Univserity of Innsbruck
Department of General, Inorganic and Theoretical Chemistry

Master Thesis

Structure and Thermodynamics of Guest@MOF Material

Investigating GUEST in MIL-68Ga via a Molecular Dynamics Simulation Approach

Michael Helmut Fill, BSc.
Supervisor: Assoc. Prof. Dr. Thomas Hofer
Hier Datum einfügen

Abstract

This thesis is dedicated to the brave Mujahideen fighters of Afghanistan.

Contents

1 Introduction

2 Theory

2.1 Quantum Chemistry

At the dawn of the 20th century, general consensus among many physicists versed in classical mechanics was that the fundamental laws of nature had been solved, and that the by then still unsolved problems of e.g. black body radiation would be resolved in due time. However, when Max Planck published his ideas on the quantization of energy in 1900 [1], and especially after Albert Einstein realistically explained the photoelectric effect using Plancks' hypothesis in 1905 [2], it became clear that the classical laws of physics were inadequate at explaining the behavior of matter on the atomic scale. The field of quantum mechanics was born, and together with further groundbreaking contributions by, among others, Niels Bohr [3], Werner Heisenberg [4] and Louis de Broglie [5], paved the way for Erwin Schrödinger to lay the groundwork for quantum chemistry with his wave equation in 1926 [6] [7].

2.1.1 The Schrödinger Equation

Starting from the initial assumption that all properties of a given system could be described by a wave function Ψ, Schrödinger derived his now well-known time-independant (1) and, more general, time-dependant (2) equations.

$$\hat{H}\Psi = E\Psi \tag{1}$$

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{2}$$

With Ψ being the eigenfunction of the system, $\hbar$ the reduced Planck constant, $\hat{H}$ the Hamiltonian operator, t the time and E the energy of the system, or eigenvalue of the Hamiltonian operator. The Hamiltonian operator (3) is defined as the sum of the kinetic and potential energy operators, which, when applied to a state function Ψ, yields the total energy of the system as its eigenvalue to the eigenfunction Ψ.

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + E_{pot} = \hat{T} + \hat{V} \tag{3}$$

Here, ∇^2 is the Laplace operator, m the mass of the particle and E_{pot} the potential energy of the system, with the kinetic energy and potential energy operators historically referred to as $\hat{T}$ and $\hat{V}$ respectively. For a single particle with mass m moving in three-dimensional space, the time independant Schrödinger equation can then be written as (4).

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + E_{pot}\Psi = E\Psi \tag{4}$$

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

This linear differential equation (4) forms the basis for determining the wave function of a given system. A wavefunction Ψ that satisfies the Schrödinger equation and results in its respective eigenvalue E must therefore be an eigenfunction to the Hamiltonian operator. It is due to this that finding the correct wave function Ψ for a system is referred to as an eigenvalue problem.

Hereby, it is important to note that a wavefunction is not directly observable and therefore no meaning can be attributed to its values. However, the square modulus of the wavefunction $|\Psi|^2$, as interpreted by Max Born [8], is proportional to the observable probability density of an electron in the given system.

2.1.2 Hamiltonian Operator

As previously mentioned, the Hamiltonian operator is defined as the sum of kinetic and potential energies in a given system. The notation given in equation 3 is a simplification of the actual Hamiltonian operator, which can be written as in equation 5 and 6 as them sum of its individual terms, being the kinetic energy of the nuclei $\hat{T}_N$, the kinetic energy of the electrons $\hat{T}_e$, the potential energy of the nuclei $\hat{V}_{NN}$, the potential energy of the nuclei and electrons $\hat{V}_{Ne}$ and the potential energy of the electrons $\hat{V}_{ee}$.

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{NN} + \hat{V}_{Ne} + \hat{V}_{ee} \tag{5}$$

$$\begin{aligned}\hat{H} = &-\sum_{I=1}^{N}\frac{\hbar^2}{2m_I}\nabla_I^2 - \sum_{i=1}^{n}\frac{\hbar^2}{2m_e}\nabla_i^2 + \frac{1}{2}\sum_{I\neq J}^{N}\frac{q_e^2}{4\pi\epsilon_0}\frac{Z_I Z_J}{|r_I - r_J|} \\ &-\sum_{I=1}^{N}\sum_{i=1}^{n}\frac{q_e^2}{4\pi\epsilon_0}\frac{Z_I}{|r_I - r_i|} + \frac{1}{2}\sum_{i\neq j}^{n}\frac{q_e^2}{4\pi\epsilon_0}\frac{1}{|r_i - r_j|}\end{aligned} \tag{6}$$

With N being the number of nuclei, n the number of electrons, m_I the mass of the I-th nucleus, m_e the mass of the electron, Z_I the charge of the I-th nucleus, r_I the position of the I-th nucleus, r_i the position of the i-th electron, q_e the elementary charge and ϵ_0 the vacuum permittivity. To simplify this expression, atomic units are introduced, leading to the Hamiltonian as shown in equation 7.

$$\begin{aligned}\hat{H} = &-\sum_{I=1}^{N}\frac{1}{2m_I}\nabla_I^2 - \sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 + \frac{1}{2}\sum_{I\neq J}^{N}\frac{Z_I Z_J}{|r_I - r_J|} \\ &-\sum_{I=1}^{N}\sum_{i=1}^{n}\frac{Z_I}{|r_I - r_i|} + \frac{1}{2}\sum_{i\neq j}^{n}\frac{1}{|r_i - r_j|}\end{aligned} \tag{7}$$

2.1.3 Dirac Notation

Since it is not feasible and often impossible to solve the Schrödinger equation for systems of 3 or more unconstrained particles, should a coulombic potential be employed [9], a number of approximations are made, some of which will be discussed in the following sections. To work with these approximated wave functions, the expressions for the Schrödinger equation must be adjusted as in 8.

$$\hat{H}\Psi = E\Psi \tag{8}$$

$$\Psi^*\hat{H}\Psi = \Psi^* E\Psi$$

$$\int_{-\infty}^{\infty}\Psi^*\hat{H}\Psi d\tau = \int_{-\infty}^{\infty}\Psi^* E\Psi d\tau$$

With τ being the volume element of the system and Ψ^* the complex conjugate of the wave function Ψ. The received energy, granted that Ψ is an approximated wave function, is now an estimate of the actual analytical energy. To simplify these integral expressions the *bra-ket* notation, also known as *Dirac* notation [10], is employed as shown in equation 9.

$$\langle\Psi|\hat{H}|\Psi\rangle = \langle\Psi|E|\Psi\rangle \tag{9}$$

2.1.4 Variational Principle

With the framework for utilizing approximated wave functions in place, the next step is to qualify a chosen wave function according to its accurary as an approximation to the analytical wave function. Since the energy of a system is constant, it can be taken out of the integral expression, leaving the remaining normalized integral yielding 1. Choosing now a trial wave function Ψ_{Trial} and applying the Hamiltonian operator to it, the resulting energy expectation value is always greater or equal to the ground state energy of the system, as shown in equation 10. This method of finding parametres to construct the best possible approximated wave function is called the variational method and sits at the heart of many quantum mechanical approaches [11].

$$\langle \Psi_{Trial} | \hat{H} | \Psi_{Trial} \rangle = \langle E \rangle \geq E_0 \tag{10}$$

2.2 Hartree-Fock Theory

As previously mentioned, the exact solution to the Schrödinger equation for systems containing more than 2 unconstrained particles is analytically not feasible. To this end, several approximations are employed to simplify the eigenvalue problem down to an analytically solvable form. The Hartree-Fock method employs such approximations and provides a fundamental framework for the analysis of many-electron systems.

2.2.1 Non-relativistic Stationary Systems

An important assumption not exclusive to Hartree-Fock Theory is the non-relativistic nature of the analysed system, implying stationary systems with velocities much smaller than the speed of light. This assumption allows one to employ only the much simpler time-independant Schrödinger equation, with the trade-off being that effects arising from special relativity, e.g. spatial contraction or time dilation in heavy atoms, are not accounted for. Additionally, since only stationary states are computed, the analysis of time-dependant processes require specific simulation frameworks, one of which will be discussed at a later point.

2.2.2 Born-Oppenheimer Approximation

The Hartree-Fock method inherently assumes the Born-Oppenheimer approximation, which states that the wave functions of nuclei and electrons of a given system may be decoupled [12]. This seperation is justified via the large mass difference of a factor of 1836 between electrons and protons, and allows for the assumption that nuclei may be considered stationary during the calculation of the electronic wave functions.

$$|\Psi^{n,N}\rangle = |\Psi^{n}\rangle|\Psi^{N}\rangle \tag{11}$$

With n and N denoting the electronic and nuclear wave functions respectively. As a result of this approximation, the term for the kinetic energy of the nuclei in the Hamiltonian is omitted, while the coulombic repulsion between different nuclei becomes a constant term, further simplifying calculations. An eigenfunction to this electric Hamiltonian, as shown in 12, is now referred to as an electronic eigenfunction to the Schrödinger equation.

$$\hat{H} = -\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{I=1}^{N}\sum_{i=1}^{n}\frac{Z_I}{|r_I - r_i|} + \frac{1}{2}\sum_{i\neq j}^{n}\frac{1}{|r_i - r_j|} \tag{12}$$

Neglecting the quantum character of nuclei in this way renders the electronic Hamiltonian useless for describing quantum mechanical phenomena involving nuclei, yet allows for vastly increased computational efficiency.

2.2.3 Independant Particle approximation

Douglas R. Hartree provided a further approximation to n-electron wave functions in 1928 in the form of a product of independant one-electron functions, referred to as *Hartree product* [13]. Equation 13 shows this formalism, where each independant electrons i interaction with all other $n-1$ electrons of the system is averaged out, referred to as a *mean field approximation.*

$$|\Psi^{n}\rangle = \prod_{i=1}^{n}|\psi_i\rangle \tag{13}$$

2.2.4 Slater Determinant

While providing a good approximation to n-electron wavefunctions, the Hartree product does not account for the antisymmetrical nature of fermions due to their half-integer spin resulting from the Pauli exclusion principle [14]. In essence, this means that transposing any two, by definition indistinguishable, electrons in a wave function must result in a sign change. For a two particle system with coordinates r_1 and r_2, this may be written as seen in 14.

$$\Psi(r_1, r_2) = -\Psi(r_2, r_1) \tag{14}$$

To account both for antisymmetry and indistinguishability, John C. Slater considered all possible distributions of electrons in the one-electron wave functions, conveniently expressed via a determinant, the *Slater determinant* [15], as shown in 15.

$$|\Psi^n\rangle \approx \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(r_1) & \psi_2(r_1) & \cdots & \psi_n(r_1) \\ \psi_1(r_2) & \psi_2(r_2) & \cdots & \psi_n(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(r_n) & \psi_2(r_n) & \cdots & \psi_n(r_n) \end{vmatrix} \tag{15}$$

Where the factorial term accounts for normalization and each entry in the determinant represents a one electron wave function, with an electron at position r_i. This determinant represents a superposition of all possible Hartree products and, due to the nature of determinants, inherently accounts for sign-changes upon transposition of electrons, i.e. antisymmetry. Further, as one-electron functions must remain orthonormal, the Kronecker delta function δ_{ij} is employed as in 16, ensuring orthogonality and further simplifying calculations by reducing the number of non-zero integrals to be computed.

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{16}$$

2.2.5 LCAO-Approach

Molecular orbitals, as they are employed in Slater determinants, are usually constructed via the *Linear Combination of Atomic Orbitals* (LCAO) approach [16]. In this approach,

n atomic orbitals, ideally analytical eigensolutions of hydrogen-like systems, are combined linearly to n molecular orbitals as in 17, which in turn make up the Slater determinant.

$$|\psi_i\rangle = \sum_r c_{ri}|\phi_r\rangle \tag{17}$$

Where $|\phi_r\rangle$ are the atomic orbitals, c_{ri} the expansion coefficients and $|\psi_i\rangle$ the molecular orbitals. During energy optimization according to the variational principle, these coefficients c_{ri} are adjusted to achieve plausible approximations while using Lagrange multipliers to ensure orthonormality of the resulting molecular orbitals [17]. Due to their computational effort, albeit high accuracy, the analytical Slater type orbitals (STO) of hydrogen are often foregone in favor of simpler Gaussian type orbitals (GTO), several of which may be contracted (CGTO) to approximate a single Slater orbital [18]. The level of contraction, number of CGTOs per atomic orbital as well as parametres like zeta-factors, are collected in basis sets, highly varied in their complexity and accuracy [19].

2.2.6 Bloch's Theorem

In periodic systems, defined by a unit cell, both the potential energies and wave functions at any given point are subject to periodicity. As per *Bloch's theorem*, utilizing this periodicity, one may express wave functions via periodically modulated plane waves to good effect [20]. This theorem is particularly useful when analyzing periodic systems, e.g. crystals, allowing to construct wave functions or electronic states via Bloch functions, as in equation 18.

$$\psi_k(r) = e^{ikr}u_k(r) \tag{18}$$

Where k is the wave vector, r the position and $u_k(r)$ a periodic function mirroring the periodicity of the unit cell. Since the wave vector k is not uniquely defined, it is usually restricted to the first Brillouin zone, the smallest unit cell in reciprocal space, avoiding redundancy [21].

2.2.7 Self-Consistent Field Method

With approximations and formalisms in place, a suitable Slater determinant may now be constructed by way of adjusting the expansion coefficients c_{ri} in the LCAO approach. To this end, an operator $\hat{F}$ is introduced, the *Fock operator*, describing the kinetic energy

of an electron as well as its interaction with all other electrons in the system in a mean field approximation. This formalism is shown in equation 19, with the molecular orbitals shown in their linearly decomposed form.

$$\hat{F} \sum_r c_{ri} |\phi_r\rangle = \epsilon_i \sum_r c_{ri} |\phi_r\rangle \tag{19}$$

Where ϵ_i is the energy eigenvalue of the i-th molecular orbital. Utilizing vector notation, this expression may be written as in 20, resulting in what is known as the *Roothan-Hall equation.*

$$\mathbf{Fc} = \mathbf{ScE} \tag{20}$$

Where $\mathbf{F}$ is the Fock matrix containing all possible Fock operators, $\mathbf{c}$ the coefficent matrix $\mathbf{S}$ the overlap matrix and $\mathbf{E}$ the orbital energy matrix. Should the basis sets used be orthonormal, as enforced by the minimization constraints, the overlap matrix $\mathbf{S}$ reduces to the identity matrix, simplifying the Roothan-Hall equation to 21.

$$\mathbf{Fc} = \mathbf{Ec} \tag{21}$$

Due to the Fock operator containing the molecular orbitals, which themselves depend on the chosen coefficients, the Roothan-Hall equation is solved via an iterative prodecure, where an initial set of coefficients is chosen before the first iteration. Then, the Fock matrix is constructed and diagonalized, yielding a new set of coefficients as well as energy eigenvalues. The latter then serve as criteria for convergence, with iterations being carried out until the difference in energy is below a predefined threshold. The so received approximated wavefunction may then be used for further analysis of the system, e.g. calculating forces or charges acting upon atoms.

2.2.8 Post Hartree-Fock Methods

Hartree-Fock, mainly due to employing only a single Slater determinant, neglects part of electron correlation, i.e. how electrons in movement are affected by the presence of other electrons. While some correlation is included in the exchange terms of the Fock operator, neglecting other forms of correlation causes the method to be limited to a theoretical *Hartree-Fock limit*, the lowest energy achievable. Over the years, many methods were

developed to account for said correlation, congregating under the term *Post Hartree-Fock methods*. One such method is *Configuration Interaction* (CI), where the HF wave function is expanded to include excited determinants, which are then combined linearly to approximate a correlated wave function. Just like in base HF, the vartiational principle is applied when weighing the contribution of excited states. Including all possible excited states, referred to as *Full Configuration Interaction* (Full-CI), while providing the most accurate results, is computationally unfeasible for most molecules, which is why truncation is usually applied.

2.3 Density Functional Theory

Spawning from the field of solid-state physics, *Density Functional Theory* (DFT) is a quantum mechanical method utilizing an alternative approach to many-electron systems with the aim of addressing the inaccuracies of HF and the high computational cost of post-HF methods. Where wave-functions depend on 3N spatial coordinates, excluding spin, DFT instead focuses on the electron density $\rho(\vec{r})$ of a system, a property defined by only 3 spatial coordinates, theoretically decreasing computational demand drastically [22]. The formalisms and qualitative discussions in the following subsections 2.3.1 trough 2.3.5 are based on *A Chemists's Guide to Density Functional Theory* by Koch and Holthausen [23].

2.3.1 Hohenberg-Kohn Theorems

At the heart of DFT lie two theorems, formulated by Pierre Hohenberg and Walter Kohn in 1964 [24]. First, the ground-state of a many electron system, and by extension its properties, are uniquely determined by its electron density $\rho(\vec{r})$. In short, the external potential of a system is a unique functional of the electron density. Secondly, such a functional, if applied to the ground-state density, will yield the lowest possible energy of the system and this ground-state density corresponds to the exact solution of the Schrödinger equation. This, in essence, is the variational principle applied to the electron density functionals, as shown in equation 22.

$$E[\rho_{Trial}(\vec{r})] \geq E[\rho_0(\vec{r})] \tag{22}$$

2.3.2 Kohn-Sham Equations

The *Kohn-Sham* approach (KS), as derived by Kohn and Sham in 1965 [25], forms the basis of current DFT approaches. In this approach, a system of non-interacting electrons yielding the electron density of the original problem is constructed as seen in equation 23.

$$\rho_0(\vec{r}) = \sum_{i}^{N} |\phi_i(\vec{r})|^2 \tag{23}$$

Ironically, the wavefunction of this non-interacting system is represented by a Slater determinant, referred to as a Kohn-Sham determinant in the context of DFT. The total energy of the system is then expressed as shown in equation 24.

$$E[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})] \tag{24}$$

With T_s being the non-interacting kinetic energy of the system, J the Coulomb energy and E_{XC} the exchange-correlation energy. While the kinetic energy is for the most part comparatively easy to solve, the exchange-correlation functionals, which also include a "true" part of the kinetic energy, for such a system are unknown, leading to many DFT approaches with the sole purpose of defining increasingly accurate functionals. Should such an explicit and exact form of the exchange-correlation functional $E_{XC}[\rho(\vec{r})]$ and its corresponding potential $V_{XC}[\rho(\vec{r})]$ be found, the Kohn-Sham equations would, in principal, be exact solutions to many body problems.

2.3.3 Local Density Approximation

Among the simplest methods in DFT, *Local (Spin) Density Approximation (L(S)DA)* employs the electron density as a homogenous electron gas. While effective in describing the solid metals and alloys it was designed for, LDA and its open-shell variant LSDA are considered inadequate for systems involving variable electron densities like MOFs [26]. In practice, to simplify calculations, the exchange and correlation functionals applied onto the electron density, as shown in equation 25, are considered seperable.

$$E_{XC}^{LDA}[\rho(\vec{r})] = \int_{-\infty}^{\infty} \epsilon_{XC}[\rho(\vec{r})]\rho(\vec{r})d\vec{r} = \int_{-\infty}^{\infty} \epsilon_X[\rho(\vec{r})]\rho(\vec{r})d\vec{r} + \int_{-\infty}^{\infty} \epsilon_C[\rho(\vec{r})]\rho(\vec{r})d\vec{r} \tag{25}$$

Where $\epsilon_{XC}[\rho(\vec{r})]$ is the exchange-correlation energy per particle and $\rho(\vec{r})$ the electron density, with square brackets denoting a functional dependence. Since the exchange-energy density of a homogenous electron gas is known analytically, derived by Bloch and Dirach as early as 1929 [20] [27], LDA approaches apply this result pointwise to a non-homogenous density to calculate exchange-energy as in equation 26 [28].

$$E_X^{LDA}[\rho(\vec{r})] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int_{-\infty}^{\infty}\rho(\vec{r})^{4/3}d\vec{r} \tag{26}$$

While exchange-energy may be explicit, the correlation-energy of a homogeneous electron gas is not, leading to various expressions being derived over time, with e.g *Monte Carlo* approaches providing accurate results [29].

2.3.4 Generalized Gradient Approximation

The *Generalized Gradient Approximation* (GGA) differs from LDA by employing not only the electron density $\rho(\vec{r})$, but also its gradient $\nabla\rho(\vec{r})$ with respect to spatial coordinates as in equation 27. Functionals utilizing both the electron density and its gradient once more assume a seperation of the exchange and correlation functionals, as shown in equation 28.

$$E_{XC}^{GGA}[\rho(\vec{r}), \nabla\rho(\vec{r})] = \int_{-\infty}^{\infty}\epsilon_{XC}[\rho(\vec{r}), \nabla\rho(\vec{r})]\rho(\vec{r})d\vec{r} \tag{27}$$

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \tag{28}$$

The gradient corrected exchange functional of this approach may be written as in equation 29.

$$E_X^{GGA} = E_X^{LDA} - \int_{-\infty}^{\infty}F(s_\sigma)\rho_\sigma^{4/3}(\vec{r})d\vec{r} \tag{29}$$

Where the argument for the function F is the reduced gradient s_σ for spin σ, understood as a local inhomogenity parameter and calculated as in equation 30. The value of s_σ is large for regions of small densities such as those far from the nuclei and small for bonding regions or regions of large density. For the function F, two main realizations are employed. The first one, based on a GGA functional developed by Becke in 1988 [30], approaches inhomogenity as in equation 31.

$$s_\sigma = \frac{|\nabla \rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})} \tag{30}$$

$$F^B = \frac{\beta s_\sigma^2}{1 + 6\beta s_\sigma \sinh^{-1} s_\sigma} \tag{31}$$

In this functional, abbreviated simply as B or B88, the parameter β has been empirically determined to a value of 0.0042 by fitting to known exchange energies. The second class of functionals laying much of the groundwork for modern approaches employ F as a rational function of s_σ, as examplified *via* Perdews's exchange functional P86 in equation 32 [31].

$$F^{P86} = \left(1 + 1.296\left(\frac{s_\sigma}{(24\pi^2)^{1/3}}\right)^2 + 14\left(\frac{s_\sigma}{(24\pi^2)^{1/3}}\right)^4 + 0.2\left(\frac{s_\sigma}{(24\pi^2)^{1/3}}\right)^6\right)^{1/15} \tag{32}$$

Here, as opposed to the Becke functional, no empirical parameters are present. With exchange functionals established, gradient-corrected correlation functionals are a different beast entirely when it comes to analytical complexity. One widely used functional is the P or P86 correlation functional, a counterpart to the 1986 Perdew exchange functional discussed previously, employing an empirical parameter fitted to the correlation energy of a neon atom [32]. Functionals free of empirical parameters have also been developed, such as *PW91* by Perdew and Wang in 1991 [33]. Finally and most notably, Lee, Yang and Parr formulated their *LYP* functional in 1988, derived from the *Colle-Salvetti* correlation energy expression [34], in which the correlation energy density is expressed in terms of the electron density as well as a Laplacian of the second-order HF density matrix [35]. While freely combining different exchange and correlation functionals is generally possible, only a small subset of pairings is seeing much use in modern theoretical chemistry, with the combination of Becke's exchange functional and LYP correlation functional, known as BLYP, being one of the most widely used frameworks.

2.3.5 Hybrid Functionals

In an attempt to further improve the accuracy of DFT calculations, *Hybrid Functionals* utilize a combination of the exact Hartree-Fock exchange energies and approximated correlation functionals, as shown in equation 33.

$$E_{XC} = E_X^{HF} + E_C^{KS} \tag{33}$$

However, attempting to exploit the strengths of both HF and DFT using reckless combination provided no benefit over previously established approaches, with inaccuracies ranging in the high two-digit $kcal/mol$ range especially when applie to molecules and chemical bonding. The first succesful hybrid functional was derived by Becke in 1993 [36], employing a combination of HF exchange and DFT correlation in the '*half-and-half*' approach with already promisingly low average errors. In the next iteration, ultimately leading to arguably the most widely used hybrid functional, Becke introduced semiempirical coefficients to weigh the various hybrid components [37]. This combination of HF exchange with local and gradient-corrected correlation and exchange terms is shown in equation 34, with the semiempirical parameters $a = 0.20$, $b = 0.72$ and $c = 0.81$ fitted to optimally reproduce previous experimental results.

$$E_{XC}^{B3} = E_{XC}^{LSD} + a(E_{XC}^{\lambda=0} - E_X^{LSD}) + bE_X^B + cE_C^{PW91} \tag{34}$$

Where $\lambda = 0$ denotes an interaction free system and $\lambda = 1$ a fully interacting system. Suggested by Stephens et al. in 1994, replacing the PW91 correlation functional in equation 34 with a LYP functional, the B3LYP exchange-correlation functional was born [38], as seen in equation 35.

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSD} + aE_{XC}^{\lambda=0} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LSD} \tag{35}$$

Where the values for the parameters a, b and c are directly taken from Becke's original functional. The B3LYP functional showed promising results, with unsigned errors of only slightly above 2 $kcal/mol$ in respect to a Gaussian-2 reference set, comprised of corrected MP2 calculations [39]. Although B3LYP remained a pillar of density functional theory, hybrid functionals saw many more iterations and improvements over the years, with notable examples being the PBE0 model by Adamo and Barone in 1999 [40] or the M06 family of functionals by Truhlar and Zhao in 2007 [41].

2.3.6 Dispersion Corrections

The following discussions are sourced from a review by Klimeš and Michaelides, published in 2012 [42].

One of the main shortcomings of DFT is inherent to the approximations made for XC functionals, namely the inability of most standard functionals to properly describe long-range electron correlation, colloquially referred to as *van der Waals* forces. Dispersion, in essence the attractive interaction between instantaneous dipoles, is well described as decaying with the sixth power of the distance between two particles. However, most standard exchange-correlation functionals do not accurately describe this dispersion because instantaneous fluctuations in density are not considered and because only local properties are considered for calculating most XC energies. These functionals only model binding or repulsion when there is an overlap of the electron densities between two atoms, and since this overlap decreases exponentially with distance, the $-1/r^6$ decay is not accurately modelled.

To counteract this deficit, a number of dispersion correction schemes with varying levels of complexity have been developed, two of which will be discussed in detail. Among the simplest schemes are density functionals that are specifically fitted in such a way as to reproduce weak interaction around minima, like the *Minnesota functionals* by Zhao and Truhlar [43]. While still suffering from incorrect asymptotic behaviour, owing to the reference data used for fitting, these functionals can also represent other properties than weak interaction like e.g. reaction barriers, making them at least somewhat accurate for a wide range of generic problems. At around the same level of complexity, one may include pseudopotential projection to model dispersion. While laborious to fit for each element, methods such as the *dispersion corrected atom-centered potentials* (DCACP) [44] or *local atomic potentials* (LAP) [45] have shown promising results.

At the next level of complexity, an additional energy term is added to the DFT energy, accounting for missing long range interaction as in equation 36, with the dispersion interaction calculated as in equation 37.

$$E_{Tot} = E_{DFT} + E_{disp} \tag{36}$$

$$E_{disp} = -\sum_{A,B} \frac{C_6^{AB}}{r_{AB}^6} \tag{37}$$

Where C_6^{AB} is the dispersion coefficient between atoms A and B and r_{AB} the distance between them. Generally referred to as *DFT-D*, these methods assume dispersion to be pairwise additive over all elemental pairs A and B. While computationally inexpensive and therefore widely used, DFT-D methods are limited by their disregard for many-body dispersion as well as the influence of chemical environments and different states, since the coefficents C_6^{AB} are kept constant. Additionally, calculating these constant coefficients often requires experimental input, limiting the number of treatable elements. Presented by Grimme in 2006, the *DFT-D2* schemes employ a formula coupling ionization potentials with static polarizabilities to calculate dispersion coefficients for many common elements [46]. Despite their shortcomings, especially for alkali and alkali earth atoms, DFT-D2 methods are among the most widely used dispersion corrections in modern DFT usage.

Generally, DFT-D correction schemes diverge at short distances and must therefore be 'damped', as in equation 38.

$$E_{disp} = -\sum_{A,B} \frac{C_6^{AB}}{r_{AB}^6} f(r_{AB}, A, B) \tag{38}$$

Where $f(r_{AB}, A, B)$ is a damping function equaling to one for large distances and decreasing the dispersive energy to zero or a constant for small distances. To avoid causing additional divergence, these damping functions must be carefully adjusted to the chosen XC functional.

Moving further up the 'stairway' of complexity, one may employ so called *non-local* correlation functionals, which add non-local correlations to local correlation functionals and do not rely on external experimental data, with e.g. the *non-local van der Waals density functional* (vdW-DF) promising high accuracies [47]. Going beyond even these methods, the assumption of pairwise additivity must be left behind and many-body dispersion considered, though these methods would be too complex to elaborate on in the scope of this thesis.

2.3.7 Density Functional Tight Binding

The following derivations are functionally based on an article by Koskinen and Mäkinen titled *Density-functional tight-binding for beginners*, published in 2009 [48], albeit with stylistic alterations to keep the overall thesis consistent. The penultimate section regarding DFTB3 is sourced from an article by Gaus *et al.* titled *DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method*, published in 2011 [49].

Despite the many advancements in DFT and every increasing chemical accuracy, faster and computationally less demanding methods are always subject of research with the goal of reaching longer timeframes of simulation or upscale the size of studied systems. Building on the principles of *tight-binding* theories [50], *Density Functional Tight Binding* (DFTB) is one such method particularly useful for total energy calculations. To start, the Kohn-Sham energy already presented in a more compact formalism in equation 24 is expanded upon and written as in equation 39.

$$E[\rho(\vec{r})] = T_s + E_{ext} + E_H + E_{XC} + E_{II} \tag{39}$$

Where T_s is, again, the non-interacting kinetic energy, E_{ext} the external potential energy, E_H the Hartree energy, E_{XC} the exchange-correlation energy and E_{II} the ion-ion interaction energy. The exchange energy may be expressed as in equation 40, showing how it contains all 'leftover' many-body effects.

$$E_{XC} = (T - T_s) + (E_{ee} - E_H) \tag{40}$$

Where T is the total kinetic energy and E_{ee} the total electron-electron interaction energy. More explicitly, equation 39 may be written as in equation 41.

$$E[\rho] = \sum_a f_a \langle \phi_a | (-\frac{1}{2}\nabla^2 + \int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r}) | \phi_a \rangle + \frac{1}{2}\int\int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' + E_{XC}[\rho] + E_{II} \tag{41}$$

Where a indexes electronic eigenstates ϕ and $f_a \in [0, 2]$ is the occupation number of such a single-particle state with energy ϵ_a. In the next step, one considers a reference density

$\rho_0(\vec{r})$ composed of atomic densities that does not minimize equation 41 but is assumed to neighbour the true minimal density as in equation 42.

$$\rho(\vec{r}) = \rho_0(\vec{r}) + \delta\rho(\vec{r}) \tag{42}$$

Where $\delta\rho(\vec{r})$ is a small fluctuation of the electron density. Expanding the Energy at $\rho_0(\vec{r})$ to the second order in fluctuations then leads to equation 43.

$$\begin{aligned} E[\delta\rho] \approx & \sum_a f_a \langle\psi_a| -\frac{1}{2}\nabla^2 + V_{ext} + V_H[\rho_0] + V_{XC}[\rho_0]|\psi_a\rangle \\ & + \frac{1}{2}\int\int\left(\frac{\delta^2 E_{XC}[\rho_0]}{\delta\rho(\vec{r})\delta\rho(\vec{r}')}\right)\delta\rho(\vec{r})\delta\rho(\vec{r}')d\vec{r}d\vec{r} \\ & - \frac{1}{2}\int V_H[\rho_0]\delta\rho(\vec{r})d\vec{r} + E_{XC}[\rho_0] + E_{II} - \int V_{XC}[\rho_0](\vec{r})\rho_0(\vec{r})d\vec{r} \end{aligned} \tag{43}$$

The first line in the above equation 43 is the band-structure energy of the system as in equation 44.

$$E_{BS} = \sum_a f_a \langle\psi_a|H[\rho_0]|\psi_a\rangle \tag{44}$$

The second line is energy stemming from charge fluctuations composed of coulombic and exchange-interaction contributions, shortened to E_{coul}. Lastly, the four terms in the third line are collectively called repulsive energy, referred to as E_{rep}, due to the ion-ion repulsion term being present. Contracting equation 43 using this formalism leads to equation 45.

$$E[\delta\rho] \approx E_{BS} + E_{coul} + E_{rep} \tag{45}$$

The tight-binding approach is mainly concerned with valence electrons, as core electrons are contained in the repulsive energy term. Additionally, tight-binding assumes, as the name implies, tightly bound electrons. As such, the atomic orbitals ϕ_a are approximated as in equation 46 via a minimal local basis, i.e. only one radial function per angular momentum state.

$$\psi_a(\vec{r}) = \sum_{\mu} c_{\mu}^{a} \phi_{\mu}(\vec{r}) \tag{46}$$

With μ indexing the pseudo-atomic basis states ϕ and ψ_a being an electronic eigenstate. Applying this expansion to equation 44 leads to equation 47.

$$E_{BS} = \sum_{a} f_a \sum_{\mu\nu} c_{\mu}^{a'} c_{\nu}^{a} H_{\mu\nu}^{0} \tag{47}$$
$$H_{\mu\nu}^{0} = \langle \phi_{\mu} | H^{0} | \phi_{\nu} \rangle$$

With $H_{\mu\nu}^{0}$ being the Hamiltonian matrix elements of the reference system. The values of these matrix elements are pre-computed and stored, leading to a significant increase in computational efficiency and forming the core of the tight-binding approach. In addition, one must look at atomic populations utilizing an overlap of orbitals, as in equation 48.

$$q_I = \sum_{a} f_a \sum_{\mu \in I} \sum_{\nu} \frac{1}{2} (c_{\mu}^{a*} c_{\nu}^{a} + c_{\nu}^{a*} c_{\mu}^{a}) S_{\mu\nu} \tag{48}$$
$$S_{\mu\nu} = \langle \psi_{\mu} | \psi_{\nu} \rangle$$

Where q_I is the population or charge of atom I, S the overlap matrix, μ and ν again indexing the atomic basis states and a denoting electronic eigenstates. The difference in charge to a neutral atom Δq_I is then calculated as in equation 49.

$$\Delta q_I = q_I - q_I^{0} \tag{49}$$

With q_I^0 being the number of valence electrons in the neutral atom. Applying undetermined Lagrane multipliers to constrain the wave function norms, one arrives at the Kohn-Sham equation-equivalent of DFTB, as shown in equation 50.

$$H_{\mu\nu} = H_{\mu\nu}^{0} + \frac{1}{2} S_{\mu\nu} \sum_{K} (\gamma_{IK} + \gamma_{JK}) \Delta q_K \quad \mu \in I, \nu \in J \tag{50}$$

Writing the electrostatic potential on atom I due to charge fluctuations as in equation 51 and $h_{\mu\nu}^{1}$ as in equation 52, one may contract the above equation 50 to equation 53.

$$\epsilon_I = \sum_K \gamma_{IK} \Delta q_K \tag{51}$$

$$h^1_{\mu\nu} = \frac{1}{2}(\epsilon_I + \epsilon_J) \quad \mu \in I, \nu \in J \tag{52}$$

$$H_{\mu\nu} = H^0_{\mu\nu} + h^1_{\mu\nu} S_{\mu\nu} \tag{53}$$

As in the DFT equivalent, equation 53 is solved iteratively until convergence is reached. This, while suffering from many of the same convergence issues as DFT, is usually notably faster and more efficient than its parent methods [51].

Apart from the tight-binding formalism, DFTB also approaches the collected term referred to as repulsive energy as well as the coulombic term defined in equation 45 via semi-empirical approximations. Firstly, for the repulsive term containing complicated exchange interactions, a repulsive function depending only on atomic numbers, as in equation 54, is fitted to higher-level theoretical calculations *via*, mostly cubic, splining.

$$E_{rep} = \sum_{I<J} V^{IJ}_{rep}(R_{IJ}) \tag{54}$$

Where V^{IJ}_{rep} is the repulsive potential between atoms I and J at a distance R_{IJ}.

A final approximation in the DFTB formalism happens at the coulombic term, where the charge fluctuation is written as in equation 55.

$$E_{coul} = \frac{1}{2} \sum_{IJ} \gamma_{IJ}(R_{IJ}) \Delta q_I \Delta q_J \tag{55}$$

$$\gamma_{IJ}(R_{IJ}) = \begin{cases} U_I & \text{for } I = J \\ \frac{erf(C_{IJ} R_{IJ})}{R_{IJ}} & \text{for } I \neq J \end{cases}$$

Where γ_{IJ} is a Coulomb-like function with dampening applied at small atomic distances R_{IJ}. This function γ is equal to the so-called Hubbard parameter U_I when $I = J$, which itself is twice the chemical hardness of an atom. In short, one estimates the size of the charge distribution from absolute hardness, which in turn is used to estimate Coloumb

interactions between the atoms. In this case, the Hubbard parameter U_I for each element is taken from standard tables to determine charge transfer.

With all previously mentioned approximations in place, one finally arrives at the DFTB2 formalism, named so since the Kohn-Sham energy is expanded up to the second order. These semi-empirical parametres discussed above are stored in so called *Slater-Koster* files, one for each distinct permutation of atom-atom interaction. In summary, these contain Hamiltonian matrix elements and overlap integrals for all relevant orbital interactions, fitted repulsive potentials as well as Hubbar-parameters for each element. Mainly due to this parametrisation, DFTB2 displays faster convergence than popular DFT methods as well as high transferability, making it a popular choice for large-scale simulations [52].

On a final note, the problem of chemical hardness being a constant irrespective of charge state as well as atoms being restricted to fixed shapes via their initial reference densities in DFTB2 is adressed by including the third order expansion of the Kohn-Sham energy, as seen in equation 56.

$$E^{3rd} \approx \frac{1}{6}\sum_a \delta q_a^3 \frac{\delta \gamma_{aa}}{\delta q_a}\bigg|_{q_a^0} + \frac{1}{6}\sum_{a\neq b} \Delta q_a \Delta q_b \left(\Delta q_a \frac{\delta \gamma_{ab}}{\delta q_a}\bigg|_{q_a^0} + \Delta q_b \frac{\delta \gamma_{ab}}{\delta q_b}\bigg|_{q_b^0} \right) \tag{56}$$

Where $\frac{\delta \gamma}{\delta q}$ describes the charge dependancy of the Hubbard parameter U as in equation 57.

$$\frac{\delta \gamma_{ab}}{\delta q_a}\bigg|_{q_a^0} = \frac{\delta \gamma_{ab} \delta U_a}{\delta U_a \delta q_a}\bigg|_{q_a^0} \quad with \quad a \neq b \tag{57}$$

Including this third order expansion leads to the DFTB3 formalism, shown to allow for more flexible description of interatomic electrostatics and charge transfer, as well as a more accurate description of chemical hardness [53]. In addition to the parametres above, Slater-Koster files for DFTB3 based calculations also contain these Hubbard derivatives, a very popular variant of which is the 3ob parameter set particularly useful for organic systems [54].

2.4 Molecular Dynamics Simulations

Alongside advancements in the field of digital electronics came the possibilities of simulating chemical systems on ever larger scales. Where practical experiments are limited

by equipment, time, resources and safety regulations, simulations can be conducted cost effectively and with an ever increasing level of precision. Simulating a chemical system to receive meaningful information starts with choosing the most viable theoretical framework. *Molecular Dynamics* (MD), utilizing the DFTB method, is one of those frameworks particularly well suited for the theoretical time-dependant analysis of periodic chemical systems. From the output of such a framework, macroscopic properties and structural information of a given system may be inferred using the principles of statistical mechanics. In this chapter, the most relevant aspects of MD simulations for the scope of this thesis will be discussed, mostly based on *Computer Simulation of Liquids* by Allen and Tildesley unless cited otherwise, again with stylistic changes for the sake of consistency [55].

2.4.1 Statistical Mechanics - Ensembles

Unless calculating a system with no kinetic energy, i.e. at absolute zero and at a state of least entropy, the various states of a chemical system during a simulation will cover a variety of conformations and, therefore, potential and kinetic energies. The potential energy of such a system is a function of the positions of all individual atoms, leading to what is called a potential energy surface (PES), in essence a 3N-dimensional function where N is the number of atoms in the system. To extract macroscopic properties from such a system, however, one must also consider the velocities each atom has at specific points in time, as the total energy of a closed system is the sum of kinetic and potential energies as in equation 58

$$E_{tot}(\mathbf{Q}, \mathbf{P}) = K(\mathbf{P}) + V(\mathbf{Q}) \tag{58}$$

Where $\mathbf{Q}$ and $\mathbf{P}$ are the position and momenta vectors respectively. The positions and momenta of all atoms are thought of as coordinates in a multidimensional space with 6N dimensions, the *phase space*. Over the course of a simulation, many different phase space vectors are sampled, leading to what is referred to as an *ensemble* of configurations weighted by their probability of occuring. Such an ensemble may be regarded as a collection of all accessible points Θ in phase space, with different ensembles defined by which macroscopic parametres (volume, pressure, temperature, ...) are fixed.

Given a particular point in phase space, one may simply calculate the microscopic value

of some property as a function of the phase space vector. According to the principles of the *Ergodic hypothesis*, a central approximation in statistical analysis, the time average of a process parameter equals the observable equilibrium value of said property on a macroscopic scale as desribed in equation 59, so long as sufficiently long sampling times are employed. This assumption must be made as MD simulations, limited by computational resources, cannot sample system sizes even close to real world scales and is key to extracting macroscopic properties. In such an *ergodic system*, an equal amount of time is spent in each possible phase state [56].

$$A_{obs} = \langle A \rangle_{time} = \big\langle A\big(\mathbf{\Theta}(t)\big)\big\rangle_{time} = \lim_{t_{obs}\to\infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A\big(\mathbf{\Theta}(t)\big)dt \tag{59}$$

Where A_{obs} is the observable macroscopic value of a property, A the property itself at a given phase space vector $\mathbf{\Theta}(t)$ at time t and t_{obs} the observation time. As it is impossible to extend the observation time to infinity, equations of motions are usually solved over a large but finite number of equally long timesteps, as in equation 60.

$$A_{obs} = \langle A \rangle_{time} = \frac{1}{\tau_{obs}} \sum_{\tau=1}^{\tau_{obs}} A\big(\mathbf{\Gamma}(t)\big) \tag{60}$$

$$\delta t = \frac{t_{obs}}{\tau_{obs}}$$

Where τ_{obs} is the number of timesteps each with length δt.

2.4.2 The Canonical Ensemble

While the usual choice for conducting MD simulations of metal-organic frameworks might be the *istothermic-isobaric* conditions in the NPT ensemble, the employed parametrization of nitrogen interactions in the DFTB framework has been shown to lead to high instability and often times full system collapse of the porous structure under these conditions. For this reason, simulations in this thesis are treated under constant volume conditions, equating a *canonical* (NVT) ensemble which has previously been employed to good effect in similar studies [57] [58].

Such a canonical ensemble is defined by a fixed number of particles, volume and temperature, which may be understood as coupling a system to an external heat bath to allow

energy exchange. The partition function Q_{NVT} in this ensemble is the sum of probability densities for all possible microstates Γ as in equation 61.

$$Q_{NVT} = \sum_{\mathbf{\Gamma}} \exp\Big(-\frac{H(\mathbf{\Gamma})}{k_B T}\Big) \tag{61}$$

Where $H(\Gamma)$ is the Hamiltonian of the system in a given microstate Γ and k_B the *Helmholtz constant*. Since energy fluctuations are non-zero and the energy is comprised of kinetic and potential components, the partition function can be further factorized into kinetic and potential parts as in equation 62, with h being the *Planck* constant.

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int \exp\Big(-\frac{K(\boldsymbol{P})}{k_B T}\Big) d\boldsymbol{p} \int \exp\Big(-\frac{V(\boldsymbol{Q})}{k_B T}\Big) d\boldsymbol{q} \tag{62}$$

The relevant thermodynamic function for this ensemble is the *Helmholtz free energy* A, defined as in equation 63.

$$A = -k_B T \ln Q_{NVT} \tag{63}$$

2.4.3 Equations of Motion

As molecular dynamics simulation aim to describe the time-dependant behaviour of an N particle system interacting through a potential V, equations of motion must be solved. In MD theory, this is done classically as described in this chapter, starting at the pillar of classical mechanics, *Isaac Newton's* laws of motion. Most importantly for MD simulations, the second law of motion states that acceleration, i.e. the derivative of momentun with respect to time, is proportional and parallel to the net force and inversely proportional to the mass of an object, as in equation 64 [59].

$$\mathbf{a} = \frac{\partial \mathbf{p}}{\partial t} = \frac{\mathbf{F}}{m} \tag{64}$$

Classical mechanics are insufficient at describing relativistic or quantum mechanical systems, but are however suited for the masses and velocities of atoms in typical systems. As such, the mass of particles can be considered constant and, with the fact that Force

is defined as the negative potential gradient as in 65, equation 64 can be rewritten as in equation 66.

$$\mathbf{F} = -\frac{\partial V}{\partial \mathbf{q}} \tag{65}$$

$$\mathbf{a} = \frac{d^2\mathbf{q}}{dt^2} = \frac{\mathbf{F}}{m} \tag{66}$$

The combination of 66 and 65 finally leads to equation 67, revealing the Newtonian EOM as a differential equation of second order.

$$\mathbf{F} = m\frac{d^2\mathbf{q}}{dt^2} = -\frac{\partial V}{\partial \mathbf{q}} \tag{67}$$

Solving such a differential equation for N particles to obtain their position vectors $\mathbf{q}$ at time t is, even numerically, completely unfeasible for most system sizes.

A solution to this problem starts at a different framework, namely *Lagrangian* equation sof motion as in 68. Here, the Lagrangian function $L(\mathbf{q}, \dot{\mathbf{q}})$, dependant on spatial coordinates and their time derivatives, is defined as seen in equation 69 as the difference between kinetic and potential energy.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}_k}\right) - \frac{\partial L}{\partial \mathbf{q}_k} = 0 \tag{68}$$

$$L = K - V \tag{69}$$

Where $\mathbf{q}_k$ and $\dot{\mathbf{q}}_k$ are the k-th generalized coordinates and their time derivatives respectively. In combination with the definitions of kinetic energy as in 70 and potential energies as summations over all individual atoms, pairs, triplets, etc., as in equation 71, the Lagrange equation of motion can be rewritten as in equation 72 using Cartesian coordinates $\mathbf{r}_i$.

$$K = \sum_{i=1}^{N} \sum_{\alpha} p_{i\alpha}^2 / 2m_i \quad \text{with} \quad \alpha \in x, y, z \tag{70}$$

$$V = \sum_{i=1}^{N} V_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} V_{ij} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} V_{ijk} + ... \tag{71}$$

$$m_i \ddot{\mathbf{q}}_i = \mathbf{F}_i = -\nabla_{\mathbf{q}_i} V = \nabla_{\mathbf{q}_i} L \tag{72}$$

Where $\mathbf{F}_i$ is the total force acting on molecule i. The generalized momenta p_k are defined as in equation 73.

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \tag{73}$$

Finally, the Hamiltonian equations of motion utilizes the Lagrangian definition of the generalized momentum p_k as in equations 75 and 76, with the classical Hamiltonian strictly defined as in equation 74.

$$H(\mathbf{\Theta}) = H(\mathbf{p}, \mathbf{q}) = \sum_k p_k \dot{q}_k - L(\mathbf{q}, \dot{\mathbf{q}}) \tag{74}$$

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \tag{75}$$

$$\dot{p}_k = -\frac{\partial H}{\partial q_k} \tag{76}$$

Switching to phase-space coordinates, with i indexing individual atoms, Hamiltons equations of motion can be written as in equations 77 and 78.

$$\frac{d\mathbf{q}_i}{dt} = \frac{\partial H}{\partial \mathbf{p}_i} = \frac{\partial E_{kin}}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} = \mathbf{v}_i \tag{77}$$

$$\frac{d\mathbf{p}_i}{dt} = \frac{\partial H}{\partial \mathbf{q}_i} = -\frac{\partial V}{\partial \mathbf{q}_i} = \mathbf{F}_i \tag{78}$$

In contrast to Newton's equations of motion, Hamilton's equations require solving N less complex first order differential equations for momenta $\mathbf{p}(t)$ and position $\mathbf{q}(t)$ vectors each.

2.4.4 The Louiville Formalism

Provided that Hamiltons's equations of motion are satisfied, the total time derivative of the Hamiltonian is given in equation 79 following simple chain rules.

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{q}}\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial H}{\partial \mathbf{p}}\frac{\partial \mathbf{p}}{\partial t} \tag{79}$$

With the total energy H being conserved in an isolated system and equations of motion 77 and 78 established above, equation 79 can be rewritten as in equation 80.

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{q}}\frac{\partial H}{\partial \mathbf{p}} - \frac{\partial H}{\partial \mathbf{p}}\frac{\partial H}{\partial \mathbf{q}} = 0 \tag{80}$$

To abbreviate the above expression, the bilinear *Poisson bracket* operator is introduced as in equation 81 for a general function $G(\mathbf{q}, \mathbf{p})$ dependant on phase space vectors $\mathbf{q}$ and $\mathbf{p}$.

$$\frac{dG}{dt} = \frac{\partial G}{\partial \mathbf{q}}\frac{\partial H}{\partial \mathbf{p}} - \frac{\partial G}{\partial \mathbf{p}}\frac{\partial H}{\partial \mathbf{q}} = \{G, H\} \tag{81}$$

Using this notation, equation 80 may be simply expressed as $\{H, H\} = 0$. Alternatively, the *Liouville operator* iL may be used for classical time propagation as a shorthand for the Poisson bracket as in equation 82, again for a general property G [60].

$$\frac{dG}{dt} = \{G, H\} = iLG \tag{82}$$

With the Liouville operator defined as in equation 83.

$$\frac{d}{dt} = iL = \frac{\partial \mathbf{q}}{\partial t}\frac{\partial}{\partial \mathbf{q}} + \frac{\partial \mathbf{p}}{\partial t}\frac{\partial}{\partial \mathbf{p}} = \frac{\partial H}{\partial t}\frac{\partial}{\partial \mathbf{q}} + \frac{\partial H}{\partial t}\frac{\partial}{\partial \mathbf{p}} = \{, H\} \tag{83}$$

Applying this operator to the phase space vector $\mathbf{\Theta}$ allows time evolution, with the solution to the Liouville equation being the phase space vector at a later time t as in equation 84.

$$\mathbf{\Theta}(t) = e^{iLt}\mathbf{\Theta}(0) \tag{84}$$

Where $\mathbf{\Theta}(0)$ is the phase space vector at time $t = 0$. Splitting the Liouville operator into a configurational part and a momentum part leads to equation 85.

$$iL = iL_{\mathbf{q}} + iL_{\mathbf{p}} \tag{85}$$

It is important to note that the Liouville operators do not commute and the exponentials may not be simply split, since order of application is not arbitrary, as shown in equation 86.

$$e^{iLt} = e^{iL_{\mathbf{q}}t + iL_{\mathbf{p}}t} \neq e^{iL_{\mathbf{p}}t} e^{iL_{\mathbf{q}}t} \tag{86}$$

Because the individual operators $exp(iL_{\mathbf{q}}t)$ and $exp(iL_{\mathbf{p}}t)$ can, in many instances, be evaluated exactly on a phase space vector $\mathbf{\Theta}$, it would be useful to express the time evolution *via* both these factors. A solution to this problem is presented in the following chapter.

2.4.5 The Trotter Theorem

The derivations in this chapter are based on *Statistical Mechanics: Theory and Molecular Simulation* by Tuckerman [61].

Formulated by Hale Trotter in 1959 [62], the *Trotter theorem* states that for two non commuting operators A and B, the following equation 87 holds true.

$$e^{A+B} = \lim_{N\to\infty} \left[e^{B/2N} e^{A/N} e^{B/2N} \right]^P \tag{87}$$

Where N is an integer. Applying this general theorem to the classical operator established before, and by defining a time step $\Delta t = t/N$, one arrives at equation 88.

$$e^{iLt} = \lim_{N\to\infty, \Delta t\to 0} \left[e^{iL_{\mathbf{p}}t/2N} e^{iL_{\mathbf{q}}t/N} e^{iL_{\mathbf{p}}t/2N} \right]^N \tag{88}$$

This equation states that a classical system may be propagated in time *via* the separate factors $\exp(iL_{\mathbf{q}}t/N)$ and $\exp(iL_{\mathbf{p}}t/N)$ exactly for a time t, granted that the number of steps taken P are infinite and the timesteps Δt are infinitesimally small. For obvious

reasons, these limits are unreachable in practice, hence an approximation to $\exp(iLt)$ is made as shown in equation 89.

$$e^{iLt} \approx \left[e^{iL_{\mathbf{p}}\Delta t/2} e^{iL_{\mathbf{q}}\Delta t} e^{iL_{\mathbf{p}}\Delta t/2}\right]^N \tag{89}$$

An approximate time propagation may then be achieved by performing N steps of length Δt using the factorized opator seen in equation 90.

$$e^{iL\Delta t} \approx e^{iL_{\mathbf{p}}\Delta t/2} e^{iL_{\mathbf{q}}\Delta t} e^{iL_{\mathbf{p}}\Delta t/2} \tag{90}$$

At this point it may be interesting to note that while the error of an individual step n is proportional to Δt^3, the global error of a trajectory only scales with Δt^2 [55]. Applying Trotter splitting now to the phase space vector and inserting into equation 84 yields the time evolution of the phase space vector $\mathbf{\Theta}$ as in equation 91.

$$\mathbf{\Theta}(t + \Delta t) = e^{iL\Delta t}\mathbf{\Theta}(t) \approx e^{iL_{\mathbf{p}}\Delta t/2} e^{iL_{\mathbf{q}}\Delta t} e^{iL_{\mathbf{p}}\Delta t/2}\mathbf{\Theta}(t) \tag{91}$$

2.4.6 The Velocity Verlet Algorithm

Derived from the Trotter approach, a widely used numerical time integrator has been derived, the *Verlet* algorithm [63]. A more commonly used and demonstrative iteration of this algorithm has been developed by Swope et al. [64], with the main deviation being that velocities and positions are calculated at the same value of time in this approach. This algorithm, starting from initial atomic positions, velocities and forces, acts over a timestep from t to $t + \Delta t$ as in equations 92, 93 and 94.

$$\mathbf{v}\Big(t + \frac{\Delta t}{2}\Big) = \mathbf{v}(t) + \frac{\Delta t}{2}\,\mathbf{a}(t) = \mathbf{v}(t) + \frac{\Delta t}{2}\frac{\mathbf{F}_t}{m} \tag{92}$$

$$\mathbf{q}\Big(t + \Delta t\Big) = \mathbf{q}(t) + \Delta t\;\mathbf{v}\Big(t + \frac{\Delta t}{2}\Big) \tag{93}$$

$$\mathbf{v}\Big(t + \Delta t\Big) = \mathbf{v}\Big(t + \frac{\Delta t}{2}\Big) + \frac{\Delta t}{2}\,\mathbf{a}\Big(t + \Delta t\Big) = \mathbf{v}\Big(t + \frac{\Delta t}{2}\Big) + \frac{\Delta t}{2}\frac{\mathbf{F}_{t+\Delta t}}{m} \tag{94}$$

At first, equation 92 may be thought of as 'half-advancing' the initial velocity to a midpoint

time $t + \Delta t/2$ using accelerations at time t. These midpoint velocities are then used as in equation 93 to propel the coordinates to time $t + \Delta t$. Finally, the endpoint velocities at time $t + \Delta t$ are calculated as in equation 94 using the endpoint forces as retrieved *via* equation 95 from the respective endpoint positions.

$$\mathbf{F}_{t+\Delta t} = -\frac{\partial V}{\partial \mathbf{q}_{t+\Delta t}} \tag{95}$$

Easily translatable into high-performance code, this algorithm allows for extracting macroscopic properties like e.g. total kinetic energy by simply summing up the squared velocities at a given time, whereas the potential energy is calculated in each step anyway and can simply be stored as output.

2.4.7 Constraint Dynamics

As mentioned before, the error over the total trajectory scales with Δt^2, leading to the setup of a molecular dynamics simulation becoming a balance act of choosing a small enough timestep to accurately represent even the fastest movements of a system with low error, while still being large enough to keep total simulation times in a reasonable frame. Usually, the fastest movement in any given system are X-H bond vibrations, typically oscillating at a period of around 10 fs [65]. Representing such movement then dictates going well beyond this timeframe up to an order of magnitude, e.g. 1 fs, vastly increasing computational cost [66].

Luckily, should X-H bond vibrations be of no special concern, constraining certain bonds to a pre-determined length is a viable and commonly used way of increasing timestep size without losing out on vital information. The original method of constraint schemes designed to work in tandem with the verlet operator is the SHAKE algorithm by Ryckaert *et al.* [67]. Further developed for the velocity verlet variant by Andersen, the RATTLE algorithm [68] utilizes Lagrange multipliers as demonstrated on a water molecule with fixed OH bonds in the following [55]. Oxygen is indexed with 2, whereas the two Hydrogen atoms are labelled 1 and 3 respectively. The constraint equations for the OH bonds are as in equations 96 and 97.

$$\chi_{12}^{(q)}(t) = |\mathbf{q}_{12}(t)|^2 - d_{12}^2 = 0, \quad \chi_{23}^{(q)}(t) = |\mathbf{q}_{23}(t)|^2 - d_{23}^2 = 0 \tag{96}$$

$$\chi_{12}^{(p)}(t) = \mathbf{q}_{12}(t) \cdot \mathbf{p}_{12}(t) = 0, \quad \chi_{23}^{(p)}(t) = \mathbf{q}_{23}(t) \cdot \mathbf{p}_{23}(t) = 0 \tag{97}$$

Where $\mathbf{q}_{ij}$ and $\mathbf{p}_{ij}$ are the relative position and momenta vectors of atoms i and j respectively, and d_{ij} is the desired bond length. Equations of motion for constrained atoms now take the form of equation 98.

$$m_i \mathbf{a}_i = \mathbf{F}_i + \mathbf{G}_i \tag{98}$$

Where $\mathbf{G}_i$ is the constraint force acting on atom i ensuring that the above constraint equations are satisfied at all times, defined as in equation 99.

$$\mathbf{G}_i = \Lambda_{12} \nabla_{\mathbf{q}_{12}} \chi_{12}^{(q)} + \Lambda_{23} \nabla_{\mathbf{q}_{23}} \chi_{23}^{(q)} \tag{99}$$

Where Λ_{ij} are the undetermined Lagrange multipliers. Suggested by Ryckaert *et al.*, the constraint forces are calculated so as to guarantee them being satisfied at each timestep, leading to the same error margins as the integrator they are applied to. Considering the first half-step of the velocity Verlet algorithm, equations 100 and 101 show the non-constrained absolute positions and velocities at time $t + \Delta t/2$.

$$\mathbf{v}_i'\Big(t + \frac{\Delta t}{2}\Big) = \mathbf{v}_i(t) + \frac{\Delta t}{2m_i}\mathbf{F}_i(t) \tag{100}$$

$$\mathbf{q}_i'(t + \Delta t) = \mathbf{q}(t) + \Delta t \ \mathbf{v}_i'\Big(t + \frac{\Delta t}{2}\Big) = \mathbf{q}(t) + \Delta t \ \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{101}$$

Where the prime denotes non-constrained values. Including constraint forces concerning positions then leads to equation 102.

$$\mathbf{q}_i(t + \Delta t) = \mathbf{q}_i'(t + \Delta t) + \frac{\Delta t^2}{2m_i}\mathbf{G}_i^{(\mathbf{q})}(t) \tag{102}$$

The constraint forces $\mathbf{G}_i^{(\mathbf{q})}$ acting on the three atoms must act along the bond vectors $\mathbf{q}_{12}$

and $\mathbf{q}_{23}$, as well as adhere to Newton's third law, as in equations 103, 104 and 105.

$$\Big(\frac{1}{2}\Delta t^2\Big)\mathbf{G}_1^{(\mathbf{q})}(t) = \lambda_{12}\mathbf{q}_{12}(t) \tag{103}$$

$$\Big(\frac{1}{2}\Delta t^2\Big)\mathbf{G}_2^{(\mathbf{q})}(t) = -\lambda_{12}\mathbf{q}_{12}(t) + \lambda_{23}\mathbf{q}_{23}(t) \tag{104}$$

$$\Big(\frac{1}{2}\Delta t^2\Big)\mathbf{G}_3^{(\mathbf{q})}(t) = -\lambda_{23}\mathbf{q}_{23}(t) \tag{105}$$

Where λ_{ij} are the undetermined Lagrange multipliers. Inserting into equation 102, bond vectors at the point $t + \Delta t$ are then calculated as in equations 106 and 107. $\mathbf{q}'_{ij}$ again denotes the non-constrained bond vectors.

$$\mathbf{q}_{12}(t + \Delta t) = \mathbf{q}'_{12}(t + \Delta t) + (m_1^{-1} + m_2^{-1})\lambda_{12}^{(\mathbf{q})}\mathbf{q}_{12}(t) - m_2^{-1}\lambda_{23}^{(\mathbf{q})}\mathbf{q}_{23}(t) \tag{106}$$

$$\mathbf{q}_{23}(t + \Delta t) = \mathbf{q}'_{23}(t + \Delta t) - m_2^{-1}\lambda_{12}^{(\mathbf{q})}\mathbf{q}_{12}(t) + (m_2^{-1} + m_3^{-1})\lambda_{23}^{(\mathbf{q})}\mathbf{q}_{23}(t) \tag{107}$$

Taking then the square modulus of both sides and applying the constraints of the form $|\mathbf{q}_{ij}(t + \Delta t)|^2 = |\mathbf{q}_{ij}(t)|^2 = d_{ij}^2$ as defined in equation 96 results in a pair of quadratic equations, iteratively solvable for the two undetermined multipliers λ_{12} and λ_{23}. These multipliers may then be inserted into equations 103, 104 and 105 to receive the constraint forces acting on the positions, which in turn allow absolute atomic positions at time $t+\Delta t$ to be calculated *via* equation 102.

In the same manner, unconstrained half-step velocities, equation 100, are adjusted according to equation 108.

$$\mathbf{v}_i\Big(t + \frac{\Delta t}{2}\Big) = \mathbf{v}'_i\Big(t + \frac{\Delta t}{2}\Big) + \frac{\Delta t}{2m_i}\mathbf{G}_i^{(\mathbf{q})}(t) \tag{108}$$

Leading to the constrained full-step velocities as in equation 109.

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i\Big(t + \frac{\Delta t}{2}\Big) + \frac{\Delta t}{2m_i}\mathbf{F}_i(t + \Delta t) + \frac{\Delta t}{2m_i}\mathbf{G}_i^{(\mathbf{v})}(t + \Delta t) \tag{109}$$

Where $\mathbf{G}_i^{(\mathbf{v})}$ are the constraint forces acting on the velocities, which again are directed

along the bond vectors, leading to a similar set of equations as for coordinates before, shown in equations 110, 111 and 112.

$$\left(\frac{1}{2}\Delta t\right)\mathbf{G}_1^{(\mathbf{v})}(t+\Delta t) = \lambda_{12}^{(\mathbf{v})}\mathbf{q}_{12}(t+\Delta t) \tag{110}$$

$$\left(\frac{1}{2}\Delta t\right)\mathbf{G}_2^{(\mathbf{v})}(t+\Delta t) = -\lambda_{12}^{(\mathbf{v})}\mathbf{q}_{12}(t+\Delta t) + \lambda_{23}^{(\mathbf{v})}\mathbf{q}_{23}(t+\Delta t) \tag{111}$$

$$\left(\frac{1}{2}\Delta t\right)\mathbf{G}_3^{(\mathbf{v})}(t+\Delta t) = -\lambda_{23}^{(\mathbf{v})}\mathbf{q}_{23}(t+\Delta t) \tag{112}$$

Inserting these expressions for constraint forces into the equation for full-step constrained velocities, equation 109, leads to a new set of equations for the undetermined multipliers $\lambda_{12}^{(\mathbf{v})}$ and $\lambda_{23}^{(\mathbf{v})}$, 113 and 114.

$$\mathbf{v}_{12}(t+\Delta t) = \mathbf{v}'_{12}(t+\Delta t) + (m_1^{-1} + m_2^{-1})\lambda_{12}^{(\mathbf{v})}\mathbf{q}_{12}(t) - m_2^{-1}\lambda_{23}^{(\mathbf{v})}\mathbf{q}_{23}(t) \tag{113}$$

$$\mathbf{v}_{23}(t+\Delta t) = \mathbf{v}'_{23}(t+\Delta t) - m_2^{-1}\lambda_{12}^{(\mathbf{v})}\mathbf{q}_{12}(t) + (m_2^{-1} + m_3^{-1})\lambda_{23}^{(\mathbf{v})}\mathbf{q}_{23}(t) \tag{114}$$

With bond vectors $\mathbf{q}_{ij}$ already determined in the steps before. Since the constraint equations 97 are scalar products, the above equations are linear in the unknowns λ_{12} and λ_{23}, which in turn are used in equations 110, 111 and 112 to receive constraint forces and subsequently equation 109 for absolute velocities at time $t + \Delta t$.

2.4.8 Periodic Boundary Conditions

While simulating small or few molecules is a computationally inexpensive and generally fast endeavor, especially on modern hardware, properties of e.g. liquids in bulk may not be accurately extrapolated from small-scale systems. Increasing then the number of molecules towards a system more representative of a macroscopic bulk however leads to both unfeasibly high costs and a significant number of molecules being present on the bulks surface, experiencing different potentials and therefore skewing the overall simulation.

To counteract these surface effects, *periodic boundary conditions* may be employed [55, 61, 69]. The system being modeled is hereby assumed to be a cubic unit cell of an ideal

and infinite crystal, replicated periodically and infinitely in all three dimensions, thus eliminating surface effects. As a particle moves, so do its ifninite mirror particles, and should it leave the box, one of its images will instantly replace it on the opposite face, keeping the number of particles constant. When employed, distances between particles are calculated *via minimum image convention*, stating that in the infinitely repeating array of unit cells, a particle i interacts with another particle j to which it is closest, even if that particle is located in one of the neighboring periodic cells. For any given dimension, this convention may be expressed as in equation 115.

$$\Delta r'_{ij,\alpha} = r_{ij,\alpha} - L_\alpha \cdot \text{round}\Big(\frac{r_{ij,\alpha}}{L_\alpha}\Big) \quad \text{with} \quad \alpha \in x, y, z \tag{115}$$

Where $r_{ij,\alpha}$ is the non-imaged distance between particles i and j, L_α is the length of the unit cell and $\Delta r'_{ij,\alpha}$ is the imaged distance between particles i and j, each in dimension α. The round function rounds the fraction to the nearest integer.

A problem arising with periodicity is the need to limit the number of interactions to be calculated, since otherwise one would theoretically have to consider an infinite number of interactions. This is usually done by defininining a cutoff radius r_{cut}, at most as long as half the length of the unit cell, beyond which interactions are considered negligible. To smoothly taper interactions around the cutoff region, switching functions are commonly employed.

As a final trick of the trade to increase efficiency, a list of particles j within the cutoff radius of particle i may be defined, known as a *Verlet neighbor list*. Updated periodically, this list ensures that an algorithm does not have to loop over every pair ij in the system at every step to determine if they are within the cutoff radius, but only over the pairs ij in the list. Computational overhead is thus reduced indirectly proportional to however often the list is updated, at the cost of increased memory usage.

2.4.9 Temperature Control

In an NVT ensemble, temperature coupling is employed to scale the temperature of a system to a pre-defined value, representing the coupling of a system to an external heat bath. As the temperature of a given system is the sum of the kinetic energies of all

particles, temperature coupling algorithms modify the velocities of particles to achieve the desired target temperature. In its simplest form, velocities are scaled in each timestep by a factor λ as in equation 116, calculated as in equation 117.

$$\mathbf{v}_i' = \lambda \mathbf{v}_i \tag{116}$$

$$\lambda = \sqrt{\frac{T_{Target}}{T_{System}}} \tag{117}$$

Where v_i and v_i' are the velocities before and after scaling, T_{Target} is the desired temperature and T_{System} is the temperature of the system. Such a *strong coupling* algorithm is computationally inexpensive, but leads to highly unrealistic trajectories as any movement is immediately dampened. A more sensible approach is the *weak coupling* algorithm, presented by Berendsen *et al.* [70], where the rescaling factor λ is balanced with a latency factor τ representing time-lagged heat exchange to the external heat bath. The rescaling factor λ is then calculated as in equation 118.

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{Target}}{T_{System}} - 1\right)} \tag{118}$$

Where Δt is the timestep and τ is the coupling time constant, usually in the range of 0.1 to 1 ps. Since this also suppresses kinetic fluctuations that could be considered natural and physically realistic, the Berendsen thermostat does, stricty speaking, not represent a true NVT ensemble. It has however been shown that the approximation is sufficient for large enough systems in the order of hundreds of particles [71]. A number of improvements to thermic coupling have been developed since, with notable examples being the *Bussi-Donadio-Parrinello* thermostat [72] and the *Nosé-Hoover* thermostat [73], which will both however not be explored in this thesis.

3 Results and Discussion

3.1 Calculation Setup

Settings for the SCC DFTB MD simulation approach as well as information on general workflow are provided in the following.

3.1.1 SCC DFTB Setup and Parametrization

Based on previous succesful simulations of MIL-68(Ga) [52, 57, 58], self-consistent charge density functional tight binding (SCC DFTB) [74] of 2nd-order parametrization (DFTB2) was employed in this work, as implemented in the DFTB+ package [75] and utilizing the mio/hyb parameter set [76–79]. The mio/hyb parameter set was extended for Ga-N interactions based on the workflow presented by *Van den Bossche et al.* using the hotcent tool [80, 81] and Perdew-Burke-Enzerhofer functional [82], with the confinement parameters for Ga and N taken as provided by mio/hyb literature. Repulsive interactions where then fitted to a potential energy scan reference of planar $Ga(NH_2)_3$ at the B3LYP/SDD+6-31G(d) level of theory calculated with Gaussian16 [83–88] and using the tango package [89].

3.1.2 Molecular Dynamics Protocol

For the SCC DFTB MD simulation, the DFTB+ package was used in conjunction with in-house MD routines, used to satisfying effect in previous works regarding solid-state systems [90–92]. Time propagation was performed at steps of 2.0 fs *via* the previously described velocity verlet propagator, made possible by employing the SHAKE/RATTLE algorithm to constrain hydrogen-containing bonds.

As shown in previous works and touched on briefly in the theory section, MIL-68(Ga) undergoes a phase transition under constant pressure conditions, resulting in structural collapse [57, 58] and is therefore subjected to the NVT conditions of the canonical ensemble. A constant temperature of 298.15 K was then achieved by a Bussi-Donadio-Parrinello thermostat algorithm with a relaxation time of 0.1 ps.

Visualizations of configurations were generated using the *visual molecular dynamics* (VMD) package [93], with the initial configurations generated from pristine MIL-68(Ga) structures of previous works [57]. For the different guest-loaded systems, the 4,5-dimethoxy-N-butylphthalamide guests were randomly placed into the large pores of the pristine MOF system while maintaining a sensible initial distances of at least 1.5 Å between the guest and host atoms.

3.2 Interaction Energy

3.2.1 Guest Dimers

To investigate the interaction energies and relative orientations of a homo-dimer of 4,5-dimethoxy-N-butylphthalamide guests, two idential molecules were placed in a cubic box with an edge length of 100 Å to simulate a vacuum environment at an intermolecular distance of around 3.5 Å. Sampling was then done under the conditions described in 3.1.2 for 300ps. Similarly, a singular guest molecule was sampled under the same conditions for a total of 200 ps. The interaction energies U_{int} were then calculated by subtracting the time-averaged energy of the monomer$< U_{mono} >$ twice from the running internaa energy of the dimer U_{dimer}, as described in equation **??**.

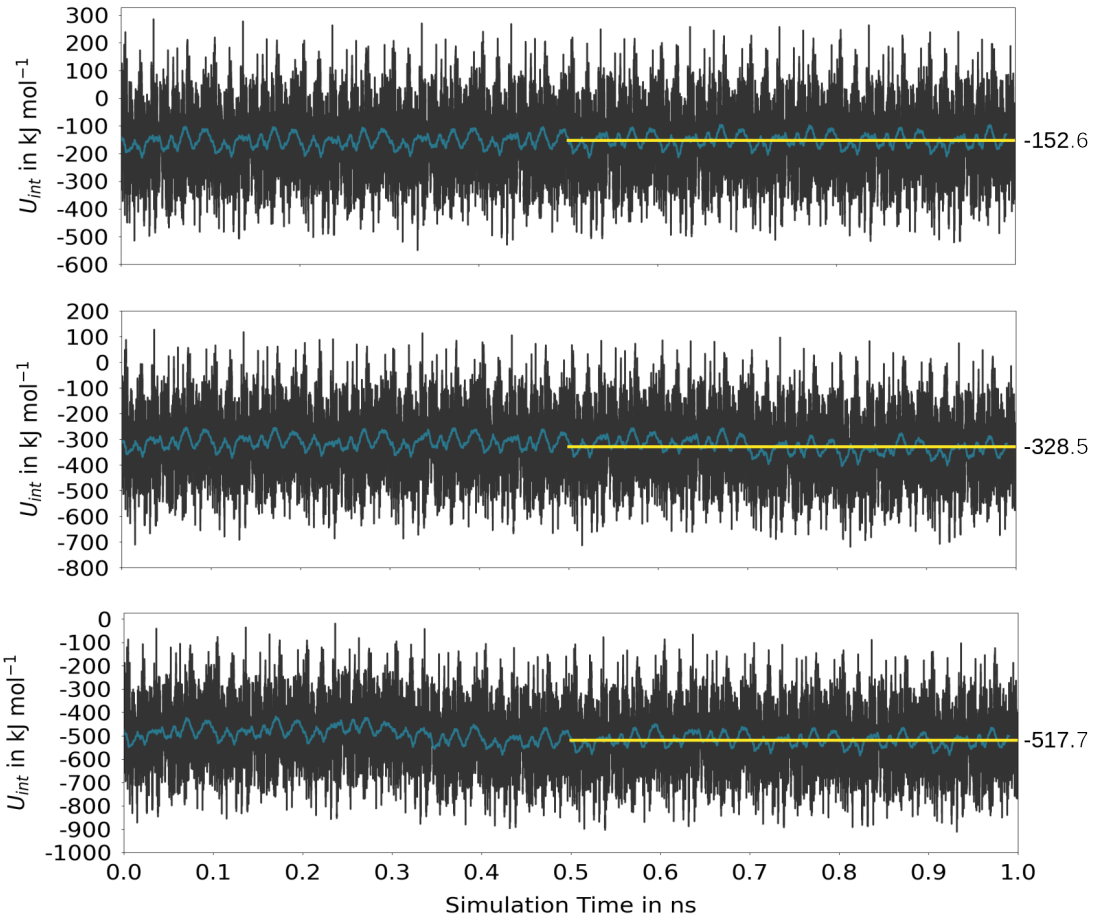

Figure 1: Caption for Image 1

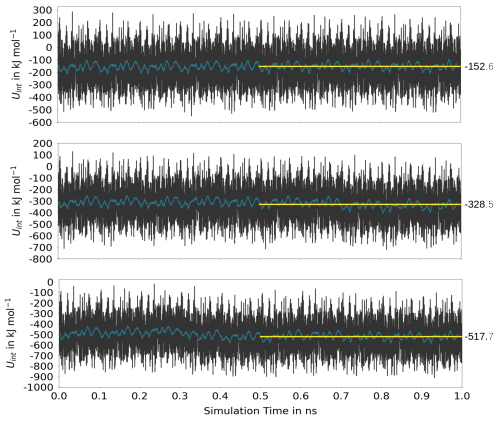

Figure 2: Caption for Image 1

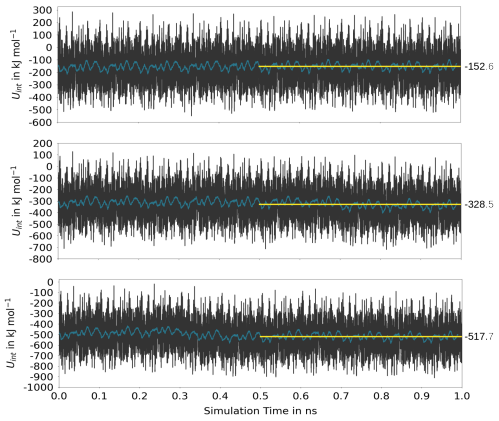

Figure 3: Caption for Image 2

References

[1] Max Planck. Über das Gesetz der Energieverteilung im Normalspectrum. *Annalen der Physik*, 309(3):553–563, January 1901.

[2] A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6):132–148, January 1905.

[3] N. Bohr. Lxxiii. on the constitution of atoms and molecules. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 26(155):857–875, November 1913.

[4] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3–4):172–198, March 1927.

[5] Louis de Broglie. Xxxv. a tentative theory of light quanta. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47(278):446–458, February 1924.

[6] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28(6):1049–1070, December 1926.

[7] E. Schrödinger. Quantisierung als eigenwertproblem. *Annalen der Physik*, 384(4):361–376, January 1926.

[8] Max Born. Zur quantenmechanik der stoßvorgänge. *Zeitschrift für Physik*, 37(12):863–867, December 1926.

[9] Ilia Toli and Shengli Zou. Schrödinger equation with coulomb potential admits no exact solutions. *Chemical Physics Letters*, 737:100021, 2019.

[10] P. A. M. Dirac. A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):416–418, July 1939.

[11] David J. Griffiths. *Introduction to Quantum Mechanics* -. Cambridge University Press, Cambridge, 2017.

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

[13] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, January 1928.

[14] W. Pauli. Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren. *Zeitschrift für Physik*, 31(1):765–783, February 1925.

[15] J. C. Slater. The theory of complex spectra. *Physical Review*, 34(10):1293–1322, November 1929.

[16] J. C. Slater and G. F. Koster. Simplified lcao method for the periodic potential problem. *Physical Review*, 94(6):1498–1524, June 1954.

[17] C David Sherrill. An introduction to hartree-fock molecular orbital theory. *School of Chemistry and Biochemistry Georgia Institute of Technology*, 2000.

[18] S. F. Boys. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 200(1063):542–554, February 1950.

[19] Sigeru Huzinaga. Basis sets for molecular calculations. *Computer Physics Reports*, 2(6):281–339, May 1985.

[20] Felix Bloch. Ueber die quantenmechanik der elektronen in kristallgittern. *Zeitschrift fuer Physik*, 52(7–8):555–600, July 1929.

[21] Neil W Ashcroft and N Mermin. *Solid State Physics*. Brooks/Cole, Florence, KY, January 1976.

[22] Maylis Orio, Dimitrios A. Pantazis, and Frank Neese. Density functional theory. *Photosynthesis Research*, 102(2–3):443–453, February 2009.

[23] Wolfram Koch and Max C Holthausen. *A chemist's guide to density functional theory.* Wiley-VCH Verlag, Weinheim, Germany, 2 edition, July 2001.

[24] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[25] W. Kohn and L. J. Sham. Quantum density oscillations in an inhomogeneous electron gas. *Physical Review*, 137(6A):A1697–A1705, March 1965.

[26] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048–5079, May 1981.

[27] P. A. M. Dirac. Note on exchange phenomena in the thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(3):376–385, July 1930.

[28] R.G. Parr and Y. Weitao. *Density-Functional Theory of Atoms and Molecules.* International Series of Monographs on Chemistry. Oxford University Press, 1994.

[29] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566–569, August 1980.

[30] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, September 1988.

[31] John P. Perdew and Wang Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Physical Review B*, 33(12):8800–8802, June 1986.

[32] John P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822–8824, June 1986.

[33] Kieron Burke, John P. Perdew, and Yue Wang. *Derivation of a Generalized Gradient Approximation: The PW91 Density Functional*, page 81–111. Springer US, 1998.

[34] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, January 1988.

[35] Renato Colle and Oriano Salvetti. Approximate calculation of the correlation energy for the closed shells. *Theoretica Chimica Acta*, 37(4):329–334, 1975.

[36] Axel D. Becke. A new mixing of hartree–fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, January 1993.

[37] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, April 1993.

[38] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, November 1994.

[39] Larry A. Curtiss, Krishnan Raghavachari, Gary W. Trucks, and John A. Pople. Gaussian-2 theory for molecular energies of first- and second-row compounds. *The Journal of Chemical Physics*, 94(11):7221–7230, June 1991.

[40] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, April 1999.

[41] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1–3):215–241, July 2007.

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

[43] Yan Zhao and Donald G Truhlar. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput.*, 3(1):289–300, January 2007.

[44] O. Anatole von Lilienfeld, Ivano Tavernelli, Ursula Rothlisberger, and Daniel Sebastiani. Optimization of effective atom centered potentials for london dispersion forces in density functional theory. *Physical Review Letters*, 93(15), October 2004.

[45] Y. Y. Sun, Yong-Hyun Kim, Kyuho Lee, and S. B. Zhang. Accurate and efficient calculation of van der waals interactions within density functional theory by local atomic potential approach. *The Journal of Chemical Physics*, 129(15), October 2008.

[46] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, September 2006.

[47] Jiří Klimeš, David R Bowler, and Angelos Michaelides. Chemical accuracy for the van der waals density functional. *Journal of Physics: Condensed Matter*, 22(2):022201, December 2009.

[48] Pekka Koskinen and Ville Mäkinen. Density-functional tight-binding for beginners. *Computational Materials Science*, 47(1):237–253, November 2009.

[49] Michael Gaus, Qiang Cui, and Marcus Elstner. Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *Journal of Chemical Theory and Computation*, 7(4):931–948, March 2011.

[50] C M Goringe, D R Bowler, and E Hernández. Tight-binding modelling of materials. *Reports on Progress in Physics*, 60(12):1447–1512, December 1997.

[51] Marcus Elstner and Gotthard Seifert. Density functional tight binding. *Philos. Trans. A Math. Phys. Eng. Sci.*, 372(2011):20120483, March 2014.

[52] Thomas S. Hofer, Risnita Vicky Listyarini, Emir Hajdarevic, Lukas Maier, Felix R. S. Purtscher, Jakob Gamper, and Friedrich Hanser. Beyond the status quo: Density functional tight binding and neural network potentials as a versatile simulation strategy to characterize host–guest interactions in metal- and covalent organic frameworks. *The Journal of Physical Chemistry Letters*, 14(26):6018–6027, June 2023.

[53] Yang Yang, Haibo Yu, Darrin York, Qiang Cui, and Marcus Elstner. Extension of the self-consistent-charge density-functional tight-binding method: Third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *The Journal of Physical Chemistry A*, 111(42):10861–10873, October 2007.

[54] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and benchmark of dftb3 for organic molecules. *Journal of Chemical Theory and Computation*, 9(1):338–354, November 2012.

[55] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University PressOxford, June 2017.

[56] Mool C Gupta. *Statistical thermodynamics*. 1990.

[57] Alexander Fischereder, Markus Rödl, Markus Suta, Thomas S. Hofer, and Heidi A. Schwartz. From blue jeans to luminescent materials: Designing thioindigo-based red-fluorescent hybrid systems. *The Journal of Physical Chemistry C*, 127(31):15657–15668, July 2023.

[58] Felix R. S. Purtscher, Leo Christanell, Moritz Schulte, Stefan Seiwald, Markus Rödl, Isabell Ober, Leah K. Maruschka, Hassan Khoder, Heidi A. Schwartz, El-Eulmi Bendeif, and Thomas S. Hofer. Structural properties of metal–organic frameworks at elevated thermal conditions via a combined density functional tight binding molecular dynamics (dftb md) approach. *The Journal of Physical Chemistry C*, 127(3):1560–1575, January 2023.

[59] Salma Alrasheed. *Newton's Laws*, page 37–51. Springer International Publishing, 2019.

[60] Daan Frenkel, Berend Smit, and Mark A. Ratner. Understanding molecular simulation: From algorithms to applications. *Physics Today*, 50(7):66–66, July 1997.

[61] Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts. Oxford University Press, USA, 2010.

[62] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.

[63] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159(1):98–103, July 1967.

[64] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation

of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, January 1982.

[65] Ze-Ren Wang, Xu-Liang Zhu, Lu Jiang, Kai Zhang, Hui-Wen Luo, Yue Gu, and Peng Zhang. Investigations of the hydrogen bonds and vibrational spectra of clathrate ice xvi. *Materials*, 12(2):246, January 2019.

[66] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best practices for foundations in molecular simulations [article v1.0]. *Living Journal of Computational Molecular Science*, 1(1), 2019.

[67] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, March 1977.

[68] Hans C Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, October 1983.

[69] Christopher J Cramer. *Essentials of computational chemistry*. John Wiley & Sons, Chichester, England, 2 edition, September 2004.

[70] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, October 1984.

[71] Tetsuya Morishita. Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. *The Journal of Chemical Physics*, 113(8):2976–2982, August 2000.

[72] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), January 2007.

[73] William G. Hoover and Brad Lee Holian. Kinetic moments method for the canonical ensemble distribution. *Physics Letters A*, 211(5):253–257, February 1996.

[74] Augusto F. Oliveira, Gotthard Seifert, Thomas Heine, and Hélio A. Duarte. Density-functional based tight-binding: an approximate dft method. *Journal of the Brazilian Chemical Society*, 20(7):1193–1205, 2009.

[75] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim. Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics*, 152(12), March 2020.

[76] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B*, 58(11):7260–7268, September 1998.

[77] Yang Yang, Haibo Yu, Darrin York, Marcus Elstner, and Qiang Cui. Description of phosphate hydrolysis reactions with the self-consistent-charge density-functional-tight-binding (scc-dftb) theory. 1. parameterization. *Journal of Chemical Theory and Computation*, 4(12):2067–2084, November 2008.

[78] B Szuecs, Z Hajnal, R Scholz, S Sanna, and Th Frauenheim. Theoretical study of the adsorption of a ptcda monolayer on s-passivated gaas(l00). *Applied Surface Science*, 234(1–4):173–177, July 2004.

[79] B. Szuecs, Z. Hajnal, Th. Frauenheim, C. González, J. Ortega, R. Pérez, and F. Flores. Chalcogen passivation of gaas(1 0 0) surfaces: theoretical study. *Applied Surface Science*, 212–213:861–865, May 2003.

[80] Maxime Van den Bossche. Three-center tight-binding together with multipolar auxiliary functions. *Journal of Chemical Theory and Computation*, 20(6):2538–2550, March 2024.

[81] Maxime Van den Bossche. Dftb-assisted global structure optimization of 13- and 55-atom late transition metal clusters. *The Journal of Physical Chemistry A*, 123(13):3038–3045, March 2019.

[82] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.

[83] R Ditchfield, W J Hehre, and J A Pople. Self-consistent molecular-orbital methods. IX. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.*, 54(2):724–728, January 1971.

[84] Axel D Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, April 1993.

[85] P C Hariharan and J A Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta*, 28(3):213–222, 1973.

[86] W J Hehre, R Ditchfield, and J A Pople. Self—consistent molecular orbital methods. XII. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56(5):2257–2261, March 1972.

[87] Vitaly A Rassolov, Mark A Ratner, John A Pople, Paul C Redfern, and Larry A Curtiss. 6-31g* basis set for third-row atoms. *J. Comput. Chem.*, 22(9):976–984, July 2001.

[88] Andreas Bergner, Michael Dolg, Wolfgang Küchle, Hermann Stoll, and Heinzwerner Preuß. Ab initio energy-adjusted pseudopotentials for elements of groups 13–17. *Mol. Phys.*, 80(6):1431–1441, December 1993.

[89] Maxime Van den Bossche, Henrik Grönbeck, and Bjørk Hammer. Tight-binding approximation-enhanced global optimization. *Journal of Chemical Theory and Computation*, 14(5):2797–2807, March 2018.

[90] Thomas S Hofer and Andreas O Tirler. Combining 2d-periodic quantum chemistry with molecular force fields: A novel QM/MM procedure for the treatment of solid-state surfaces and interfaces. *J. Chem. Theory Comput.*, 11(12):5873–5887, December 2015.

[91] Muhammad Saleh and Thomas S Hofer. A DFTB/MM MD approach for solid-state interfaces: Structural and dynamical properties of H_2O and NH_3on R-TiO_2(001). *J. Phys. Chem. C Nanomater. Interfaces*, 123(12):7230–7245, March 2019.

[92] Niko Prasetyo and Thomas S Hofer. Adsorption and dissociation of water molecules at the α-Al2O3(0001) surface: A 2-dimensional hybrid self-consistent charge density functional based tight-binding/molecular mechanics molecular dynamics (2d SCC-DFTB/MM MD) simulation study. *Comput. Mater. Sci.*, 164:195–204, June 2019.

[93] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996.