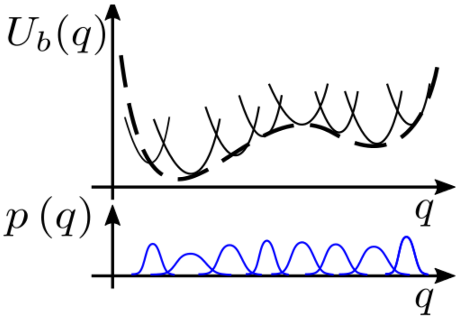
$$F(V) - F(V_{ref}) = \int_{V_{ref}}^{V} \frac{\partial F(V')}{\partial V'} dV' = - \int_{V_{ref}}^{V} \langle P \rangle_{V'} dV' \tag{2.42}$$

where $\langle P \rangle_{V'}$ is the ensemble average of the pressure in the system fixed at a volume V'. Also here the problem of sampling all the microstates arises. This technique assumes that the complete ensemble at a given volume V' is efficiently sampled, so the average pressure $\langle P \rangle_{V'}$ is accurately determined. However, the system can get stuck in a local minimum again, for which a different average pressure $\langle P \rangle_{V'}$ is obtained. More advanced techniques will offer a solution to the problem.

2.7 Enhanced sampling

The equilibrium state of a system connected with a heat bath is the state with minimal Helmholtz free energy F (see equation 2.25). When a simulation is performed at low temperature, the system fluctuates around this optimal state. As a consequence, the states far away from this equilibrium are very rarely visited, especially if they are separated from the equilibrium by a high free energy barrier. The enhanced sampling methods proposed here add a bias potential $U_b(q)$ as a function of the collective variable q to the free energy profile, forcing the system to visit states which are otherwise not visited sufficiently in normal MD simulations [73].

2.7.1 Umbrella sampling



In an umbrella simulation, the bias potential is added a priori to the system and is not altered during the MD simulation. The biased partition function $Q_{(r,p)b}$ in the $NV\mathbf{h}_0T$ ensemble reads:

$$Q_{(r,p)b} \;=\; \frac{1}{h^{3N}}\int e^{-\beta\left(H(\vec{r}^N,\vec{p}^N)+U_b(C(\vec{r}^N,\vec{p}^N))\right)}d\vec{r}^N d\vec{p}^N \tag{2.43}$$

and $Q_{(r,p)b}(q)$ is given by:

$$Q_{(r,p)b}(q) \;=\; \frac{1}{h^{3N}}\int e^{-\beta\left(H(\vec{r}^N,\vec{p}^N)+U_b(q)\right)}\delta\Big(C(\vec{r}^N,\vec{p}^N)-q\Big)d\vec{r}^N d\vec{p}^N \tag{2.44}$$

The biased free energy $F_b(q)$ follows from Equation 2.36 and 2.39 (again, the constant by which $F_b(q)$ can be shifted is not taken into account):

$$\begin{aligned} F_b(q) &= -k_BT\ln[Q_{(r,p)b}(q)] && (2.45)\\ &= -k_BT\ln[Q_{(r,p)}(q)e^{-\beta U_b(q)}] && (2.46)\\ &= F(q)+U_b(q) && (2.47)\\ \to F(q) &= F_b(q)-U_b(q) && (2.48)\\ &= -k_BT\ln[p_b(q)]-U_b(q) && (2.49)\end{aligned}$$

The free energy profile of the system can hence be reconstructed by correcting the biased free energy $F_b(q)$ (for example, calculated with the histogram method) for the applied bias potential (see Equation 2.49).

The bias potential U_b is usually harmonic in q, i.e. $U_b(q) = K\big(q(\vec{r}^N, \vec{p}^N) - q_0\big)^2$, with q_0 the rest value and K the force constant. If K is large, the system is sampled very narrowly around the minimum q_0. A common technique consists of performing a set of k umbrella simulations differing in $q_0^{(k)}$, with $q_0^{(k)}$ evenly distributed over a range of values one is interested in. It is necessary that the neighbouring umbrellas overlap: they need to contain a common set of macrostates that are visited sufficiently. Subsequently, a global free energy profile spanning different umbrellas is reconstructed using the weighted histogram analysis method (WHAM)[74].

Consider a system for which the free energy as a function of q is a priori known. One can then perform a biased simulation to obtain an equal sampling over the range of the collective variable q. The bias potential is taken as the negative of the know free energy profile:

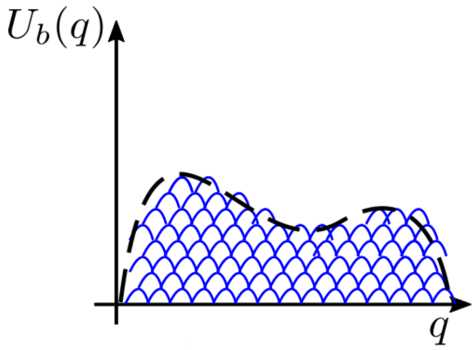
$$U_b(q) = -F(q) \tag{2.50}$$

This leads to all states q having an equal probability $p_b(q)$ (see Equation 2.36):

$$p_b(q) \propto e^{-\beta F_b(q)} = e^{-\beta\big(F(q)+U_b(q)\big)} = 1 \tag{2.51}$$

Hence, all states as a function of q are evenly sampled.

2.7.2 Metadynamics



Metadynamics [75, 76] is a sampling technique that adds a bias potential $U_b(t, q)$ to the system that changes during the MD simulation. It alters the energy expression by adding a history-dependent bias potential as a function of a chosen collective variable q. This collective variable can for example be an interatomic distance, dihedral angle or coordination number [77]. A metadynamics simulation is said to be converged when the final histogram as a function of q, constructed with the simulation steps after the last change to the bias potential, is sufficiently flat in the region of interest. The final result of this technique places the system in a flattened free energy profile, in which all states (for different q) are thus evenly sampled.

This technique allows for the straightforward reconstruction of the free energy profile along the collective variable, by taking the negative of the converged bias potential. Since the histogram is flat after conversion, the biased probability of visiting a given macrostate is constant $p(q) = \frac{1}{\Omega}$ with Ω the width of the sampled region of q. Again, as in Equation 2.49, the biased free energy becomes $F_b(q) = F(q) + U_b(q)$. From the histogram method, Equation 2.39, it follows that $F_b(q)$ is constant:

$$F_b(q) = -k_B T \ln p_b(q) = k_B T \ln \Omega = c \tag{2.52}$$

Combining this with the expression for the biased free energy $F_b(q) = F(q) + U_b(q)$:

$$\begin{aligned} c &= U_b(q) + F(q) && (2.53) \\ F(q) &= -U_b(q) + c && (2.54) \end{aligned}$$

This is the final expression for the estimate of the free energy $F(q)$ once the metadynamics simulation has converged.

Evolution of the bias potential

An expression of the bias potential $U_b(t, q)$ during the metadynamics simulation is needed. This expression is formed by summing over Gaussian hills, $h_i e^{\frac{-(q-q_i)^2}{2w_i^2}}$, where h_i and w_i are the height and the width of the Gaussian respectively. A metadynamics step consists of a simulation with a length on the order of picoseconds, after which (at time t_i) a Gaussian hill is added to the bias potential. In this work, the width w_i and height h_i of the Gaussian are specified a priori (so they become w and h respectively). The bias potential at time t during the simulation then becomes:

$$U_b(t, q) = \sum_{i|t_i<t} h e^{\frac{-(q-q_i)^2}{2w^2}} \tag{2.55}$$

Further calculations require the partial derivative of $U_b(t,q)$ with respect to q. Straightforward calculation shows:

$$\frac{\partial U_b(t,q)}{\partial q} = -\sum_{i|t_i<t} \frac{h}{w^2}(q-q_i)e^{\frac{-(q-q_i)^2}{2w^2}} \tag{2.56}$$

Unit cell component as collective variable

In this work, the collective variables for metadynamics are components of the unit cell. The collective variable q is described as a function of the unit cell parameters. The derivative of the bias potential $U_b(q)$ with respect to the transposed unit cell matrix $\mathbf{h}^T$ yields [78]:

$$\frac{\partial U_b(q)}{\partial \mathbf{h}^T} = \mathbf{\Xi}_b^T \cdot \mathbf{h}^{-1} \tag{2.57}$$

with $\mathbf{\Xi}_b$ the virial tensor due to the bias potential U_b and the inverse of the unit cell matrix $\mathbf{h}^{-1}$. The previous equation is transformed into

$$\begin{aligned}
\mathbf{\Xi}_b &= \mathbf{h}^T \cdot \frac{\partial U_b(q)}{\partial \mathbf{h}} && (2.58)\\
&= \mathbf{h}^T \cdot \Big(\frac{\partial U_b(q)}{\partial q}\frac{\partial q}{\partial \mathbf{h}}\Big) && (2.59)\\
&= \mathbf{h}^T \cdot \Big(\frac{\partial U_b(q)}{\partial q}\frac{\partial q}{\partial (\mathbf{h}_s \cdot \mathbf{R})}\Big) && (2.60)\\
&= \mathbf{h}^T \cdot \Big(\frac{\partial U_b(q)}{\partial q}\frac{\partial q}{\partial \mathbf{h}_s} \cdot \mathbf{R}^{-1}\Big) && (2.61)
\end{aligned}$$

Where $\mathbf{h}_s$ is the symmetrized unit cell matrix, $\mathbf{R} = \mathbf{UV}$ the rotation matrix making the unit cell matrix symmetric (see Subsection 4.1.1) and $\frac{\partial U_b(q)}{\partial q}$ is calculated in Equation 2.56. The contribution to the virial tensor due to the bias potential is implemented by adding $\mathbf{\Xi}_b$:

$$\mathbf{\Xi} \rightarrow \mathbf{\Xi} + \mathbf{\Xi}_b \tag{2.62}$$

Chapter 3

Exploration of the metastable states of MIL-53(Al)

3.1 The need for implicit sampling of the unit cell shape

Two free energy profiles are constructed as a function of the unit cell volume of MIL-53(Al) by thermodynamic integration of the average (isotropic) pressure over the unit cell volume (see equation 2.42). For the first free energy profile, the average pressure is determined by fixed-volume simulation in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble. For the calculation of the second free energy profile, the average pressure is determined by simulations in the $NV\mathbf{h}_0T$ ensemble. Comparison of the profiles shows the effect of the implicit sampling of the unit cell shape fluctuations (first simulation set) versus the not sampling of the unit cell shape (second simulation set). The obtained free energy profiles serve as a reference to which further profiles are compared.

3.1.1 Implicitly sampling the unit cell shape fluctuations

For the technique described in Subsection 2.6.2, the isotropic pressure $\langle P(V)\rangle$, averaged over the ensemble at a fixed volume V, is needed. To determine this average isotropic pressure, 160 independent simulations in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble are performed, with volumes ranging from 740 to 1535 Å^3 in steps of 5 Å^3. Initial structures for the system at these volumes are obtained by performing a simulation in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, where the unit cell transitions from the lp ($\approx$ 1500 Å^3) to the np ($\approx$ 800 Å^3) phase. Detailed information on the set of simulations can be found in Table 3.1.

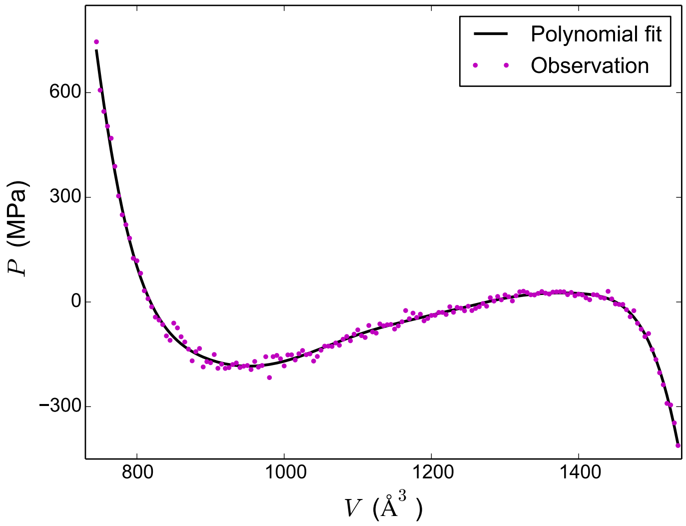


Figure 3.1: **The NV($\boldsymbol{\sigma}_\mathbf{a} = \mathbf{0}$)T ensemble**. The pressure $\langle P(V)\rangle$ as a function of the unit cell volume V for MIL-53(Al). MD results obtained from 160 simulations in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, with volumes ranging from 740 to 1535 Å^3 in steps of 5 Å^3. The solid line is an 11th order polynomial fit.

The obtained averages of the pressure can be found in Figure 3.1. An 11th order polynomial is fitted through these points. This polynomial provides a good description of the pressure profile, because it is of odd order and no overfitting issues arise [46].

$$P(V) = \sum_{i=0}^{11} c_i (V - V_{\mathrm{ref}})^i \tag{3.1}$$

Table 3.1: Simulations performed in Subsection 3.1.1.

Ensemble	$NP(\boldsymbol{\sigma}_a = \mathbf{0})T$	$NV(\boldsymbol{\sigma}_a = \mathbf{0})T$
Timestep	0.5 fs	0.5 fs
Simulation time	75 ps	400 ps
Temperature	300 K	300 K
Pressure	10^6 pascal	–
Starting volume	1496 Å^3	740 – 1535 Å^3
Description	Transition lp→np	Fixed volume simulations
Goal	Obtain initial structures	Obtain ensemble average pressure

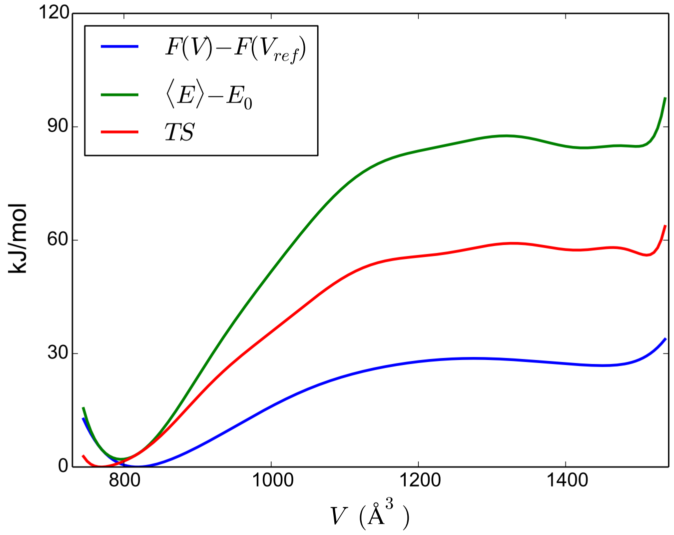


Figure 3.2: **The NV($\boldsymbol{\sigma_a} = 0$)T ensemble**. **Blue**: Free energy profile as a function of the unit cell volume V for MIL-53(Al), obtained by thermodynamic integration of the polynomial fit to the pressure $\langle P(V)\rangle$ (see Figure 3.1). $F(V_{\text{ref}})$ is chosen to be the minimum at $V \approx 820$Å^3. **Green**: 12th order polynomial fit to the average energy $\langle E\rangle - E_0$ for 160 different volume simulations ranging from 740 to 1535 Å^3 in steps of 5 Å^3. The constant E_0 is chosen such that the minimum of the entropy curve $TS = \langle E\rangle - E_0 - \big(F - F(V_{\text{ref}})\big)$ (**Red**) is zero (see Equation 2.25).

Equation 2.42 can then be integrated analytically:

$$F(V) - F(V_{\text{ref}}) = -\sum_{i=0}^{11} \frac{c_i}{i+1}(V - V_{\text{ref}})^{i+1} \tag{3.2}$$

The resulting free energy profile is shown in Figure 3.2. $F(V_{\text{ref}})$ is chosen to be the minimum of the free energy around 820 Å^3. It is observed that MIL-53(Al) exhibits bistable behavior in the volume, corresponding with experimental observations described in Subsection 1.2.1. The free energy barrier between the lp local minimum and the maximum of the transition region found here ($V = 1275$ Å^3) is 1.9 kJ/mol. However, the relative stability of these states does not agree with experiment (see Subsection 1.2.1): the lp minimum ($V = 1450$ Å^3) is found to be higher in free energy than the global minimum in the np, by an amount of 26.8 kJ/mol. The second quantity that is calculated is the average internal energy $\langle E\rangle$ of each fixed volume (12th order polynomial is fitted). Since the internal energy is determined up to a constant, a constant E_0 is subtracted from this value, chosen such that the minimum of $TS = \langle E \rangle - E_0 - \big(F - F(V_{\text{ref}})\big)$ is zero (see Equation 2.25).

3.1.2 Not sampling the unit cell shape fluctuations

To get an estimate of the contribution of the unit cell shape fluctuations on the previously obtained free energy profile, the method is repeated. However, now not only the volume V, but also the unit cell shape $\mathbf{h}_0$ is kept fixed during the simulations, i.e. a different ensemble is used. Previously, the ensemble describing the set of microstates was the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, while now it is the $NV\mathbf{h}_0T$ ensemble. An appropriate choice for the shape at a given volume in which the unit cell is fixed is an average over the unit cell components obtained in the previous $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ simulations. Due to a bistability in shape (see next section), additional measures were taken to obtain a physically meaningful initial structure: the average configuration of the longest existing stable state. At each time during the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ simulation, the unit cell has the same volume V. However, because of the nonlinearity of the volume, the average unit cell matrix $\langle \mathbf{h} \rangle$ does not necessarily have the same volume V. To account for this, the average unit cell matrix $\langle \mathbf{h} \rangle$ is scaled with a constant c so the product corresponds to a unit cell with volume V:

$$V = \det(c\langle \mathbf{h} \rangle) \tag{3.3}$$

A summary of the simulations can be found in Table 3.2.

The obtained pressures as a function of the volume can be found in Figure 3.3. Again, a polynomial is fitted and the pressure is integrated analytically to finally obtain the free energy profile in Figure 3.4. The lp minimum is found to be at $V = 1470$ Å^3, which is 8.7 kJ/mol less stable than the np. The top of the transition region lies at $V = 1190$ Å^3, with

Table 3.2: Simulations performed in Subsection 3.1.2.

Ensemble	$NV\mathbf{h}_0T$
Timestep	0.5 fs
Simulation time	400 ps
Temperature	300 K
Pressure	–
Starting volume	740 – 1535 Å^3
Description	Fixed unit cell simulations
Goal	Obtain ensemble average pressure

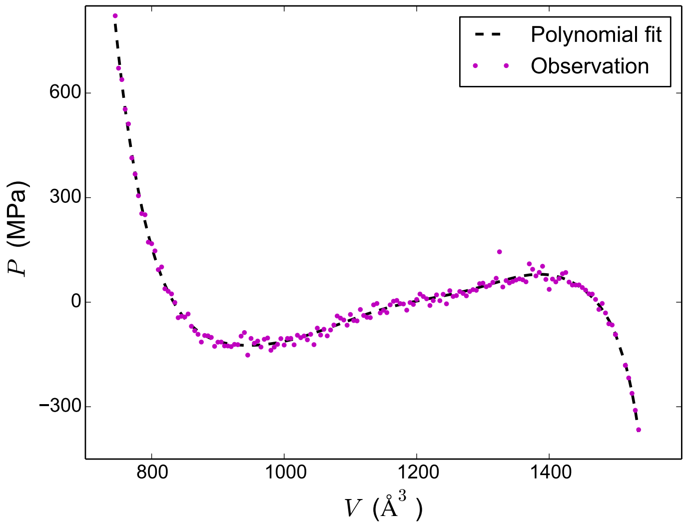


Figure 3.3: **The NVh$_0$T ensemble**. The pressure $\langle P(V)\rangle$ as a function of the unit cell volume V for MIL-53(Al). MD results obtained from 160 simulations in the $NV\mathbf{h}_0T$ ensemble, with volumes ranging from 740 to 1535 Å^3 in steps of 5 Å^3. The dashed line is an 11th order polynomial fit.

a barrier height of 7.7 kJ/mol with respect to the lp phase.

3.1.3 Comparison

The free energy profiles obtained from sampling in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ and the $NV\mathbf{h}_0T$ ensemble are compared in Figure 3.5. Fixing the unit cell shape, hence not sampling the fluctuations associated with the unit cell shape, has a large effect on the free energy profile: the free energy difference between the lp and np state lowers from 26.8 to 8.7 kJ/mol, an underestimation of around 20 kJ/mol. The lp-to-np transition has a higher barrier, rising from 1.9 to 7.7 kJ/mol. The top of the transition barrier shifts towards lower volume, from V = 1275 Å^3 to V = 1190 Å^3. A slight shift towards a higher stable volume of the np state is also observed for the $NV\mathbf{h}_0T$ ensemble.

The reason for this difference in free energy is further investigated, by comparing both the average energies (Figure 3.6) and the resulting entropies (Figure 3.7). No considerable difference between the average energies of the two ensembles is found, which may lead to the assumption there is some evidence to assume that the fixed unit cell structure is a good choice for obtaining adequate energy estimates. As a consequence, the difference between

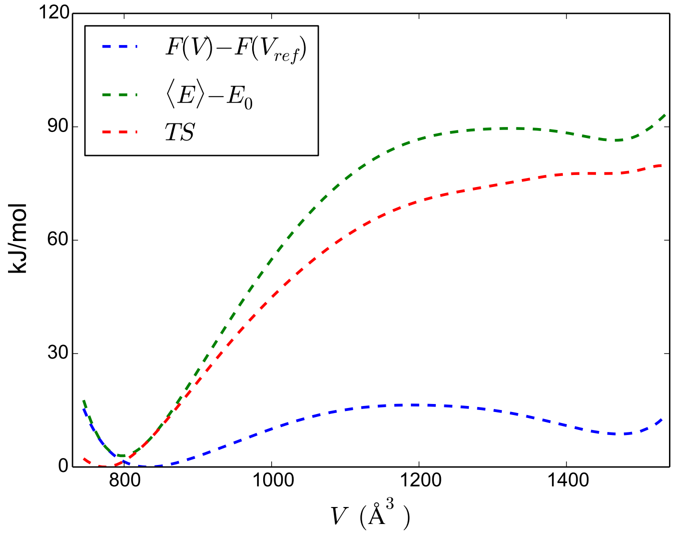


Figure 3.4: **The NV$\mathbf{h_0}$T ensemble**. **Blue**: Free energy profile as a function of the unit cell volume V for MIL-53(Al), obtained by thermodynamic integration of the polynomial fit to the pressure $\langle P(V)\rangle$ (see Figure 3.1). $F(V_{\mathrm{ref}})$ chosen to be the minimum at $V \approx 820$Å^3. **Green**: 12th order polynomial fit to the average energy $\langle E\rangle - E_0$ for 160 different volume simulations ranging from 740 to 1535 Å^3 in steps of 5 Å^3. The constant E_0 is chosen such that the minimum of the entropy curve $TS = \langle E\rangle - E_0 - \big(F - F(V_{\mathrm{ref}})\big)$ (**Red**) is zero (see Equation 2.25).

the two free energy profiles is due to a difference in entropy, as seen in Figure 3.7.

The entropy is a measure of configuration freedom in the system. Both curves show a higher entropy in the lp than in the np state. Note that both entropy curves have previously been shifted by a constant, so absolute entropies can not be compared, only the form of the curves (and especially the np-lp entropy difference). Fixing the unit cell shape removes the entropy contribution of the five corresponding degrees of freedom. The entropy difference between the np and lp becomes larger when neglecting unit cell shape fluctuations. This leads to the conclusion that the contribution to the entropy, due to the unit cell shape fluctuations, is larger in the np than in the lp structure.

In conclusion, the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble is proposed as the most suitable method to construct a free energy profile as a function of the volume. Neglecting unit cell shape fluctuations, as is the $NV\mathbf{h}_0T$ ensemble, does not render the correct free energy profile. These unit cell shape fluctuations primarily contribute to the entropy term TS, which explains the free energy difference between both methods. However, the $NV\mathbf{h}_0T$ ensemble

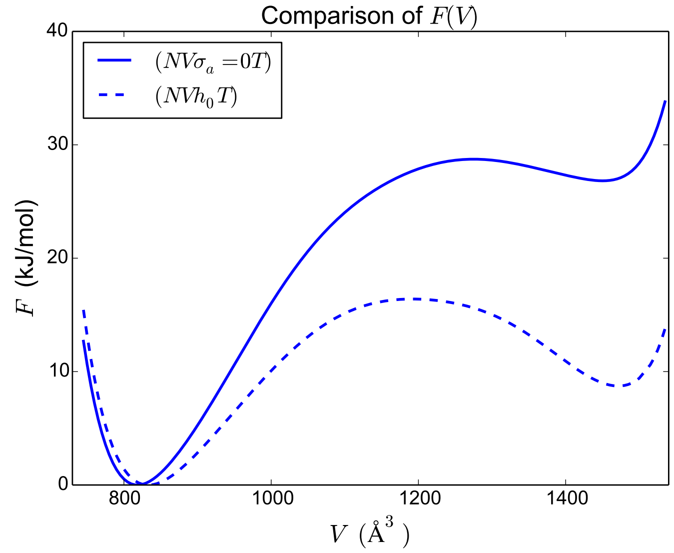


Figure 3.5: Comparison of the free energy profiles $F(V)$ as a function of the unit cell volume V for both ensembles, the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble (solid line) and the $NV\mathbf{h}_0T$ ensemble (dashed line).

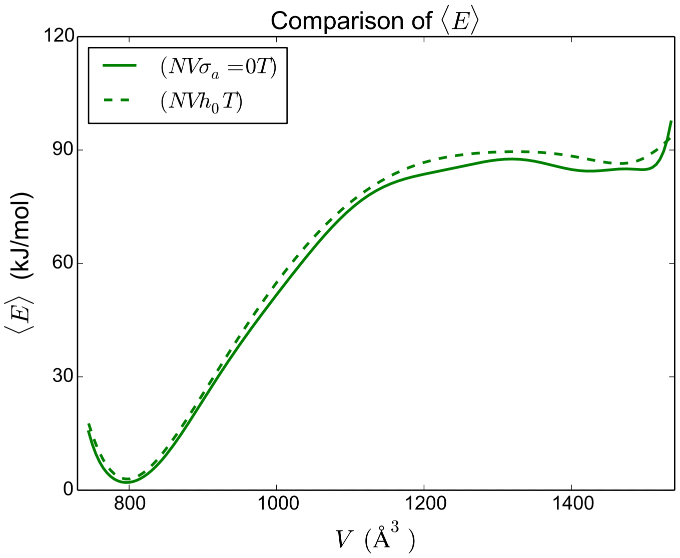


Figure 3.6: Energy profile as a function of the unit cell volume V for the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble (solid line) and $NV\mathbf{h}_0T$ ensemble (dashed line).

does give roughly correct energies for the system as a function of V.

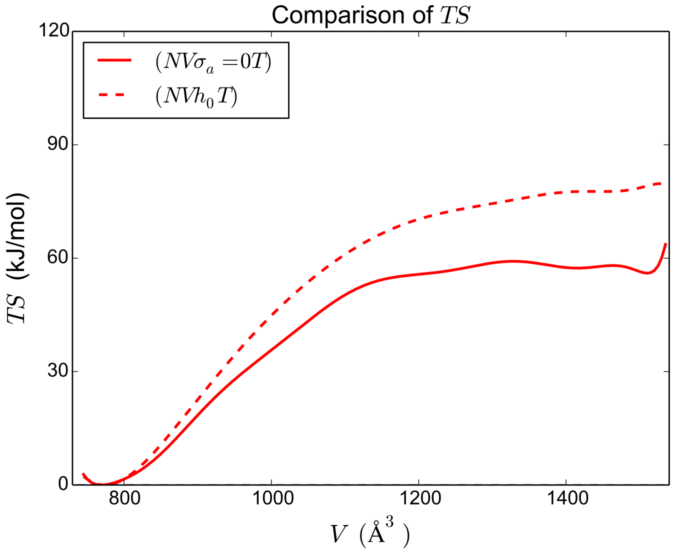


Figure 3.7: TS profile as a function of the unit cell volume V for the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble (solid line) and the $NV\mathrm{h}_0T$ ensemble (dashed line). Both entropy curves have previously been shifted by a constant.

3.2 The detection of a bistability in shape

During some of the 160 $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ simulations performed in Subsection 3.1.1, the unit cell underwent a sudden change in shape. This involved only the symmetric cell matrix component $a_y = b_x$. An illustrative example at a unit cell volume of 1080 Å^3 is shown in Figure 3.8. The (small) y component of the unit cell vector $\vec{a}$ and the (small) x component of the unit cell vector $\vec{b}$ switch signs, both from positive to negative, which we will name a "$a_y = b_x$ transition". At volumes smaller than 1080 Å^3, this transition is not observed during the simulations, while this transition occurs frequently at large volumes but with a smaller amplitude (see Figure 3.9).

In the lp state, the component $a_y = b_x$ is almost zero which corresponds to an unskewed unit cell, while in all other states $a_y = b_x$ is non zero, and hence indicates a slightly skewed unit cell. This corresponds to a shift in the stacking of the different layers in the plane perpendicular to the vector $\vec{a} \times \vec{b}$ of MIL-53(Al). At small volumes the shift in the stacking of the layers is larger (more skewed unit cell), because the two layers are closer together.

There are two possible reasons as to why the transition did not occur at small volumes. This can either be because there is a large free energy barrier between these states, or

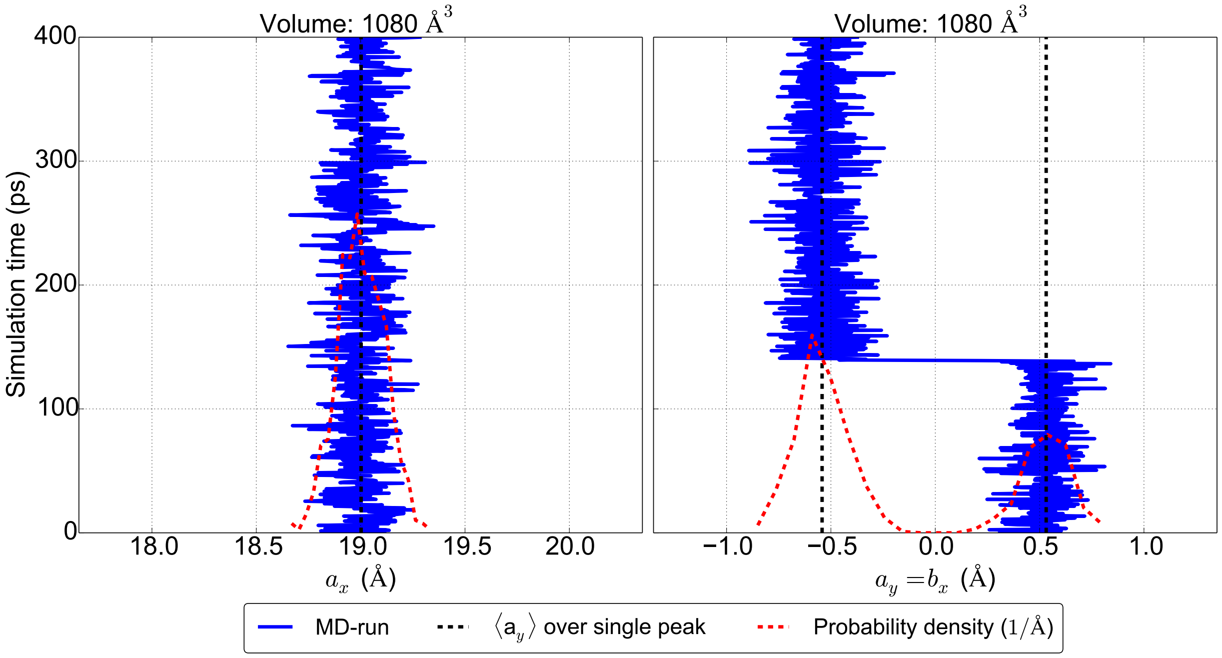


Figure 3.8: Illustration of an $a_y = b_x$ transition in an $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ simulation at $V = 1080$ Å^3. The component $a_y = b_x$ flips sign (right subfigure), while no other cell component varies (left subfigure, one component shown as an example). The blue line shows the evolution of the components throughout the simulation, the red dotted line illustrates the density of states as a function of the cell component, and the black dashed line indicates ithe average of that component over a single density peak.

because the second stable state does not exist at small volumes. With the set of simulations performed in Section 3.1, the second stable state is certainly not sampled at low volumes. Therefore, more advanced sampling methods (discussed in Section 2.7) are proposed. One way of sampling the negative $a_y = b_x$ region better, is by letting the system evolve freely between small and large volumes. If the system at low volumes does not visit the state because of a large free energy barrier at the $a_y = b_x = 0$ region, it should follow a different path to this negative $a_y = b_x$ state. This path consist of coming from a positive $a_y = b_x$ in the np phase and lowering the volume, to visit the lp phase. Next, the transition from the positive to the negative $a_y = b_x$ state is made, subsequently the volume is lowered in the negative $a_y = b_x$ state. This path avoids the free energy barrier at small volumes (it is represented in Figure 3.10), if it exists. This approach is used in Subsection 3.2.2.

Another way to sample the $a_y = b_x$ transition is by performing metadynamics in the direction of the transition. This is performed in Subsection 3.3. This can be either done in one dimension (the $a_y = b_x$ direction) while the unit cell volume remains fixed, or in two

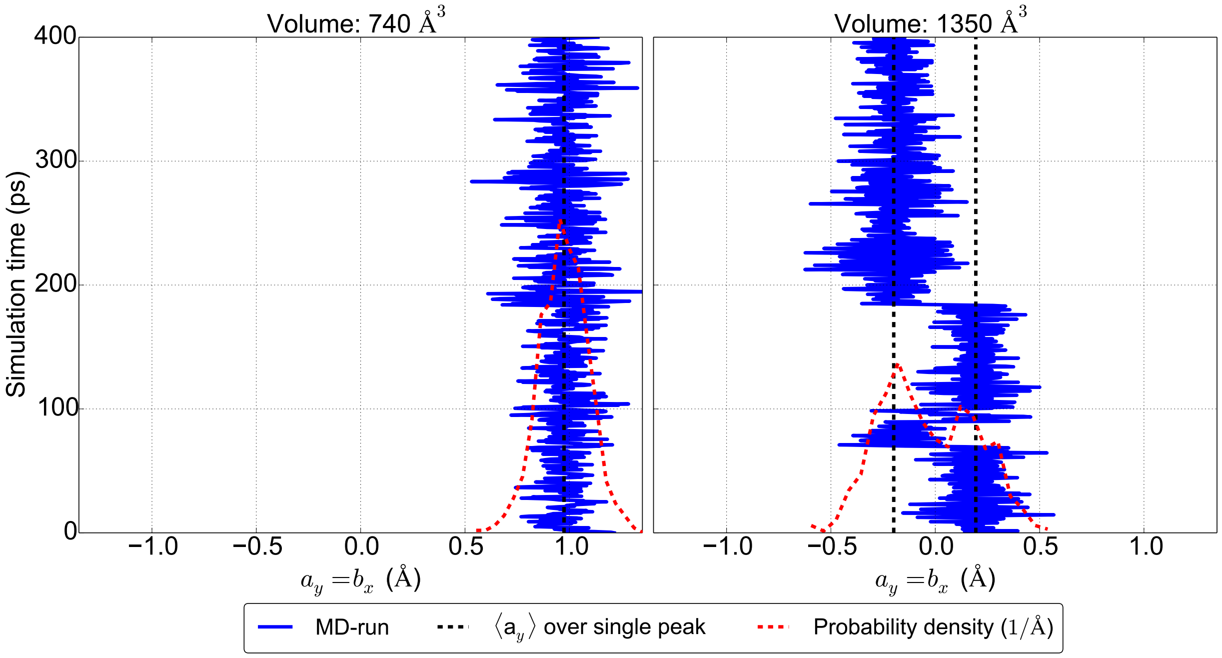


Figure 3.9: Illustration of the frequency of $a_y = b_x$ transitions in an $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble simulation at two different volumes. The blue line shows the evolution of the components throughout the simulation, the red dotted line shows the density of the state, and the black dashed line shows the average of that component over a single density peak.

dimensions (the volume V and the $a_y = b_x$ direction).

3.2.1 Note on average unit cell structures

Special attention needs to be paid to every method that involves averaging the unit cell components to obtain a mean unit cell structure. Because of the $a_y = b_x$ transition, averaging over a fixed volume simulation of e.g. 1080 Å^3 will result in cell components $|a_y| = |b_x| \approx 0$ Å. However, as seen in Figure 3.8, the component b_x of the system is never 0 Å, but fluctuates around ± 0.6 Å. Hence, fixing the unit cell structure in a state $b_x \approx 0$ Å is not optimal.

A generally applicable solution to this problem is provided. The nature of the problem is that the distribution of the cell component is not unimodal. Returning to the example in Figure 3.8, the density of states (which are distributions of the cell component) are shown in red. It is clear that averaging over a_x results in a correct structure, because the mean and median of the distribution agree, unlike in the bimodal distribution of b_x.

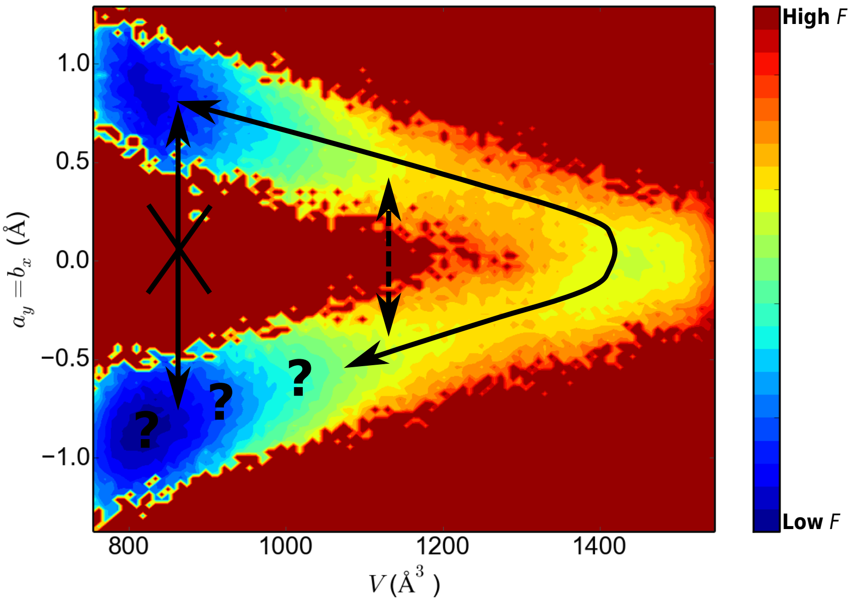


Figure 3.10: Free energy profile of a fictituous system and the path it can follow to get from positive to negative $a_y = b_x$ in the np phase. In the np phase this transition was not seen, while at intermediate volumes it happens rarely, and it happens frequently in the lp phase.

The most straightforward solution is to average the component over one of the peaks. A consequence of this method is that only one of two states of the system is represented. If the two states are equivalent in terms of configuration and free energy, no difference can be found between these two average states. This method is applied to all unit cell components that show bistable behavior, but for this to work, an automatic detection mechanism is constructed.

The detection of this type of distributions is done with a parametric statistical test: Hartigan's dip test for unimodality [79]. Firstly, to interpret the results from this test correctly, an introduction to parametric statistics is written in Appendix A. After performing the dip test on the simulation sets for all unit cell components, it is seen that only the $a_y = b_x$ component shows the bistable behavior. The unit cell is averaged over the longest set of positive or negative $a_y = b_x$.

3.2.2 Explicit sampling: The extended umbrella simulation

To obtain information about the two different $a_y = b_x$ states, a set of four independent 400 ps extended umbrella simulations with a timestep of 0.5 fs are performed in the

$NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble at $P = 0$ Pa and $T = 300$ K. The bias potential $U_b(V)$ flattens the free energy surface over the complete volume range between the lp and np state. This bias potential is the negative of the free energy $-F(V)$ as a function of the volume, as obtained in section 3.1.1. The bias potential counteracts the free energy barrier between the np and lp phase, and hence promotes large fluctuations in the volume, and it is expected that it will sample volumes between the np and lp state uniformly. Only the data in the volume range of interest (750− 1550 Å^3) have been used. Other data have been taken out of the dataset, resulting in a total data set of ≈ 1 ns simulation time.

Since the $a_y = b_x$ transition is not observed at smaller volumes, it is necessary to force the system in larger volume states to make a transition happen, and then letting it evolve into smaller volumes. In this way, it is expected that both $a_y = b_x$ states at smaller volumes are adequately sampled. By constructing a 2D histogram of the data (Figure 3.11), it is observed that the negative $a_y = b_x$ state exists and that it is frequented by the system. From now on the two states, $a_y = b_x > 0$ and $a_y = b_x < 0$, are specified as the positive and negative branch respectively.

With this 2D histogram and knowledge of the bias potential $U_b(V)$, a corresponding free energy profile can be constructed (see Equation 2.49). The free energy $F(V, a_y = b_x)$ obtained here is a function of both V and $a_y = b_x$. The bias potential $U_b(V)$ is only a function of the volume V, and constant in $a_y = b_x$. The constant F_b is chosen such that the global minimum of the profile (in this case, the np of the negative branch) is zero. The profile is shown in Figure 3.12. It is seen that the two branches generally have a slightly different free energy, that for example in the np minimum this difference is around 2 kJ/mol. Visual comparison of the system in the two branches in the np renders no structural differences which could result in a different free energy.

To answer the question whether both branches are equivalent, further investigation is needed. Firstly, it is investigated whether the two branches are sufficiently sampled. Next, for the two branches, the free energy as a function of volume, and as a function of the branch is compared. To find structural differences, the point groups of the np structures for both branches are calculated. Next, the two np structures are optimized at 0 K, and the energies E of the two states are compared.

Sampling of both states

Inspection of the simulation results shows a large number of volume fluctuations, going from the lp to the np state of the system. These volume fluctuations are expected to

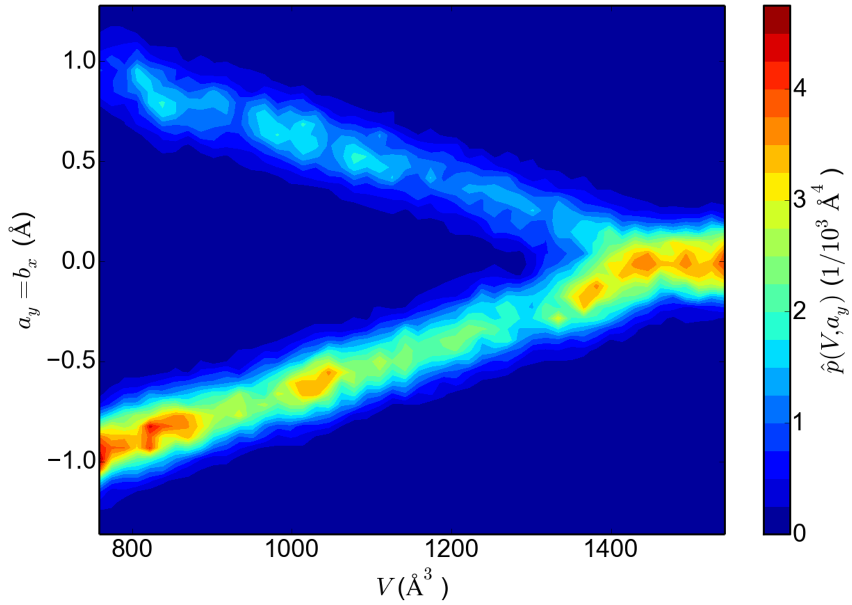


Figure 3.11: 2D histogram of the set of four extended umbrella simulations. The two components shown here are the unit cell volume V and the unit cell component $a_y = b_x$. The quantity plotted is the estimated probability density $p(V, a_y = b_x) = h(V, a_y = b_x)/NS$, with $h(V, a_y = b_x)$ the 2D histogram of the simulation set as a function of the volume V and the $a_y = b_x$ component, N the total number of observations and S the surface of the histogram bin (here in Å^4).

result in a better sampling of the bistable $a_y = b_x$ state previously observed, thanks to the path indicated in Figure 3.10. It is also seen that large and numerous fluctuations are obtained in the $a_y = b_x$ component. In conclusion, it is expected that both the lp and np volume states, as well as both the positive and negative $a_y = b_x$ states are sufficiently sampled during the set of four extended umbrella simulations. The previously not observed state $a_y = b_x \leq -0.7$ Å is visited. If the system exists in the positive $a_y = b_x$ state at low volume, the transition to the negative $a_y = b_x$ state at low volumes is induced in a three step process (see Figure 3.10): first there is a lp-to-np transition, secondly there is a $a_y = b_x$ transition from positive to negative $a_y = b_x$, and thirdly there is a lp-to-np transition to small volumes. In Figure 3.11 it is shown that the negative $a_y = b_x$ branch is indeed frequented, and even more remarkable, that it is frequented more than the positive branch.

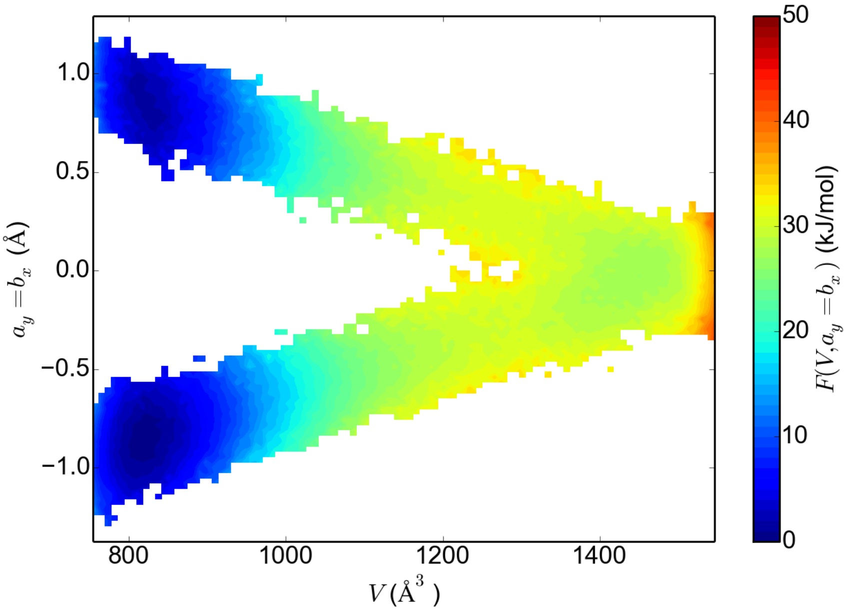


Figure 3.12: The 2D free energy profile $F(V, a_y = b_x)$ as a function of the unit cell volume V and unit cell component $a_y = b_x$, calculated from Figure 3.11 with Equation 2.49. The histogram bins which have not been visited, have not been converted to a free energy (those bins are white).

Free energy profile of both branches

The free energy $F(V, B)$ as a function of the volume V and the branch $B \in \{+, -\}$ can be calculated using the histogram method for an umbrella simulation (see Subsection 2.7.1). The branch $B = +$ is defined as the set of states where $a_y = b_x > 0$, the branch $B = -$ as the set of states where $a_y = b_x < 0$. It is an important detail to note that the point $a_y = b_x = 0$ is not included in the definition of the branches, but practically the $a_y = b_x$ component will never be exactly zero. Here the collective variables are (V, B), describing the volume of the system, and whether the system has a positive or negative $a_y = b_x$ component. Since the bias potential $U_b(V)$ does not depend on the branch B, but only on the volume V, Equation 2.49 becomes:

$$F(V, B) = -k_B T \ln[p_b(V, B)] - U_b(V) \tag{3.4}$$

where the probability density of the state (V,B) is estimated by the 2D histogram $h(V,B)$ with 50 bins:

$$p_b(V,B) = \frac{h(V,B)}{NS} \qquad \text{with} \quad N = \sum_{B\in\{-,+\}} \frac{1}{V_{\max}-V_{\min}} \int_{V_{\min}}^{V_{\max}} h(V,B)dV \tag{3.5}$$

$$\text{and} \quad S \quad \text{histogram bin surface} \tag{3.6}$$

N specifies the total number of states for the system with a volume component between $V_{\min}$ and $V_{\max}$. $V_{\min}$ and $V_{\max}$ specify the volume range for which the free energy is calculated.

The results of this calculation are shown in Figure 3.13. The difference of local probability $p(V,B)$ between the branches at a given volume is translated to a difference in free energy between both curves. Both curves also have a slightly different local minimum in the lp phase (=1450 Å^3 for the + branch, =1420 Å^3 for the − branch), and a different local maximum of the lp-to-np transition barrier (=1244 Å^3 for the + branch, =1300 Å^3 for the − branch).

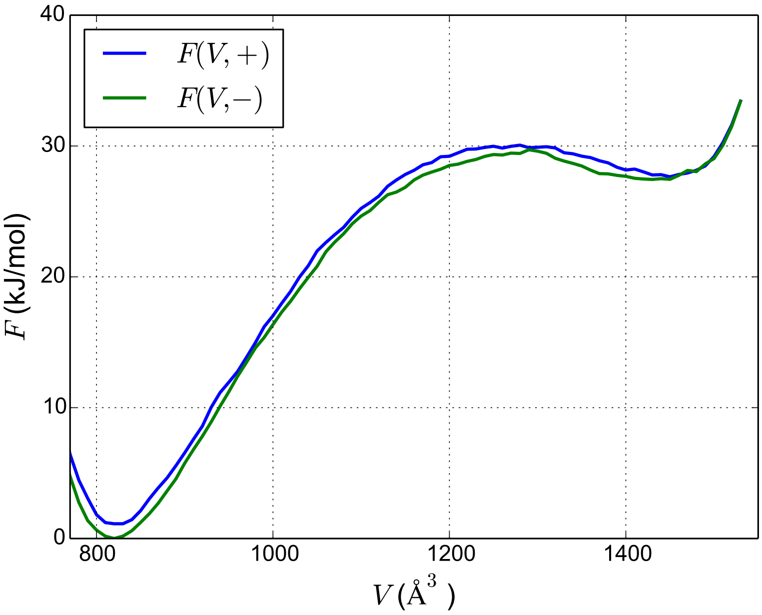


Figure 3.13: The free energy $F(V,B)$ as a function of the unit cell volume V and the branch B. This free energy is calculated by Equation 3.4, where F_b is chosen such that the minimum of the green curve is zero. Each curve depends on the free energy as a function of the volume $F(V)$ (see Subsection 3.1.1) and on the local estimated probability $p(V,B)$ of the extended umbrella simulation.

Space group of branches in the np phase

To compare the two states in the np volume range, their space group is calculated. For this, the average unit cell structure is calculated for both $a_y = b_x$ positive as negative. Calculations were performed by the program PLATON [80]. Within the error margin implemented in the software, both cell structures have the same space group P21/c (No. 14), experiments (at non-zero temperature) report a space group C2/2 (No. 15) [42].

Inspection of the structures shows that linkers are show a slight rotation around their axis, with reference to the configuration where the benzenes are parallel to the linker planes. Linkers from neighboring planes are rotated in the opposite direction. This configuration is thought to be more favorable in energy because of two effects: rotation of the linkers is favorable for the hydrogen repulsion between linkers from the same linker plane, and that opposite rotation of linkers from neighboring planes results in a configuration resembling a more favorable T-shaped interaction of the benzene [81]. However, the linker rotation remains limited because it is strongly influenced by the constraints imposed by the periodic structure of the MOF [53].

Optimization of np state for both branches

Two snapshots of the structures in the np states were optimized, one in the positive branch, the other in the negative branch. Optimization of a structure means finding a neighbouring minimum in energy. For the optimization the conjugate gradient method is used, and the complete unit cell matrix varies freely (both V as $\mathbf{h}_0$). Both obtained optimal np unit cell volumes are $V = 793.05$ Å^3, and the energies E of both unit cells are also equal to within numerical noise. The obtained unit cell matrices are

$$\mathbf{h}_{\text{np},+} = \begin{pmatrix} 19.51 \text{ Å} & 1.00 \text{ Å} & -0.07 \text{ Å} \\ 1.00 \text{ Å} & 12.98 \text{ Å} & 0.20 \text{ Å} \\ -0.07 \text{ Å} & 0.20 \text{ Å} & 6.29 \text{ Å} \end{pmatrix} \tag{3.7}$$

$$\mathbf{h}_{\text{np},-} = \begin{pmatrix} 19.51 \text{ Å} & -1.00 \text{ Å} & -0.07 \text{ Å} \\ -1.00 \text{ Å} & 12.98 \text{ Å} & -0.20 \text{ Å} \\ -0.07 \text{ Å} & -0.20 \text{ Å} & 6.29 \text{ Å} \end{pmatrix} \tag{3.8}$$

As expected, both structures are skewed and have opposite $a_y = b_x$ components. With this structure corresponds an angle between the $\vec{a}$ and $\vec{b}$ unit cell vectors of γ=97°, in very rough accordance with the angle reported experimentally of 114° at 295 K [42]. Surprisingly, also the $b_z = c_y$ components are seen to be opposite. It is not observed during the MD simulations that there is a correlation between the $a_y = b_x$ and $b_z = c_y$, and this difference is not further investigated, the details are given in Appendix B.

Conclusion

In conclusion, the negative $a_y = b_x$ state exists at small volumes. The fact that no transition at low volumes was previously observed, is due to a free energy barrier between both states, the height of which can not be determined from this analysis. Information about the occupation of the $a_y = b_x$ branches is obtained, and a difference is found between the relative occupation of the branch pair. This difference in free energy is around $1 - 2$ kJ/mol, and it is investigated whether this is due to poor sampling or whether it is an inherent physical phenomenon. No difference is found in the space group, nor in the energies of the optimal configuration of the np minima, pointing towards the equivalence of both branches. This statement is further studied with metadynamics.

3.3 Metadynamics

Three sets of metadynamics simulations are performed, each having a different collective variable. The first two simulations both have a single collective variable (1D metadynamics), one with the volume V (Subsection 3.3.1), and the other the cell component $a_y = b_x$ at fixed volume V (Subsection 3.3.2). The latter gives detailed information about the relative stability of both branches, at a discrete set of volumes. The last set of metadynamics simulations has two collective variables (2D metadynamics): the volume V and cell component $a_y = b_x$ (Subsection 3.3.3). All simulations have a timestep of 0.5 fs and are performed at a temperature of 300 K.

3.3.1 Explicit sampling: variable volume

Using the metadynamics technique described in Subsection 2.7.2, ten independent metadynamics simulations are performed in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, where the unit cell can vary freely under a pressure P of 0 Pa. The collective variable q is chosen to be the volume V, and each simulation has a length of 2.2 ns and contains 3000 metadynamics steps. In the first 1.2 ns of the simulation, a Gaussian hill with a height of 1 kJ/mol and a width of 50 Å^3 is added to the profile each 0.6 ps. In the last 1 ns, hills are placed each 1 ps and have a height of 0.5 kJ/mol and a width of 25 Å^3. From these simulations, the

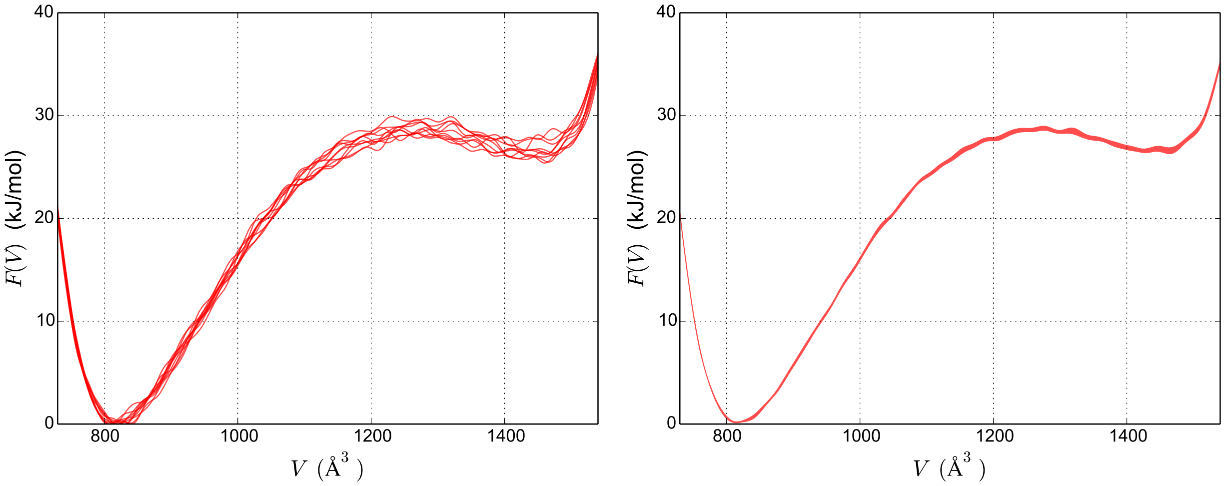


Figure 3.14: The free energy profiles $F(V)$ as a function of the unit cell volume V obtained by the 1D metadynamics simulations in the volume direction. Left: 10 free energy profiles obtained by 10 independent metadynamics simulations. Right: The average $F(V)$ and with the linewidth denoting $\pm 1\mathrm{SE}_{F(V)}$ standard error.

mean free energy profile $F(V)$ is calculated. The standard error on this profile is estimated by $\mathrm{SE}_V = \frac{\sigma_{F(V)}}{\sqrt{9}}$ with σ the standard deviation of the ten datapoints at each volume.

The resulting free energy profiles are shown in Figure 3.14. The lp phase minimum at $V = 1427$ Å^3 lies at 26.6 ± 0.2 kJ/mol with respect to the np phase minimum at 815 Å^3. The top of the free energy barrier at $V = 1275$ Å^3 has a height of 28.8 ± 0.3 kJ/mol. These results correspond with the values obtained in Subsection 3.2.

3.3.2 Explicit sampling: fixed volume, variable shape

Ten fixed volume metadynamics simulations are performed in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble for each unit cell volume in the set [780, 930, 1080, 1260, 1450] Å^3. The symmetric unit cell component a_y is chosen as the collective variable q of the system. The partial derivative of q with respect to the symmetric unit cell components $\mathbf{h}_s$ becomes:

$$\frac{\partial q}{\partial \mathbf{h}_s} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.9}$$

Together with Equation 2.56 and given a unit cell $\mathbf{h}$, the biased virial tensor from Equation 2.61 can be calculated. An alternative choice for q would be $\frac{a_y+b_x}{2}$, but this makes for a heavier calculation while the results stay the same.

The metadynamics profiles are the result of a three stage simulation, with each stage having a simulation length of 1 ns. In the first and second stage, a metadynamics step was performed every 0.5 ps, resulting in a total of 4000 Gaussian hills added to the bias potential. The first stage contains 2000 hills of width 0.05 Å and of height 0.3 kJ/mol. Because of their height, the hills rapidly fill up the free energy basins, while they approximate the characteristics of the profile. The second stage contains 2000 narrower and lower hills (a width of 0.025 Å and a height of 0.1 kJ/mol) to improve on the narrow and steep characteristics of the profile. The last stage performed a metadynamics step every 2 ps, hence 500 hills are added in this stage (with the same height and width as in the previous stage). This longer time between metadynamics steps serves to capture slower relaxation effects.

The results of the simulations are shown in Figure 3.15. At low volumes, a high free energy barrier is found between the positive and negative $a_y = b_x$ state. The barrier gradually disappears when the unit cell volume increases. This behavior is probably due to the interaction of the linkers. At low volumes, the linkers are close together, such that shifting the layers over each other requires a higher energy.

With the free energy curves $F(V, a_y = b_x)$, the difference between the free energy of the branches at a given volume can be calculated. At a given volume V, the probability $p(V, q)$ of visiting the state (V, q) is calculated (with q the unit cell component $a_y = b_x$):

$$p(V,q) = ce^{-\beta F(V,q)} \tag{3.10}$$

The normalisation constant c is written explicitly. The probability of visiting a given branch $B = [q_1, q_2]$ is the sum (or in continuous equivalent, the integral normalized over the integration range) of the probabilities $p(V, q)$ for which $q \in B$.

$$p(V,B) = \int_{q_1}^{q_2} p(V,q)dq = c\int_{q_1}^{q_2} e^{-\beta F(V,q)}dq \tag{3.11}$$

The integration is performed over $[q_1, q_2]$ equal to $[-\infty, 0]$ for $B = -$ and $[0, \infty]$ for $B = +$. In practice, integration ranges are chosen to be $[-2.0\ \text{Å}, 0]$ for $B = -$ and $[0, 2.0\ \text{Å}]$ for $B = +$, because these regions include every contribution to the free energy profile for $a_y = b_x$. The free energy becomes (see Equation 2.39):

$$F(V,B) = -k_BT\ln\left[p(V,B)\right] = -k_BT\ln\left[\int_{q_1}^{q_2} e^{-\beta F(V,q)}dq\right] + k_BT\ln\left[c\right] \tag{3.12}$$

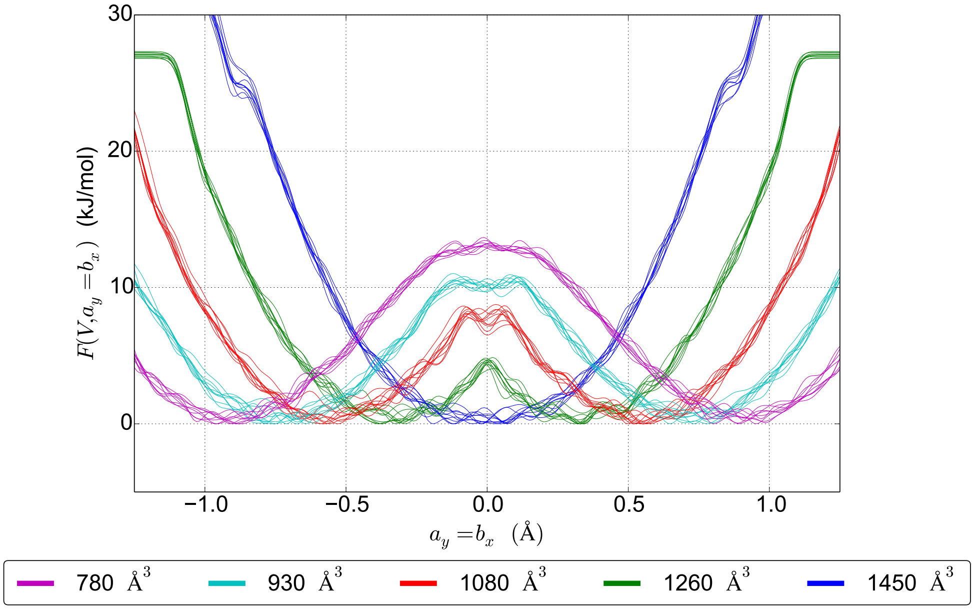


Figure 3.15: Free energy curves $F(V,a_y=b_x)$ at a discrete set of unit cell volumes V and as a function of unit cell component $a_y=b_x$, obtained by performing $a_y=b_x$ metadynamics at fixed unit cell volumes 780, 930, 1080, 1260 and 1450 Å^3. The curves are shifted such that the minimum always lies at 0 kJ/mol.

When calculating the free energy difference between the positive and negative branch $\Delta F(V)=F(V,+)-F(V,-)$, the constant c disappears from the equation:

$$\begin{aligned}\Delta F(V) &= F(V,+)-F(V,-) && (3.13)\\ &= -k_BT\left[\ln\left[\int_0^{|q_{\max}|} e^{-\beta F(V,-)}dq\right]-\ln\left[\int_{-|q_{\max}|}^{0} e^{-\beta F(V,+)}dq\right]\right] && (3.14)\\ &= -k_BT\ln\left[\frac{\int_0^{|q_{\max}|} e^{-\beta F(V,-)}dq}{\int_{-|q_{\max}|}^{0} e^{-\beta F(V,+)}dq}\right] && (3.15)\end{aligned}$$

This expression is used to calculate the difference in free energy between the positive and negative branch, for every metadynamics simulation. Hence for every volume, ten values of ΔF are obtained. These serve to get a mean difference, and an estimate of the standard error $\mathrm{SE}_{\Delta F}=\frac{\sigma_{\Delta F}}{\sqrt{n-1}}$ on this mean difference. The same statistical procedure is used to estimate the height of the free energy barrier $h(V)$.

The results of this calculation are shown in Table 3.3. It is observed that $\Delta F = 0$ for all volumes, within an error of $1\text{SE}_{\Delta F(V)}$. Hence, there is no evidence to assume that both branches are inequivalent in terms of free energy. The free energy barrier between the two branches is observed to be non existent in the lp phase, consistent with earlier results, see Figure 3.12. The free energy barrier in the np phase has a height of 13 kJ/mol.

For each volume, the mean $F(V, a_y = b_x)$ and its standard error $\text{SE}(V, a_y = b_x)$ are calculated and shown in Figure 3.16. The colored regions are each the mean of the 10 curves at a given volume, plus or minus two SE.

The five mean free energy curves $F(V, a_y = b_x)$ are shifted by a constant F_{0V} such that for each of the volumes $V \in \{780, 930, 1080, 1260, 1450\}$ Å^3, the probability p of visiting the state $(V, a_y = b_x)$ is proportional to $e^{-\beta(F(V,a_y=b_x)+F_{0V})}$. It follows

$$F(V, a_y = b_x) + F_{0V} = -k_B T \ln \Big[p(V, a_y = b_x) \Big] \tag{3.16}$$

and for the $F(V)$ curve obtained in Subsection 3.1.1 (Equation 2.39):

$$F(V) = -k_B T \ln \Big[p(V) \Big] \tag{3.17}$$

Integrating the probabilities $p(V, a_y = b_x)$ over a_y in the complete range in which p is non zero, yields $p(V)$:

$$p(V) = \int_{-|q|}^{|q|} p(V, a_y = b_x) da_y \tag{3.18}$$

Where $|q| \approx 2$ Å for practical purposes. The following relation between the $F(V)$ and the

Table 3.3: The differences in free energy $\Delta F(V)$ between both branches and the height of the free energy barrier $h(V)$, both with estimated standard error.

Volume (Å^3)	$\Delta F(V)$ (kJ/mol)	$\text{SE}_{\Delta F(V)}$ (kJ/mol)	$h(V)$ (kJ/mol)	$\text{SE}_{h(V)}$ (kJ/mol)
780	0.10	0.14	13.23	0.08
930	0.09	0.12	10.42	0.06
1080	-0.12	0.12	8.23	0.10
1260	-0.02	0.12	4.57	0.08
1450	0.07	0.06	/	/

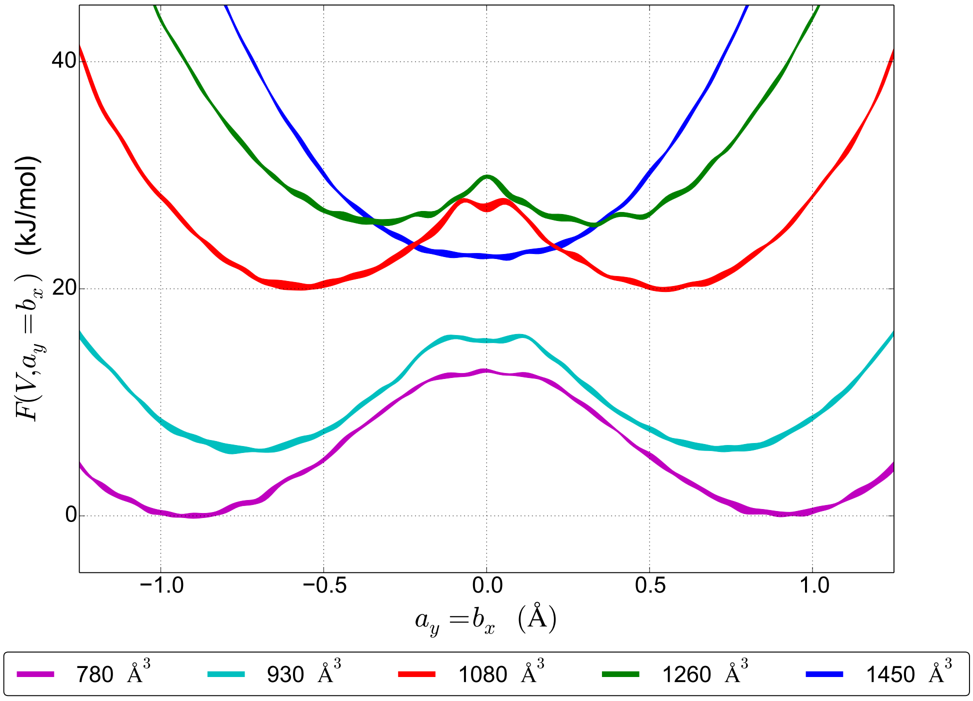


Figure 3.16: Estimate of the free energy curve $F(V, a_y = b_x)$ for five different unit cell volumes 780, 930, 1080, 1260 and 1450 Å^3. The curves are shifted with respect to each other by a constant, calculated in Equation 3.23. The linewidth denotes the region $F(V, a_y = b_x) \pm 2\ \mathrm{SE}_{F(V,a_y=b_x)}$.

$F(V, a_y = b_x)$ curves is found:

$$\begin{aligned} F(V) &= -k_B T \ln \Big[p(V)\Big] && (3.19)\\ &= -k_B T \ln \left[\int_{-|q|}^{|q|} p(V, a_y = b_x) da_y\right] && (3.20)\\ &= -k_B T \ln \left[e^{-\beta F_{0V}} \int_{-|q|}^{|q|} e^{-\beta F(V,a_y=b_x)} da_y\right] && (3.21)\\ &= F_{0V} - k_B T \ln \left[\int_{-|q|}^{|q|} e^{-\beta F(V,a_y=b_x)} da_y\right] && (3.22) \end{aligned}$$

With this equation, the difference between the constants F_{0V} by which the $F(V, a_y = b_x)$

curves are shifted, is calculated:

$$F_{0V_2} - F_{0V_1} = F(V_2) - F(V_1) + k_B T \ln \left[\frac{\int_{-|x|}^{|x|} e^{-\beta F(V_2, a_y = b_x)} da_y}{\int_{-|x|}^{|x|} e^{-\beta F(V_1, a_y = b_x)} da_y} \right] \tag{3.23}$$

3.3.3 Explicit sampling: variable volume and shape

The collective variable q in this set of simulations will be two-dimensional: $q = (V, a_y = b_x)$. The mathematics described in Subsection 2.7.2 remain valid. The simulated ensemble is $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$, with $P = 0$ Pa and $T = 300$ K. This method serves to construct a 2D free energy surface $F(V, a_y = b_x)$. The 2D metadynamics simulation has a total simulation time of 4.32 ns. The obtained free energy profile contains 7200 Gaussian hills of height 1 kJ/mol placed every 0.6 ps, 0.05 Å wide in the $a_y = b_x$ direction, and 100 Å^3 in the V direction.

Inspecting the obtained 2D free energy profile $F(V, a_y = b_x)$ (Figure 3.17), indicates that the branch structure of the 2D profile is correctly described, with two branch minima in the np phase, and one local minimum in the lp phase. The lp phase minimum lies approximately $26 - 28$ kJ/mol higher than the np phase minima.

The profile $F(V, a_y = b_x)$ is integrated into a 1D profile, calculated by the same method as Equation 3.22.

$$F(V) = F_0 - k_B T \ln \left[\int_{-|q|}^{|q|} e^{-\beta F(V, a_y = b_x)} da_y \right] \tag{3.24}$$

The result of this calculation is shown in Figure 3.18. The lp phase minimum at $V = 1467$ Å^3 lies at 27.9 ± 0.2 kJ/mol with respect to the np phase minimum at a volume of 807 Å^3. The top of the free energy barrier at $V = 1277$ Å^3 has a height of 30.3 ± 0.2 kJ/mol.

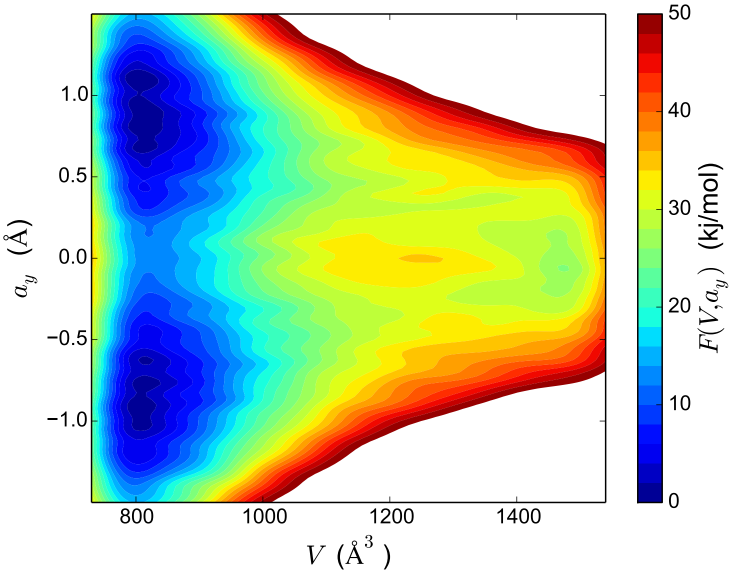


Figure 3.17: 2D free energy profile $F(V, a_y = b_x)$ as a function of the unit cell volume V and the unit cell component $a_y = b_x$, obtained by performing 2D metadynamics in the components $(V, a_y = b_x)$.

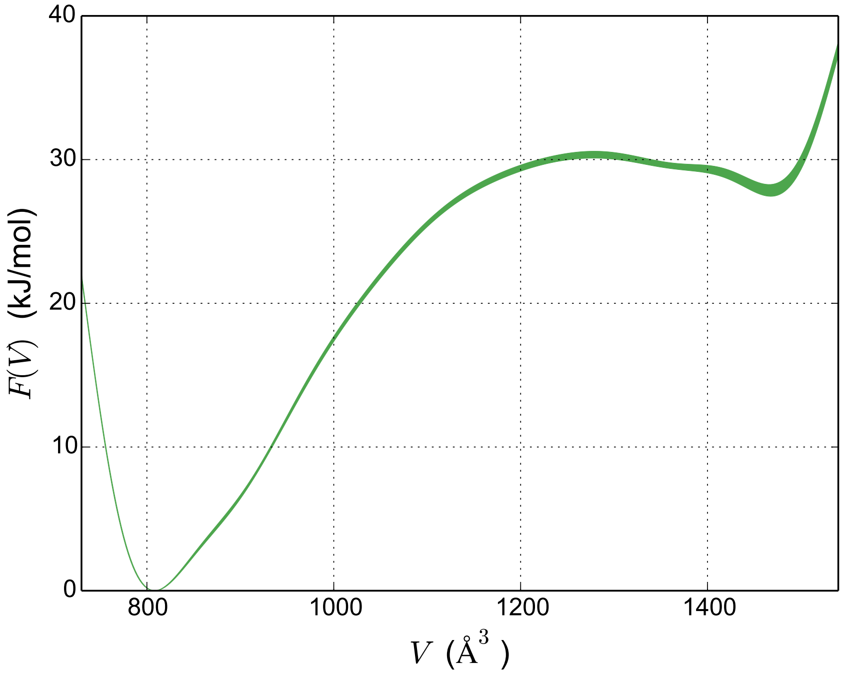


Figure 3.18: Free energy profile $F(V)$ as a function of the unit cell volume V obtained by integrating the $F(V, a_y = b_x)$ profile obtained with 2D metadynamics. The linewidth of the curve denotes the region $F(V) \pm 2\mathrm{SE}_{F(V)}$.

3.4 Comparison of the sampling methods

The results of the different sampling methods discussed in the previous sections, thermodynamic integration, extended umbrella sampling and metadynamics are compared and discussed. We distinguish between free energy profiles as a function of the volume alone (Subsection 3.4.1), as a function of the component $a_y = b_x$ at fixed volumes (Subsection 3.4.2) and as a function of both unit cell volume and $a_y = b_x$ component (Subsection 3.22).

For ease of the discussing, the free energy profiles $F(V)$ and $F(V, a_y = b_x)$ are named after the method that is used for their construction. The free energy profiles as a function of unit cell volume are named $F_{\text{therm}}(V)$ for the thermodynamic integration method (Subsection 3.1.1), $F_{\text{1Dmeta}}(V)$ for the 1D metadynamics in the volume direction of Subsection 3.3.1 and $F_{\text{2Dmeta}}(V)$ for the profile obtained by integration of a 2D metadynamics profile (Subsection 3.3.3). For the discrete set of free energy profiles at fixed volumes, in the direction of $a_y = b_x$, the naming becomes $F_{\text{1Dmeta}}(V, a_y = b_x)$ for the 1D metadynamics performed at fixed volumes with $a_y = b_x$ as collective variable (Subsection 3.3.2). Finally, the two sets of 2D free energy profiles are named $F_{\text{ext}}(V, a_y = b_x)$ for the extended umbrella simulation performed in Subsection 3.2.2, and $F_{\text{2Dmeta}}(V, a_y = b_x)$ for the 2D metadynamics (Subsection 3.3.3).

3.4.1 Free energy profile as a function of volume

As observed in Figure 3.19, $F_{\text{therm}}(V)$ and $F_{\text{1Dmeta}}(V)$ both render equivalent results. The volumes of the lp phase minimum are 1450 Å^3 and 1427 Å^3, the free energy differences with respect to the np phase minimum 26.8 kJ/mol and 26.6±0.2 kJ/mol, respectively. $F_{\text{2Dmeta}}(V)$ however, shows a lp phase volume of 1467 Å^3, and a free energy difference of 27.9 ± 0.2 kJ/mol.

This free energy difference is explained by the fact that at a given moment in the np phase, the thermodynamic integration method as well as the 1D metadynamics simulation sample only one of the two existing branches in the shape space, hence these simulations do not detect a degeneracy in the shape space. The 2D metadynamics simulation, however, flattens the free energy barrier between both branches, so both branches are sampled at the same time. In contrast, in the lp phase, no free energy barrier exists between both branches, and the thermodynamic integration method and 1D metadynamics simulation, as well as the 2D metadynamics simulation sample both branches.

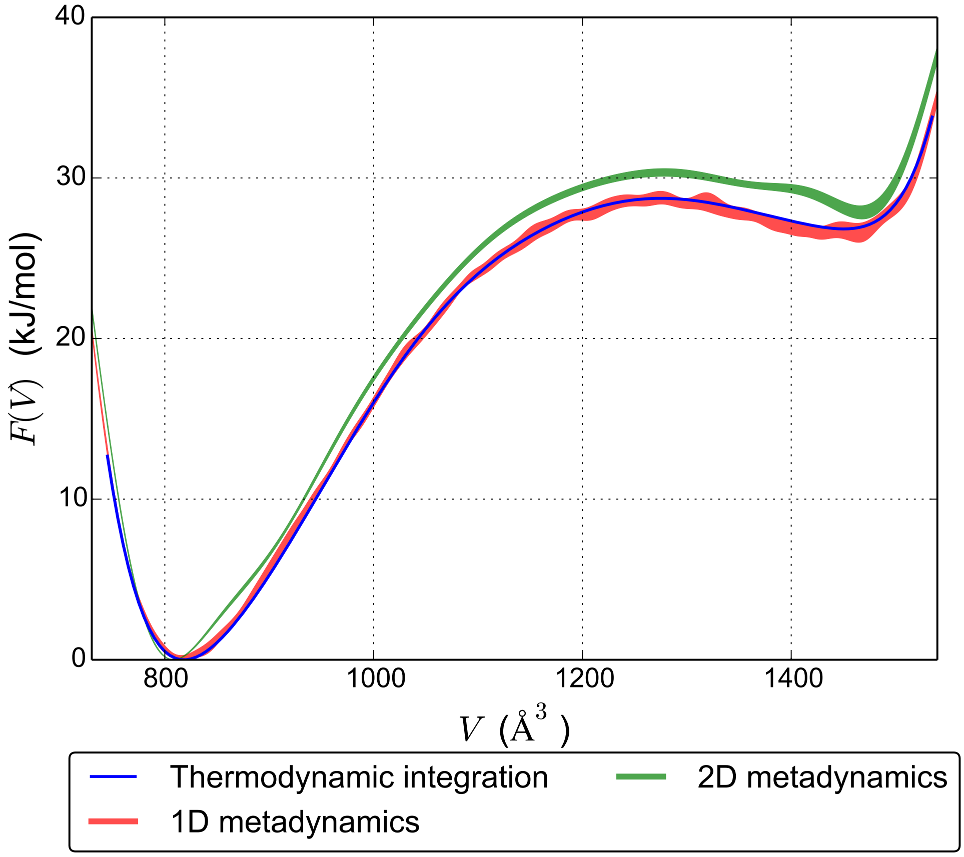


Figure 3.19: The free energy profiles $F(V)$ as a function of the unit cell volume V obtained by the different sampling methods. The volume V is collective variable of the 1D metadynamics simulation, described in Subsection 3.3.1. The linewidth of the 1D and 2D metadynamics curves denotes the region $F(V) \pm 2\mathrm{SE}_{F(V)}$.

As a result, because the thermodynamic integration method and 1D metadynamics simulation do not detect a degeneracy in the np phase, but do in the lp phase, the free energy minimum of the np phase must by corrected by a factor $-k_BT\ln 2$. This correction factor has a value of -1.73 kJ/mol, and results in a higher free energy difference between the lp and np phase. The comparison of original and corrected values is shown in Figure 3.20. Now all values correspond to each other within reasonable error limits.

Proof of concept, correction term $\mathbf{k_B T \ln(2)}$

The bistable shape structure of MIL-53(Al) contains two $a_y = b_x$ branches ($B = \pm$) that are separated by a large free energy barrier in the np phase, while they merge in the lp phase. Because of this structure, simulations that do not overcome the free energy barrier

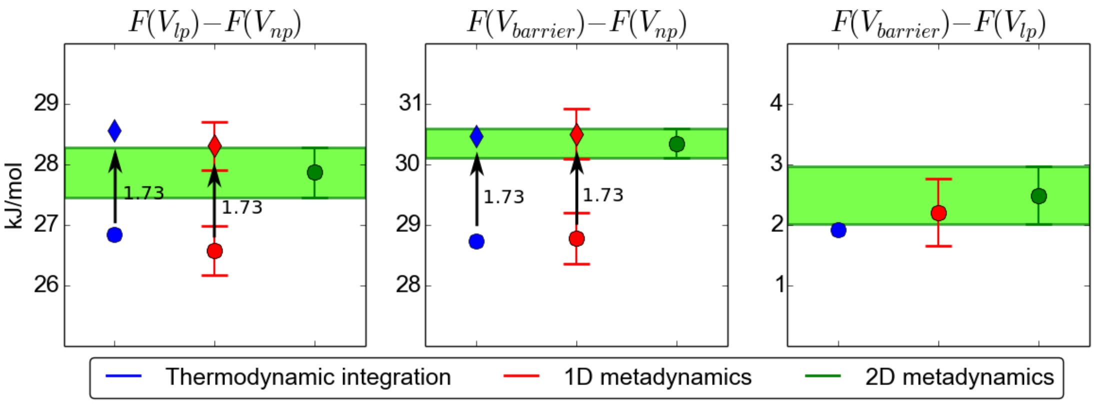


Figure 3.20: Values of the free energy difference between the lp phase and np phase, the height of the free energy barrier from the np phase and the height of the free energy barrier from the lp phase. The original values are denoted by dots, the values for which $F(V_{np})$ is corrected by a factor $-k_BT\ln 2 = -1.73$ kJ/mol are denoted by diamonds. Error bars denotes the region of ±2SE.

in the np phase, sample only one branch at low volumes. At large volumes, however, both branches can be sampled.

Assume that both branches are completely equivalent, so the law of total probability renders for the probability of visiting a volume V in branch $\pm$:

$$p(V|+) = p(V|-) = \frac{1}{2}\Big(p(V|+) + p(V|-)\Big) = \frac{1}{2}p(V) \tag{3.25}$$

Suppose a simulation is performed that only samples the positive branch at V_{np}, but samples both branches at V_{lp}. For the system, the estimated probability $\hat{p}(V)$ of visiting a volume state V (and where the hat denotes the fact that this is an incorrect estimate) is given by:

$$\begin{aligned} \hat{p}(V_{\text{np}}) &= p(V_{\text{np}}|+) = \frac{1}{2}p(V_{\text{np}}) && (3.26)\\ \hat{p}(V_{\text{lp}}) &= p(V_{\text{lp}}|+) + p(V_{\text{lp}}|-) = p(V_{\text{lp}}) && (3.27)\end{aligned}$$

Now, from Equation 2.39, the true free energy $F(V)$ as a function of the unit cell volume, and the difference between the lp and np phase $\Delta F_{\text{np-lp}}$ are given by:

$$\begin{aligned} F(V) = &\; -k_BT\ln\Big(p(V)\Big) = \; -k_BT\ln\Big(p(V|+) + p(V|-)\Big) && (3.28)\\ \Delta F_{\text{lp-np}} = &\; F(V_{\text{lp}}) - F(V_{\text{np}}) = \; -k_BT\ln\Big(p(V_{\text{lp}})\Big) + k_BT\ln\Big(p(V_{\text{np}})\Big) && (3.29)\end{aligned}$$

For the estimated free energy $\hat{F}(V)$ and the difference between the lp and np phase $\Delta\hat{F}_{\text{np-lp}}$

however:

$$\begin{aligned}
\hat{F}(V) &= -k_B T \ln\Big(\hat{p}(V)\Big) && (3.30)\\
\Delta\hat{F}_{\text{lp-np}} &= -k_B T \ln\Big(\hat{p}(V_{\text{lp}})\Big) + k_B T \ln\Big(\hat{p}(V_{\text{np}})\Big) && (3.31)\\
&= -k_B T \ln\Big(p(V_{\text{lp}})\Big) + k_B T \ln\Big(\frac{1}{2}p(V_{\text{np}})\Big) && (3.32)\\
&= \Delta F_{\text{lp-np}} - k_B T \ln 2 && (3.33)
\end{aligned}$$

Hence it is seen that the correct $\Delta\hat{F}_{\text{lp-np}}$ is found by accounting for the missed degeneracy in the np phase, and so lowering the free energy in the np phase by $k_B T \ln 2$. Note that if in both the np as in the lp only one branch is visited, both correction terms cancel in the calculation of $\Delta\hat{F}_{\text{lp-np}}$.

The discussion here holds for a system with a twofold degeneracy in two volume states, np and lp, at which in one state this degeneracy is sampled, and it the other state it is not. Expanding this reasoning to the complete free energy profile as a function of the volume is not straightforward.

3.4.2 Free energy profile as a function of shape direction

The free energy curves $F_{\text{1Dmeta}}(V, a_y = b_x)$ are compared to data from the 2D profiles $F_{\text{ext}}(V, a_y = b_x)$ and $F_{\text{2Dmeta}}(V, a_y = b_x)$. The 2D free energy profile $F_{\text{ext}}(V, a_y = b_x)$ is constructed by using the histogram method, with 100 histogram bins in the volume direction. The values of the free energy as a function of $a_y = b_x$ at the volumes [780, 930, 1080, 1260, 1450] Å^3 are obtained by a linear interpolation of the data at $a_y = b_x$ between two histogram bin centers $V_{1,\text{bin}}$ and $V_{2,\text{bin}}$:

$$F_{\text{ext}}(V, a_y = b_x) = \frac{(V - V_{1,\text{bin}})}{V_{2,\text{bin}} - V_{1,\text{bin}}} F_{\text{ext}}(V_{2,\text{bin}}, a_y = b_x) + \frac{(V_{2,\text{bin}} - V)}{V_{2,\text{bin}} - V_{1,\text{bin}}} F_{\text{ext}}(V_{1,\text{bin}}, a_y = b_x) \tag{3.34}$$

with $V_{1,\text{bin}} < V < V_{2,\text{bin}}$. The datasets are shifted by a constant such that the minimum at $V = 1450$ Å^3 is in line with the previous profiles. The results are shown in Figure 3.21.

There is good agreement between the free energy obtained by the three different simulation methods. The agreement with the histogram method is best in the wells of the free energy, because of the better sampling in those regions. For the umbrella simulation data at low volumes, there is respectively an under- and overestimation of the free energy at negative and positive $a_y = b_x$. This is due to the fact that both branches were not evenly

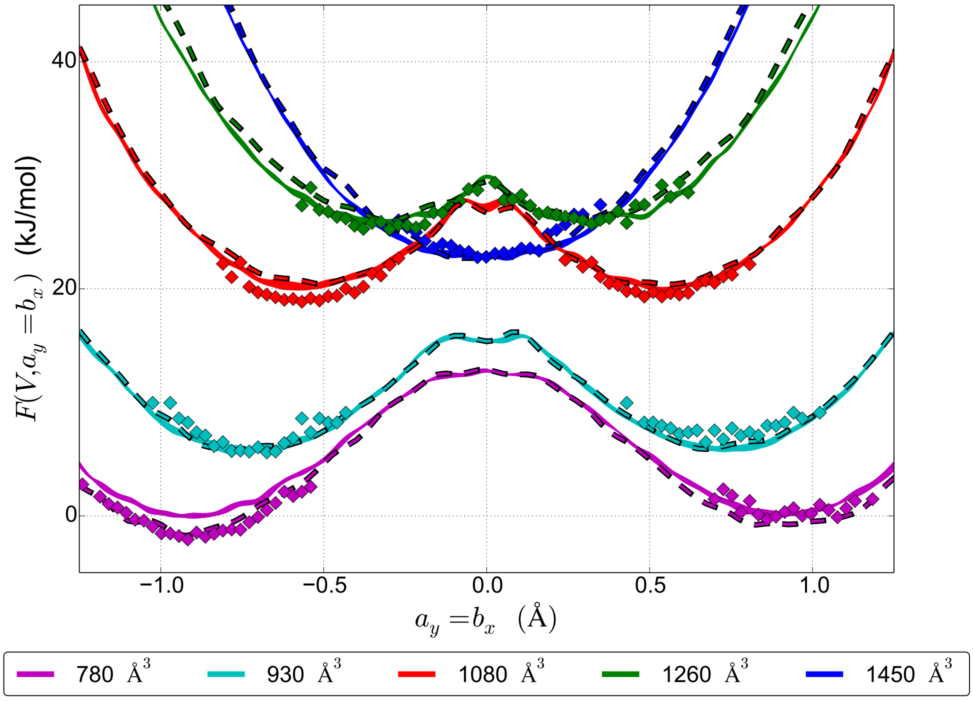


Figure 3.21: Solid lines: Estimate of the free energy curve $F_{1\mathrm{Dmeta}}(V, a_y = b_x)$ (see Subsection 3.3.2). The linewidth denotes the region ±2 $\mathrm{SE}_{F_{1\mathrm{Dmeta}}(V,a_y=b_x)}$. Dashed lines: Cross-sections of the free energy curve $F_{2\mathrm{Dmeta}}(V, a_y = b_x)$ (see Subsection 3.3.3). Diamonds: $F_{\mathrm{ext}}(V, a_y = b_x)$, linear interpolations between bins of $F(V, a_y = b_x)$ of the 2D free energy profile (see Equation 3.34).

frequented, see Figure 3.11, probably due to statistical fluctuations. Keep in mind that the data of the extended umbrella simulation is the result of a linear interpolation between data of two histogram bin centers, and that additional accuracy may be obtained by using more advanced methods.

Because of the large fluctuations and the lack of data for the free energy barrier between two wells, the extended umbrella simulation combined with the histogram method is considered the least precise method for obtaining detailed information of the free energy profile of the $a_y = b_x$ direction at certain volumes. For a discrete set of volumes, the 1D metadynamics simulation with collective variable $a_y = b_x$ exhibits the highest accuracy and fastest convergence. However, if information about the complete space $(V, a_y = b_x)$ is required, 2D metadynamics has reasonable accuracy and convergence.

3.4.3 Free energy profile as a function of the volume and shape direction

Finally, two methods to construct 2D free energy profiles in the collective variable $(V, a_y = b_x)$ are compared.

The first method consists of performing an extended umbrella simulation (see Subsection 3.2.2), with a bias potential that is the negative of the free energy profile obtained in Subsection 3.1.1. With the data from this umbrella simulation, a 2D histogram is constructed, which is transformed to a 2D free energy profile. The flaws of this method are clearly shown in the left panel of Figure 3.22: the free energy can only be calculated for regions that are visited by the system. Since no bias is applied in the $a_y = b_x$ direction, the probability for the system to visit a state at higher free energy decreases exponentially with the height of that free energy. Hence, large areas of the free energy surface remain unsampled. Even in regions that are visited, statistical noise limits the precision. The statistical noise increases dramatically for states with higher free energy. This is not straightforward to see on the figure because the contourplot has levels of height 2 kJ/mol.

The second free energy profile (left panel of Figure 3.22), is obtained by performing 2D metadynamics with the collective variable $(V, a_y = b_x)$ (see Subsection 3.3.3). The correspondence with the extended umbrella simulation is clear. Because of the metadynamics method, the 2D free energy profile of the system is flattened after sufficient simulation time, and every state $(V, a_y = b_x)$ is evenly sampled. Hence, the accuracy of the method at a certain state $(V, a_y = b_x)$ does not depend on the height of the free energy in this state. This method provides detailed information about the height of the free energy barrier between the two branches $\pm|a_y|$, at every volume.

The 2D metadynamics simulation is preferred over the extended umbrella simulation to determine 2D free energy profiles, because of its complete sampling of the 2D space.

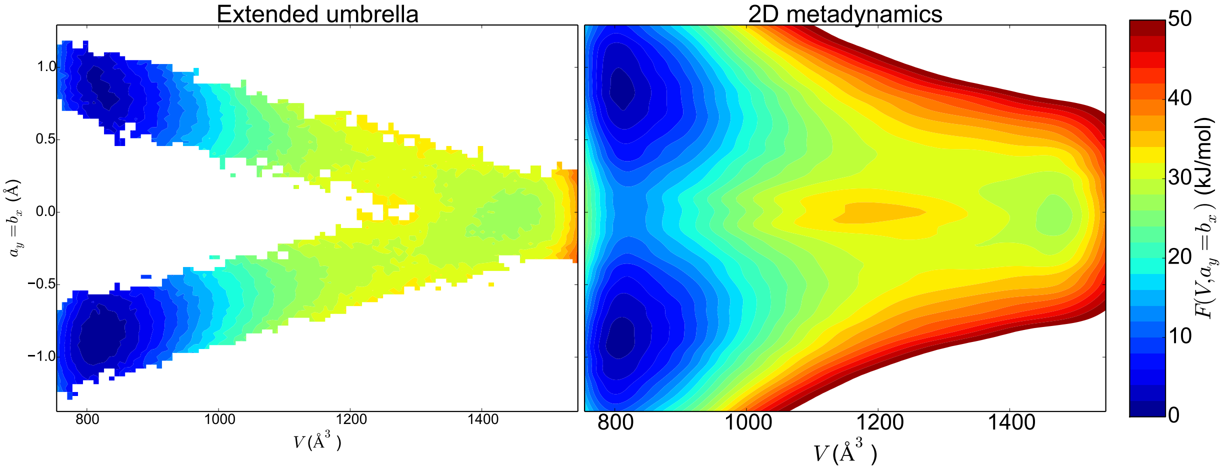


Figure 3.22: 2D free energy profiles $F(V, a_y = b_x)$. Left: profile obtained by application of the histogram method on an extended umbrella biased simulation (see Subsection 3.2.2). Right: profile obtained by performing 2D metadynamics with collective variables $(V, a_y = b_x)$ (see Subsection 3.3.3). Both profiles are shifted such that their global minimum is zero. The white space in the extended umbrella profile denotes the regions on which no information is obtained, the white space in the 2D metadynamics profile shows free energies higher than 50 kJ/mol (a more complete version is shown in Figure 3.17).

Chapter 4

Principal component analysis

Principal component analysis (PCA) is a statistical procedure developed by Pearson in 1901 [82] to convert a set of observations of correlated variables into an uncorrelated set of values. In this work, it is used as a method to accurately analyse the important coordinate fluctuations observed in an MD simulation [83]. These fluctuations are described as function of the eigenmodes. The coordinates describing the eigenmodes of a molecular system are for example the positions of the nuclei $\vec{r_\alpha} = [r_{\alpha x}, r_{\alpha y}, r_{\alpha z}]^T$ (where $\alpha = 1...N$), and/or the unit cell tensor components h_{ij}, $i, j = 1...3$.

PCA is often used in the analysis of protein simulations [84]. More recently, it has also been employed to calculate configurational entropies [85, 86, 87]. Another application of PCA is to return a set of essential coordinates on which metadynamics simulations can be performed [88].

4.1 Preparing data for analysis

Before performing PCA, the data obtained from a simulation must be transformed into an appropriate form.

4.1.1 Rotation of the unit cell

Unit cells have nine independent components, of which three are associated with the orientation of the reference system. This orientation does not influence the unit cell volume, unit cell shape, nor the nuclear positions, it corresponds to a global rotation. There is no coupling between this rotation and the other characteristics of the unit cell, hence the physics remains unchanged. To eliminate these rotations, all unit cells and nuclear positions are rotated back in the same way. This means that the system rotation is adjusted every timestep.

The unit cell will always be rotated in such a way that the unit cell matrix is symmetric. Since the nuclear positions are written in Cartesian coordinates, they must also rotate together with the unit cell. Therefore, for every timestep, an expression for the rotation tensor between the unit cell and the new symmetric unit cell needs to be determined. This can be carried out efficiently using singular value decomposition.

Singular value decomposition

Singular value decomposition (SVD) expands an $m\times n$-matrix $\mathbf{h}$ (in our case $m = n = 3$) into a product of three different matrices: a unitary $m \times m$-matrix $\mathbf{U}$, a diagonal $m \times n$-matrix $\mathbf{S}$ and a unitary $n \times n$-matrix $\mathbf{V}$. This matrix $\mathbf{h}$ is written as

$$\mathbf{h} = \mathbf{USV} = \mathbf{USU}^T\mathbf{UV} = \mathbf{h}_s\mathbf{R} \tag{4.1}$$

$$\text{with} \qquad \mathbf{h}_s = \mathbf{USU}^T \qquad \text{and} \qquad \mathbf{R} = \mathbf{UV} \tag{4.2}$$

As seen in the last part of the equation, expressions for the symmetric cell tensor $\mathbf{h}_s$ and the rotation matrix $\mathbf{R}$ are obtained. During the further discussion, any reference to the cell tensor $\mathbf{h}$ should be interpreted as the symmetric cell tensor $\mathbf{h}_s$, unless mentioned explicitly.

$$\mathbf{h} \to \mathbf{h}_s \tag{4.3}$$

Next, the positions of the nuclei are rotated in accordance with the unit cell.

Nuclear positions

The position of a nucleus α can be represented by its Cartesian coordinates $(x_\alpha, y_\alpha, z_\alpha)$ or by its fractional coordinates $(s_{\alpha a}, s_{\alpha b}, s_{\alpha c})$ along the lattice vectors. The relation between these two representations is given by:

$$\vec{r}_\alpha = x_\alpha\vec{1}_x + y_\alpha\vec{1}_y + z_\alpha\vec{1}_z = s_{\alpha a}\vec{a} + s_{\alpha b}\vec{b} + s_{\alpha c}\vec{c} = \vec{s}_\alpha\mathbf{h} \tag{4.4}$$

Analogously to the unit cell above, the nuclei need to be rotated appropriately to a set of positions $\vec{r}_{R\alpha}$ such that their fractional coordinates $\vec{s}_\alpha$ remain unaltered after rotation of the unit cell. This is done by right multiplying the nuclear Cartesian positions $\vec{r}_\alpha$ with the inverse of the rotation matrix $\mathbf{R}^{-1}$, see Equation 4.1:

$$\vec{r}_{R\alpha} = \vec{s}_\alpha\mathbf{h}_s = \vec{s}_\alpha\mathbf{h}\mathbf{R}^{\mathbf{-1}} = \vec{r}_\alpha\mathbf{R}^{\mathbf{-1}} \tag{4.5}$$

4.1.2 Coordinates

To determine the eigenmodes of the system, the following notation for the components of the unit cell tensor **h** is introduced:

$$\mathbf{h} = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix} = \begin{pmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} & \mathrm{h}_{13} \\ \mathrm{h}_{21} & \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{31} & \mathrm{h}_{32} & \mathrm{h}_{33} \end{pmatrix} \tag{4.6}$$

The PCA performed in this work serves to obtain information on the primary motions/fluctuations of the unit cell and nuclei, and to identify the coupling between them. Therefore the relative set of coordinates consists of:

(i) The symmetric unit cell tensor **h** (of the double unit cell, expanded in the $\vec{b}$ direction), written in Voigt notation as a vector $\vec{\mathrm{h}}$ (6 components):

$$\mathbf{h} = \begin{pmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} & \mathrm{h}_{13} \\ \mathrm{h}_{12} & \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{13} & \mathrm{h}_{23} & \mathrm{h}_{33} \end{pmatrix} \rightarrow \vec{\mathrm{h}} \triangleq \begin{pmatrix} \mathrm{h}_1 \\ \mathrm{h}_2 \\ \mathrm{h}_3 \\ \mathrm{h}_4 \\ \mathrm{h}_5 \\ \mathrm{h}_6 \end{pmatrix} = \begin{pmatrix} \mathrm{h}_{11} \\ \mathrm{h}_{22} \\ \mathrm{h}_{33} \\ 2\mathrm{h}_{23} \\ 2\mathrm{h}_{13} \\ 2\mathrm{h}_{12} \end{pmatrix} \tag{4.7}$$

(ii) The set of appropriately rotated nuclear Cartesian positions $\vec{r}_R$ ($3N = 456$ components):

$$\vec{r}_R = \begin{pmatrix} r_{R1x} \\ r_{R1y} \\ r_{R1z} \\ . \\ . \\ r_{RNz} \end{pmatrix} \tag{4.8}$$

By merging the two sets of coordinates, the complete coordinate set $\vec{q}$ for PCA, containing $6 + 3N$ independent components, is obtained:

$$\vec{q} = \Big(\mathrm{h}_1, ..., \mathrm{h}_6, r_{R1x}, ..., r_{RNz}\Big)^T = \begin{pmatrix} \vec{\mathrm{h}} \\ \vec{r}_R \end{pmatrix} \tag{4.9}$$

4.1.3 Covariance matrix and eigenmodes

During the simulations, the different coordinates will fluctuate or change in magnitude. PCA constructs a covariance matrix **C** of these fluctuations, with as reference the average position of the coordinates, $\langle\vec{q}\rangle$. This results in the following covariance matrix:

$$C_{ij} = \langle(q_i - \langle q_i\rangle)(q_j - \langle q_j\rangle)\rangle \tag{4.10}$$

The units of all the components in this covariance matrix are per definition those of q^2, namely Å^2. The diagonalization of this covariance matrix results in eigenvectors $\vec{v}^{(l)}$ that specify mutually orthogonal directions of fluctuations. The corresponding eigenvalues $\lambda^{(l)}$ are the variances associated with the motion in the directions $\vec{v}^{(l)}$. The square root of the eigenvalues, $\sqrt{\lambda^{(l)}}$, is the standard deviation, and is a measure of the fluctuation in the direction of the eigenvector $\vec{v}^{(l)}$. After diagonalizing, $\mathbf{C}$ is written as:

$$\mathbf{C} = \sum_{l=1...3N+6} \lambda^{(l)}\vec{v}^{(l)}(\vec{v}^{(l)})^T \leftrightarrow C_{ij} = \sum_{l=1...3N+6} \lambda^{(l)} v_i^{(l)} v_j^{(l)} \tag{4.11}$$

Every eigenmode $\vec{v}^{(l)}$ is normalized to 1, and is hence, per definition, dimensionless. The eigenvalues $\lambda^{(l)}$ have dimension Å^2. Also, the eigenmodes are ordered in such a way that $\lambda^{(1)} \geq \lambda^{(2)} \geq \cdots \geq \lambda^{(6+3N)}$, such that the first eigenmode has the largest covariance. Note that all modes are per definition linear in the set of input coordinates. Since the unit cell matrix and Cartesian nuclear positions are used here, all motions represented by PCA are linear combinations of nuclear coordinates and unit cell components. Linker rotations and other non linear motions are expected to be badly represented by this coordinate set. Linear approximations will hold as long as the amplitude of the nonlinearity is sufficiently small.

4.1.4 Visualization of eigenmodes

To inspect the effect of different eigenmodes $\vec{v}^{(l)}$, these modes are visualized. Each mode represents a direction in the space of $\vec{q} = \left(\mathrm{h}_1, ..., \mathrm{h}_6, r_{R1x}, ..., r_{RNz}\right)^T$, but is dimensionless and normalized, as stated earlier. Therefore, all modes are multiplied by a scaling factor A (with the dimension Å) for visualization. This scaling factor is kept small to be able to neglect nonlinear contributions to the motion. The motion is referred to the average of the vector $\vec{q}$ used in the calculation of the covariance (Equation 4.10). To visualize the mode, the following shift in positions is used:

$$\vec{q}^{(l)}_{\pm A} = \langle\vec{q}\rangle \pm A\vec{v}^{(l)} \tag{4.12}$$

4.1.5 Critical note

It is important to discuss the limitations of PCA and of the methods performed further to quantify and describe the eigendirections.

Firstly, PCA is a descriptive method that returns eigendirections by diagonalizing the covariance matrix. This covariance matrix depends on the coordinate set chosen to describe the system, the simulation time, the simulation range, etc. [83, 89]. If another coordinate set is chosen, another set of perpendicular modes will be returned. Also, the modes are ordered according to their variance and this variance also depends on the coordinate set. The obtained PCA modes are hence a result of these choices: the coordinate set is taken as the set of cell matrix components and Cartesian nuclear positions (both having unit Å), to accurately describe the coupling of nuclear motion with the unit cell.

This choice of coordinates also induces another problem: PCA makes a linear approximation of the motion. As a result, rotations are not described accurately. However, as long as these rotations are small, the linear approximation results in fairly good results. For visualization and characterization, this means that the amplitude of the modes should ideally be kept small. If one is specifically interested in rotations, a different choice of coordinate set should be made (e.g. angles and distances [90]).

4.2 Analysis of the simulation set

PCA was performed on a set of three umbrella simulations, in which a different bias potential U_b was applied (see Table 4.1 and Figure 4.1). It is expected that PCA provides an accurate description of the fluctuations in the unit cell shape, as well as the nuclear movements associated with them.

Table 4.1: Simulations performed for PCA.

Ensemble	$NP(\boldsymbol{\sigma}_a = \mathbf{0})T$	$NP(\boldsymbol{\sigma}_a = \mathbf{0})T$	$NP(\boldsymbol{\sigma}_a = \mathbf{0})T$
Timestep	0.5 fs	0.5 fs	0.5 fs
Simulation time	400 ps	400 ps	400 ps
Temperature	300 K	300 K	300 K
Pressure	0 MPa	0 MPa	0 MPa
Volume	≈ 820 Å^3	≈ 1450 Å^3	$770 - 1550$ Å^3
Bias potential	parabola	parabola	inverse FEP
V_0	820 Å^3	1450 Å^3	–
K	$10^{-4} k_B T$	$10^{-4} k_B T$	–
Description	np umbrella	lp umbrella	extended umbrella

The largest 20 standard deviations $\sqrt{\lambda^{(l)}}$ obtained by the analysis are shown in Figure 4.2. The extended simulation contains the largest eigenvalue, because the system was sampled in a large range, corresponding to a large standard deviation. Important to note is that the fluctuations around the lp phase are more pronounced compared to the fluctuations in the np phase , which can be explained by the larger unit cell size. Hence, there is relatively more space available for nuclear motions in the lp phase.

The cell component sizes of the first 20 eigenmodes are shown in Figure 4.3. This figure indicates differences between the unit cell components of the largest eigenmodes in the lp and np phases. While there are only two modes in the lp phase that have a large component in a_x and c_z (modes 1 and 10), in correspondence with previous research [24], there are numerous such examples in the np phase. Also the $2c_x = 2a_z$ component is unaltered by most modes in the lp, in contrast to the np. The b_y component, which lies in the direction of the Al chain, remains quasi constant in all modes. Hence, the direction of the aluminum oxide chain can be considered a rigid direction.

4.2.1 Volume change

The volume change related to the different eigenmodes is studied. For each eigenmode l the volume change is quantified. This is performed by taking the mean coordinates throughout the simulation as the reference state, applying a small perturbation in the direction of the eigenmode, and calculating the obtained volume change.

As already introduced in Equation 2.2, the volume of the unit cell is calculated from the determinant of the unit cell tensor. To get an expression for an infinitesimal volume change dV as a function of the infinitesimal unit cell tensor changes $d\mathrm{h}_{ij}$, the chain rule is applied:

$$dV = \sum_{i,j=1}^{3} \frac{\partial(\det(\mathbf{h}))}{\partial \mathrm{h}_{ij}} d\mathrm{h}_{ij} \tag{4.13}$$

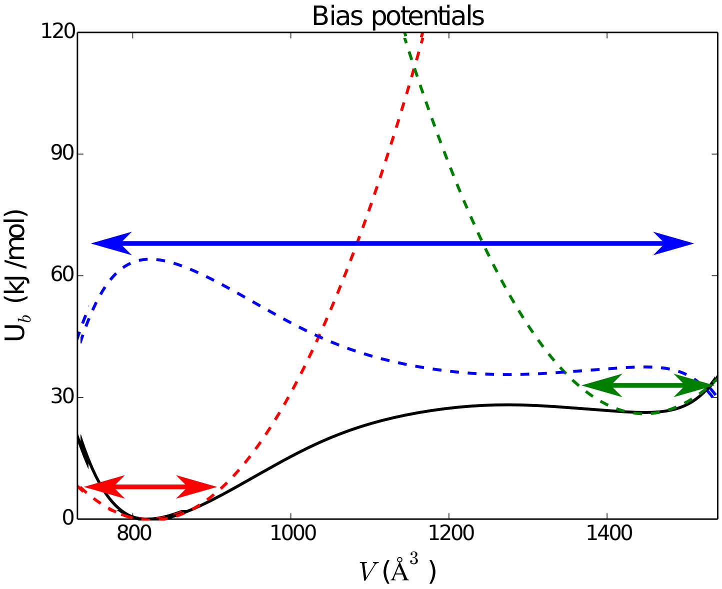


Figure 4.1: Bias potentials U_b applied to the simulations studied in Table 4.1.

Since V can be written as

$$\begin{aligned}
V = \det(\mathbf{h}) &= \begin{vmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} & \mathrm{h}_{13} \\ \mathrm{h}_{21} & \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{31} & \mathrm{h}_{32} & \mathrm{h}_{33} \end{vmatrix} && (4.14) \\
&= (-1)^{1+1}\mathrm{h}_{11} \begin{vmatrix} \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{32} & \mathrm{h}_{33} \end{vmatrix} + (-1)^{1+2}\mathrm{h}_{12} \begin{vmatrix} \mathrm{h}_{21} & \mathrm{h}_{23} \\ \mathrm{h}_{31} & \mathrm{h}_{33} \end{vmatrix} && (4.15) \\
&\quad +(-1)^{1+3}\mathrm{h}_{13} \begin{vmatrix} \mathrm{h}_{21} & \mathrm{h}_{22} \\ \mathrm{h}_{31} & \mathrm{h}_{32} \end{vmatrix}
\end{aligned}$$

and equivalent for other rows, the expression for dV yields, with "%" the modulo sign:

$$\begin{aligned}
dV &= \begin{vmatrix} \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{32} & \mathrm{h}_{33} \end{vmatrix} d\mathrm{h}_{11} - \begin{vmatrix} \mathrm{h}_{21} & \mathrm{h}_{23} \\ \mathrm{h}_{31} & \mathrm{h}_{33} \end{vmatrix} d\mathrm{h}_{12} + \cdots + \begin{vmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} \\ \mathrm{h}_{21} & \mathrm{h}_{22} \end{vmatrix} d\mathrm{h}_{33} && (4.16) \\
&= \begin{vmatrix} \mathrm{h}_{22} & \mathrm{h}_{23} \\ \mathrm{h}_{32} & \mathrm{h}_{33} \end{vmatrix} d\mathrm{h}_{11} + \begin{vmatrix} \mathrm{h}_{23} & \mathrm{h}_{21} \\ \mathrm{h}_{33} & \mathrm{h}_{31} \end{vmatrix} d\mathrm{h}_{12} + \cdots + \begin{vmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} \\ \mathrm{h}_{21} & \mathrm{h}_{22} \end{vmatrix} d\mathrm{h}_{33} && (4.17) \\
&= \sum_{i,j=1}^{3} \begin{vmatrix} \mathrm{h}_{(i\%3+1)(j\%3+1)} & \mathrm{h}_{(i\%3+1)((j+1)\%3+1)} \\ \mathrm{h}_{((i+1)\%3+1)(j\%3+1)} & \mathrm{h}_{((i+1)\%3+1)((j+1)\%3+1)} \end{vmatrix} d\mathrm{h}_{ij} && (4.18) \\
&= \sum_{i,j=1}^{3} (\mathrm{h}_{\mathrm{vol}})_{ij} d\mathrm{h}_{ij} && (4.19)
\end{aligned}$$

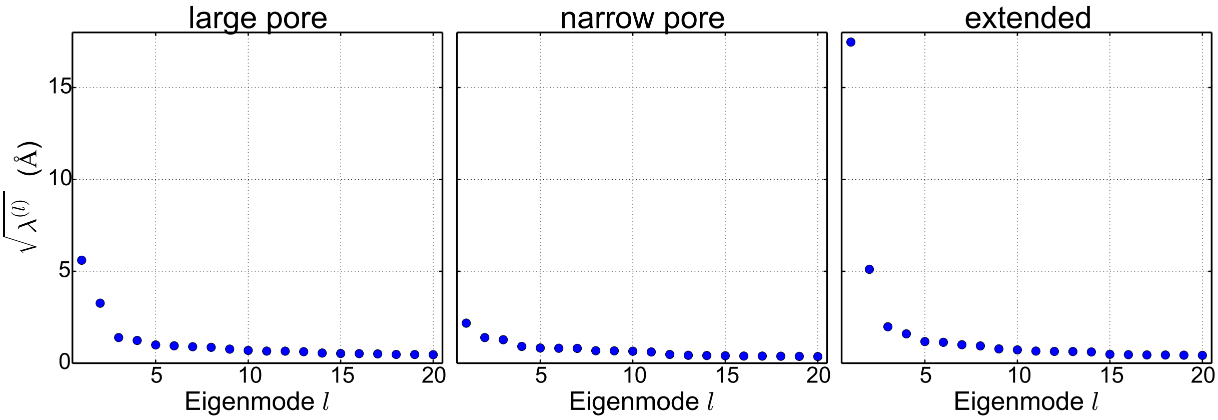


Figure 4.2: The square root of the eigenvalue $\lambda^{(l)}$ for the first 20 eigenmodes $\vec{v}^{(l)}$. This is a measure of the fluctuation magnitude.

The matrix $\mathbf{h}_{\mathrm{vol}}$, having the dimension of Å^2, is defined by the last equation. This equation can be interpreted as a projection of the unit cell change $d\mathbf{h}$ on the direction of volume change $\mathbf{h}_{\mathrm{vol}}$. This matrix is evaluated for a reference unit cell, the average unit cell matrix $\langle \mathbf{h} \rangle$. For example, the element (12) of $\mathbf{h}_{\mathrm{vol}}$ renders

$$(\mathrm{h}_{\mathrm{vol}})_{12} = \begin{vmatrix} \mathrm{h}_{23} & \mathrm{h}_{21} \\ \mathrm{h}_{33} & \mathrm{h}_{31} \end{vmatrix} = \mathrm{h}_{23}\mathrm{h}_{31} - \mathrm{h}_{21}\mathrm{h}_{33} \tag{4.20}$$

Since all unit cells and unit cell shape changes are symmetric (see section 4.1.1), they can be transformed to the same Voigt notation as seen in equation 4.7:

$$d\mathbf{h} = \begin{pmatrix} d\mathrm{h}_{11} & d\mathrm{h}_{12} & d\mathrm{h}_{13} \\ d\mathrm{h}_{12} & d\mathrm{h}_{22} & d\mathrm{h}_{23} \\ d\mathrm{h}_{13} & d\mathrm{h}_{23} & d\mathrm{h}_{33} \end{pmatrix} \rightarrow d\vec{\mathrm{h}} = \begin{pmatrix} d\mathrm{h}_{11} \\ d\mathrm{h}_{22} \\ d\mathrm{h}_{33} \\ 2d\mathrm{h}_{23} \\ 2d\mathrm{h}_{13} \\ 2d\mathrm{h}_{12} \end{pmatrix} \tag{4.21}$$

$$\mathbf{h}_{\mathrm{vol}} = \begin{pmatrix} (\mathrm{h}_{\mathrm{vol}})_{11} & (\mathrm{h}_{\mathrm{vol}})_{12} & (\mathrm{h}_{\mathrm{vol}})_{13} \\ (\mathrm{h}_{\mathrm{vol}})_{12} & (\mathrm{h}_{\mathrm{vol}})_{22} & (\mathrm{h}_{\mathrm{vol}})_{23} \\ (\mathrm{h}_{\mathrm{vol}})_{13} & (\mathrm{h}_{\mathrm{vol}})_{23} & (\mathrm{h}_{\mathrm{vol}})_{33} \end{pmatrix} \rightarrow \vec{\mathrm{h}}_{\mathrm{vol}} = \begin{pmatrix} (\mathrm{h}_{\mathrm{vol}})_{11} \\ (\mathrm{h}_{\mathrm{vol}})_{22} \\ (\mathrm{h}_{\mathrm{vol}})_{33} \\ 2(\mathrm{h}_{\mathrm{vol}})_{23} \\ 2(\mathrm{h}_{\mathrm{vol}})_{13} \\ 2(\mathrm{h}_{\mathrm{vol}})_{12} \end{pmatrix} \tag{4.22}$$

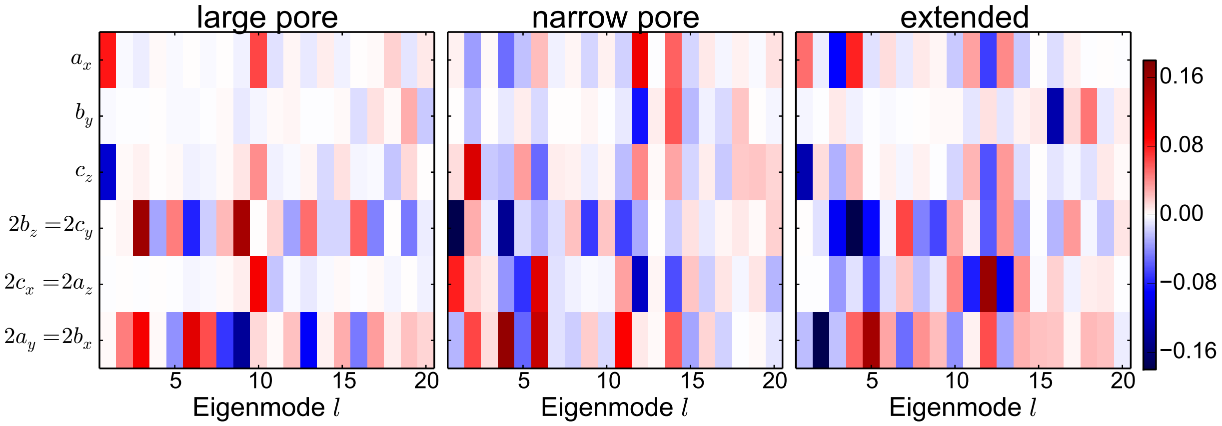


Figure 4.3: Visualization of the cell component sizes of the eigenmodes $\vec{v}^{(l)}_{1...6}$ for $l = 1, .., 20$ for the three different biased simulations. A multiplication of an eigenmode $\vec{v}^{(l)}$ by -1 leaves the orthonormality of the modes invariant.

Since the volume depends only on $\vec{h}_{\text{vol}}$, the last vector is used to construct a unit vector in the direction of the maximum volume change $\vec{e}_{\text{vol}}$ in the basis of $\vec{q}$ ($6 + 3N$ components, see Equation 4.9) with only the first six components nonzero and equal to $\vec{\mathrm{h}}_{\text{vol}}$.

$$\vec{e}_{\text{vol}} = \frac{1}{||\vec{\mathrm{h}}_{\text{vol}}||}\begin{pmatrix} \vec{\mathrm{h}}_{\text{vol}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{1}{||\vec{\mathrm{h}}_{\text{vol}}||}\begin{pmatrix} (\mathrm{h}_{\text{vol}})_{11} \\ (\mathrm{h}_{\text{vol}})_{22} \\ (\mathrm{h}_{\text{vol}})_{33} \\ 2(\mathrm{h}_{\text{vol}})_{23} \\ 2(\mathrm{h}_{\text{vol}})_{13} \\ 2(\mathrm{h}_{\text{vol}})_{12} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{4.23}$$

The dot product of the normalized $\vec{e}_{\text{vol}}$ with an eigenvector $\vec{v}^{(l)}$ quantifies the normalized change of volume present in the eigenvector $\vec{v}^{(l)}$:

$$(c^{(l)}_{\text{vol}})^2 = (\vec{v}^{(l)} \cdot \vec{e}_{\text{vol}})^2 \tag{4.24}$$

Alternatively, we can use $\vec{e}_{\text{vol}}$ to compute the finite volume change $\Delta V^{(l)}$ caused by a displacement $A\vec{v}^{(l)}$, in the direction of $\vec{v}^{(l)}$ and with amplitude A.

$$\Delta V^{(l)} = A\vec{v}^{(l)} \cdot ||\vec{\mathrm{h}}_{\text{vol}}||\vec{e}_{\text{vol}} \tag{4.25}$$

$\Delta V^{(l)}$ specifies the difference in volume between the state $\langle \vec{q} \rangle$ and $\langle \vec{q} \rangle + Av^{(l)}$. The scaling factor is chosen to be $A = 1$ Å, the eigenmode $\vec{v}^{(l)}$ and the projection vector $\vec{e}_{\text{vol}}$ are dimensionless, the norm $||\vec{\mathrm{h}}_{\text{vol}}||$ has the dimension Å^2, so $\Delta V^{(l)}$ is expressed in Å^3. The choice of the scaling factor results in a maximum change of unit cell vector component of about 0.18 Å, as seen from the component sizes in Tables C.1 and C.2.

4.2.2 Component size associated with shape change

The physical unit cell has a total of six degrees of freedom. An expression for the direction associated with the volume change is provided by $\vec{e}_{\text{vol}}$. The additional five degrees of freedom are not associated with volume change, but rather with a change in cell shape. These directions $\vec{e}_{\text{shape}}$ are found by considering components perpendicular to the volume direction. These five vectors are also independent of the nuclear positions, such that all nuclear components are zero, are normalized and mutually perpendicular. For a given shape direction k the following conditions are met:

$$(\vec{e}_{\text{shape}})_k \perp \vec{e}_{\text{vol}} \tag{4.26}$$

$$(\vec{e}_{\text{shape}})_k \perp (\vec{e}_{\text{shape}})_{k'} \qquad \forall k' \neq k \tag{4.27}$$

$$||(\vec{e}_{\text{shape}})_k|| = 1 \tag{4.28}$$

For every eigenmode l, the components in the shape direction k (found by projection of $\vec{v}_{(l)}$ on $\vec{e}_{\text{shape }k}$) are squared and summed over k. This gives the size of the shape-changing component in the eigendirection l:

$$(c^{(l)}_{\text{shape}})^2 = \sum_{k=1}^{5} (\vec{v}^{(l)} \cdot (\vec{e}_{\text{shape}})_k)^2 \tag{4.29}$$

4.2.3 Component size associated with nuclear movements

The space of the nuclear positions is $3N$ dimensional. The $3N$ projection vectors $(\vec{e}_{\text{nuc}})_l$ need to be perpendicular to $\vec{e}_{\text{vol}}$ and the five $(\vec{e}_{\text{shape}})_k$. The most straightforward expression for a set of $3N$ perpendicular vectors $(\vec{e}_{\text{nuc}})_m$ $(m = 1, .., 3N)$ of size $6 + 3N$ is to set the only nonzero component at place $6 + m$ equal to 1. The size of the nuclear movement component $(c^{(l)}_{\text{nuc}})^2$ of component l is found in the same way as in the previous section:

$$(c^{(l)}_{\text{nuc}})^2 = \sum_{m=1}^{3N} (\vec{v}^{(l)} \cdot (\vec{e}_{\text{nuc}})_m)^2 \tag{4.30}$$

Because the set of $6+3N$ unit vectors $\vec{e}_{\mathrm{vol}}$, $(\vec{e}_{\mathrm{shape}})_k$ and $(\vec{e}_{\mathrm{nuc}})_m$ forms a complete basis, the following validation can be performed:

$$\begin{aligned} 1 &= (\vec{v}^{(l)}\cdot\vec{e}_{\mathrm{vol}})^2+\sum_{k=1}^{5}(\vec{v}^{(l)}\cdot(\vec{e}_{\mathrm{shape}})_k)^2+\sum_{m=1}^{3N}(\vec{v}^{(l)}\cdot(\vec{e}_{\mathrm{nuc}})_m)^2 && (4.31)\\ &= (c_{\mathrm{vol}}^{(l)})^2+(c_{\mathrm{shape}}^{(l)})^2+(c_{\mathrm{nuc}}^{(l)})^2 && (4.32)\end{aligned}$$

for each component l.

4.2.4 Component size associated with internal nuclear movements

A capital remark has to be made. The size of the component associated with nuclear movement $(c_{\mathrm{nuc}}^{(l)})^2$ is not very informative. It shows the size of the complete variance of the nuclear movements, but it is expected that the nuclei move together with the fluctuations of the unit cell, and thus large cell fluctuations result in very large variances in the nuclear positions. More local and more interesting nuclear movements, such as linker rotations or kneecap motions, are completely overshadowed by this effect. However, there is a way to correct for the correlation between the cell fluctuations and nuclear movements, by correcting the motion of the nuclei for the change in unit cell.

The displacement the nuclei undergo due to a change of unit cell can be calculated by keeping the fractional coordinates of the nuclei constant while the unit cell undergoes a transition. The initial position $\vec{r}_\alpha$ of a nucleus α is the mean position obtained in PCA $\langle\vec{r}_\alpha\rangle$. The fractional coordinates $\vec{s}_\alpha$ with respect to the initial unit cell $\mathbf{h}=\langle\mathbf{h}\rangle$ (see Equation 4.4) are given by:

$$\vec{s}_\alpha=\vec{r}_\alpha\mathbf{h}^{-1} \tag{4.33}$$

Next, a deformation of this unit cell in the direction of the components of an eigendirection $\vec{v}^{(l)}$ is applied to this unit cell (as a function of a scaling factor A):

$$\mathbf{h}'(A)=\begin{pmatrix} \mathrm{h}_{11}+Av_1^{(l)} & \mathrm{h}_{12}+\frac{1}{2}Av_6^{(l)} & \mathrm{h}_{13}+\frac{1}{2}Av_5^{(l)}\\ \mathrm{h}_{12}+\frac{1}{2}Av_6^{(l)} & \mathrm{h}_{22}+Av_2^{(l)} & \mathrm{h}_{23}+\frac{1}{2}Av_4^{(l)}\\ \mathrm{h}_{13}+\frac{1}{2}Av_5^{(l)} & \mathrm{h}_{23}+\frac{1}{2}Av_4^{(l)} & \mathrm{h}_{33}+Av_3^{(l)} \end{pmatrix} \tag{4.34}$$

The fractional coordinates $\vec{s}_\alpha$ are transformed to new Cartesian coordinates $\vec{r'}_\alpha(A)$, in the new unit cell $\mathbf{h}'(A)$:

$$\vec{r'}_\alpha(A)=\vec{s}_\alpha\mathbf{h}'(A) \tag{4.35}$$

The displacement $\vec{d}_\alpha(A)$ of nucleus α due to the unit cell deformation is $\vec{d}_\alpha(A) = \vec{r'}_\alpha(A) - \vec{r}_\alpha$. Merging the displacement of every nucleus into a complete displacement vector $\vec{d}(A) = \left(\vec{d}_1^T(A), ..., \vec{d}_N^T(A)\right)^T$ of length $3N$, an expression is obtained for the complete set of displacements of the nuclei due to the unit cell displacement. This can be compared to the total displacement of the nuclei $A\vec{v}^{(l)}_{7...6+3N}$, to obtain the displacement of the nuclei independent of the unit cell displacement:

$$\vec{r}_{\text{nuc,int}}(A) = A\vec{v}^{(l)}_{7...6+3N} - \vec{d}(A) \tag{4.36}$$

The vector $\vec{r}_{\text{nuc,int}}(A)$ has the dimension Å, because it specifies the internal motion of the atoms if the system undergoes a displacement $A\vec{v}^{(l)}$. The dimensionless variant is obtained by dividing by this scaling factor:

$$\vec{v}^{(l)}_{\text{nuc,int}} = \frac{1}{A}\vec{r}_{\text{nuc,int}}(A) \tag{4.37}$$

The size of this fluctuation is projected on the set of $\vec{e}_{\text{nuc }k}$ as in Equation 4.30:

$$(c^{(l)}_{\text{nuc,int}})^2 = \sum_{k=1}^{3N} (\vec{v}^{(l)}_{\text{nuc,int}} \cdot (\vec{e}_{\text{nuc }k})_{7\cdots 6+3N})^2 \tag{4.38}$$

The results of this set of calculations are shown in Figure 4.4. Only the first 20 modes ($l = 1, ..20$) are plotted, because these have the highest eigenvalue (the highest fluctuation). The cardinality of the set of modes that change the cell shape and/or volume, and the set of modes that do not, are of comparable size. Some modes have a large internal nuclear motion, which indicates that the nuclear motion is decoupled from the unit cell changes for these eigenmodes. However, overall, there is a large coupling between the nuclear motions and the cell volume and shape changes. Therefore it is suggested that to appropriately sample these nuclear motions, the cell shape needs to be allowed to fluctuate. This leads to the conclusion that the $NV\mathbf{h}_0T$ ensemble, in which the unit cell is kept fixed, is not suited for the sampling of all nuclear motions. A small remark on the component $(c^{(l)}_{\text{nuc,int}})^2$ has to be made. If it is zero, during motion in the direction of the eigenmode, the fractional coordinates of the nuclei do not change. If it is equal to $(c^{(l)}_{\text{nuc}})^2$, the motion of the nuclei is completely decoupled from the motion of the unit cell.

4.2.5 Variance in volume

During the simulation, the volume will fluctuate. A certain variance is associated with this fluctuation, and can be expressed as:

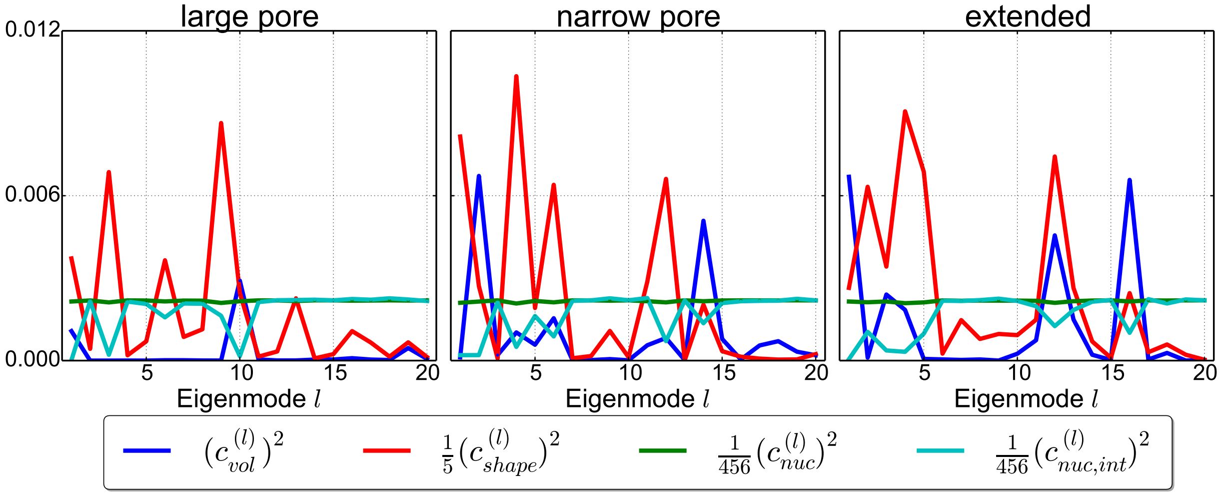


Figure 4.4: The component sizes $(c^{(l)}_{\mathrm{vol}})^2$, $(c^{(l)}_{\mathrm{shape}})^2$, $(c^{(l)}_{\mathrm{nuc}})^2$ and $(c^{(l)}_{\mathrm{nuc,int}})^2$ for modes $l = 1, .., 20$ in the set of biased simulations (lp, np and extended bias). Terms are scaled by their dimension (1 for volume, 5 for shape, and 456 for the nuclei) for visualization.

$$\sigma_V^2 = \langle (V - \langle V \rangle)^2 \rangle \approx \sum_{l=1\ldots 3N+6} \lambda^{(l)} (||\vec{\mathrm{h}}_{\mathrm{vol}}||\vec{v}^{(l)} \cdot \vec{e}_{\mathrm{vol}})^2 \equiv \sum_{l=1\ldots 3N+6} (\sigma_V^{(l)})^2 \tag{4.39}$$

The second step is an approximation, because the volume is a non linear function of the unit cell vector components. This corresponds to calculating $\Delta V^{(l)}$ with a scaling factor $A = \sqrt{\lambda^{(l)}}$, taking the square and summing over all modes. The dimension of σ_V^2 is $[V]^2 = \text{Å}^6$. The quantity σ_V is a measure of the standard deviation of the volume. The quantity $\sigma_V^{(l)}$ is hence straightforwardly interpreted as the amount of volume fluctuation caused by the variation in the eigenmode l.

In the case of MIL-53(Al), the fluctuations $\sigma_V^{(l)}$ for the set of simulations are shown in Figure 4.5. From this figure can be concluded that, one mode is responsible for the majority of the volume fluctuation in each simulation (the second in the np, the first in the lp and extended simulation). This mode is labeled the lp-to-np transition. The size of the volume fluctuations depends on the bias applied in the volume, but one is only interested in the relative sizes of the fluctuations for a given bias. This is the reason why the volume fluctuation in the extended bias simulation is by far the largest, because the applied bias has flattened the free energy profile as a function of the volume in the lp-to-np range.

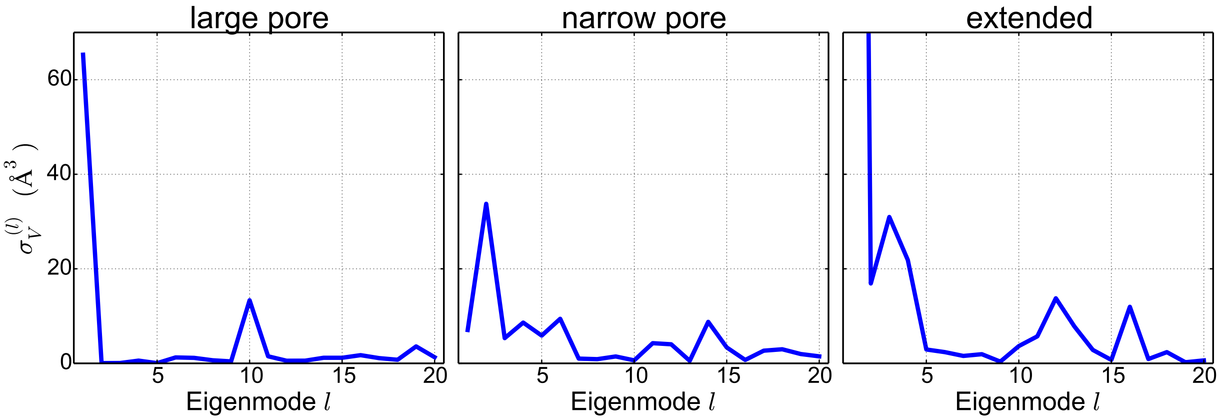


Figure 4.5: Volume fluctuation $\sigma_V^{(l)}$ per eigenmode l as calculated in Equation 4.39 for the first 20 eigenmodes ($l = 1, .., 20$), interpreted as the size of the fluctuation in the volume per mode.

4.3 Discussion of eigenmodes present in MIL-53(Al)

Combining all the information obtained in the previous sections, a description of the most important eigenmodes present in the simulation set is provided. These modes are characterized by the displacement they induce. Some modes are characteristic for the lp state of the system, others for the np or the transition region, but most of the eigenmodes have a counterpart (or are a linear combination of a small number of other modes) in every volume region. For example, finding the counterparts of one mode l of an lp umbrella simulation (lp) in an np umbrella simulation (np), is done by projecting this lp mode on the complete set of np modes $l' = 1, .., 6 + 3N$:

$$c_{\mathrm{lp,np}}^{(ll')} = \vec{v}_{\mathrm{lp}}^{(l)} \cdot \vec{v}_{\mathrm{np}}^{(l')} \qquad (l, l' = 1, ..., 6 + 3N) \tag{4.40}$$

The results of these calculations are visualized in Figure 4.6. This set of calculations aids the interpretation of the correspondence between eigenmodes in different simulations.

The description and the quantification will only be done for the first 20 eigenmodes $\vec{v}^{(l)}$ with $l = 1, .., 20$, i.e. the set of 20 modes with the largest eigenvalues. The cell component sizes are shown in Figure 4.3, the fluctuation sizes $\sigma_V^{(l)}$ in Figure 4.5. Next, the component sizes associated with cell volume $(c_{\mathrm{vol}}^{(l)})^2$, cell shape $(c_{\mathrm{shape}}^{(l)})^2$, nuclear Cartesian $(c_{\mathrm{nuc}}^{(l)})^2$ and nuclear internal motion $(c_{\mathrm{nuc,int}}^{(l)})^2$ are shown in Figure 4.4. Finally, the set of projections between the different simulation sets is shown in Figure 4.6. These data are used to give the following characterization of the eigenmodes:

- **Lp-to-np transition**: The modes $\vec{v}_{\mathrm{lp}}^{(1)}$, $\vec{v}_{\mathrm{ext}}^{(1)}$ and $\vec{v}_{\mathrm{np}}^{(2)}$ are the ones associated with

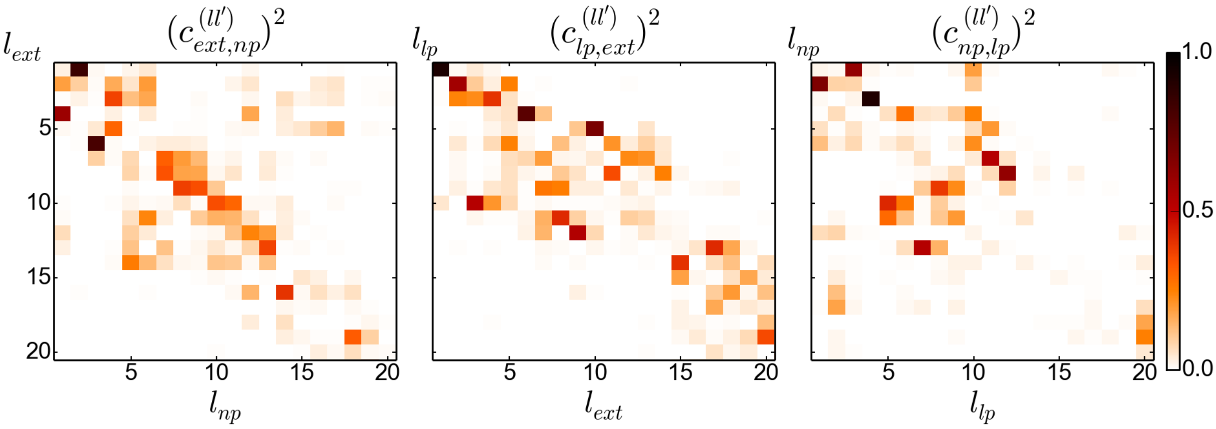
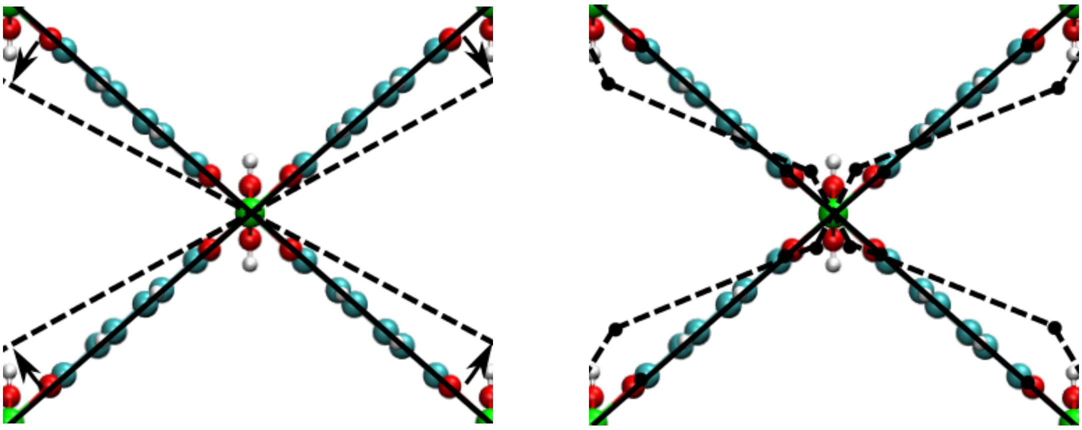


Figure 4.6: The set of projections of $(c^{(ll')}_{\text{lp,ext}})^2$ (left), $(c^{(ll')}_{\text{np,ext}})^2$ (middle) and $(c^{(ll')}_{\text{lp,np}})^2$ (right) for $l, l' = 1, .., 20$. The mixing of the modes remains fairly small, because the majority of the modes are a linear combination of a maximum of six or seven modes from another simulation. Also, the structure of the matrices is fairly diagonal, so the ordening of the modes remains roughly the same.

the lp-to-np transition (this is the first mode with a very large volume fluctuation). In the lp phase no internal nuclear movement is observed, so it is assumed that there is little to no kneecap motion present. Since nuclei follow the unit cell motion during the cell change, the lp-to-np transition in the lp can be characterized as a winerack motion (see Figure 4.7a). In the np phase, a nonnegligible component size is associated with internal nuclear motion. After visual inspection, it is concluded that the lp-to-np transition in the np phase contains a kneecap motion (see Figure 4.7b). In the extended bias simulation the volume fluctuation component $\sigma^{(1)}_{V,ext}$ is bigger than the volume fluctuation in the lp ($\sigma^{(1)}_{V,lp}$) and np ($\sigma^{(2)}_{V,np}$) because of the parabolic bias potential applied in this lp/np simulation pair. The correspondence of the eigendirection between the lp/np and extended simulations is good, as seen in Figure 4.6.

- **Linker rotations**: Mode 2 from the lp and extended umbrella simulation is observed to be the linker rotation around the linker axis (see Figure 4.8b). This motion was found to be very suppressed in the np simulation, because linkers are more closely packed together, which makes it harder for them to rotate. In the lp, the variation in the volume and shape of the unit cell is zero, so these rotations are not associated with changes in cell components. However, in the extended umbrella simulation, the rotations of the linkers are associated with changes in unit cell shape, resulting in

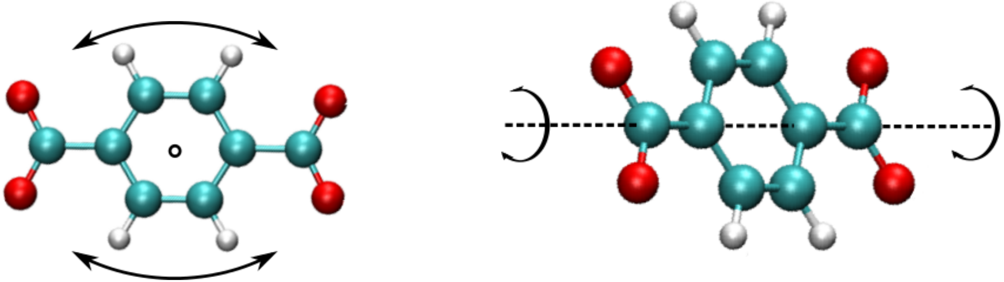


(a) The winerack motion. (b) The kneecap motion.

Figure 4.7: Modes that result in the lp-to-np transition.

lower agreement between both modes.

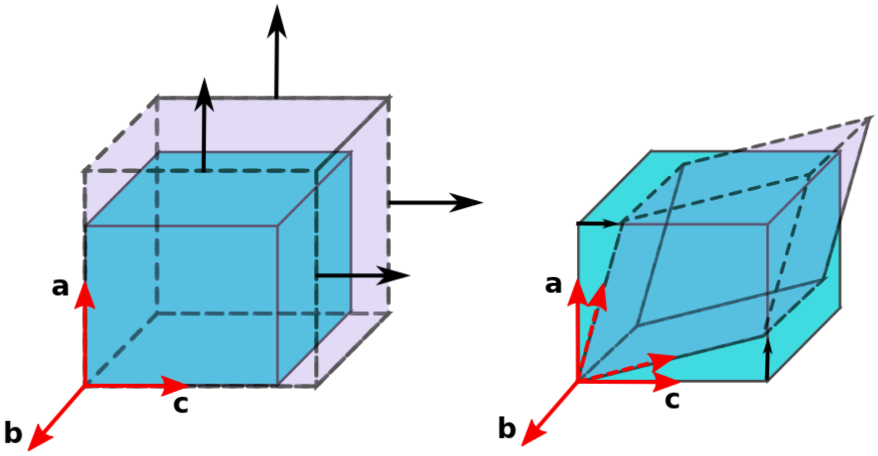
- **The (b_x,$2b_z$) cell skewing**: Mode 3 in the lp simulation, a linear combination of modes 2, 3 and 4 in the extended simulation, corresponds to mode 1 in the np simulation. It results in a skewing of the unit cell in the direction perpendicular to one plane of linkers (an example of cell skewing is given in Figure 4.9b). No volume change is associated with this movement, only a cell shape change in the ($b_x = a_y$,$2b_z = 2c_y$) direction. The internal nuclear motion is very small. Small linker rotations in the plane were observed (see Figure 4.8a), but this is mostly due to cell shape change, not due to internal nuclear motion.
- **The symmetric linker twist**: Mode 4 in the lp simulation, corresponding mostly with mode 6 in the extended simulation and corresponding very well with mode 3 in the np simulation, is the symmetric linker twist, as shown in Figure 4.10a. This mode represents a rotation of the linkers around the aluminum node chains. All linkers rotate in the same direction. Very small changes happen in the cell components, so most of the nuclear motions are independent of the cell.
- **The antisymmetric linker twist**: Mode 5 in the lp simulation, corresponding with mode 10 in extended and mode 10 and 11 in np simulation, is the antisymmetric linker twist, as shown in Figure 4.10b. There is a rotation of one half of the linkers around the aluminum node chains. Linkers connected to the same two aluminums rotate in the same direction. The other set of linkers rotate around their long axis. Small changes happen in the cell components, so the vast majority of the nuclear motions



(a) Linker rotation in the plane.

(b) Linker rotation around the main axis.

Figure 4.8: Linker rotations.



(a) The cell expansion.

(b) Cell skewing.

Figure 4.9: Examples of unit cell motion.

are independent of the cell.

- **The antisymmetric linker twist 2**: Mode 6 in the lp simulation, corresponding with modes 5 and 11 in the extended and modes 4 and 10 in the np simulation, is a variant of the antisymmetric linker twist, as shown in Figure 4.10b and described above. The difference between this mode and the previous lies in the fact that more motion is associated with linker rotation around their axis, and less in the rotation of linkers around the aluminum chain. More movement is associated with cell shape.

- **The aluminum chain parallel shift**: Mode 7 in the lp simulation, corresponding with mode 12 in extended and mode 13 in the np simulation, is a motion associated with movement of the aluminum chains with respect to each other, parallel to the

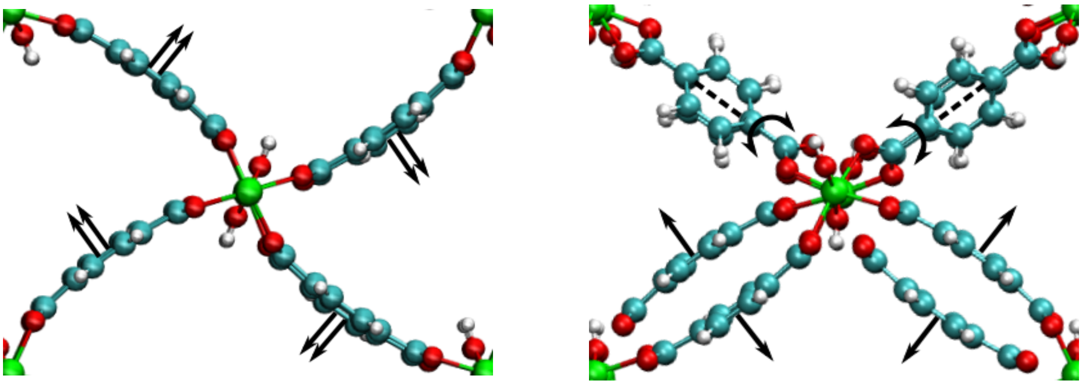

(a) The symmetric linker twist. (b) The antisymmetric linker twist.

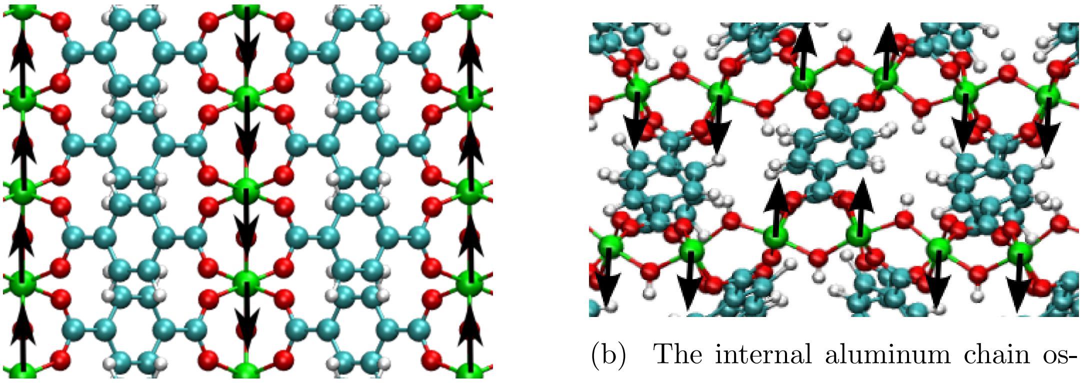
Figure 4.10: Rotation of linkers about the Al nodes.

direction of the chain (see Figure 4.11a). Because of this motion, linkers rotate both in their plane and about their axis. No volume change nor big cell shape changes are associated with this motion.

- **The aluminum chain parallel shift with extra linker rotation**: Mode 8 in the lp simulation, corresponding with mode 11 in extended and mode 9 in the np simulation, is a variant of the aluminum chain parallel shift described above. In this variant, more motion is associated with the linker rotation around the linker axis.

- **The internal aluminum chain oscillation**: Mode 9 in the lp simulation, corresponding with modes 7 and 8 in the extended simulation and modes 9 and 11 in the np simulation, is one where the aluminum nuclei move with respect to each other inside the oxide chain. As shown in Figure 4.11b, this motion happens in pairs of Al atoms. Linker rotations are also present, mostly around the linker axis. Some very big cell shape changes are associated with this motion, in the ($a_y = b_x$,$-b_z = -c_y$) direction. No volume change is observed.

- **The cell expansion**: Mode 10 in the lp simulation, corresponding with mode 3 in extended and partially with modes 1, 4 and 6 in the np simulation, is the expansion of the unit cell (see Figure 4.9a). The length of the aluminum chain remains unaltered, but the framework is stretched in the $\vec{a}$ and $\vec{c}$ direction. This results in a big volume change, and some cell shape change. The atoms do not have internal motion, since their fractional coordinates remain fixed. However, it is not the winerack motion because both the a_x and c_z component of the cell move in the same direction.



(a) The aluminum chain parallel shift.

(b) The internal aluminum chain oscillation.

Figure 4.11: Motion of the aluminum chains.

- **The aluminum chain perpendicular shift**: Mode 5 in the np simulation, corresponding mostly with mode 14 in the extended simulation and with mode 11 in the lp simulation, is another shift of the aluminum chain, this time perpendicular to the direction of the chain, in the direction of the unit cell vector $\vec{c}$ (see Figure 4.12). It is associated with a small volume change, a small change in cell shape, but mostly with a large internal nuclear motion.

- **The aluminum chain perpendicular shift 2**: Mode 7 in the np simulation, corresponding mostly with modes 7 and 8 in the extended and with mode 11 in the lp simulation, is a variant of the aluminum chain perpendicular shift described above, in the direction of the unit cell vectors $\vec{a} \perp \vec{c}$. The cell does not change shape nor volume, so all nuclear motions are independent of the cell.

- **The aluminum chain parallel shift with extra linker rotation 2**: Mode 8 in the np simulation, corresponding with modes 7, 8 and 9 in the extended simulation and corresponding very well with mode 12 in the lp simulation, is a variant of the aluminum chain parallel shift. In this variant, more motion is associated with linker rotation around the linker axis. During this motion, the cell remains unchanged.

An overview the correspondence of the eigenmodes and the different motions described in the previous discussion is given in Table 4.2.

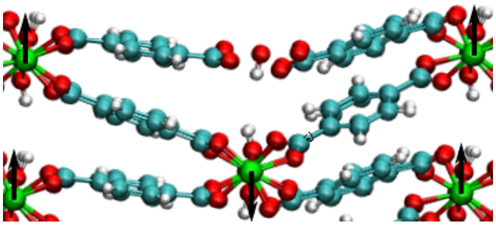


Figure 4.12: The aluminum chain perpendicular shift

Table 4.2: Summary of the motions associated with the 20 largest eigenmodes

l	**lp umbrella**	**np umbrella**	**extended umbrella**
1	lp-to-np transition	$(b_x, 2b_z)$ cell skewing	lp-to-np transition
2	linker rotations	lp-to-np transition	linker rotations
3	$(b_x, 2b_z)$ cell skewing	symm. linker twist	cell expansion $(b_x, 2b_z)$ cell skewing
4	symm. linker twist	asymm. linker twist 2 cell expansion	$(b_x, 2b_z)$ cell skewing asymm. linker twist 2
5	asymm. linker twist	Al ch. ⊥ shift cell expansion	asymm. linker twist 2
6	asymm. linker twist 2	cell expansion	symm. linker twist
7	Al ch. || shift	Al ch. ⊥ shift 2	int. Al ch. oscillation Al ch. ⊥ shift 2
8	Al ch. || shift + linker rot.	Al ch. || shift + linker rot. 2	int. Al ch. oscillation Al ch. ⊥ shift 2
9	int. Al ch. oscillation	Al ch. || shift + linker rot. int. Al ch. oscillation	Al ch. || shift + linker rot. int. Al ch. oscillation
10	cell expansion	asymm. linker twist asymm. linker twist 2	asymm. linker twist

4.4 The free energy contribution of unit cell fluctuations in the harmonic approximation

In this section, the contribution of the unit cell fluctuations to the free energy of the system in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble is calculated in the harmonic approximation, because this provides us with an explicit analytical formula. The free energy of the system in this ensemble is defined as:

$$F = -k_B T \ln Q_{(r,p,\mathbf{h})} \tag{4.41}$$

This partition function $Q_{(r,p,\mathbf{h})}$, also defined in Equation 2.34, is an integral over the symmetric unit cell matrix $\mathbf{h}$ (six independent components) [72], containing the partition function $Q_{(r,p)} \equiv Q_{(r,p)}(\mathbf{h})$ from Equation 2.8:

$$Q_{(r,p,\mathbf{h})} = \int \exp\big[-\beta P_{\text{ext}} \det(\mathbf{h})\big] Q_{(r,p)}(\mathbf{h}) \frac{1}{\det(\mathbf{h})^2} d\mathbf{h} \tag{4.42}$$

The partition function $Q_{(r,p)}(\mathbf{h})$ can be written as a function of the free energy surface $F(\mathbf{h})$ in the $NV\mathbf{h}_0T$ ensemble (see Equation 2.28):

$$Q_{(r,p)}(\mathbf{h}) = e^{-\beta F(\mathbf{h})} \tag{4.43}$$

This free energy surface is now harmonically expanded around a local equilibrium of the unit cell matrix, $\mathbf{h}_e$ (with volume V_e). This expansion will lead to the analytical evaluation of 4.42. For readability, the notation is switched to the Voigt convention (see Equation 4.7):

$$\mathbf{h}_e \rightarrow \vec{\mathrm{h}}_e \tag{4.44}$$

$$\mathbf{h} \rightarrow \vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}} \tag{4.45}$$

Firstly, the local harmonical approximation of the free energy surface $F(\vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}})$ around the local minimum is constructed in resemblance to techniques that perform normal mode analysis by calculating the Hessian matrix. The expansion for $F(\vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}})$ shows the stiffness tensor $\mathbf{C}_{\text{stiffness}}$ in $\vec{\mathrm{h}}_e$ (see Section D.2),

$$F(\vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}}) = \frac{1}{2} V \vec{\epsilon}^{\,T} \mathbf{C}_{\text{stiffness}} \vec{\epsilon} = \frac{1}{2} V_e \Delta\vec{\mathrm{h}}^T \mathbf{L}^T \mathbf{C}_{\text{stiffness}} \mathbf{L} \Delta\vec{\mathrm{h}} \tag{4.46}$$

with $\mathbf{L}$ a 6×6 matrix depending only on $\mathbf{h}_e$ and calculated in Equation D.34.

Secondly, the terms with $\det(\mathbf{h})$ in the integrand of the partition function $Q_{(r,p,\mathbf{h})}$ in Equation 4.42 can also be expanded in a Taylor series to second order in $\Delta\vec{\mathrm{h}}$. This allows to evaluate the integral in Equation 4.42 based on well-known Gaussian-related integrals, as proven in Appendix D. The result is an analytical expression for $Q_{(r,p,\mathbf{h})}$ which depends only on the temperature T, the external pressure P_{ext}, the unit cell matrix in equilibrium $\mathbf{h}_e$, and the stiffness tensor $\mathbf{C}_{\text{stiffness}}$ in this equilibrium state. The derived formula is given in Equation D.55 in the Appendix.

Illustrative example

Because of the limited timeframe, the calculations are performed on the np and lp phase of a previously studied system, MIL-47(V). The equilibrium unit cell matrix $\mathbf{h}_e$ and stiffness tensor $\mathbf{C}_{\text{stiffness}}$ were calculated quantummechanically at a temperature of 0 K in Ref. [91]. With these data we obtain a value of 6.3 kJ/mol for the free energy difference due to unit cell fluctuations at a temperature of 300 K. This means that in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, the fluctuations of the unit cell around the lp phase have a larger free energy than those around the np phase. The true free energy difference between the lp and np phase in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble is obtained by also incorporating the free energy difference due to nuclear vibrations at 300 K, which amounts to -13.3 kJ/mol, and the difference in ground state energy of -1.3 kJ/mol, both reported in Ref. [91]. As a result, the free energy difference between the lp and np phase at 300 K and zero pressure in the $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble is estimated to be -8.3 kJ/mol. In conclusion, the lp phase of MIL-47(V) is the stable phase under these conditions.

Outlook

There is a relation between the components of the covariance matrix $\mathbf{C}$ introduced in Equation 4.10 and the free energy surface $F(\mathbf{h})$. On the one hand, if the free energy surface is harmonic, the probability of visiting a state $\mathbf{h}$ follows a perfect multivariate Gaussian distribution: $p(\vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}}) \propto e^{-\beta F(\vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}})}$ (see Equation 2.37). On the other hand, in PCA, the construction of the covariance matrix from an $NP(\boldsymbol{\sigma}_a = \mathbf{0})T$ simulation is done by fitting a multivariate Gaussian distribution $e^{-1/2\Delta\vec{q}^T\mathbf{C}^{-1}\Delta\vec{q}}$ to the data. This suggests that the stiffness tensor $\mathbf{C}_{\text{stiffness}}$ in Equation 4.46 could be replaced by the inverse of a 6×6 covariance matrix of the unit cell components $\vec{h}$. It will be interesting to test this approach.

Further, the true strength of having an analytical expression for calculating the partition function, is that it can be chosen to not integrate the expression over certain unit cell components. In principle, this would result in a continuous surface of the free energy contribution of the unit cell fluctuations, as a function of these unit cell components. It should then be possible to estimate the free energy contribution of individual unit cell components. However, the mathematics become tedious, and it is not sure how these calculations must be interpreted. Also, it is not yet clear how to construct the covariance matrix or stiffness tensor as a function of the unit cell coordinates that are left out of the integration of the partition function. Additional research is required.

Chapter 5

Conclusions

More efficient simulation and analysis methods of molecular systems save computational scientists very costly simulation time. This work is a thorough investigation of the complete and efficient sampling of the unit cell shape space, for the construction of an accurate and precise free energy profile describing the breathing behavior of the flexible MIL-53(Al) crystal. Explicit sampling of the degrees of freedom related to the unit cell shape yields a better understanding of their influence on the calculated free energy profile. Finally, the most suitable methods for particular applications are proposed.

The need for sampling the unit cell shape is demonstrated by constructing free energy profiles as a function of the volume. The free energy is calculated by thermodynamic integration of the average pressure, as determined by a set of simulations at well-chosen, fixed volumes. This method is repeated twice, once for the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble, and once for the $NV\mathbf{h}_0T$ ensemble. In the first ensemble, the unit cell shape varies freely while the anisotropic contribution to the stress is on average zero. In the second ensemble, the unit cell shape space is not sampled, because it is fixed at a given $\mathbf{h}_0$. It is observed that not sampling the unit cell shape fluctuations results in an underestimation of the free energy in the large-pore phase of around 20 kJ/mol, demonstrating the great need for a correct sampling of the unit cell shape.

The existence of two stable phases of MIL-53(Al) is demonstrated, corresponding with a narrow-pore phase at low volume (ca. 807 Å^3) and a large-pore phase at high volume (ca. 1467 Å^3). In addition, a bistability of the unit cell shape in the $a_y = b_x$ direction is detected. At intermediate volumes (starting at around 1080 Å^3) the unit cell undergoes a transformation between a positive $a_y = b_x$ branch to a negative $a_y = b_x$ branch, labelled the "$a_y = b_x$ transition". By performing an extended umbrella simulation, it is seen that at low volumes, these two branches exist at $a_y = b_x = \pm 1$ Å, and merge together to $a_y = b_x = 0$ Å at volumes near the lp volume. The lack of transitions at these low volumes leads to the conclusion that a high free energy barrier between both branches exists at these low volumes, which disappears at large volumes. However, the extended umbrella simulation shows a different relative stability of the branches. After additional research, it is concluded that this is a statistical fluctuation, rather than an inherent physical phenomenon.

More in particular, the additional research consisted of 1D metadynamics simulations both in the volume as in the $a_y = b_x$ direction at certain well-chosen fixed volumes are performed. The volume metadynamics simulations show complete correspondence with the free energy profile obtained by thermodynamic integration of the average pressure in the $NV(\boldsymbol{\sigma}_a = \mathbf{0})T$ ensemble. From the set of $a_y = b_x$ metadynamics simulations it is concluded that there is insufficient evidence to assume that the $a_y = b_x$ branches are inequivalent. This method also samples the free energy barrier between the branches, and a clear quantification of the height and structure of this barrier is given. The free energy profiles of the 1D metadynamics simulation are compared to the data from the extended umbrella simulation, and both methods render equivalent results in the $a_y = b_x$ wells of the free energy profiles (this is the region sampled by the extended metadynamics simulation), while 1D metadynamics has a much larger overall precision.

The last advanced sampling method used in this work is 2D metadynamics, with the volume V and the unit cell component $a_y = b_x$ as the collective variables, rendering a 2D free energy profile in the $(V, a_y = b_x)$ coordinate set. A 1D free energy profile as a function of the volume is constructed by integration of the previous 2D profile with respect to the $a_y = b_x$ direction. The 2D free energy profile shows the $a_y = b_x$ branch structure, and in addition, it quantifies the free energy barrier between the two branches at every volume. It also corresponds with the 2D profile obtained from the extended umbrella simulation, although this last is inprecise and does not contain information about the free energy barriers between the branches. Hence, from the set of methods tested in this work, 2D metadynamics is most suited for the construction of 2D free energy profiles. The 1D free energy profile as a function of the unit cell volume, stemming from this 2D profile, is compared with the other free energy profiles obtained by performing thermodynamic integration or 1D metadynamics. It is seen that the np minimum of the 2D metadynamics simulation is lowered by $k_B T \ln 2 = 1.73$ kJ/mol with respect to the lp minimum. It is concluded that 2D metadynamics is the only method that samples both branches in the np phase, and that this results in the difference of free energy. Hence, 2D metadynamics renders the correct 1D free energy profile in the unit cell volume. The free energy profiles of the other methods that fail to sample both branches at the np (thermodynamic integration, 1D metadynamics), need to be corrected at volumes around the np phase by a term $-k_BT\ln 2$. Correcting the free energy profile at intermediate states between the np and lp is not straightforward, and additional analysis is required.

In this work, because of the detection of the $a_y = b_x$ transition, the sampling of the unit cell shape was focussed on the $a_y = b_x$ direction. Although not detected in the simulations, the existence of additional degeneracies in unit cell shape space is not excluded. Future work can explicitly scan the other unit cell shape directions to give a decisive answer regarding the existence of additional degeneracies. It is expected this scanning can be done by performing additional (2D) metadynamics simulations in unit cell shape space, especially in the other two off-diagonal components of the unit cell, $a_z = c_x$ and $b_z = c_y$.

Finally, an appropriate qualitative description of the collective fluctuations observed for MIL-53(Al) is obtained. These fluctuations are characterized by their contribution to cell volume, cell shape, nucleic and internal nucleic motion. A corresponding quantitative analysis is provided in Appendix C. The coupling between the changes in the unit cell volume and shape on the one hand, and the nucleic motion on the other hand is described. It is concluded that a large set of collective nucleic motion is associated with a change in the unit cell shape. While some unit cell shape changes contain no internal nuclear motion, reflected in the fact that the fractional coordinates of nuclei remain fixed during the motion, most unit cell shape changes correlate with nuclear motion. Hence, to correctly sample the collective fluctuations of the nuclei, there is a great need to include the cell fluctuations in MD simulations. Again, the $NV\mathbf{h}_0T$ ensemble, in which the unit cell is kept fixed, is proven to be insufficient to simulate the collective nuclear fluctuations present in the system. One of the important results of the discussion of the modes is that the lp-to-np transition does not contain any kneecap motion in the lp phase, in contrast to the np phase. Also, linker rotation around the linker axis in the np state is very suppressed in comparison to in the lp state.

In conclusion, to construct a correct free energy profile as a function of the volume for MIL-53(Al), it is necessary to sample the unit cell shape. A degeneracy is found in the $a_y = b_x$ direction of the symmetric unit cell in the np phase of the system, undetected by methods that only sample local fluctuations in the unit cell shape space, such as thermodynamic integration and 1D metadynamics in the volume direction. Of the methods described in this work, 2D metadynamics is the only method that can explicitly sample the unit cell shape space. This has an impact on the construction of the correct free energy profiles of similar systems. It should be checked if these systems show a similar branched structure, and the method of construction of the free energy profile should be adapted. The free energy differences between certain discrete states (the np and lp minima and the top of the np-to-lp transition barrier) obtained by thermodynamic integration or 1D metadynamics in the volume direction, must be corrected by an additive factor $k_B T \ln(2)$, resulting from lowering the np minimum.

Appendix A

Introduction to parametric statistical tests

Parametric versus nonparametric statistics

In parametric statistics, it is assumed that the data (observations or measurements) are drawn from a population that follows a given probability distribution (based on a fixed set of parameters). This contrasts with nonparametric statistics, where no assumptions about the underlying probability distribution of the population are made. On the one handn parametric statistics are more accurate than nonparametric statistics, given the assumptions on the underlying distribution are correct. On the other hand, a nonparametric test is more robust, i.e. they have a smaller chance of failing. In this work, because of their simplicity and ease of implementation, parametric statistics are used.

Statistical tests

A statistical test generally tries to reject a null hypothesis. This is a hypothesis about the underlying mechanism that generates an observation or measurement (e.g. "This measurement is drawn from a distribution which has a mean of zero"). The assumption is one about the distribution of the measurements under the null hypothesis, the null distribution. In this example this is the distribution centered about zero.

The test statistic

In the case that the observation or measurement can not be calculated as a number, a test statistic is constructed. This test statistic is a relation between an observation or a measurement and a number. The precise nature of the test statistic is chosen freely, but ideally summarizes the characteristic described in the null hypothesis well, and has a large deviation from the median (in comparison to the distribution of the test statistic, given the null hypothesis and assumptions on the distribution) when the null hypothesis does not hold.

The test

The test is carried out as follows: first, under the null hypothesis and a given distribution, a large set of "fake" test data are generated. For each observation of the test data, the test statistic can be calculated. This test statistic will have some variation, and the complete distribution of this test statistic can be obtained (the test distribution).

The p-value and its interpretation

For each value of the test statistic, the p-value (the probability of observing a value as extreme as this one under the null hypothesis) can be calculated. The p-value is calculated by summing over the number of the less extreme observations, divided by the complete number of observations. Subsequently, the test statistic for the true observation is calculated, and its location in the test distribution is located. The p-value of this observation is the probability that this observation arises given the fact that the null hypothesis is true. If the p-value is low, the null hypothesis can be rejected, keeping in mind that there is always a probability that the null hypothesis generates an observation as extreme as the one observed. In other words, the p-value gives the probability of a false rejection of the null hypothesis.

Hartigan's dip test

In Hartigan's dip test of unimodality [79], the unimodality of an empirical distribution is tested. The null hypothesis states that the distribution is unimodal, and the test statistic is chosen to be the dip statistic. This statistic is the maximum difference between the empirical distribution and the unimodal distribution that minimizes this maximum difference. The appropriate null distribution is uniform, and the dip is asymptotically larger for the uniform than for any distribution in a wide class of unimodal distributions. The distribution of the test statistic is determined by sampling from the uniform distribution.

Appendix B

Correlations between unit cell components

The flexibility observed in MIL-53(Al) is caused by its overall framework structure, and not by independent variation of the bond/linker lengths. This results in large variations of unit cell components not being independent from each other. The relations between these components are summarized in this chapter.

The dependencies between the unit cell components are shown for the complete range of volumes known in MIL-53(Al),i.e. the range between the np and lp structure. The extended umbrella simulation introduced in Section 3.2.2 is able to sample this complete range. The work here is an extra analysis of the data that were produced during that simulation, by constructing histograms for different unit cell components. Note that these histograms are obtained from a biased simulation, for which the bias is the negative free energy profile $-F(V)$.

In total, seven properties related to the unit cell are under investigation: the volume V, and the six symmetric unit cell components a_x, $a_y = b_x$, $a_z = c_x$, b_y, $b_z = c_y$ and c_z. This means that there are a total of $6 + 5 + 4 + 3 + 2 + 1 = 21$ unique 2D histograms to be constructed.

The volume V

The first component of interest is the volume V. It is straightforward to measure this unit cell component which describes the lp-to-np transition experimentally. The dependency of the complete set of symmetric unit cell components on the volume is illustrated in Figure B.1. The component c_z (Subfigure B.1f) undergoes the largest change when the structure transforms from the np to the lp phase, it almost triples in size. The fluctuation of this component is small in the np phase, and increases towards the lp phase. The same holds for the fluctuation of component a_x, which drops in size when the structure transitions to the lp (Subfigure B.1a). The $a_y = b_x$ component shows the bistable behavior investigated at length in Section 3.2 (Subfigure B.1b). The component b_y undergoes no changes until it reaches the lp structure, where it increases by about 0.5 Å (Subfigure B.1d). The components $a_z = c_x$ and $b_z = c_y$ remain constant during the lp-to-np transition, only their fluctuation shows a slight drop in the lp phase (Subfigure B.1c and B.1e). This analysis of the histograms of the cell components and the volume is confirmed by the PCA performed on the extended umbrella simulation, as seen in the first eigenmode in Table C.3.

The component a_x

In Figure B.2, the 2D histograms of the component a_x and the five remaining components are shown. A small a_x corresponds to the lp state, a large a_x to the np state of MIL-53(Al). Again, the bistable behavior is observed, see Subfigure B.2a. In Subfigures B.2b-d, the components remained constant, with an increase in fluctuations at large a_x. The c_z component in Subfigure B.2e decreases sharply with a_x.

The component $a_y = b_x$

The 2D histograms $a_y = b_x$ and the four remaining components are shown in Figure B.3. The $a_y = b_x$ contains two branches, which are ±1 Å in the np phase and 0 Å in the lp phase. During the extended umbrella simulation, the negative branch was more frequented than the positive branch. Both the $a_z = c_x$ (Subfigure B.3a) and the $b_z = c_y$ (Subfigure B.3c) component show no dependency on $a_y = b_z$. The small increase in b_y at $a_y = b_x = 0$ Å (Subfigure B.3b) corresponds to the lp phase. The branch structure in the c_z coordinate is visible in Subfigure B.3d.

The components $a_z = c_x$, b_y and $b_z = c_y$

In Figure B.4 the 2D histograms of the $a_z = c_x$ component and the three remaining components are shown. The components b_y, $b_z = c_y$ and c_z show no dependency on $a_z = c_x$. In Figure B.5 the 2D histograms of respectively b_y and $b_z = c_y$ (a), b_y and c_z (b), $b_z = c_y$ and c_z (c) are shown. The small increase of b_y at large c_z corresponds to the lp phase.

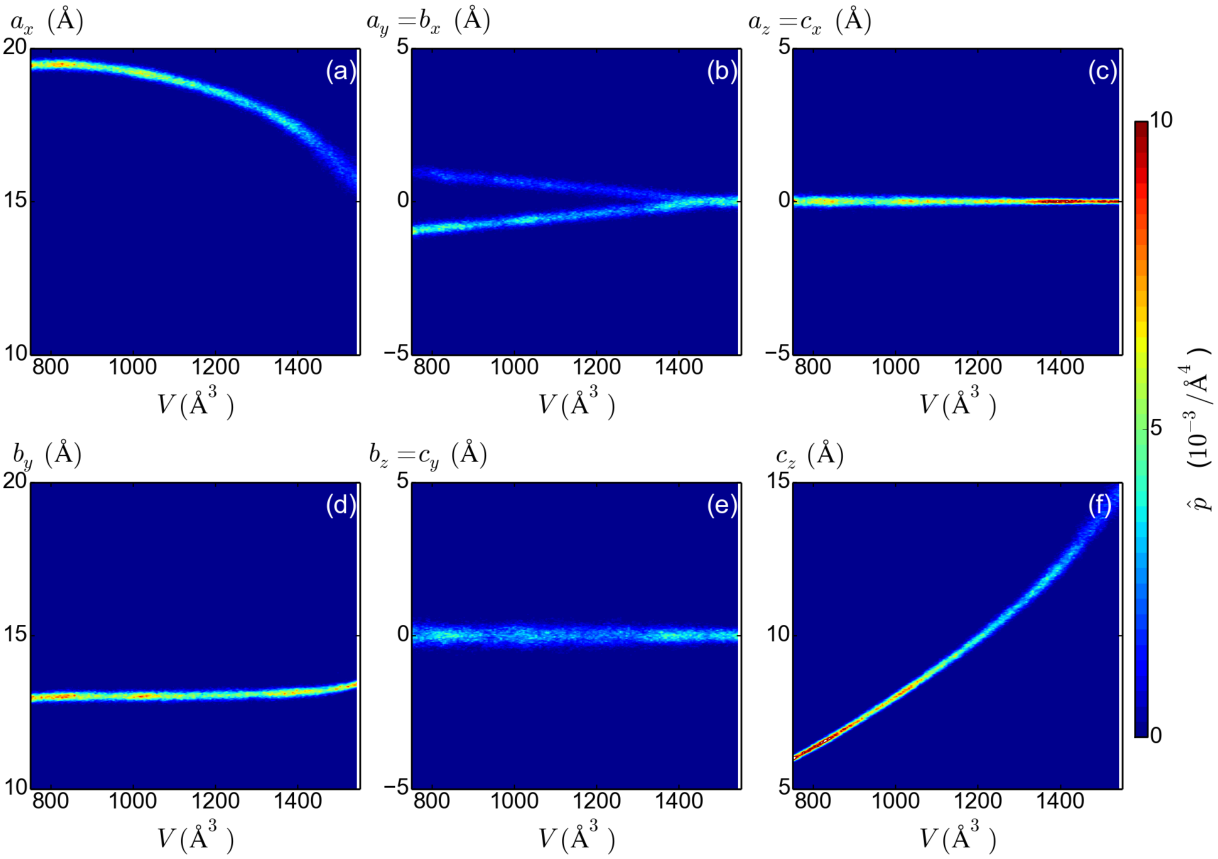


Figure B.1: Set of 2D histograms of the volume V and the symmetric unit cell components. The estimated probability density is $\hat{p} = h/NS$, with h the 2D histogram of the volume V and a unit cell component, N the total number of observations and S the surface of the histogram bin (here Å^4).

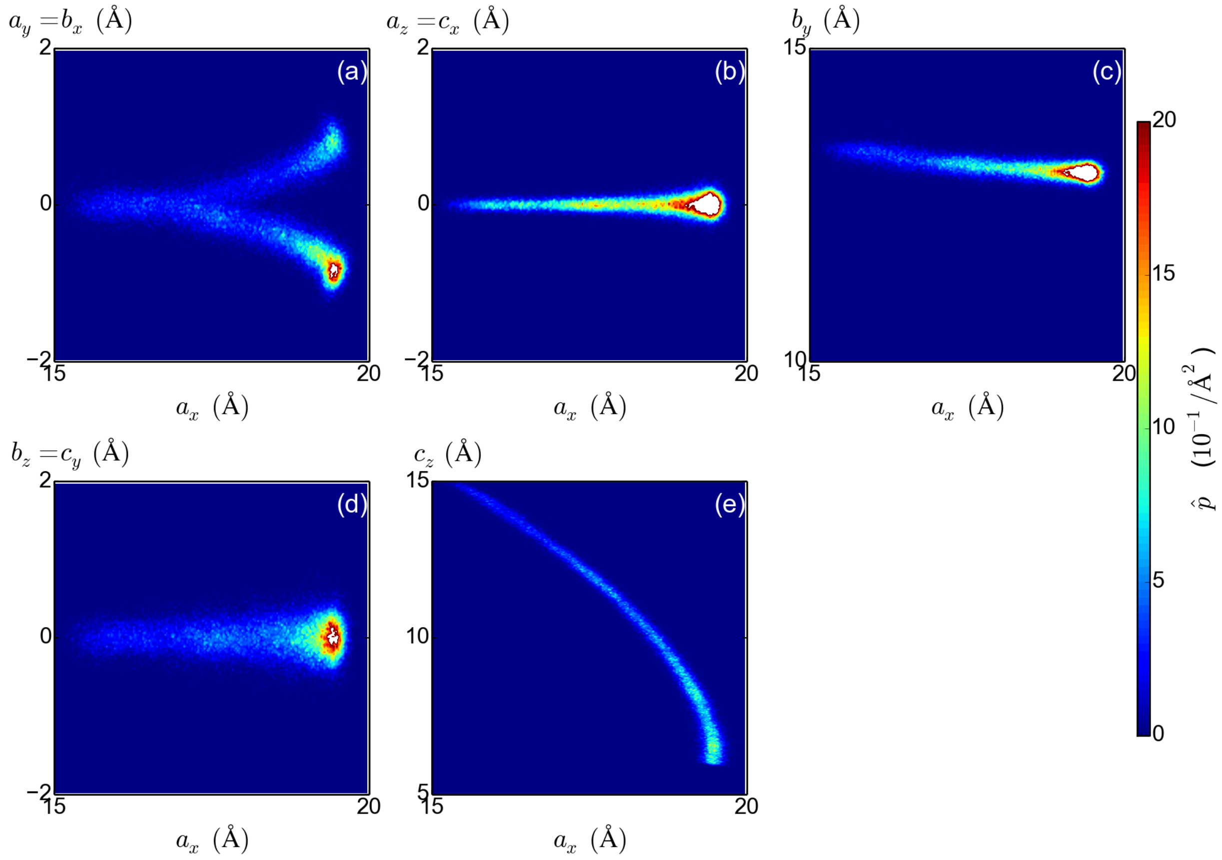


Figure B.2: 2D histograms of the symmetric unit cell component a_x and the five remaining symmetric unit cell components. The estimated probability density is $\hat{p} = h/NS$, with h the 2D histogram of two unit cell components, N the total number of observations and S the surface of the histogram bin (here Å^2). White regions in the figure denote probability densities higher than the range of the colorbar.

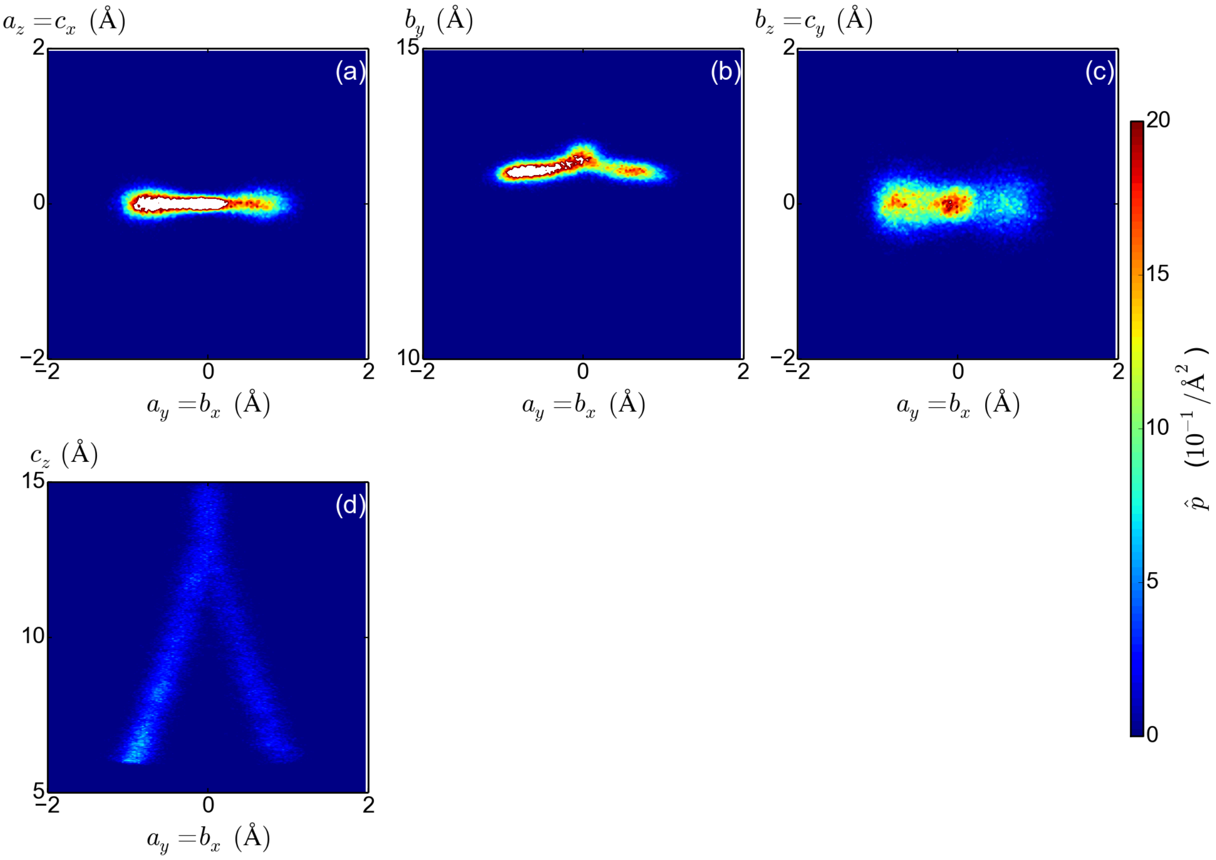


Figure B.3: Set of 2D histograms of the symmetric unit cell component $a_y = b_x$ and the four remaining symmetric unit cell components. The estimated probability density is $\hat{p} = h/NS$, with h the 2D histogram of two unit cell components, N the total number of observations and S the surface of the histogram bin (here Å^2). White regions in the figure denote probability densities higher than the range of the colorbar.

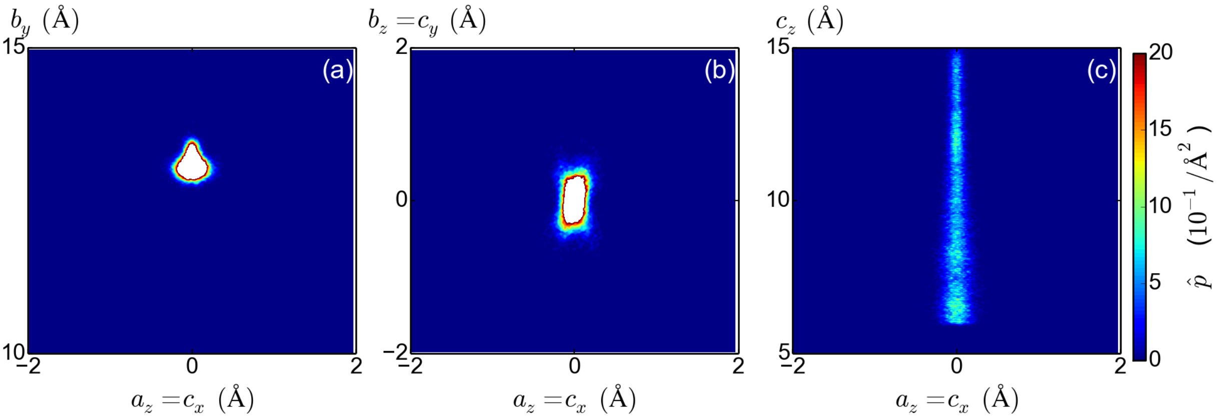


Figure B.4: Set of 2D histograms of the symmetric unit cell component $a_z = c_x$ and the three remaining symmetric unit cell components. The estimated probability density is $\hat{p} = h/NS$, with h the 2D histogram of two unit cell components, N the total number of observations and S the surface of the histogram bin (here Å^2). White regions in the figure denote probability densities higher than the range of the colorbar.

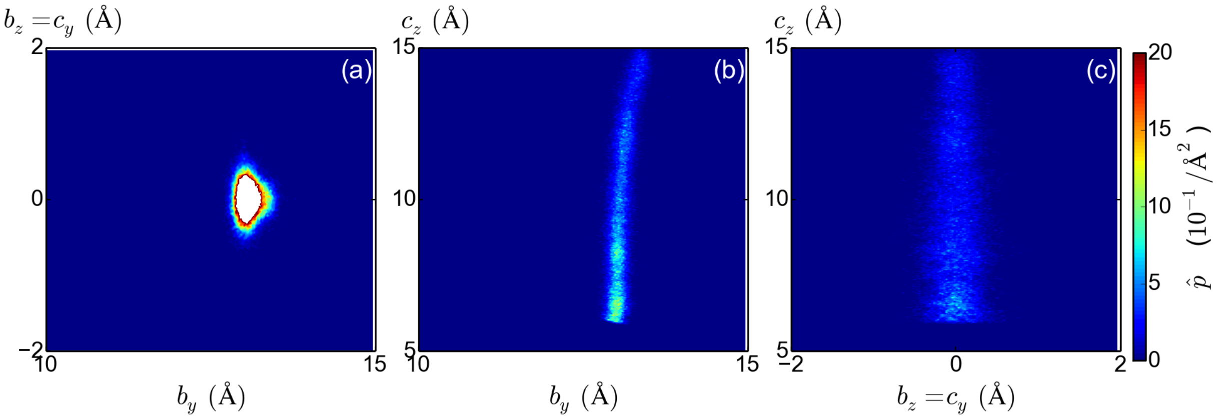


Figure B.5: (a-b): Set of 2D histograms of the symmetric unit cell component b_y and the two remaining symmetric unit cell components. (c): 2D histogram of the symmetric unit cell component $b_z = c_y$ and the last remaining symmetric unit cell components c_z. The estimated probability density is $\hat{p} = h/NS$, with h the 2D histogram of two unit cell components, N the total number of observations and S the surface of the histogram bin (here Å^2). White regions in the figure denote probability densities higher than the range of the colorbar.

Appendix C

Quantitative description of the eigenvalues and unit cell components of the eigenmodes

In addition to the PCA performed in chapter 4, the eigenvalues and cell component sizes of the largest 20 eigenmodes are given in this appendix. Note that these values are obtained by analysis of the fluctuations in simulations of a double unit cell of MIL-53(Al), expanded in the $\vec{b}$ direction. Tables 1-3 show the modes from the lp, np and extended umbrella simulation respectively. The eigenvalues $\lambda^{(l)}$ have the dimension Å^2, while the sizes of the eigenvector components are dimensionless, since the eigenvectors are normalized.

Table C.1: Eigenmodes identified in the umbrella lp simulation

l	$\lambda^{(l)}$ (Å^2)	a_x	b_y	c_z	$2b_z = 2c_y$	$2c_x = 2a_z$	$2a_y = 2b_x$
1	31.39	0.083	-0.002	-0.113	0.001	-0.000	0.003
2	10.66	-0.002	-0.001	0.003	0.003	0.000	0.046
3	1.94	-0.006	-0.001	0.005	0.158	-0.001	0.097
4	1.53	0.002	0.001	0.000	-0.031	0.001	0.002
5	0.99	-0.002	-0.002	0.003	0.045	-0.000	-0.038
6	0.90	0.001	-0.002	-0.005	-0.078	-0.004	0.110
7	0.80	-0.002	-0.001	-0.003	-0.017	-0.003	0.063
8	0.75	0.000	0.001	0.002	0.025	0.002	-0.071
9	0.59	-0.005	-0.008	0.009	0.152	0.006	-0.141
10	0.48	0.065	-0.004	0.040	0.001	0.096	0.013
11	0.43	-0.010	0.002	-0.005	0.014	-0.021	0.002
12	0.43	0.003	0.003	-0.002	-0.034	-0.002	0.022
13	0.39	-0.008	-0.001	0.003	0.051	-0.008	-0.093
14	0.31	0.006	-0.001	-0.013	-0.014	-0.002	0.005
15	0.28	-0.005	0.003	0.011	-0.015	0.002	0.029
16	0.27	0.001	-0.012	-0.004	0.055	-0.001	-0.047
17	0.26	0.003	0.010	-0.002	-0.044	0.002	0.035
18	0.23	0.015	0.002	-0.021	0.001	-0.004	0.006
19	0.22	-0.010	0.028	0.014	-0.047	0.000	0.022
20	0.21	0.008	-0.019	0.000	-0.007	-0.006	0.014

Table C.2: Eigenmodes identified in the umbrella np simulation

l	$\lambda^{(l)}$ (Å^2)	a_x	b_y	c_z	$2b_z = 2c_y$	$2c_x = 2a_z$	$2a_y = 2b_x$
1	4.75	0.004	0.001	0.012	-0.183	0.080	-0.025
2	1.95	-0.035	-0.022	0.115	0.029	0.014	0.065
3	1.63	0.003	0.001	-0.018	-0.006	0.007	0.012
4	0.83	-0.051	-0.007	-0.024	-0.144	-0.038	0.165
5	0.69	-0.022	0.006	0.035	-0.013	-0.072	-0.054
6	0.66	0.023	-0.013	-0.053	-0.026	0.112	0.130
7	0.65	-0.005	0.001	0.006	-0.013	-0.010	-0.011
8	0.46	0.003	0.001	0.005	0.022	-0.005	-0.018
9	0.45	-0.014	0.003	-0.006	-0.070	-0.003	0.018
10	0.42	0.005	-0.001	0.003	0.020	-0.004	-0.013
11	0.37	-0.016	-0.007	-0.023	-0.067	0.033	0.093
12	0.23	0.101	-0.083	0.041	-0.031	-0.119	0.007
13	0.18	-0.001	0.006	0.002	0.001	0.004	0.002
14	0.17	0.056	0.060	0.032	-0.017	-0.062	0.057
15	0.17	-0.015	-0.024	-0.014	0.011	0.020	-0.031
16	0.15	0.008	-0.005	0.007	-0.008	-0.018	0.015
17	0.15	-0.003	-0.013	-0.019	0.004	0.009	-0.016
18	0.14	0.005	0.020	0.019	0.008	-0.003	-0.001
19	0.14	-0.002	0.001	0.021	0.002	0.009	0.002
20	0.13	0.010	-0.004	0.015	0.016	-0.026	-0.009

Table C.3: Eigenmodes identified in the extended umbrella simulation

l	$\lambda^{(l)}$ (Å^2)	a_x	b_y	c_z	$2b_z = 2c_y$	$2c_x = 2a_z$	$2a_y = 2b_x$
1	305.38	0.051	-0.003	-0.129	0.000	-0.000	-0.023
2	26.10	-0.007	-0.000	0.013	-0.010	-0.001	-0.177
3	3.93	-0.089	0.008	-0.028	-0.095	-0.035	-0.022
4	2.55	0.077	-0.005	0.024	-0.194	-0.010	0.054
5	1.39	-0.010	-0.001	-0.001	-0.087	-0.055	0.154
6	1.29	0.012	-0.002	0.005	-0.004	-0.012	0.031
7	1.01	-0.008	-0.001	-0.004	0.065	0.021	-0.052
8	0.89	0.008	0.001	0.005	-0.044	-0.021	0.039
9	0.61	0.003	0.002	-0.000	-0.065	-0.009	0.023
10	0.53	-0.019	0.002	-0.014	0.032	0.043	-0.038
11	0.43	0.033	-0.009	0.026	0.006	-0.079	0.005
12	0.41	-0.068	0.010	-0.061	-0.058	0.161	0.063
13	0.40	0.041	-0.009	0.036	0.037	-0.096	-0.032
14	0.37	-0.019	0.004	-0.012	-0.030	0.039	0.028
15	0.23	0.000	0.008	0.001	-0.000	-0.003	0.021
16	0.21	-0.012	-0.132	-0.009	-0.026	0.017	0.020
17	0.20	0.005	0.013	-0.003	0.036	-0.005	0.003
18	0.20	-0.000	0.049	-0.011	-0.004	0.006	0.025
19	0.19	0.001	-0.008	0.008	-0.021	0.008	0.022
20	0.18	-0.001	0.006	0.002	0.009	-0.000	-0.007

Appendix D

Derivation of the analytical partition function

An expression for the partition function $Q_{\mathbf{h}}$ describing the ensemble with anisotropic unit cell fluctuations is proposed by Martyna, Tobias and Klein [72]:

$$Q_{\mathbf{h}} = \int d\mathbf{h} \exp\big[-\beta P_{\text{ext}} \det(\mathbf{h})\big] Q(\mathbf{h}) \det(\mathbf{h})^{1-d} \tag{D.1}$$

Where d is the dimension of the system (in this case $d = 3$). It is assumed that all states of the unit cell occur with an equal a priori probability, weighted by the Boltzmann factor $Q(\mathbf{h})$.

$$Q(\mathbf{h}) = \int\limits_{\Omega(\mathbf{h})} d\vec{r}^N d\vec{p}^N \exp\big[-\beta H(\vec{r}^N, \vec{p}^N)\big] \tag{D.2}$$

with all nuclear positions and velocities integrated in the phase space $\Omega(\mathbf{h})$, given a unit cell $\mathbf{h}$. As is introduced in Equation 2.2, the volume V of the system is defined as the determinant of the unit cell matrix $\mathbf{h}$. Here $\mathbf{h}$ is defined to be symmetric, and the same Voigt notation as in Equation 4.7 is used.

In this chapter an effort is made to expand this partition function into an analytical expression. This analytical expression can be helpful to derive properties of the unit cell. Certain assumptions/approximations are made, and some Gaussian integrals need to be calculated.

D.1 Derivation of the appropriate Gaussian integral

The following standard integral expression, with $\mathbf{x}$ an n-dimensional vector, and $\mathbf{A}$ a real symmetric $n \times n$-dimensional matrix, which does not depend on $\mathbf{x}$:

$$\int_{\mathbb{R}^n} \exp\Big(-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\Big) d\mathbf{x} = \sqrt{\frac{(2\pi)^n}{\det(\mathbf{A})}} \tag{D.3}$$

This expression is used to derive a more general integral I, with $\mathbf{B}$ an $n \times n$-matrix and $\mathbf{J}$ another n-dimensional vector (also independent of $\mathbf{x}$).

$$I = \int_{\mathbb{R}^n} \mathbf{x}^T\mathbf{B}\mathbf{x} \exp\Big(-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{J}^T\mathbf{x}\Big) d\mathbf{x} \tag{D.4}$$

Firstly, the real symmetric matrix $\mathbf{A}$ is diagonalized to $\mathbf{D}$, with unitary matrix $\mathbf{O}$. The coordinates $\mathbf{x}$ are transformed into coordinates $\mathbf{y}$:

$$\mathbf{D} = \mathbf{O}^T\mathbf{A}\mathbf{O} \tag{D.5}$$

$$\mathbf{A}^{-1} = \mathbf{O}\mathbf{D}^{-1}\mathbf{O}^T \tag{D.6}$$

$$\mathbf{x} = \mathbf{O}\mathbf{y} \tag{D.7}$$

The expression inside the exponential becomes:

$$\begin{aligned}
-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{J}^T\mathbf{x} &= -\frac{1}{2}\mathbf{y}^T\mathbf{O}^T\mathbf{A}\mathbf{O}\mathbf{y} + \mathbf{J}^T\mathbf{O}\mathbf{y} && \text{(D.8)}\\
&= -\frac{1}{2}(\mathbf{y}^T\mathbf{D}\mathbf{y} - 2\mathbf{J}^T\mathbf{O}\mathbf{y}) && \text{(D.9)}\\
&= -\frac{1}{2}\Big[\Big(\mathbf{y}^T - \mathbf{J}^T\mathbf{O}\mathbf{D}^{-1}\Big)\mathbf{D}\Big(\mathbf{y} - (\mathbf{J}^T\mathbf{O}\mathbf{D}^{-1})^T\Big)\\
&\quad -\mathbf{J}^T\mathbf{O}\mathbf{D}^{-1}\mathbf{D}(\mathbf{J}^T\mathbf{O}\mathbf{D}^{-1})^T\Big] && \text{(D.10)}\\
&= -\frac{1}{2}\Big(\mathbf{y}^T - \mathbf{J}^T\mathbf{O}\mathbf{D}^{-1}\Big)\mathbf{D}\Big(\mathbf{y} - \mathbf{D}^{-1}\mathbf{O}^T\mathbf{J}\Big)\\
&\quad +\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J} && \text{(D.11)}
\end{aligned}$$

Note that $\mathbf{D} = \mathbf{D}^T$ for the diagonal matrix $\mathbf{D}$ (and so also $\mathbf{D}^{-1} = \mathbf{D}^{-T}$). From Equation D.9 to D.10 it is used that $\mathbf{J}^T\mathbf{O}\mathbf{y} = (\mathbf{J}^T\mathbf{O}\mathbf{y})^T$ because this expression is a scalar. In the last step Equation D.6 was used. Transforming to another coordinate set

$$\mathbf{z} = \mathbf{y} - \mathbf{D}^{-1}\mathbf{O}^T\mathbf{J} \tag{D.12}$$

the term in the exponential becomes:

$$-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{J}^T\mathbf{x} \;=\; -\frac{1}{2}\mathbf{z}^T\mathbf{D}\mathbf{z}+\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J} \tag{D.13}$$

Now the integral I becomes, first transforming from coordinates $\mathbf{x}$ to $\mathbf{y}=\mathbf{O}^T\mathbf{x}$

$$I=\int\limits_{\mathbb{R}^n}\mathbf{y}^T\mathbf{O}^T\mathbf{B}\mathbf{O}\mathbf{y}\exp\Big(-\frac{1}{2}\mathbf{y}^T\mathbf{D}\mathbf{y}+\mathbf{J}^T\mathbf{O}\mathbf{y}\Big)d\mathbf{y} \tag{D.14}$$

note that $d\mathbf{x}=d\mathbf{y}$ because the coordinates are only rotated and the Jacobian of a rotation is 1. Switching to the coordinate $\mathbf{z}$:

$$\int\limits_{\mathbb{R}^n}(\mathbf{z}^T+\mathbf{J}^T\mathbf{O}\mathbf{D}^{-1})\mathbf{O}^T\mathbf{B}\mathbf{O}(\mathbf{z}+\mathbf{D}^{-1}\mathbf{O}^T\mathbf{J})\exp\Big(-\frac{1}{2}\mathbf{z}^T\mathbf{D}\mathbf{z}+\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J}\Big)d\mathbf{z} \tag{D.15}$$

Now it is important to note that only even powers of $\mathbf{z}$ contribute to the integral, because odd powers vanish, for example:

$$\begin{aligned}\int\limits_{\mathbb{R}^n} z_1\exp\Big(-\frac{1}{2}\mathbf{z}^T\mathbf{D}\mathbf{z}\Big)d\mathbf{z} &= \int\limits_{\mathbb{R}^n} z_1\prod_{i=1}^{n}\exp\Big(-\frac{1}{2}D_{ii}z_i^2\Big)dz_i && \text{(D.16)}\\ &= \int\limits_{-\infty}^{\infty} z_1\exp\Big(-\frac{1}{2}D_{11}z_1^2\Big)dz_1\cdot\int\limits_{\mathbb{R}^{n-1}}\prod_{i=2}^{n}\exp(-\frac{1}{2}D_{ii}z_i^2)dz_i\\ &= 0\cdot\prod_{i=2}^{n}\int\limits_{-\infty}^{\infty}\exp\Big(-\frac{1}{2}D_{ii}z_i^2\Big)dz_i && \text{(D.17)}\end{aligned}$$

The integral disappears because its integrand is odd and evaluated between symmetric limits. With this information, the integral with a quadratic prefactor in $\mathbf{z}$ can be solved. The only terms that contribute to the integral are the constant and the quadratic terms z_i^2, which are multiplied by the diagonal elements of $\mathbf{O}^T\mathbf{B}\mathbf{O}$ (because the odd terms z_iz_j

disappear):

$$\int_{\mathbb{R}^n} \mathbf{z}^T\mathbf{O}^T\mathbf{B}\mathbf{O}\mathbf{z} \quad \exp\Big(-\tfrac{1}{2}\mathbf{z}^T\mathbf{D}\mathbf{z}\Big) \quad d\mathbf{z}$$

$$= \int_{\mathbb{R}^n} \sum_{i=1}^{n} z_i(\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii}z_i \quad \prod_{j=1}^{n} \exp\Big(-\frac{1}{2}z_jD_{jj}z_j\Big)dz_j \tag{D.18}$$

$$= \sum_{i=1}^{n} \quad (\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii} \quad \int_{\mathbb{R}^n} z_i^2 \prod_{j=1}^{n} \exp\Big(-\frac{1}{2}z_jD_{jj}z_j\Big)dz_j \tag{D.19}$$

$$= \sum_{i=1}^{n} \quad (\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii} \quad \int_{-\infty}^{\infty} z_i^2 \exp\Big(-\frac{1}{2}D_{ii}z_i^2\Big)dz_i \cdot \prod_{j\neq i}^{n} \int_{-\infty}^{\infty} \exp\Big(-\frac{1}{2}z_jD_{jj}z_j\Big)dz_j \tag{D.20}$$

$$= \sum_{i=1}^{n} \quad (\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii} \quad \sqrt{\frac{2\pi}{D_{ii}}}\frac{1}{D_{ii}} \cdot \sqrt{\frac{(2\pi)^{n-1}}{\prod\limits_{j\neq i}^{n} D_{jj}}} \tag{D.21}$$

$$= \sum_{i=1}^{n} \quad \tfrac{1}{D_{ii}}(\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii} \quad \sqrt{\frac{(2\pi)^{n}}{\prod\limits_{j=1}^{n} D_{jj}}} \tag{D.22}$$

Going from D.20 to D.21 two types of very famous onedimensional standard integrals are used, namely $\int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2}dx = \sqrt{\frac{2\pi}{a}}$ and $\int_{-\infty}^{\infty} x^2e^{-\frac{1}{2}ax^2}dx = \sqrt{\frac{2\pi}{a}}\frac{1}{a}$. The expression of the general integral I becomes:

$$I = \exp\Big(\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J}\Big) \int_{\mathbb{R}^n} \Big[\mathbf{J}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{J} + \mathbf{z}^T\mathbf{O}^T\mathbf{B}\mathbf{O}\mathbf{z}\Big] \exp\Big(-\frac{1}{2}\mathbf{z}^T\mathbf{D}\mathbf{z}\Big)d\mathbf{z} \tag{D.23}$$

$$= \exp\Big(\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J}\Big) \sqrt{\frac{(2\pi)^{n}}{\prod\limits_{j=1}^{n} D_{jj}}} \Big[\mathbf{J}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{J} + \sum_{i=1}^{n} \frac{(\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii}}{D_{ii}}\Big] \tag{D.24}$$

Or because $\prod\limits_{i=1}^{n} D_{ii} = \det(\mathbf{D}) = \det(\mathbf{A})$ and $D_{ii} = \lambda_i$, with λ_i the ith eigenvalue of $\mathbf{A}$:

$$I = \int_{\mathbb{R}^n} \mathbf{x}^T\mathbf{B}\mathbf{x} \exp\Big(-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{J}^T\mathbf{x}\Big)d\mathbf{x}$$

$$= \exp\Big(\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J}\Big)\sqrt{\frac{(2\pi)^n}{\det\mathbf{A}}} \quad \Big[\mathbf{J}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{J} + \sum_{i=1}^{n} \tfrac{1}{\lambda_i}(\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii}\Big] \tag{D.25}$$

D.2 Expansion of the partition function around a local equilibrium

The partition function $Q_{\mathbf{h}}$ is expanded about a local energy minimum of the unit cell $\mathbf{h}_e$:

$$\mathbf{h} = \mathbf{h}_e + \Delta\mathbf{h} \equiv \vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}} \quad \text{(Voigt)} \tag{D.26}$$

Since $\mathbf{h}_e$ is a constant, it is also seen that $d\mathbf{h} \equiv d\Delta\vec{\mathrm{h}}$. As a quadratic approximation of the change in energy upon a disturbance of the system out of equilibrium, the Boltzmann factor $Q(\mathbf{h})$ can be written as a function of the strain energy density $U(\boldsymbol{\epsilon})$, defined in Equation D.27. This energy density is a function of the strain $\boldsymbol{\epsilon}$ ($\vec{\epsilon}$ in Voigt notation), containing the same information as the deformation of the unit cell $\Delta\vec{\mathrm{h}}$ (for small deformations).

$$U(\boldsymbol{\epsilon}) = \int \boldsymbol{\sigma} \cdot d\boldsymbol{\epsilon} = \frac{1}{2}\vec{\epsilon}^T \cdot \mathbf{C}_{\text{stiffness}} \cdot \vec{\epsilon} \tag{D.27}$$

With $\boldsymbol{\epsilon}$ the infinitesimal dimensionless strain, written in Voigt notation (a 6 × 1 matrix),

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ & \epsilon_{yy} & \epsilon_{yz} \\ & & \epsilon_{zz} \end{pmatrix} \Leftrightarrow \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix} = \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ 2\epsilon_{yz} \\ 2\epsilon_{xz} \\ 2\epsilon_{xy} \end{pmatrix} \tag{D.28}$$

and $\mathbf{C}_{\text{stiffness}}$ the stifness tensor. This tensor of rank 4 remains constant in the elastic region and its elements are called the elastic constants. These constants specify the stress in the material under a certain strain, in energy per volume or pressure (force per unit area). In Voigt notation ($xx \rightarrow 1$, $yy \rightarrow 2$, $zz \rightarrow 3$, $yz = zy \rightarrow 4$, $xz = zx \rightarrow 5$ and $xy = yx \rightarrow 6$), $\mathbf{C}_{\text{stiffness}}$ is written as a 6 × 6 symmetric matrix with 21 independent elements:

$$\mathbf{C}_{\text{stiffness}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ & & C_{33} & C_{34} & C_{35} & C_{36} \\ & & & C_{44} & C_{45} & C_{46} \\ & & & & C_{55} & C_{56} \\ & & & & & C_{66} \end{pmatrix} \tag{D.29}$$

Consider a deformation of the unit cell $\mathbf{h}$ under strain, resulting in a unit cell matrix $\mathbf{h}' = \mathbf{h} \cdot (\mathbf{1} + \boldsymbol{\epsilon})$. The matrix $\boldsymbol{\epsilon}$ can hence be written as a function of the unit cell deformation

$\Delta\mathbf{h}$:

$$\begin{aligned} \mathbf{h}' = \mathbf{h}\cdot(\mathbf{1}+\boldsymbol{\epsilon}) &= \mathbf{h}+\Delta\mathbf{h} && \text{(D.30)}\\ \mathbf{h}\cdot\boldsymbol{\epsilon} &= \Delta\mathbf{h} && \text{(D.31)}\\ \boldsymbol{\epsilon} &= \mathbf{h}^{-1}\cdot\Delta\mathbf{h} && \text{(D.32)} \end{aligned}$$

This formula is converted to Voigt notation:

$$\vec{\epsilon} = \mathbf{L}\cdot\Delta\vec{\mathrm{h}} \tag{D.33}$$

With

$$\mathbf{L} = \begin{pmatrix} (\mathbf{h}^{-1})_{11} & 0 & 0 & 0 & \frac{1}{2}(\mathbf{h}^{-1})_{13} & \frac{1}{2}(\mathbf{h}^{-1})_{12} \\ 0 & (\mathbf{h}^{-1})_{22} & 0 & \frac{1}{2}(\mathbf{h}^{-1})_{23} & 0 & \frac{1}{2}(\mathbf{h}^{-1})_{12} \\ 0 & 0 & (\mathbf{h}^{-1})_{33} & \frac{1}{2}(\mathbf{h}^{-1})_{23} & \frac{1}{2}(\mathbf{h}^{-1})_{13} & 0 \\ 0 & 0 & 2(\mathbf{h}^{-1})_{23} & (\mathbf{h}^{-1})_{22} & (\mathbf{h}^{-1})_{12} & 0 \\ 2(\mathbf{h}^{-1})_{13} & 0 & 0 & 0 & (\mathbf{h}^{-1})_{33} & (\mathbf{h}^{-1})_{23} \\ 0 & 2(\mathbf{h}^{-1})_{12} & 0 & (\mathbf{h}^{-1})_{13} & 0 & (\mathbf{h}^{-1})_{11} \end{pmatrix} \tag{D.34}$$

This translates to $Q(\mathbf{h}) = \exp\big[-\frac{1}{2}\beta V \Delta\vec{\mathrm{h}}^T \mathbf{L}^T \mathbf{C}_{\text{stiffness}} \mathbf{L} \Delta\vec{\mathrm{h}}\big]$ with $\mathbf{C}_{\text{stiffness}}$ the stiffness tensor and V the volume of the unit cell (see Equation D.29). The partition function becomes

$$Q_{\mathbf{h}} = \int\limits_{\Delta\vec{\mathrm{h}}} d\Delta\vec{\mathrm{h}} \exp\Big[-\beta P_{\text{ext}} V - \frac{1}{2}\beta V \Delta\vec{\mathrm{h}}^T \mathbf{L}^T \mathbf{C}_{\text{stiffness}} \mathbf{L} \Delta\vec{\mathrm{h}}\Big] V^{-2} \tag{D.35}$$

This expression for the partition function will be expanded to second order in $\Delta\vec{\mathrm{h}}$, using the Taylor formula:

$$\begin{aligned} V &= \det(\mathbf{h}) && \text{(D.36)}\\ &= V_e + \sum_{i=1}^{6} \Delta \mathrm{h}_i \left[\frac{\partial \det(\mathbf{h})}{\partial \Delta \mathrm{h}_i}\right]_{\vec{\mathrm{h}}_e} + \sum_{i=1}^{6}\sum_{j=1}^{6} \Delta \mathrm{h}_i \Delta \mathrm{h}_j \left[\frac{\partial \det(\mathbf{h})}{\partial \Delta \mathrm{h}_i \partial \Delta \mathrm{h}_j}\right]_{\vec{\mathrm{h}}_e} + \dots \\ &\equiv V_e + \mathbf{M}^T \Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T \mathbf{N} \Delta\vec{\mathrm{h}} + \dots && \text{(D.37)} \end{aligned}$$

D.2.1 Derivatives of V

For this expansion, the derivatives of V with respect to $\Delta\vec{\mathrm{h}}$ (or equivalent, $\vec{\mathrm{h}}$) need to be calculated. Firstly, an expression for the determinant as a function of $\vec{\mathrm{h}}$ is written.

$$
\begin{aligned}
V = \det(\mathbf{h}) &= \begin{vmatrix} \mathrm{h}_1 & \frac{1}{2}\mathrm{h}_6 & \frac{1}{2}\mathrm{h}_5 \\ \frac{1}{2}\mathrm{h}_6 & \mathrm{h}_2 & \frac{1}{2}\mathrm{h}_4 \\ \frac{1}{2}\mathrm{h}_5 & \frac{1}{2}\mathrm{h}_4 & \mathrm{h}_3 \end{vmatrix} \qquad \text{(D.38)} \\
&= \mathrm{h}_1\mathrm{h}_2\mathrm{h}_3 + \frac{1}{4}(\mathrm{h}_4\mathrm{h}_5\mathrm{h}_6 - \mathrm{h}_1\mathrm{h}_4^2 - \mathrm{h}_2\mathrm{h}_5^2 - \mathrm{h}_3\mathrm{h}_6^2) \qquad \text{(D.39)}
\end{aligned}
$$

The partial derivative of the determinant with respect to $\Delta\vec{\mathrm{h}}$ is the same as the partial derivative to $\vec{\mathrm{h}}$ because $\vec{\mathrm{h}} = \vec{\mathrm{h}}_e + \Delta\vec{\mathrm{h}}$. It follows, using notation $\left[\frac{\partial \det(\mathbf{h})}{\partial \Delta\vec{\mathrm{h}}}\right]_{\vec{\mathrm{h}}_e, i} = \left[\frac{\partial \det(\mathbf{h})}{\partial \Delta\mathrm{h}_i}\right]_{\vec{\mathrm{h}}_e}$:

$$
\mathbf{M} = \left[\frac{\partial \det(\mathbf{h})}{\partial \Delta\vec{\mathrm{h}}}\right]_{\vec{\mathrm{h}}_e} = \left[\frac{\partial \det(\mathbf{h})}{\partial \vec{\mathrm{h}}}\right]_{\vec{\mathrm{h}}_e} = \begin{pmatrix} \mathrm{h}_2\mathrm{h}_3 - \frac{1}{4}\mathrm{h}_4^2 \\ \mathrm{h}_1\mathrm{h}_3 - \frac{1}{4}\mathrm{h}_5^2 \\ \mathrm{h}_2\mathrm{h}_1 - \frac{1}{4}\mathrm{h}_6^2 \\ \frac{1}{4}\mathrm{h}_5\mathrm{h}_6 - \frac{1}{2}\mathrm{h}_1\mathrm{h}_4 \\ \frac{1}{4}\mathrm{h}_4\mathrm{h}_6 - \frac{1}{2}\mathrm{h}_2\mathrm{h}_5 \\ \frac{1}{4}\mathrm{h}_5\mathrm{h}_4 - \frac{1}{2}\mathrm{h}_3\mathrm{h}_6 \end{pmatrix} \tag{D.40}
$$

And for the second derivatives:

$$
\mathbf{N} = \left[\frac{\partial \det(\mathbf{h})}{\partial \Delta\vec{\mathrm{h}} \partial \Delta\vec{\mathrm{h}}}\right]_{\vec{\mathrm{h}}_e} = \begin{pmatrix} 0 & \mathrm{h}_3 & \mathrm{h}_2 & -\frac{1}{2}\mathrm{h}_4 & 0 & 0 \\ \mathrm{h}_3 & 0 & \mathrm{h}_1 & 0 & -\frac{1}{2}\mathrm{h}_5 & 0 \\ \mathrm{h}_2 & \mathrm{h}_1 & 0 & 0 & 0 & -\frac{1}{2}\mathrm{h}_6 \\ -\frac{1}{2}\mathrm{h}_4 & 0 & 0 & -\frac{1}{2}\mathrm{h}_1 & \frac{1}{4}\mathrm{h}_6 & \frac{1}{4}\mathrm{h}_5 \\ 0 & -\frac{1}{2}\mathrm{h}_5 & 0 & \frac{1}{4}\mathrm{h}_6 & -\frac{1}{2}\mathrm{h}_2 & \frac{1}{4}\mathrm{h}_4 \\ 0 & 0 & -\frac{1}{2}\mathrm{h}_6 & \frac{1}{4}\mathrm{h}_5 & \frac{1}{4}\mathrm{h}_4 & -\frac{1}{2}\mathrm{h}_3 \end{pmatrix} \tag{D.41}
$$

The expression V^{-2} can be expanded to second order in (small) $\Delta\vec{\mathrm{h}}$:

$$V^{-2} = \Big[\det(\mathbf{h})\Big]^{-2} \tag{D.42}$$

$$= \Big[V_e + \mathbf{M}^T\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\mathbf{N}\Delta\vec{\mathrm{h}} + \dots\Big]^{-2} \tag{D.43}$$

$$= V_e^{-2}\Big[1 + \frac{\mathbf{M}^T}{V_e}\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\frac{\mathbf{N}}{V_e}\Delta\vec{\mathrm{h}} + \dots\Big]^{-2} \tag{D.44}$$

$$\begin{aligned} &= V_e^{-2}\Big[1 - 2\Big(\frac{\mathbf{M}^T}{V_e}\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\frac{\mathbf{N}}{V_e}\Delta\vec{\mathrm{h}} + \dots\Big) \\ &\qquad +3\Big(\big(\frac{\mathbf{M}^T}{V_e}\Delta\vec{\mathrm{h}}\big)^2 + \dots\Big) + \dots\Big] \end{aligned} \tag{D.45}$$

$$\begin{aligned} &\approx \frac{1}{V_e^2}\Big[1 + \Big(-2\frac{\mathbf{M}^T}{V_e}\Big)\Delta\vec{\mathrm{h}} \\ &\qquad +\Delta\vec{\mathrm{h}}^T\Big(-2\frac{\mathbf{N}}{V_e} + 3\frac{\mathbf{M}\mathbf{M}^T}{(V_e)^2}\Big)\Delta\vec{\mathrm{h}}\Big] \end{aligned} \tag{D.46}$$

D.2.2 Combining the formulas

Inserting Equations D.37 and D.46 in Equation D.35, the expression for the partition function $Q_{\mathbf{h}}$, to second order in $\Delta\vec{h}$ becomes:

$$\begin{aligned} Q_{\mathbf{h}} &\approx \int d\Delta\vec{\mathrm{h}}\; e^{-\beta P_{\mathrm{ext}}\left(V_e + \mathbf{M}^T\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\mathbf{N}\Delta\vec{\mathrm{h}}\right) - \frac{1}{2}\beta V_e\Delta\vec{\mathrm{h}}^T\mathbf{L}^T\mathbf{C}_{\mathrm{stiffness}}\mathbf{L}\Delta\vec{\mathrm{h}}} \\ &\qquad \frac{1}{V_e^2}\Big[1 + \Big(-2\frac{\mathbf{M}^T}{V_e}\Big)\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\Big(-2\frac{\mathbf{N}}{V_e} + 3\frac{\mathbf{M}\mathbf{M}^T}{V_e^2}\Big)\Delta\vec{\mathrm{h}}\Big] \end{aligned} \tag{D.47}$$

$$\begin{aligned} &= \frac{e^{-\beta P_{\mathrm{ext}}V_e}}{V_e^2}\int d\Delta\vec{\mathrm{h}}\Big[1 + \Delta\vec{\mathrm{h}}^T\Big(-2\frac{\mathbf{N}}{V_e} + 3\frac{\mathbf{M}\mathbf{M}^T}{V_e^2}\Big)\Delta\vec{\mathrm{h}}\Big] \\ &\qquad e^{-\beta P_{\mathrm{ext}}\mathbf{M}^T\Delta\vec{\mathrm{h}} + \Delta\vec{\mathrm{h}}^T\left(\frac{1}{2}\beta V_e\mathbf{L}^T\mathbf{C}_{\mathrm{stiffness}}\mathbf{L} - \beta P_{\mathrm{ext}}\mathbf{N}\right)\Delta\vec{\mathrm{h}}} \end{aligned} \tag{D.48}$$

$$\tag{D.49}$$

because only even powers in the prefactor of the integral contribute to the expression. To simplify the notation, some new constants are defined, which are independent of Δh:

$$K \equiv \frac{e^{-\beta P_{\mathrm{ext}}V_e}}{V_e^2} \tag{D.50}$$

$$\mathbf{A} \equiv \frac{1}{2}\beta V_e\mathbf{L}^T\mathbf{C}_{\mathrm{stiffness}}\mathbf{L} - \beta P_{\mathrm{ext}}\mathbf{N} \tag{D.51}$$

$$\mathbf{J} \equiv -\beta P_{\mathrm{ext}}\mathbf{M} \tag{D.52}$$

$$\mathbf{B} \equiv -2\frac{\mathbf{N}}{V_e} + 3\frac{\mathbf{M}\mathbf{M}^T}{V_e^2} \tag{D.53}$$

Note that $\mathbf{A}$ is symmetric, because both $\mathbf{L}^T\mathbf{C}_{\text{stiffness}}\mathbf{L}$ and $\mathbf{N}$ are symmetric, and so $\mathbf{A}$ can be diagonalized. The expression for the partition function becomes

$$\begin{aligned} Q_{\mathbf{h}} &= K\Bigg[\int \exp\left(\Delta\vec{\mathrm{h}}^T\mathbf{A}\Delta\vec{\mathrm{h}} + \mathbf{J}^T\Delta\vec{\mathrm{h}}\right)d\Delta\vec{\mathrm{h}} \\ &\quad + \int \Delta\vec{\mathrm{h}}^T\mathbf{B}\Delta\vec{\mathrm{h}}\exp\left(\Delta\vec{\mathrm{h}}^T\mathbf{A}\Delta\vec{\mathrm{h}} + \mathbf{J}^T\Delta\vec{\mathrm{h}}\right)d\Delta\vec{\mathrm{h}}\Bigg] \qquad \text{(D.54)} \\ &= K\frac{(2\pi)^3}{\sqrt{\det\mathbf{A}}}\exp(\frac{1}{2}\mathbf{J}^T\mathbf{A}^{-1}\mathbf{J})\left[1 + \left(\mathbf{J}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{J} + \sum_i^n \frac{(\mathbf{O}^T\mathbf{B}\mathbf{O})_{ii}}{\lambda_i}\right)\right] \qquad \text{(D.55)} \end{aligned}$$

Bibliography

[1] H. Furukawa et al. "The chemistry and applications of metal-organic frameworks". In: *Science* 341.6149 (2013), p. 1230444.

[2] M. Eddaoudi et al. "Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage". In: *Science* 295.5554 (2002), pp. 469–472.

[3] S. Kitagawa and M. Kondo. "Functional Micropore Chemistry of Crystalline Metal Complex-Assembled Compounds." In: *Bulletin of the Chemical Society of Japan* 71.8 (1998), pp. 1739–1753.

[4] H Li et al. "Design and synthesis of an exceptionally stable and highly porous metal-organic framework". In: *Nature* 402.November (1999), pp. 276–279. ISSN: 00280836.

[5] S. S.-Y. Chui et al. "A chemically functionalizable nanoporous material [Cu3 (TMA) 2 (H2O) 3] n". In: *Science* 283.5405 (1999), pp. 1148–1150.

[6] K. Barthelet et al. "A breathing hybrid organic–inorganic solid with very large pores and high magnetic characteristics". In: *Angewandte Chemie* 114.2 (2002), pp. 291–294.

[7] H. Li et al. "Establishing microporosity in open metal-organic frameworks: Gas sorption isotherms for Zn (BDC)(BDC= 1, 4-benzenedicarboxylate)". In: *Journal of the American Chemical Society* 120.33 (1998), pp. 8571–8572.

[8] O. M. Yaghi et al. "Reticular synthesis and the design of new materials". In: *Nature* 423.6941 (2003), pp. 705–714.

[9] O. M. Yaghi. *Crystalline metal-organic microporous materials.* US Patent 5,648,508. 1997.

[10] T. Loiseau et al. "A Rationale for the Large Breathing of the Porous Aluminum Terephthalate (MIL-53) Upon Hydration". In: *Chemistry–A European Journal* 10.6 (2004), pp. 1373–1382.

[11] K. S. Park et al. "Exceptional chemical and thermal stability of zeolitic imidazolate frameworks". In: *Proceedings of the National Academy of Sciences* 103.27 (2006), pp. 10186–10191.

[12] R. L. Martin et al. "In silico design of porous polymer networks: High-throughput screening for methane storage materials". In: *Journal of the American Chemical Society* 136.13 (2014), pp. 5006–5022.

[13] C. M. Simon et al. "The materials genome in action: identifying the performance limits for methane storage". In: *Energy & Environmental Science* 8.4 (2015), pp. 1190–1199.

[14] Y. Bao et al. "Computational Design of Metal–Organic Frameworks with High Methane Deliverable Capacity". In: *Bulletin of the American Physical Society* (2016).

[15] F.-X. Coudert. "Responsive metal-organic frameworks and framework materials: Under pressure, taking the heat, in the spotlight, with friends". In: *Chemistry of Materials* 27.6 (2015), pp. 1905–1916. ISSN: 15205002.

[16] G. Férey and C. Serre. "Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences". In: *Chemical Society Reviews* 38.5 (2009), pp. 1380–1399.

[17] I. Beurroies et al. "Using pressure to provoke the structural transition of metal–organic frameworks". In: *Angewandte Chemie International Edition* 49.41 (2010), pp. 7526–7529.

[18] D. Dubbeldam et al. "Exceptional negative thermal expansion in isoreticular metal–organic frameworks". In: *Angewandte Chemie* 119.24 (2007), pp. 4580–4583.

[19] J.-R. Li, R. J. Kuppler, and H.-C. Zhou. "Selective gas adsorption and separation in metal–organic frameworks". In: *Chemical Society Reviews* 38.5 (2009), pp. 1477–1504.

[20] P. Horcajada et al. "Porous metal-organic-framework nanoscale carriers as a potential platform for drug delivery and imaging". In: *Nature Materials* 9.2 (2010), pp. 172–178.

[21] R. Kitaura et al. "Porous Coordination-Polymer Crystals with Gated Channels Specific for Supercritical Gases". In: *Angewandte Chemie International Edition* 42.4 (2003), pp. 428–431.

[22] F. Salles et al. “Multistep N2 Breathing in the Metal- Organic Framework Co (1, 4-benzenedipyrazolate)”. In: *Journal of the American Chemical Society* 132.39 (2010), pp. 13782–13788.

[23] P. K. Allan et al. “Pair distribution function-derived mechanism of a single-crystal to disordered to single-crystal transformation in a hemilabile metal–organic framework”. In: *Chemical Science* 3.8 (2012), pp. 2559–2564.

[24] A. U. Ortiz et al. “Anisotropic Elastic Properties of Flexible Metal-Organic Frameworks: How Soft are Soft Porous Crystals?” In: *Physical Review Letters* 109.19 (2012), p. 195502.

[25] J. M. Ogborn et al. “Supramolecular mechanics in a metal-organic framework”. In: *Chem. Sci.* 3 (10 2012), pp. 3011–3017.

[26] W. Li et al. “Negative linear compressibility of a metal–organic framework”. In: *Journal of the American Chemical Society* 134.29 (2012), pp. 11940–11943.

[27] P. Serra-Crespo et al. “Experimental evidence of negative linear compressibility in the MIL-53 metal–organic framework family”. In: *CrystEngComm* 17.2 (2015), pp. 276–280.

[28] J.-C. Tan et al. “Quantum mechanical predictions to elucidate the anisotropic elastic properties of zeolitic imidazolate frameworks: ZIF-4 vs. ZIF-zni”. In: *CrystEngComm* 17.2 (2015), pp. 375–382.

[29] J.-C. Tan et al. “Exceptionally Low Shear Modulus in a Prototypical Imidazole-Based Metal-Organic Framework”. In: *Physical Review Letters* 108.9 (2012), p. 095502.

[30] W. Zhou et al. “Hydrogen and methane adsorption in metal-organic frameworks: a high-pressure volumetric study”. In: *The Journal of Physical Chemistry C* 111.44 (2007), pp. 16131–16137.

[31] P. Ryan et al. “Computational screening of metal-organic frameworks for xenon/krypton separation”. In: *AIChE Journal* 57.7 (2011), pp. 1759–1766.

[32] A. U. Czaja, N. Trukhan, and U. Müller. “Industrial applications of metal–organic frameworks”. In: *Chemical Society Reviews* 38.5 (2009), pp. 1284–1293.

[33] J. Lee et al. “Metal–organic framework materials as catalysts”. In: *Chemical Society Reviews* 38.5 (2009), pp. 1450–1459.

[34] K. M. Taylor, W. J. Rieter, and W. Lin. “Manganese-Based Nanoscale Metal- Organic Frameworks for Magnetic Resonance Imaging”. In: *Journal of the American Chemical Society* 130.44 (2008), pp. 14358–14359.

[35] M. Yoon et al. “Proton conduction in metal–organic frameworks and related modularly built porous solids”. In: *Angewandte Chemie International Edition* 52.10 (2013), pp. 2688–2700.

[36] P. G. Yot et al. “Large breathing of the MOF MIL-47 (V IV) under mechanical pressure: a joint experimental–modelling exploration”. In: *Chemical Science* 3.4 (2012), pp. 1100–1104.

[37] C. Serre et al. “Very Large Breathing Effect in the First Nanoporous Chromium (III)-Based Solids: MIL-53 or CrIII (OH){O2C-C6H4-CO2}{HO2C-C6H4-CO2H} x H2O y”. In: *Journal of the American Chemical Society* 124.45 (2002), pp. 13519–13526.

[38] T. R. Whitfield et al. “Metal-organic frameworks based on iron oxide octahedral chains connected by benzenedicarboxylate dianions”. In: *Solid State Sciences* 7.9 (2005), pp. 1096–1103.

[39] C. Volkringer et al. “XRD and IR structural investigations of a particular breathing effect in the MOF-type gallium terephthalate MIL-53 (Ga)”. In: *Dalton Transactions* 12 (2009), pp. 2241–2249.

[40] J. P. Mowat et al. “Synthesis, characterisation and adsorption properties of microporous scandium carboxylates with rigid and flexible frameworks”. In: *Microporous and Mesoporous Materials* 142.1 (2011), pp. 322–333.

[41] Y.-X. Tan et al. “Dynamic Microporous indium (III)-4, 4-oxybis (benzoate) Framework with High Selectivity for the Adsorption of CO 2 over N 2”. In: *Chemical Communications* 47.2 (2011), pp. 770–772.

[42] Y. Liu et al. “Reversible structural transition in MIL-53 with large temperature hysteresis”. In: *Journal of the American Chemical Society* 130.35 (2008), pp. 11813–11818.

[43] C. Serre et al. “An explanation for the very large breathing effect of a metal–organic framework during CO2 adsorption”. In: *Advanced Materials* 19.17 (2007), pp. 2246–2251.

[44] J Rouquerol et al. “Recommendations for the characterization of porous solids (Technical Report)”. In: *Pure and Applied Chemistry* 66.8 (1994), pp. 1739–1758.

[45] P. G. Yot et al. “Metal–organic frameworks as potential shock absorbers: the case of the highly flexible MIL-53 (Al)”. In: *Chemical Communications* 50.67 (2014), pp. 9462–9464.

[46] S. M. J. Rogge et al. “A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks”. In: *Journal of Chemical Theory and Computation* 11.12 (2015), pp. 5583–5597. ISSN: 15499626.

[47] A. M. Walker et al. “Flexibility in a Metal–Organic Framework Material Controlled by Weak Dispersion Forces: The Bistability of MIL-53 (Al)”. In: *Angewandte Chemie International Edition* 49.41 (2010), pp. 7501–7503.

[48] F. Salles et al. “Molecular Dynamics Simulations of Breathing MOFs: Structural Transformations of MIL-53 (Cr) upon Thermal Activation and CO2 Adsorption”. In: *Angewandte Chemie International Edition* 47.44 (2008), pp. 8487–8491.

[49] L. Vanduyfhuys et al. “Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al)”. In: *Journal of Chemical Theory and Computation* 8.9 (2012), pp. 3217–3231. ISSN: 15499618.

[50] M. A. Addicoat et al. “Extension of the Universal Force Field to Metal–Organic Frameworks”. In: *Journal of chemical theory and computation* 10.2 (2014), pp. 880–891.

[51] F.-X. Coudert et al. “Thermodynamics of guest-induced structural transitions in hybrid organic-inorganic frameworks”. In: *Journal of the American Chemical Society* 130.43 (2008), pp. 14294–14302.

[52] A. Ghysels et al. “On the thermodynamics of framework breathing: A free energy model for gas adsorption in MIL-53”. In: *Journal of Physical Chemistry C* 117.22 (2013), pp. 11540–11554. ISSN: 19327447.

[53] L. Vanduyfhuys et al. “Semi-analytical mean-field model for predicting breathing in metal–organic frameworks”. In: *Molecular Simulation* 41.16-17 (2015), pp. 1311–1328.

[54] N. Metropolis and S. Ulam. “The Monte Carlo method”. In: *Journal of the American Statistical Association* 44.247 (1949), pp. 335–341.

[55] N. Metropolis et al. “Equation of state calculations by fast computing machines”. In: *The journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.

[56] B. J. Alder and T. Wainwright. “Studies in molecular dynamics. I. General method”. In: *The Journal of Chemical Physics* 31.2 (1959), pp. 459–466.

[57] L. Vanduyfhuys. “Insights in the behaviour of metal-organic frameworks through molecular modeling: from force field derivation to thermodynamic analysis”. eng. PhD thesis. Ghent University, 2015, pp. XXIX, 372. ISBN: 9789085788478.

[58] N. L. Allinger, Y. H. Yuh, and J. H. Lii. “Molecular mechanics. The MM3 force field for hydrocarbons. 1”. In: *Journal of the American Chemical Society* 111.23 (1989), pp. 8551–8566.

[59] A. K. Rappé et al. “UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations”. In: *Journal of the American chemical society* 114.25 (1992), pp. 10024–10035.

[60] S. L. Mayo, B. D. Olafson, and W. A. Goddard. “DREIDING: a generic force field for molecular simulations”. In: *Journal of Physical Chemistry* 94.26 (1990), pp. 8897–8909.

[61] P. Bultinck et al. “Critical analysis and extension of the Hirshfeld atoms in molecules”. In: *The Journal of Chemical Physics* 126.14 (2007), p. 144111.

[62] F. L. Hirshfeld. “Bonded-atom fragments for describing molecular charge densities”. In: *Theoretica Chimica Acta* 44.2 (1977), pp. 129–138.

[63] T. Verstraelen et al. *Yaff, yet another force field.* 2013.

[64] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *Journal of Chemical Physics* 126.1 (2007), pp. 1–7. ISSN: 00219606. arXiv: `arXiv:0803.4060v1`.

[65] D. Beeman. “Some multistep methods for use in molecular dynamics calculations”. In: *Journal of Computational Physics* 20.2 (1976), pp. 130–139.

[66] H. C. Andersen. “Molecular dynamics simulations at constant pressure and/or temperature”. In: *The Journal of Chemical Physics* 72.4 (1980), pp. 2384–2393.

[67] H. W. G. “Canonical Dynamics: Equilibrium Phase-Space Distributions”. In: *Phys. Rev. A: At., Mol., Opt. Phys.* 31.3 (1985), p. 1695.

[68] H. J. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

[69] T. Morishita. “Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath”. In: *The Journal of Chemical Physics* 113.8 (2000), pp. 2976–2982.

[70] G. J. Martyna, D. J. Tobias, and M. L. Klein. “Constant pressure molecular dynamics algorithms”. In: *The Journal of Chemical Physics* 101.September (1994), p. 4177. ISSN: 00219606.

[71] S. E. Feller et al. “Constant pressure molecular dynamics simulation: the Langevin piston method”. In: *The Journal of Chemical Physics* 103.11 (1995), pp. 4613–4621.

[72] G. J. Martyna, D. J. Tobias, and M. L. Klein. “Constant pressure molecular dynamics algorithms”. In: *The Journal of Chemical Physics* 101.5 (1994), pp. 4177–4189.

[73] S. Pal and K. A. Fichthorn. “Accelerated molecular dynamics of infrequent events”. In: *Chemical Engineering Journal* 74.1-2 (1999), pp. 77–83. ISSN: 13858947.

[74] S. Kumar et al. “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

[75] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566.

[76] R. Martoňák, A. Laio, and M. Parrinello. “Predicting crystal structures: the Parrinello-Rahman method revisited”. In: *Physical review letters* 90.7 (2003), p. 075503.

[77] V. Spiwok, P. Lipovová, and B. Králová. “Metadynamics in essential coordinates: free energy simulation of conformational changes”. In: *The Journal of Physical Chemistry B* 111.12 (2007), pp. 3073–3076.

[78] S. Rogge. *Afleiding en implementatie van bewegingsvergelijkingen voor isotherme en isobare Moleculaire Dynamica simulaties.* 2014.

[79] J. A. Hartigan and P. Hartigan. “The dip test of unimodality”. In: *The Annals of Statistics* 70–84 (1985).

[80] A. Spek. “Single-crystal structure validation with the program PLATON”. In: *Journal of Applied Crystallography* 36.1 (2003), pp. 7–13.

[81] M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill. “Estimates of the ab initio limit for π-π interactions: The benzene dimer”. In: *Journal of the American Chemical Society* 124.36 (2002), pp. 10887–10893.

[82] K. Pearson. “Principal components analysis”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 6.2 (1901), p. 559.

[83] M. A. Balsera et al. “Principal component analysis and long time protein dynamics”. In: *The Journal of Physical Chemistry* 100.7 (1996), pp. 2567–2572.

[84] M. Karplus and J. N. Kushick. “Method for estimating the configurational entropy of macromolecules”. In: *Macromolecules* 14.2 (1981), pp. 325–332.

[85] J. Schlitter. “Estimation of absolute and relative entropies of macromolecules using the covariance matrix”. In: *Chemical Physics Letters* 215.6 (1993), pp. 617–621. ISSN: 00092614.

[86] I. Andricioaei and M. Karplus. “On the calculation of entropy from covariance matrices of the atomic fluctuations”. In: *Journal of Chemical Physics* 115.14 (2001), pp. 6289–6292. ISSN: 00219606.

[87] U. Hensen, F. Gräter, and R. H. Henchman. “Macromolecular entropy can be accurately computed from force”. In: *Journal of Chemical Theory and Computation* 10.11 (2014), pp. 4777–4781. ISSN: 15499626.

[88] V. Spiwok, B. Králová, and I. Tvaroška. “Continuous metadynamics in essential coordinates as a tool for free energy modelling of conformational changes”. In: *Journal of molecular modeling* 14.11 (2008), pp. 995–1002.

[89] B. Hess. “Convergence of sampling in protein simulations”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 65.3 (2002), p. 031910. ISSN: 15393755.

[90] F. Sittel, A. Jain, and G. Stock. “Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates”. In: *The Journal of Chemical Physics* 141.1 (2014), p. 014111.

[91] D. E. Vanpoucke et al. “Mechanical Properties from Periodic Plane Wave Quantum Mechanical Codes: The Challenge of the Flexible Nanoporous MIL-47 (V) Framework”. In: *The Journal of Physical Chemistry C* 119.41 (2015), pp. 23752–23766.