

Ylva Os

Treatment of Large-Scale Systems with Multiscale and Fragmentation Methods

Bachelor's thesis in Theoretical Chemistry

Supervisor: Sarai Dery Folkestad

April 2022

Ylva Os

Treatment of Large-Scale Systems with Multiscale and Fragmentation Methods

Bachelor's thesis in Theoretical Chemistry
Supervisor: Sarai Dery Folkestad
April 2022

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Chemistry

Abstract

Quantum chemical calculation is currently a large and important field within chemistry, and has been more important the past few years as computational technology has advanced. However, there are still many challenges within this field, the largest one being that accurate methods are not applicable on large systems as they are too expensive in terms of computational time and memory space. This thesis will look at two different large-scale quantum methods, the quantum mechanical/molecular mechanical method and the fragmented molecular orbital method, which in different ways divide the main system into subsystems to make the calculations more affordable. Of course, this is not without problems, as all subsystems must be put together without losing important properties, and in some cases covalent bonds must be broken to achieve usable subsystems. The methods are shown to have different advantages, which mainly becomes apparent when we are interested in a specific area of the system or if the whole system is equally important to consider.

Contents

1	Introduction	3
2	Theoretical Background	3
2.1	Wave function theory	3
2.2	Hartree-Fock Theory	7
2.2.1	Variational theorem	7
2.2.2	Basis sets	7
2.2.3	Hamiltonian	7
2.2.4	The Roothaan-Hall equations	8
2.3	Electron-correlation methods	9
2.3.1	Post-Hartree-Fock methods	10
2.3.2	Kohn-Sham Density Functional Theory	13
3	Large-System Quantum Approaches	14
3.1	The QM/MM approach	14
3.1.1	Energy calculation	15
3.1.2	Challenges	16
3.1.3	Scaling	18
3.2	Fragmentation methods	18
3.2.1	Energy calculation	19
3.2.2	Challenges	20
3.2.3	Scaling	21
4	Case studies	21
4.1	HF in QM/MM	22
4.2	HF in FMO	23
5	Comparison	24
6	Conclusion	25

1 Introduction

As an alternative or addition to experimental chemistry, computational chemistry is an important asset to have when experiments are too expensive, hard to get precise or just to support experimental data. In contrast to experiments, the computational environment is very controllable, as you can set the system however you like. Generally, we divide the different chemical mechanics into quantum (QM) and molecular (MM) mechanics. The main difference between the two is that QM takes into account electronic effects and behaviour, while MM tries to generalize the behaviour of the atoms using classical mechanics and thus neglecting the existence of individual electrons. This neglect can result in inaccuracies, as the interaction between electrons are crucial to get an exact description of a molecular system. The downside is that QM calculations are very expensive, in form of computational time and memory space, and are currently restricted to small systems. To be able to do calculations on larger systems, one must manipulate the system or the calculations to reduce the cost. For example, this can be done by doing quantum calculations on the most essential part of the system and then apply the effects of the surrounding system with molecular mechanics, by using a force field. This is the way of the QM/MM approach. Another way to approach the problem is with fragmentation methods, where the system is divided into smaller systems which are calculated separately and then combined to get information regarding the whole system. This thesis will consider some of the fundamental concepts behind wave function theory and some calculation methods such as Hartree-Fock, and some electron-correlation methods. The main focus will be on the large-scale methods QM/MM and the fragmentation method Fragmented Molecular Orbitals (FMO), where we will consider some challenges and advantages with using these as an extension to the conventional electronic structure methods. We will also compare costs of the different methods, as this is the main bottle-neck of quantum calculations, in addition to look at which systems it would be most beneficial to use either a QM/MM or FMO calculation.

2 Theoretical Background

2.1 Wave function theory

Most quantum methods are based on wave function theory, which states that a quantum state of an isolated system can be totally described by a wave function, Ψ . This function can give us a variety of information regarding our system, such as the probability-distribution of a particle, as well as the expectation value of the

energy, different momenta, electric dipole moments and so on^[1]. These can be found as the expectation value of the Hermitian operators representing the corresponding observables. If we let Ω be the observable and $\hat{\Omega}$ be its Hermitian operator, the expectation value becomes^[1],

$$\langle \hat{\Omega} \rangle = \frac{\langle \Psi | \hat{\Omega} | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (2.1)$$

When using a normalized wave function, the denominator is equal to 1. This arises from the probability-distribution ($\langle \Psi | \Psi \rangle$) often decided to be equal to 1 for the guarantee of the particle's existence. It is also important that the operator is Hermitian, as this provides real eigenvalues/observables.

The Hermitian operator representing the energy of the system (E) is the Hamilton operator, often called the Hamiltonian, \hat{H} . We can let the Hamiltonian work on an unknown eigenfunction, which gives us the famous differential equation known as the time-dependent Schrödinger equation^[2],

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle, \quad (2.2)$$

with \hbar being the reduced Planck constant. If the Hamiltonian is time-independent, which is the case for free molecules, we can separate the time-dependency from the equation, resulting in the time-independent Schrödinger equation,

$$\hat{H} |\Psi\rangle = E |\Psi\rangle. \quad (2.3)$$

One of the most essential and quantum mechanically accurate approximations to this day is the Born-Oppenheimer approximation (BOA). As the weight difference between the electrons and the nuclei is very large, the movement of the nuclei can be considered to stay constant compared to the electrons. This allows us to separate the total wave function of into an electronic ($\psi_e(\mathbf{r}; \mathbf{R})$) and a nuclear ($\psi_N(\mathbf{R})$) wave function^[3]. The electronic wave function depends on the position of the electrons (\mathbf{r}) and uses the position of the nuclei (\mathbf{R}) as a parameter. The nuclear wave function only depends on the position of the nuclei^[1].

$$|\Psi(\mathbf{r}, \mathbf{R})\rangle = |\psi_e(\mathbf{r}; \mathbf{R})\rangle |\psi_N(\mathbf{R})\rangle. \quad (2.4)$$

Following the BOA we get an electronic Schrödinger equation (Equation 2.5), with a corresponding electronic Hamiltonian (\hat{H}_e) and a nuclear Schrödinger equation (Equation 2.6), with the nuclear Hamiltonian (\hat{H}_N),

$$\hat{H}_e(\mathbf{r}, \mathbf{R})|\psi_e(\mathbf{r}, \mathbf{R})\rangle = E_e(\mathbf{R})|\psi_e(\mathbf{r}, \mathbf{R})\rangle, \quad (2.5)$$

$$\left(\hat{H}_N(\mathbf{R}) + E_e(\mathbf{R})\right)|\psi_N(\mathbf{R})\rangle = E|\psi_N(\mathbf{R})\rangle. \quad (2.6)$$

As we can observe, the ability to solve the equation for the total energy relies on solving the electronic Schrödinger equation first, to find the potential energy surface (E_e), that can later be used in the nuclear Schrödinger equation. The electronic Hamiltonian contains terms describing the kinetic energy of the electrons (\hat{T}_e), the electron-electron Coulomb interactions (\hat{V}_{ee}), the potential energy between the electron and the nuclei (\hat{V}_{Ne}), and the potential energy from the nuclei-nuclei repulsion (\hat{V}_{NN}) as a constant:

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne} + \hat{V}_{NN}. \quad (2.7)$$

The nuclear Hamiltonian contains only the kinetic energy of the nuclei. In electronic structure theory, the main focus is usually to find good approximations for the electronic equation. This is because the unapproximated problem can only be solved exact for a single hydrogen atom and the hydrogen molecule, H_2^+ ^[4], in other words only single-electron systems.

The BOA is a very good approximation on closed shell systems, however, it fails whenever two or more states have very similar energy^[5]. This is because of the adiabatic coupling term in the nuclear Hamiltonian, which is inversely proportional with the energy difference of two states, causing the potential to reach infinity for states with equal energy. This happens, for instance, in electron transfers. Furthermore the BOA is invalid if the electrons are exchanged with heavier particles, such as antiprotons or muons^[5]. This is due to the assumption of size difference, meaning that for an antiproton, the nuclei will not remain consistent compared to the electron.

The wave function is often described by Slater-determinants, as they can describe multi-fermionic systems, as well as satisfy the anti-symmetry constraint^[1], and thus the Pauli-principle, by changing the sign upon the exchange of two fermions:

$$|\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)\rangle = -|\Psi(\mathbf{x}_2, \mathbf{x}_1, \dots, \mathbf{x}_N)\rangle, \quad (2.8)$$

where \mathbf{x}_i is a compound coordinate consisting of the spacial coordinate \mathbf{r}_i and the spin coordinate ω_i . A Slater-determinant is created from a linear combination of spin-orbitals ($|\phi(\mathbf{x})\rangle$), and for a generalized system we get,

$$|\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |\phi_1(\mathbf{x}_1)\rangle & |\phi_2(\mathbf{x}_1)\rangle & \cdots & |\phi_N(\mathbf{x}_1)\rangle \\ |\phi_1(\mathbf{x}_2)\rangle & |\phi_2(\mathbf{x}_2)\rangle & \cdots & |\phi_N(\mathbf{x}_2)\rangle \\ \vdots & \vdots & \ddots & \vdots \\ |\phi_1(\mathbf{x}_N)\rangle & |\phi_2(\mathbf{x}_N)\rangle & \cdots & |\phi_N(\mathbf{x}_N)\rangle \end{vmatrix}. \quad (2.9)$$

The factor $\frac{1}{\sqrt{N!}}$ is the normalizing factor, used to fulfill the requirement of a normalized N -electron wave function.

When considering electrons, one must take into account both spatial and spin properties. The spatial property describes the coordinates of an electron in space, whereas spin is an angular momentum property, proven to have an important role in electron behaviour^[6]. Following the Pauli exclusion principle, each electron can be uniquely described by a spin orbital, a function made up of a spatial and spin function.

The wave function can be described as a determinant containing both the spacial and the spin functions. It can be sufficient to only use one Slater-determinant (Hartree-Fock), however, we can also make a linear combination of two or more Slater-determinants (configuration interaction, MCSCF, coupled-cluster, MP2, and more) to improve the wave function. Generally, adding more determinants gives a more accurate result, but at a much higher computational cost, so these factors must be considered when deciding on a calculation method.

The largest challenge within the computational chemistry field is as previously stated, the cost of quantum mechanical calculations. These tend to scale heavily with the system size. Thus, when comparing the cost of a calculation, it is very convenient to compare the scaling of cost with the system size. We often denote polynomial scaling with the notation $\mathcal{O}(N^n)$, $\mathcal{O}()$ is used for the asymptotic scaling order, N is the system size, and n is the specific scaling. A system scaling with $\mathcal{O}(N^3)$, implies that doubling the system size, increases the calculation time eightfolds.

2.2 Hartree-Fock Theory

2.2.1 Variational theorem

Hartree-Fock (HF) is a variational method, meaning it parameterizes the wave function with a Slater-determinant and then minimizes the energy. Such methods follow the important variational theorem, stating that the expectation value of the energy of a normalized wave function will always be higher or equal to the actual energy, E_0 ^[7]:

$$\langle \Psi | \hat{H} | \Psi \rangle \geq E_0, \quad (2.10)$$

This is true for all variational methods, including e.g. MCSCF and CI.

2.2.2 Basis sets

In HF it is convenient to introduce basis sets, containing different basis functions, which are used in linear combinations to describe the atomic orbitals (AOs). There are different ways to define the basis functions, but we often use Gaussian-type functions as they are proven to give efficient results for larger systems^[8]. After obtaining a description of the AOs, one can use these functions in new linear combinations to create molecular orbitals (MOs).

To increase the accuracy of a HF calculation, the basis set can be expanded, meaning we increase the number of basis functions. When using an infinitely large basis set, the convergence point of the calculation is called the HF-limit energy, and following the variational principle, this will always be higher than the actual energy minimum.

2.2.3 Hamiltonian

As usual, we want to approximate the solution of the electronic Schrödinger equation, Equation 2.5. In HF we do this by using a single Slater-determinant. The MOs defining the determinant are found as eigenfunctions to the Fock-operator, which is found by collecting all one-electron operators into one term, $\hat{h}(i)$, and adding the effective two-electron term, $\hat{v}_{ee}(i)$, for the i -th electron:

$$\hat{f}(i) = \hat{h}(i) + \hat{v}_{ee}(i). \quad (2.11)$$

The difference between the Fock-operator in HF and the general Hamiltonian is that the electrostatic repulsion between the electrons, described in $\hat{v}_{ee}(i)$, is approx-

imated by a mean-field method, meaning each electron only see an average of the other electrons. This results in the $\hat{v}_{ee}(i)$ term consisting of a Coulomb (\hat{J}) and exchange (\hat{K}) term:

$$\hat{v}_{ee}(i) = \sum_{j=1}^N \left(\hat{J}_j(i) - \hat{K}_j(i) \right), \quad (2.12)$$

with N being the number of electrons, and the sum over j is the sum over all other electrons that i interacts with.

2.2.4 The Roothaan-Hall equations

The HF method is the starting point for many quantum chemical calculations, including some QM/MM and some fragmentation methods, as it is the most elementary *ab initio* approach in wave function theory^[9]. In HF we use a single Slater-determinant to describe the wave function, so for each electron in the system, we get:

$$\hat{f}(i)|\phi_i\rangle = \epsilon_i|\phi_i\rangle. \quad (2.13)$$

Combining the equations for every electron allows us to derive the Roothaan-Hall equations for closed shell systems (Equation 2.14), by introducing a basis set of AOs, which is solved to get the energy minimum of the MOs:

$$\mathbf{FC} = \mathbf{SC}\epsilon, \quad (2.14)$$

where \mathbf{F} is a matrix containing the Fock-operators ($\hat{f}(i)$) of all electrons in the system and is also depending on \mathbf{C} . \mathbf{C} is a matrix containing the AO contributions to the MOs, \mathbf{S} is an overlap matrix, describing the overlap between the different non-orthogonal AOs and ϵ is the energies of the associated MOs. As the Fock-operator depends on \mathbf{C} , one must first make a guess on the MO coefficients, and then solve the equation to get a new and improved \mathbf{C} -matrix. This new matrix is then used again to solve the equation and the process is repeated until self consistency is reached (shown in Figure 2.1). The result of a HF calculation is the energies to the different MOs. These orbital energies can then be put into the Schrödinger equation and be solved for the energy of the system.

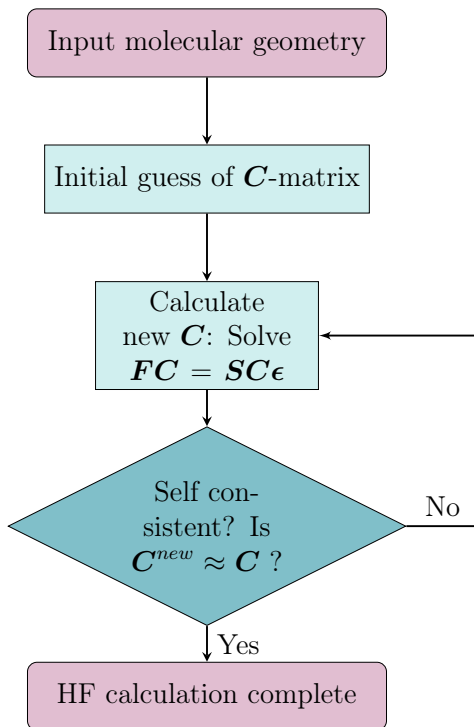


Figure 2.1: The process of running a HF calculation.

The cost of running a HF calculation depends heavily on the implementation and can vary from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2)$, depending on the evaluation of the integrals. HF is thus not a very expensive method, however, it lacks the very important terms of electron correlation, which is replaced by an approximation of the electron interactions as the Coulomb-Exchange term shown in Equation 2.12.

2.3 Electron-correlation methods

Simply put, the electron-correlation describes how the motion of an electron is affected by the presence of other electrons^[10], and is the reason the Schrödinger equation cannot be solved exactly for systems with more than one electron. As HF is explicitly lacking this property, it is incapable of accurately describing dissociation or bond formation^[4]. Instead, the correction to the electron of interest is approximated by an average of the surrounding electrons^[1]. For many-electron systems, only using HF can thus be very inaccurate. The electron-correlation energy (E_{corr}) is often defined as the difference between the exact energy (E_0) and the HF-limit energy (ϵ_0)^[7],

$$E_{corr} = E_0 - \epsilon_0 \quad (2.15)$$

To be able to get a good approximation on the correlation energy, many so-called electron-correlation methods have been developed.

2.3.1 Post-Hartree-Fock methods

Most of the electron-correlation methods are based on HF and has thus been named post-HF methods.

Configuration Interaction

The main feature of configuration interaction (CI) is the expansion of the wave function as multiple Slater-determinants (Φ), instead of just one, as used in HF^[11]. Starting with the HF ground state ($|HF\rangle$), and then including various excited states (shown in Equation 2.16 as Φ_i^a and Φ_{ij}^{ab}), CI tries to get a better approximation of the electron-correlation.

$$|\Psi_{CI}\rangle = c_0|HF\rangle + \sum_{ai} c_i^a |\Phi_i^a\rangle + \sum_{a>b, i>j} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (2.16)$$

The indices denotes the excitation from the occupied MOs i and j to the virtual MOs a and b . How many excitation sums are included is described in the name, for example, CISD include Single and Double excited determinants, whereas Full CI (FCI) includes all possible excitations for the given basis set formed by the Slater-determinants. To obtain the exact solution for a system one would need infinite basis sets, however, for logical reasons, this is not possible. The FCI solution is thus considered the exact solution for a given basis^[12].

As in HF (Equation 2.14), the ground state is found by minimizing the expectation value of the energy with respect to the CI coefficients, c , by solving the equation^[12],

$$\mathbf{H}^{CI} \mathbf{C} = \mathbf{C} \epsilon. \quad (2.17)$$

However, this time the right hand side of the equation does not include the overlap matrix, \mathbf{S} . This is because the equations are expressed in a basis containing orthogonal Slater-determinants, making the overlap matrix the identity matrix, which can thus be excluded^[12]. It is also important to note that the \mathbf{C} matrix here denotes a vector containing the CI coefficients, whereas in HF, the \mathbf{C} matrix contains the MO coefficients. Logically, the complete basis of Slater-determinants increases extremely fast and the size of the Hamiltonian expands with the number of Slater-

determinants squared. This makes it very challenging to run a Full CI calculation on larger systems.

Full CI calculation scales with $\mathcal{O}(N_{det}N^4)^{[13]}$, N_{det} being the number of Slater-determinants, which also have a scaling of $\mathcal{O}(\sum_k \binom{n_{occ}}{k} \binom{n_{vir}}{k})$, with k being the number of excited determinants, and n_{occ} and n_{vir} , being the number of occupied and virtual orbitals, respectively. Furthermore, the size of the \mathbf{H}^{CI} -matrix becomes extremely large for FCI, so that a calculation on CO using PySCF^[14] with the basis set cc-pVDZ, which includes 28 basis functions, would need 80TB of memory, which is not available on most computers.

Multiconfigurational Self-Consistent Field

In contrast to CI, multiconfigurational self-consistent field (MCSCF) minimizes the energy with respect to both the CI and the MO coefficients simultaneously, instead of just the CI coefficients, thus this method can be considered a hybrid between CI and HF^[4]. As a result of the wave function being a sum of Slater-determinants, we need an additional variational parameter to the CI wave function in Equation 2.16, that being an orbital rotation operator, $e^{-\hat{\kappa}}$ ^[15],

$$|\Psi_{MCSCF}\rangle = e^{-\hat{\kappa}} \sum_i c_i \Phi_i. \quad (2.18)$$

The scaling of MCSCF highly depends on the active space and how many determinants are present in the system. The non-CI part of MCSCF has a scaling depending of the implementation of HF, whereas the the CI part has the same scaling as normal CI^[16].

Coupled-Cluster Theory

Coupled-Cluster Theory (CC) is a solution to some of the most severe problems with CI; the lack of size-extensivity and the challenge of effectively implementing the method^[4]. Here we expand the expression of the wave function, by multiplying the exponential of the cluster-operator, \hat{T} , to some wave function^[17],

$$|\Psi\rangle = e^{\hat{T}} |\Psi_i\rangle. \quad (2.19)$$

The operator \hat{T} depends how many excitations are explicitly included as parameters in the calculation. It can be expanded as,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots, \quad (2.20)$$

where \hat{T}_1 describes single excitations, \hat{T}_2 describes double excitations, and so on. The nomenclature is similar to that of CI, meaning CCSD is the Coupled-Cluster name for the method including the Single and Double excitations. Even though CCSD only explicitly contain the single and doubly excitation operators in T , when we Taylor-expand the exponential cluster-operator working on the HF ground state ($|HF\rangle$), the result is a linear combination of determinants containing contributions of all the excited determinants,

$$e^{\hat{T}}|HF\rangle = (1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \frac{1}{2}\hat{T}_2^2 + \hat{T}_1\hat{T}_2 + \dots)|HF\rangle. \quad (2.21)$$

The outcome is an improved accuracy, without increasing the cost of the calculation. More explicitly, this means a CCSD calculation is more accurate than a CISD calculation, even though they both have a scaling of $\mathcal{O}(N^6)$ ^{[18][19]}.

Møller-Plesset Perturbation Theory

In Møller Plesset perturbation theory (MP n , where n denotes the order of perturbations added) the main concept relies on adding a small perturbation, or change, to a known and solvable problem to find an approximation of the electron-correlation to the exact unknown solution^[4]. The perturbation (\hat{V}) is added as a Rayleigh-Schrödinger perturbation (RS), Taylor expanded to a decided order (n), making MP a special case of RS. The perturbation operator for such a system is defined as the difference between the Hamiltonian (\hat{H} , seen in Equation 2.7) and the Fock Operator (\hat{f})^[20],

$$\hat{V} := \hat{H} - \hat{f} \implies \hat{H} = \hat{f} + \lambda\hat{V}. \quad (2.22)$$

The parameter λ is a real parameter describing the size of the perturbation. After getting this updated Hamiltonian, the Schrödinger equation can be solved for one order of perturbation at the time.

Trivially, the cost scales with increased order of perturbations, where an MP2 calculation has the asymptotic scaling of $\mathcal{O}(N^5)$ ^[21].

2.3.2 Kohn-Sham Density Functional Theory

In density functional theory (DFT), we take a step away from the wave function theory, and describe the energy of a system using density functions instead of wave functions^[22]. The Hohenberg-Kohn theorem states that there exist a density functional that has a minima equal to the exact ground state energy of a system^[23].

The most popular DFT method is Kohn-Sham DFT (KS-DFT), which introduces some orbitals that is used to find the electron density. The density functional (ρ) can be expressed as a sum of orbitals (Φ_i):

$$\rho = \sum_i |\Phi_i|^2. \quad (2.23)$$

The KS-DFT Hamiltonian is similar to that of HF, except with an added exchange-correlation term modifying the single electron Fock-operator^[24],

$$\hat{f}^{KS}(i) = \hat{v}_K + \hat{v}_{ext} + \hat{v}_C + \hat{v}_{xc}, \quad (2.24)$$

Consisting of the kinetic energy of an electron (\hat{v}_K), the electronic potential of the nuclei (\hat{v}_{ext}), the Coulomb interactions (\hat{v}_C), and the exchange-correlation energy (\hat{v}_{xc}).

From this we can derive the eigenvalue problem^[1],

$$\hat{f}^{KS}(i)|\Psi_i\rangle = \epsilon_i|\Psi_i\rangle. \quad (2.25)$$

After obtaining these equations, they can be solved in the same manner as HF (Figure 2.1), with an initial guess of the electron density, instead of MO coefficients. The downside of KS-DFT is that it is hard to get an accurate parameterization of the exchange-correlation from the density only, as the density is a three-coordinate property, whilst the exchange-interaction is a $4N$ -dimensional property, where N is the number of electrons. The exchange energy is a result of the anti-symmetry properties of the wave functions, and the correlation arises from the electrons having real-time interactions and wanting to avoid each other. This is an important contribution to the total energy, and without an accurate description of the exchange-interaction, which affects electrons with the same spin, properties such as Hund's rule becomes invalid.

3 Large-System Quantum Approaches

This being said, it should be obvious that it is crucial to have alternative methods able to handle larger systems with high accuracy. A few such approaches will be discussed, these being QM/MM and the fragmentational method FMO.

3.1 The QM/MM approach

The Quantum Mechanics/Molecular Mechanics (QM/MM) method is a hybrid method between quantum and molecular mechanics. This means in essence that we divide our system into two parts, where one is described by classical mechanics and one by quantum mechanics (Figure 3.1). In MM we use different kinds of force fields to describe properties such as small-amplitude vibrations, torsions, van der Waals interactions and electrostatic interactions^[25], and are often used in molecular dynamic (MD) simulations on large and complex organic-, biological-^[26], inorganic-^[27] and solid-state systems, like heterogeneous catalysis^[28]. The disadvantage of an MM approach is that it does not consider changes in the electronic structure like bond-breaking/forming, charge transfer or electronic excitation, as individual electrons are not taken into account. These properties are described by QM. QM calculations are, however, very expensive to run, as we have been discussing previously, which is why it can be beneficial to approximate parts of the system with MM.

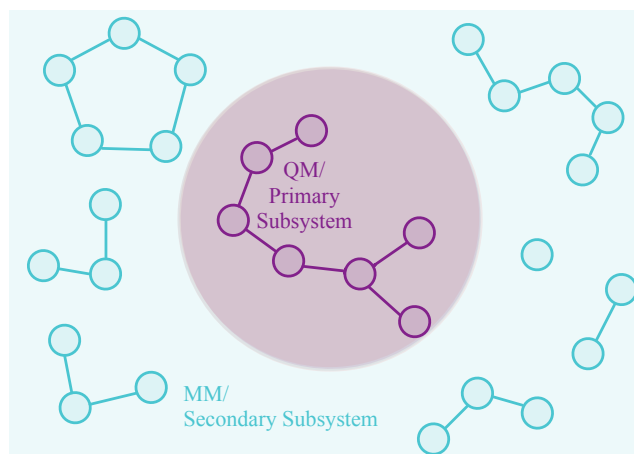


Figure 3.1: System divided by which method each part is treated with. The inner, primary subsystem and the outer secondary system is calculated by using QM and MM methods, respectively.

3.1.1 Energy calculation

When finding the total energy of a QM/MM system one must choose a way to handle the interactions between the systems. Generally, this can be done with two different schemes; additive and subtractive.

In the additive method, the energy ($E_{total}^{additive}$) is a sum of the energies from the QM subsystem ($E_{QM}^{primary}$), the MM subsystem ($E_{MM}^{secondary}$) and the interaction energy between the two ($E_{interaction}$):

$$E_{total}^{additive} = E_{MM}^{secondary} + E_{QM}^{primary} + E_{interaction}. \quad (3.1)$$

The MM calculations are only carried out on the secondary subsystem, the QM calculations on the primary subsystem and the last term is the coupling between the two systems. The additive scheme is as of today the most common method^[29].

In the subtractive method we do an MM calculation on the whole system and a QM calculation of the active site/primary subsystem. The MM calculation of the whole system and the QM calculation are added together and then the MM calculation of the primary subsystem is subtracted:

$$E_{total}^{subtractive} = E_{MM}^{system} + E_{QM}^{primary} - E_{MM}^{primary}. \quad (3.2)$$

The advantage with this method is that it is simple, with no explicit term for the coupling between the primary and secondary subsystems. However, for this to work one needs a complete set of MM parameters for the primary subsystem, which can be challenging. In addition, the coupling term is only described in MM terms, which leaves out, among other things, the precise electrostatic interactions of the electrons, as MM treats the electrostatics as point charges, which are not included in the Hamiltonian^[29].

The energy of the MM force field expression explains the bonded terms (including bond stretching, angle bending, torsion and deformations), Lennard-Jones type van der Waals interactions and Coulomb interactions between rigid point charges^[29]. Implementations of these force fields exist in software like AMBER^[26], CHARMM^[30] and GROMACS^[31]. Calculating the QM expression is in essence to solve the S.E^[32], shown in Equation 2.5. As the number of systems having an exact solution is very limited^[1], approximations of the energy is essential. In practice DFT or other semi-empirical methods are very common, as they are easy to use and quite cheap

compared to other more accurate methods, like post-HF *ab initio* methods, including MP2 and CC^[29].

In the additive scheme, finding the interaction term between the QM and MM systems can be quite challenging and expensive^[33]. As with the total energy, these interactions are additively separable into QM and MM interactions^[33]. The MM interaction is easy to calculate and is done by applying the same force field as earlier and provides a good estimate for steric effects, as well as other molecular properties described by the subtractive method earlier. In QM interaction the electronic effects are taken into account. Generally, we consider the electrostatic interactions to fully describe the electronic embedding, and it is therefore used as the only interactive force in the quantum part of the coupling term^[27]. The electrostatic interactions are interactions between charged particles^[32].

3.1.2 Challenges

Using QM/MM on a small molecule surrounded by others can be easy, however, one of the largest problems occur when a molecule is too large to be included in the primary subsystem, and a covalent bond has to be cut off between the QM and MM subsections. To make up for physically unrealistic situations, this problem can be dealt with in various ways, where the two most popular methods are link atoms or local orbitals.

Link atoms

In the link atom method, an atom, usually hydrogen^[34], is placed in the middle of a bond as shown in Figure 3.2. The link atom can also be methyl groups or pseudo-halogen atoms parameterized to be similar to the associated MM fragment^[35].

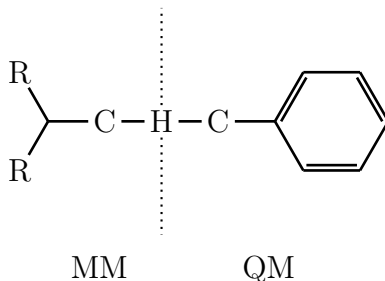


Figure 3.2: An arbitrary system showing the hydrogen link atom at the QM-MM interface.

However, using such a method results in a non-defined specific energy term, as all systems must be treated differently to make up for the energy difference provided by the link atom. In many implementations the Coulomb interactions between the

MM system and the link atom is excluded from the QM Hamiltonian. This can, however, result in an unrealistic partial charge on the link atom, due to a charge polarization on the primary subsystem^[36]. The placement of the link atom is shown to have a significant impact on the result, and it is desired to be placed as far away from the reaction in question as possible, without leaving a very negative impact on the efficiency^[37].

Using a link atom is a very simple approach, which can also lead to errors. By using a this approach the degrees of freedom will not be correct, even if the link atom is restricted to be as similar to the broken bond as possible. There will also most likely be errors in polarity. In addition, one must consider if the link atom is to interact with the MM section. If the interaction is taken into account, one might observe unrealistic phenomena. However, if this is neglected the polarization in the interface might become overwhelmed^[38]. This is because of the short bond length between the link atom and its neighbours, which is normally around 0.5 Å for a C-C bond. Thus the point charge of the original MM atom interferes with the total charge density^[25]. The last problem can be corrected by using pseudo-halogens, among some other parameterized groups instead of the normally used hydrogen^[38]. Another way to solve this is by setting the charge of the MM atom closest to the interface to zero. This can be done to the first MM atom (Z1 scheme), the first and second MM atom (Z2 scheme), and so on. The downside of this approach is that the total charge of the MM system changes, and a neutral system can become charged. To prevent this, force fields such as CHARMM, let the sum of charges over a given group be zero during the parameterization of the force field, this allows for some charge changes on specific atoms, while preventing the total charge of the system to be disturbed. Methods that preserve the total charge has been shown to more robust than the Z1/Z2 schemes^[25].

Local orbitals

The second option is to use local orbitals, for example in a local self-consistent field (LSCF). Such a field is created by treating the bonds in the QM/MM interface with a set of frozen spin-orbitals or strictly localized bond orbitals (SLBO), placed on the nearest QM atom. In contrast to the link atom approach, local orbital methods provide a QM description around the interface atoms^[33].

By being "frozen", the orbitals in question are excluded from the energy optimization, however, they still contribute to the total energy, but as they are excluded from the optimization, this contribution does not change^[38]. The energy of the

SLBOs only depend on the frontier atoms and is thus considered to be transferable between systems^[29]. This makes it possible to do the SLBO calculation on a much smaller system and then transferring it to the large QM/MM system^[38]. However, this means that the LSCF method is depending on parameters from other model studies for each new system.

As with all approximative methods, LSCF have many related methods with some deviations. One example is the generalized Hybrid orbitals (GHO) method. The difference from LSCF is in essence the placement of the localized and frozen orbitals, which are placed on the nearest MM atom. These orbitals are divided into active and auxiliary orbitals, where the active orbital is included in the SCF optimization. With this follows the advantage of not needing an assisting system to decide these parameters^[35].

However, such methods are much more technically challenging to implement, compared to the link atom method. For example, one must make sure the orbitals do not mix, by adding orthogonality constraints to the orbitals. In addition, there are some parameters depending on force fields, basis sets and the QM method chosen, that must be determined before the calculation can begin, which are not necessarily transferable between systems and can therefore be a challenge to get accurate^[29].

3.1.3 Scaling

As the MM calculation is very cheap compared to the QM calculation, the scaling is solely depending on the QM method and the size of its region.

3.2 Fragmentation methods

Fragmentation methods are an alternative where the system is also divided into smaller parts, but instead of using an MM approximation, all fragments are treated with QM. The goal is to combine the results from the different fragments to predict the properties of the whole system^[39]. Fragmentation methods include many different approaches, including the systematic molecular fragmentation (SMF), combined fragmentation method (CFM) and molecular orbital method (FMO). SMF and CFM group atoms based on bonds and similar chemical properties and combine the results into an expression based on the exclusion/inclusion principles^[40]. In comparison, the characteristics of FMO is that it includes electrostatic forces, as a field, for the whole system in each fragment^[41]. From now on, we will only discuss the FMO method.

The advantage of FMO, is that dealing with cleaved covalent bond are not as hard as seen previously the QM/MM method. The Coulomb field which is added here is almost enough to saturate the lone bonds and the other adjustments needed are easier to implement^[41].

The starting point of doing an FMO calculation is dividing the system. The specific grouping can depend on, among other things, the QM method chosen for the calculations, often used are MP2, CC, MCSCF and DFT^[41]. When working with chemical systems as electron density distributions, in comparison to spherical particles, the electron density within each fragment should be as localized as possible, so choosing the right fragmenting is crucial. For example would it be obvious to keep a benzene ring within the same fragment^[41].

3.2.1 Energy calculation

The total energy expression of a FMO system is given as^[41],

$$E = \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J). \quad (3.3)$$

The first sum is the energy of all the monomer sections (E_I), shown as I, J and K , in Figure 3.3. N is the number of fragments. The second term is the interaction between all combinations of dimers (E_{IJ}). As we can see, the energy of the individual monomers is subtracted as they are already included in the first term.

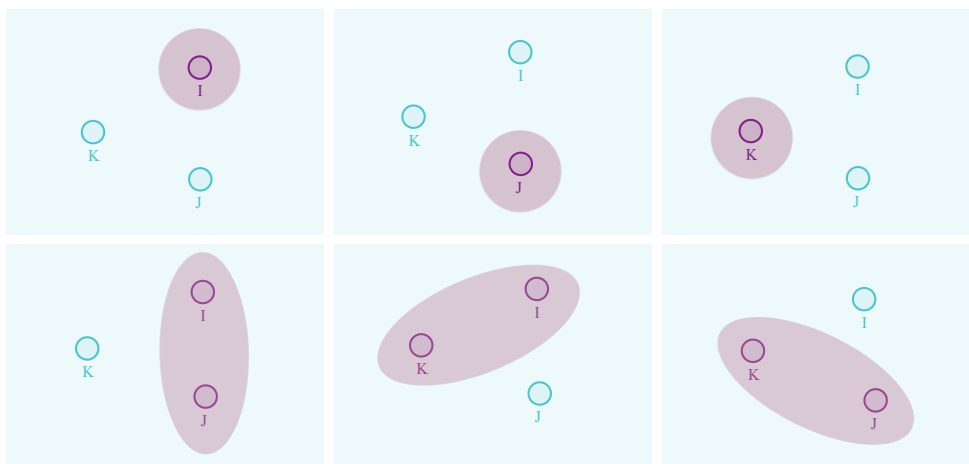


Figure 3.3: Visualisation of the different terms in Equation 3.3. I, J and K denotes different fragments and the active parts are colored purple. The first row shows the one-body terms, whereas the second row shows the two-body terms.

To increase accuracy, it is possible and proven efficient, to include a more body-

terms, and by adding a third, the energy expression will become^[42]:

$$\begin{aligned}
 E = & \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J) \\
 & + \sum_{I>J>K}^N \{ (E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) \\
 & - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I) \}.
 \end{aligned} \tag{3.4}$$

However, adding a lot more terms can make the calculations very expensive. And as the three-body term almost covers the *ab initio* properties for some systems^[42], this is in many cases not necessary. How many body-terms are present is expressed in the name as FMOX, with X being the number of body-terms. Therefore, Equation 3.3 is actually the energy expression of an FMO2 system.

3.2.2 Challenges

As with all QM methods, molecular integral calculations are a large restriction. It is therefore often beneficial to exchange accuracy with faster, but more inaccurate, semi-empirical methods^[9]. One example is combining FMO with DFT, which has been shown to be able to compute calculations on 10^6 atoms, with a time scaling on $O(N^{1.21})$ for FMO2-DFT, the same calculations computed with full SCC-DFT has a scaling of $O(N^{3.09})$ ^[43]. However, DFT is not always preferable as it fails to get a good approximation of the exchange correlation. Nonetheless, this proves that it is possible to reduce the scaling from being cubic to almost linear, using the FMO method.

As mentioned, FMO2 is often accurate enough for most systems, however, in some cases FMO3 and FMO4 might be necessary. Examples of such systems include biological systems of alanine dodecamer and chignolin^[44], and segmentation of amino acids on peptides^[45] for FMO3 and FMO4 respectively. However, the accuracy of FMO4 methods is observed to have an incremental cost of 10 times that of a corresponding FMO2 calculation, for some systems^[45]. As usual, one must therefore compare the cost of accuracy to what is needed for the system.

There is also the problem of breaking covalent bonds, even though it is not as problematic as in QM/MM, some precautions needs to be taken into account. Normally, this is done by adding the bond-electrons to one of the fragments and restrains are added to the Fock-operator to make sure the electrons does not become part of the, for instance, core electrons of the fragment they are attached to^[46]. One can also

add local orbitals, such as in QM/MM, where the occupied MOs used to describe the detached orbitals are frozen^[46].

3.2.3 Scaling

For FMO the scaling will become $\mathcal{O}(m^n f(\frac{nN}{m}))$, with $f()$ being the original scaling of the QM method chosen, m being the number of equally sized fragments and n the number of terms included in the energy calculation, so for FMO2 n is 2 and so on. To put this in some context, we can compare costs by reformulating the expression for the cost:

$$\mathcal{O}\left(m^n f\left(\frac{nN}{m}\right)\right) \implies \mathcal{O}\left(m^n \left(\frac{nN}{m}\right)^k\right) = \mathcal{O}\left(n^k m^{n-k} N^k\right) = \mathcal{O}\left(m^{n-k} N^k\right). \quad (3.5)$$

The term n^k can be neglected as it is not dependant on the system size nor the number of fragments, as k is the exponent of the original cost, meaning $k = 2$ for HF and $k = 6$ for CCSD. If we first consider a problem using HF with FMO2, where we let $k = 2$ and $n = 2$, we get a scaling of $\mathcal{O}(m^0 N^2)$. As we can see, regardless of the fragment size, FMO2 has the same scaling as a HF calculation.

If we now look at CCSD, where $k = 6$ and again $n = 2$ we will get, $\mathcal{O}(m^{-4} N^6)$, which shows that increasing m , will decrease the cost. It is, however, important to note that if one increases the number of fragments a lot, the accuracy of the calculation can be reduced. This could be compensated with by using a higher body term, such as FMO3 or higher. Another aspect to consider is fact the the level of body terms to include. If n is higher than p , adding fragments is not cost efficient, meaning that FMO6 has the same cost as CCSD and FMO7 is actually more expensive.

4 Case studies

Up until now, we have looked at a general study of the different methods, and some of their characteristics. To get a better understanding of their behavior, we will consider two case studies, where we look at the methods using HF. We will also for simplicity use restricted HF (closed shell system) and have no broken covalent bonds.

4.1 HF in QM/MM

Let us consider a simple QM/MM system, where the QM region consists of some molecular system, X , surrounded by other molecules, Y . In this case study, we will approximate the MM interaction with an electrostatic embedding, where the molecules, Y , are given partial atomic charges (PACs). To start off, we need to find an expression for the electronic Hamiltonian of the QM region. This is done by reformulation the expression for electronic Hamiltonian in Equation 2.7, to only consider the electrons in X , using the additive scheme^[29],

$$\hat{H}_e^X = \hat{T}_e^X + \hat{V}_{ee}^X + \hat{V}_{Ne}^X + \hat{V}_{NN}^X + \hat{V}_{e,MM}^{XY} + \hat{V}_{N,MM}^{XY}. \quad (4.1)$$

We have also added two terms, $V_{e,MM}^{XY}$ and $V_{N,MM}^{XY}$, which take into account the interactions between the PACs and the particles in X . $V_{e,MM}^{XY}$ is the interaction between the electrons in X and the PACs in Y , and can be written as^[29]:

$$\hat{V}_{e,MM}^{XY} = - \sum_{Ii} \frac{q_I}{|\mathbf{r}_i - \mathbf{R}_I|}. \quad (4.2)$$

Where q_I is the charge of particle I , and $|\mathbf{r}_i - \mathbf{R}_I|$ the distance between electron i and PAC I .

Naturally, $V_{N,MM}^{XY}$ thus consider the interactions between the nuclei in X and the PACs in Y . This can be written as^[29],

$$\hat{V}_{N,MM}^{XY} = - \sum_{I\zeta} \frac{q_I Z_\zeta}{|\mathbf{R}_\zeta - \mathbf{R}_I|}. \quad (4.3)$$

We must now also consider the charge of the nuclei in X , expressed as Z_ζ . Except for this, the terms are similar to that of Equation 4.2, with \mathbf{R}_ζ being the position of nuclei in X .

From here, we can set up an expression for the expectation value for our new Hamiltonian, \hat{H}_e^X , as seen previously in Equation 2.11. This gives us the Fock-operator and its eigenvalue problem^[29],

$$\hat{f}^{QM/MM}(i)|\Phi_i\rangle = \epsilon_i|\Phi_i\rangle, \quad (4.4)$$

with a modified Fock-operator including the XY terms in the one-electron operator,

$$\hat{f}^{QM/MM}(i) = \tilde{h}^X(i) + \hat{V}_{ee}^X, \quad \tilde{h}^X(i) = \hat{h}^X(i) + \hat{V}_{e,MM}^{XY} + \hat{V}_{N,MM}^{XY}, \quad (4.5)$$

which can be formulated as some Roothaan-Hall-like equations,

$$\mathbf{F}^{QM/MM} \mathbf{C} = \mathbf{S} \mathbf{C} \epsilon, \quad (4.6)$$

and the SCF calculation can begin, equally to that in Figure 2.1. As we can see the only significant deviation from the original HF calculation, is the Fock-operator, which now includes impacts from the PACs of the surroundings.

4.2 HF in FMO

Let us now consider a fragment, X , surrounded by multiple other fragments, Y . We start off by expressing the Hamiltonian for one fragment in an FMO2 system, which has to be solved for each fragment in the system^[46]:

$$\hat{H}_e^X = \hat{T}_e^X + \hat{V}_{ee}^X + \hat{V}_{Ne}^X + \hat{V}_{NN}^X + \hat{V}_{ee}^{XY} + \hat{V}_{Ne}^{XY}. \quad (4.7)$$

\hat{T}_e^X , \hat{V}_{ee}^X , \hat{V}_{Ne}^X and \hat{V}_{NN}^X , still describe interactions within fragment X . The new terms for FMO are \hat{V}_{ee}^{XY} and \hat{V}_{Ne}^{XY} . \hat{V}_{ee}^{XY} includes the interaction between the electrons in X and the electrons in Y , via the electron density in the fragments Y (ρ^Y)^[46]:

$$\hat{V}_{ee}^{XY} = \sum_{i \in X} \sum_{Y \neq X} \int \frac{\rho^Y(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}', \quad (4.8)$$

with \mathbf{r}_i and \mathbf{r}' being the positions of electrons in X and the fragment Y , respectively.

\hat{V}_{Ne}^{XY} is the interaction between the electrons in X and the nuclei in Y ^[46], and can be expressed as the charge in the nuclei (Z_I) divided by the difference coordinates of the nuclei (\mathbf{R}_I) and the different electrons (\mathbf{r}_i):

$$\hat{V}_{Ne}^{XY} = \sum_{i \in X} \sum_I \left(-\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \right). \quad (4.9)$$

\hat{V}_{ee}^{XY} and \hat{V}_{Ne}^{XY} are implemented as a modification of the Fock-operator, ($\hat{f}(i)$), as

an addition to the one-electron operator ($\hat{h}(i)$)^[46],

$$\hat{f}^{FMO}(i) = \tilde{h}(i) + \hat{V}_{ee}^X, \quad \tilde{h}(i) = \hat{h}^X(i) + \hat{V}_{ee}^{XY} + \hat{V}_{Ne}^{XY}. \quad (4.10)$$

If we introduce a basis for our system, we can again formulate the Roothaan-Hall equations^[46],

$$\mathbf{F}^{FMO}\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon, \quad (4.11)$$

with \mathbf{F}^{FMO} being the matrix containing the modified Fock-operator for FMO.

5 Comparison

After looking at these two moderately different methods, one can try to consider in which cases one is better than the other. Generally, we can say that on electronic systems, QM methods are more accurate. Based on this, one could assume that FMO would be preferable in most cases. However, is the cost of running QM on the whole system necessarily worth the extra cost, or would an approximate force field be accurate enough?

We cannot only look at the accuracy of the methods but should take into account the cost of actually running the calculation. For QM/MM the limiting factor is of course the QM calculation, which we will give the scaling $\mathcal{O}(f(N))$, where $f(N)$ is the asymptotic scaling of the method used, and N the size of the QM region. As we have seen previously in Section 3.2.3, FMO seems to be highly cost efficient for methods with an originally large scaling, as long as we do not consider very high coupling terms, such as in FMO4 or higher. For cheaper methods, such as HF, we saw that using FMO does not reduce the scaling.

From looking at the case studies, we see that the mathematical difference is the modification of the Fock-operator and thus also the Fock-matrix. In HF/MM we have additional potential energies between the electrons and nuclei in the QM section and with the electrostatic embedding in the MM region, both based on their atomic charges and the distance between them (see Equation 4.2-4.3). Similarly, FMO-HF includes the electrostatics between the electrons in a given fragment and the nuclei in the remaining fragments, using the same method (Equation 4.9). The interaction between the electrons in different fragments, however, is based on the

electron density in the secondary fragments (Equation 4.8). Another thing to note, is that when solving the RH equations in FMO-HF, one must do so for each fragment in the system, as the Fock-operator only considers one main fragment at the time. This, however, is very easy to parallelize^[41].

The main qualitative difference between FMO and QM/MM is that QM/MM is a focused method, while FMO is not. This being that QM/MM considers a small part of the system to prioritize, and optimizes the calculation to fit this area. FMO on the other hand, tries to have the same accuracy on the whole system. Due to this, one would often prefer QM/MM if the reaction or region of interest is small^[32], as it might not be very dependent on the electrons far away. Another example illustrating the focused/unfocused property is by comparing geometry optimizations. QM/MM is a very common method used for such calculations^{[47][48][49]}, however, only the primary subsection will get an accurate geometry. Geometry optimizations have also been developed using FMO-HF^{[50][51]} and FMO-MP2^[52], with sufficient accuracy, which optimizes the whole system.

However, this also depends on what properties we are interested in. If we consider a core-excitation of a single electron, the electron will be heavily shielded from the outer electrons by the valence electrons of the molecule. In this case QM/MM might be a sufficient method compared to e.g. a valence excitation. Due to the weaker shielding effect of the valence electrons, it might be necessary to apply QM to a much larger region, and FMO could be preferable. As FMO looks at all fragments of the whole system equally, the result is that one needs to do all the calculations for every fragment, in contrast to QM/MM, where you only have one subsystem of interest.

6 Conclusion

We have now looked at some of the benefits and challenges with different large-scale quantum methods, which are important to be able to do calculation on larger systems. The most important is that they are both able to reduce the cost while at the same time being able to account for electronic properties at high accuracy, depending on the conventional QM method chosen. The biggest challenge with these methods are the breakage of covalent bonds, which can result in inaccuracy and more complicated implementation, depending on the correction method. To create a hierarchy of FMO and QM/MM does not seem beneficial as they are both good methods for different kinds of systems, with QM/MM being a focused method,

while FMO is not. So the method should be chosen based on what properties one is interested in. Overall, large-scale methods seem like an important part of the future development in quantum chemistry and could solve many problems regarding the state-of-the-art bottlenecks of large-system simulations, which currently only can be solved using cheaper semi-empirical methods.

References

- [1] P. W. Atkins and Ronald Friedman. *Molecular quantum mechanics*. Oxford University Press, Oxford ; New York, 5th ed edition, 2011. ISBN 978-0-19-954142-3.
- [2] E. Schrödinger. Quantisierung als Eigenwertproblem. *Physikalisches Institut der Universität*, 384(4):361–376, 1926. ISSN 1521-3889. doi: 10.1002/andp.19263840404.
- [3] M. Born and J. R. Oppenheimer. On the Quantum Theory of Molecules. ,, 1927. URL <https://elib.bsu.by/handle/123456789/154381>.
- [4] Jacob Townsend, Justin K. Kirkland, and Konstantinos D. Vogiatzis. Chapter 3 - Post-Hartree-Fock methods: configuration interaction, many-body perturbation theory, coupled-cluster theory. In S. M. Blinder and J. E. House, editors, *Mathematical Physics in Theoretical Chemistry*, Developments in Physical & Theoretical Chemistry, pages 63–117. Elsevier, January 2019. ISBN 978-0-12-813651-5. URL <https://www.sciencedirect.com/science/article/pii/B9780128136515000036>.
- [5] Satoshi Takahashi and Kazuo Takatsuka. On the validity range of the Born-Oppenheimer approximation: A semiclassical study for all-particle quantization of three-body Coulomb systems. *The Journal of Chemical Physics*, 124(14):144101, April 2006. ISSN 0021-9606. doi: 10.1063/1.2173997. URL <https://aip.scitation.org/doi/10.1063/1.2173997>. Publisher: American Institute of Physics.
- [6] Walther Gerlach and Otto Stern. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik, Volume 9, Issue 1*, pages 349–352, March 1922.
- [7] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, June 2012. ISBN 978-0-486-13459-8.
- [8] Susi Lehtola. A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules. *International Journal of Quantum Chemistry*, 119(19):e25968, 2019. ISSN 1097-461X. doi: 10.1002/qua.25968. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.25968>.
- [9] Alexey V. Akimov and Oleg V. Prezhdo. Large-Scale Computations in Chemistry: A Bird’s Eye View of a Vibrant Field. *Chemical Reviews*, 115(12):

- 5797–5890, June 2015. ISSN 0009-2665, 1520-6890. doi: 10.1021/cr500524c. URL <https://pubs.acs.org/doi/10.1021/cr500524c>.
- [10] David P. Tew, Wim Klopper, and Trygve Helgaker. Electron correlation: the many-body problem at the heart of chemistry. *Journal of Computational Chemistry*, 28(8):1307–1320, June 2007. ISSN 0192-8651. doi: 10.1002/jcc.20581.
- [11] Christopher J. Cramer. *Essentials of computational chemistry: theories and models*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2nd ed edition, 2004. ISBN 978-0-470-09182-1 978-0-470-09181-4.
- [12] C. David Sherrill and Henry F. Schaefer. The Configuration Interaction Method: Advances in Highly Correlated Approaches. In *Advances in Quantum Chemistry*, volume 34, pages 143–269. Elsevier, 1999. ISBN 978-0-12-034834-3. URL <https://linkinghub.elsevier.com/retrieve/pii/S0065327608605328>.
- [13] Yann Garniron. Development and parallel implementation of selected configuration interaction methods. *Theoretical and/or physical chemistry. Université Toulouse*, page 255, April 2019.
- [14] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. PySCF: the Python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018. ISSN 1759-0884. doi: 10.1002/wcms.1340. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340>.
- [15] Robert E. Thomas, Qiming Sun, Ali Alavi, and George H. Booth. Stochastic Multiconfigurational Self-Consistent Field Theory. *Journal of Chemical Theory and Computation*, 11(11):5316–5325, November 2015. ISSN 1549-9618, 1549-9626. doi: 10.1021/acs.jctc.5b00917. URL <https://pubs.acs.org/doi/10.1021/acs.jctc.5b00917>.
- [16] Hefeng Wang, Sabre Kais, Alán Aspuru-Guzik, and Mark R. Hoffmann. Quantum algorithm for obtaining the energy spectrum of molecular systems. *Physical Chemistry Chemical Physics*, 10(35):5388–5393, 2008. doi: 10.1039/B804804E. URL <https://pubs.rsc.org/en/content/articlelanding/2008/cp/b804804e>.
- [17] Nevin Oliphant and Ludwik Adamowicz. Coupled-cluster method truncated at quadruples. *The Journal of Chemical Physics*, 95(9):6645–6651, November

1991. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.461534. URL <http://aip.scitation.org/doi/10.1063/1.461534>.
- [18] Kenny B. Lipkowitz, Thomas R. Cundari, and Donald B. Boyd. *Reviews in Computational Chemistry, Volume 23*. John Wiley & Sons, February 2007. ISBN 978-0-470-11643-2. Google-Books-ID: TrIRyTB8kJ0C.
- [19] Neil Shenvi, Helen van Aggelen, Yang Yang, and Weitao Yang. Tensor hypercontracted ppRPA: Reducing the cost of the particle-particle random phase approximation from $O(r^6)$ to $O(r^4)$. *The Journal of Chemical Physics*, 141(2):024119, July 2014. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.4886584. URL <http://aip.scitation.org/doi/10.1063/1.4886584>.
- [20] Chr. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review*, 46(7):618–622, October 1934. ISSN 0031-899X. doi: 10.1103/PhysRev.46.618. URL <https://link.aps.org/doi/10.1103/PhysRev.46.618>.
- [21] Hans-Joachim Werner, Frederick R. Manby, and Peter J. Knowles. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *The Journal of Chemical Physics*, 118(18):8149–8160, May 2003. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.1564816. URL <http://aip.scitation.org/doi/10.1063/1.1564816>.
- [22] Haoyu S. Yu, Shaohong L. Li, and Donald G. Truhlar. Perspective: Kohn-Sham density functional theory descending a staircase. *The Journal of Chemical Physics*, 145(13):130901, October 2016. ISSN 0021-9606. doi: 10.1063/1.4963168. URL <https://aip.scitation.org/doi/full/10.1063/1.4963168>. Publisher: American Institute of Physics.
- [23] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, November 1964. ISSN 0031-899X. doi: 10.1103/PhysRev.136.B864. URL <https://link.aps.org/doi/10.1103/PhysRev.136.B864>.
- [24] W. Kohn, A. D. Becke, and R. G. Parr. Density Functional Theory of Electronic Structure. *The Journal of Physical Chemistry*, 100(31):12974–12980, January 1996. ISSN 0022-3654, 1541-5740. doi: 10.1021/jp960669l. URL <https://pubs.acs.org/doi/10.1021/jp960669l>.
- [25] Hai Lin and Donald G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, 117

- (2):185, July 2006. ISSN 1432-2234. doi: 10.1007/s00214-006-0143-z. URL <https://doi.org/10.1007/s00214-006-0143-z>.
- [26] Ross C. Walker, Michael F. Crowley, and David A. Case. The implementation of a fast and accurate QM/MM potential method in Amber. *Journal of Computational Chemistry*, 29(7):1019–1031, May 2008. ISSN 1096-987X. doi: 10.1002/jcc.20857.
- [27] Tell Tuttle. Applications of QM/MM in inorganic chemistry. In *Spectroscopic Properties of Inorganic and Organometallic Compounds*, pages 87–110. ., July 2010. doi: 10.1039/9781849730853-00087.
- [28] Uwe Eichler, Christoph M. Kölmel, and Joachim Sauer. Combining ab initio techniques with analytical potential functions for structure predictions of large systems: Method and application to crystalline silica polymorphs. *Journal of Computational Chemistry*, 18(4):463–477, 1997. ISSN 1096-987X.
- [29] Hans Martin Senn and Walter Thiel. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009. ISSN 1521-3773. doi: 10.1002/anie.200802019. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200802019>.
- [30] Daniel A. B. de Oliveira and João B. L. Martins. QM/MM (ONIOM) Calculation on FAK/Dasatinib Docking. *Revista Processos Químicos*, 9(18): 68–72, July 2015. ISSN 1981-8521. doi: 10.19142/rpq.v9i18.259. URL http://ojs.rpqsenai.org.br/index.php/rpq_n1/article/view/259. Number: 18.
- [31] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015. ISSN 2352-7110. doi: 10.1016/j.softx.2015.06.001.
- [32] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *Journal of Chemical Theory and Computation*, 1(6):1176–1184, November 2005. ISSN 1549-9618, 1549-9626. doi: 10.1021/ct050123f. URL <https://pubs.acs.org/doi/10.1021/ct050123f>.
- [33] Carles Bo and Feliu Maseras. QM/MM methods in inorganic chemistry. *Dalton Transactions*, 0(22):2911–2919, 2008. doi: 10.1039/B718076D. Publisher: Royal Society of Chemistry.

- [34] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, 1990. ISSN 1096-987X. doi: 10.1002/jcc.540110605. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540110605>.
- [35] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *The Journal of Physical Chemistry A*, 102(24):4714–4721, June 1998. ISSN 1089-5639, 1520-5215. doi: 10.1021/jp9809890. URL <https://pubs.acs.org/doi/10.1021/jp9809890>.
- [36] Dirk Bakowies and Walter Thiel. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *The Journal of Physical Chemistry*, 100(25):10580–10594, January 1996. ISSN 0022-3654, 1541-5740. doi: 10.1021/jp9536514.
- [37] Kirsten P. Eurenium, David C. Chatfield, Bernard R. Brooks, and Milan Hodosek. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *International Journal of Quantum Chemistry*, 60(6):1189–1200, 1996. ISSN 1097-461X. doi: 10.1002/(SICI)1097-461X(1996)60:6<1189::AID-QUA7>3.0.CO;2-W.
- [38] Nicolas Ferré, Xavier Assfeld, and Jean-Louis Rivail. Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *Journal of Computational Chemistry*, 23(6):610–624, 2002. ISSN 1096-987X. doi: 10.1002/jcc.10058. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10058>.
- [39] Mark S. Gordon, Dmitri G. Fedorov, Spencer R. Pruitt, and Lyudmila V. Slipchenko. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chemical Reviews*, 112(1):632–672, January 2012. ISSN 0009-2665, 1520-6890. doi: 10.1021/cr200093j. URL <https://pubs.acs.org/doi/10.1021/cr200093j>.
- [40] Michael A. Collins, Milan W. Cvitkovic, and Ryan P. A. Bettens. The Combined Fragmentation and Systematic Molecular Fragmentation Methods. *Accounts of Chemical Research*, 47(9):2776–2785, September 2014. ISSN 0001-4842. doi: 10.1021/ar500088d. Publisher: American Chemical Society.
- [41] Dmitri G. Fedorov and Kazuo Kitaura. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method.

- The Journal of Physical Chemistry A*, 111(30):6904–6914, August 2007. ISSN 1089-5639. doi: 10.1021/jp0716740. URL <https://doi.org/10.1021/jp0716740>. Publisher: American Chemical Society.
- [42] Dmitri G. Fedorov and Kazuo Kitaura. The importance of three-body terms in the fragment molecular orbital method. *The Journal of Chemical Physics*, 120(15):6832–6840, April 2004. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.1687334. URL <http://aip.scitation.org/doi/10.1063/1.1687334>.
- [43] Yoshio Nishimoto, Dmitri G. Fedorov, and Stephan Irlé. Density-Functional Tight-Binding Combined with the Fragment Molecular Orbital Method. *Journal of Chemical Theory and Computation*, 10(11):4801–4812, November 2014. ISSN 1549-9618, 1549-9626. doi: 10.1021/ct500489d. URL <https://pubs.acs.org/doi/10.1021/ct500489d>.
- [44] Tomoki Kobori, Keitaro Sodeyama, Takao Otsuka, Yoshitaka Tateyama, and Shinji Tsuneyuki. Trimer effects in fragment molecular orbital-linear combination of molecular orbitals calculation of one-electron orbitals for biomolecules. *The Journal of Chemical Physics*, 139(9):094113, September 2013. ISSN 0021-9606. doi: 10.1063/1.4818599. URL <https://aip.scitation.org/doi/full/10.1063/1.4818599>. Publisher: American Institute of Physics.
- [45] Tatsuya Nakano, Yuji Mochizuki, Katsumi Yamashita, Chiduru Watanabe, Kaori Fukuzawa, Katsunori Segawa, Yoshio Okiyama, Takayuki Tsukamoto, and Shigenori Tanaka. Development of the four-body corrected fragment molecular orbital (FMO4) method. *Chemical Physics Letters*, 523:128–133, January 2012. ISSN 0009-2614. doi: 10.1016/j.cplett.2011.12.004. URL <https://www.sciencedirect.com/science/article/pii/S0009261411014849>.
- [46] Takeshi Nagata, Dmitri Fedorov, and Kazuo Kitaura. Mathematical Formulation of the Fragment Molecular Orbital Method. In *Mathematical Formulation of the Fragment Molecular Orbital Method*, pages 17–64. ., March 2011. doi: 10.1007/978-90-481-2853-2_2.
- [47] Thom Vreven, Keiji Morokuma, Ödön Farkas, H. Bernhard Schlegel, and Michael J. Frisch. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *Journal of Computational Chemistry*, 24(6):760–769, 2003. ISSN 1096-987X. doi: 10.1002/jcc.10156. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10156>.
- [48] T. Vreven, M. J. Frisch, K. N. Kudin, H. B. Schlegel, and K. Morokuma.

- Geometry optimization with QM/MM methods II: Explicit quadratic coupling. *Molecular Physics*, 104(5-7):701–714, March 2006. ISSN 0026-8976. doi: 10.1080/00268970500417846. URL <https://doi.org/10.1080/00268970500417846>.
- [49] Stefano Caprasecca, Sandro Jurinovich, Lucas Viani, Carles Curutchet, and Benedetta Mennucci. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *Journal of Chemical Theory and Computation*, 10(4):1588–1598, April 2014. ISSN 1549-9618. doi: 10.1021/ct500021d. URL <https://doi.org/10.1021/ct500021d>. Publisher: American Chemical Society.
- [50] Dmitri G. Fedorov, Yuri Alexeev, and Kazuo Kitaura. Geometry Optimization of the Active Site of a Large System with the Fragment Molecular Orbital Method. *The Journal of Physical Chemistry Letters*, 2(4):282–288, February 2011. doi: 10.1021/jz1016894. URL <https://doi.org/10.1021/jz1016894>. Publisher: American Chemical Society.
- [51] Dmitri G. Fedorov, Toyokazu Ishida, Masami Uebayasi, and Kazuo Kitaura. The Fragment Molecular Orbital Method for Geometry Optimizations of Polypeptides and Proteins. *The Journal of Physical Chemistry A*, 111(14): 2722–2732, April 2007. ISSN 1089-5639. doi: 10.1021/jp0671042. URL <https://doi.org/10.1021/jp0671042>. Publisher: American Chemical Society.
- [52] Takayuki Tsukamoto, Yuji Mochizuki, Naoki Watanabe, Kaori Fukuzawa, and Tatsuya Nakano. Partial geometry optimization with FMO-MP2 gradient: Application to TrpCage. *Chemical Physics Letters*, 535:157–162, May 2012. ISSN 0009-2614. doi: 10.1016/j.cplett.2012.03.046. URL <https://www.sciencedirect.com/science/article/pii/S0009261412003570>.

