An Incremental Scheme for correlation energies including expansions for occupied and virtual spaces through QM/QM embedding

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Ilyas Türkmen

aus Düsseldorf

Für meine Mutter

Abstract

Wave function based electron correlation methods are reliable and systematically improvable. However, calculations on large systems are not feasible without the introduction of additional approximations due to the powerful scaling of the related methods, especially with respect to the increasing virtual space. A very successful strategy is to exploit the local nature of the electron correlation. This allows orbital spaces to be compressed, thereby reducing computational requirements. The ordinary incremental scheme is an example for a local method that provides highly accurate correlation energies which is suitable for benchmarking and achieves an accuracy of less than 1 kcal/mol.

Although the incremental expansion of the occupied space is common and has been extensively discussed in the literature, there has been very little progress towards a local virtual space expansion or a combined occupied and virtual space expansion. This thesis highlights key ideas and features of this endeavor in the context of local virtual spaces including the formulation of paradigms and a concrete implementation of a combined occupied and virtual space expansion in the framework of the incremental scheme with embedding generated virtual orbitals. A proof of concept is provided on a small set of organic molecules, water clusters and a water complex which demonstrates that the additional virtual expansion is possible and yields accurate results within chemical accuracy while reducing the computational demands dramatically. Indeed, the presented method scales asymptotically linear. Furthermore it is presented how the already reduced virtual space can be further truncated with approximate natural orbitals which are specifically designed for the needs of the incremental scheme.

Besides, a separate formal examination of the CCSD energy of the ordinary incremental scheme is also presented, which demonstrates that the incremental expansion of the occupied space corresponds to an incremental improvement of CCSD amplitude quantities.

Kurzzusammenfassung

Wellenfunktionsbasierte Elektronenkorrelationsmethoden sind zuverlässig und systematisch verbesserungsfähig. Jedoch sind Berechnungen an großen Systemen ohne die Einführung zusätzlicher Näherungen aufgrund der starken Skalierung der entsprechenden Methoden, insbesondere im Hinblick auf den wachsenden virtuellen Raum, nicht durchführbar. Eine sehr erfolgreiche Strategie besteht darin, die lokale Natur der Elektronenkorrelation auszunutzen. Dadurch können die Orbitalräume komprimiert werden, was den Rechenaufwand verringert. Die Inkrementenmethode ist ein Beispiel für eine lokale Methode, die hochgenaue Korrelationsenergien liefert, die sich für Benchmarking eignet und eine Genauigkeit von weniger als 1 kcal/mol erreicht.

Obwohl die inkrementelle Entwicklung im besetzten Raum weit verbreitet ist und in der Literatur ausgiebig diskutiert wurde, gab es nur sehr wenige Fortschritte in Richtung einer lokalen Entwicklung des virtuellen Raums oder einer kombinierten Entwicklung des besetzten und des virtuellen Raums. In dieser Arbeit werden die wichtigsten Ideen und Merkmale dieses Unterfangens im Kontext lokaler virtueller Räume hervorgehoben, einschließlich der Formulierung von Paradigmen und einer konkreten Implementierung einer kombinierten Entwicklung des besetzten und virtuellen Raums im Rahmen der Inkrementenmethode mit Embedding generierten virtuellen Orbitalen. An einer kleinen Ansammlung organischer Moleküle, Wasserclustern und eines Wasserkomplexes wird ein Konzeptnachweis erbracht, der zeigt, dass die zusätzliche virtuelle Entwicklung möglich ist und genaue Ergebnisse mit chemischer Genauigkeit liefert, während die Rechenanforderungen drastisch reduziert werden. Tatsächlich skaliert die vorgestellte Methode asymptotisch linear. Darüber hinaus wird gezeigt, wie der bereits reduzierte virtuelle Raum mit approximierten natürlichen Orbitalen, die speziell auf die Bedürfnisse der Inkrementenmethode zugeschnitten sind, weiter verkleinert werden kann.

Außerdem wird eine separate formale Untersuchung der CCSD-Energie des gewöhnlichen inkrementellen Schemas vorgestellt, die zeigt, dass die inkrementelle Erweiterung des besetzten Raums einer inkrementellen Verbesserung der CCSD-Amplitudengrößen entspricht.

Contents

Contents

Contents

Contents

1 Introduction

Quantum chemistry has gained in popularity over time and has evolved to a common tool which is used by many chemists and physicists for investigations on all kinds of chemical problems. The methods are diverse and differ in their scope of application. This is due to distinguishing approximations at various levels of theory. Highly accurate methods, which describe the correlated motion of particles, have a very steep increase in computational requirements with the number of treated particles and their applicability is thus limited to small atomic or molecular systems. Although the progress in computer hardware made mean field methods very popular and feasible for a large range of systems, still correlation effects must be included for a proper description in order to answer many chemical questions.

However, the calculation of correlation energies for large systems is a challenging task since reliable methods as CCSD or CCSD(T) are often unfeasible due to their sixth and seventh power scaling with molecular size. Several strategies have been reported seeking for approximations that yield computation time savings combined with an acceptable accuracy loss.

A very successful class of methods follows the ideas of Sinanoğlu [1] and Nesbet [2] that electron correlation is a local phenomenon. Many so called local correlation methods have emerged from these ideas. Conceptionally, the methods divide into two sub-classes: direct and fragmentation based methods. The direct methods investigate the whole system in a single correlation calculation and associate specific local correlation spaces instead of correlating the full occupied and virtual spaces.[3–12] In fragmentation based methods, the system is partitioned into smaller units and the energies of these units are determined without performing a correlation calculation on the full system.[13–19] The common bases of local methods are local orbital spaces as for example localized molecular orbitals (LMOs) [20–22], orbital specific virtuals (OSVs) [7] and pair natural orbitals (PNOs) [23, 24].

The incremental scheme [13] counts to the latter class. The energy is expanded in terms of local occupied domains. Each occupied space tuple is correlated individually. This procedure shrinks the occupied space but does not affect the virtual space. Additional approximations have been reported with either local PAO virtual spaces [14] or domain specific basis sets [15, 25] in order to reduce the virtual space size as well. However, the virtual spaces can still be large with PAOs and increases with system size in the latter approximation.

This thesis mainly addresses the issue of large virtual spaces in the environment of the incremental scheme. A solution to this problem is formulated and implemented in terms of a local incremental expansion of the virtual space. The individual virtual spaces are generated

through QM/QM embedding. The details of the new method and a proof of concept are presented in a manuscript. Furthermore, general aspects of a combined occupied and virtual space expansion are discussed within the local correlation picture.

Besides, two additional aspects of the the incremental scheme are investigated in this thesis. The first is an analysis of the CCSD energy within the ordinary incremental scheme which reveals a bridge to other fragmentation based correlation methods as the divide expand consolidate (DEC) [16] and cluster-in-molecule (CIM) [18] methods. The second is a possible adaption of approximate natural orbitals to the incremental scheme.

2 Theory

Atomic and subatomic particles behave different than those at macroscopic scale. This has been discovered experimentally [26–30] and initiated the development of a theory for such systems, which is called quantum mechanics. The distinctive features of objects at microscopic scale include discrete values of physical observables, a duality of wave-like and particle-like behaviour and superposition of states.[31, 32] Chemical systems, as molecules, are at the scale where quantum effects are predominant and thus have to be described with the quantum theory.

An isolated quantum state is represented as a wave function, which reads in position space $\Psi(\vec{x}, t)$. This function is interpreted, if normalized, as a probability amplitude for the configuration $\{\vec{x}, t\}$, where $\vec{x}$ is an abbreviated notation for all coordinates of all particles and t is the time coordinate. The square modulus of the wave function therefore represents a n-particle probability density $\rho(\vec{x}, t) = |\Psi(\vec{x}, t)|^2$.[33] Observable physical quantities as energy, momentum or position are represented by linear and Hermitian operators. The associated values of these observables are accessed mathematically as eigenvalues of the respective operators. The eigenfunctions are quantum states, called pure states, with the specific observable values. Given a superposition of pure states, a measurement of an observable yields one of the superposed pure states and the corresponding value. Without a measurement, the time-evolution of any state, pure or not, is expressed through the Schrödinger equation.[34]

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{x}, t) = \hat{H}\Psi(\vec{x}, t) \tag{2.1}$$

The Hamilton operator $\hat{H}$ is the quantum mechanical analogue of the Hamilton function in classical physics and represents the total energy of the system. This representation uses position $\vec{r}_i$ and momentum $\vec{p}_i$ variables for each particle i. Acquiring the correspondence principle, the Hamilton operator is built by replacing the ordinary position and momentum variables by the respective operators.

$$\hat{\vec{r}}_i = \vec{r}_i \qquad \hat{\vec{p}}_i = -i\hbar\vec{\nabla}_i \tag{2.2}$$

These observables are an example for complementary variables in quantum mechanics. It is not possible to measure both properties simultaneously with an arbitrary high precision, regardless of the experimental setup. The uncertainty in the observable values, denoted by

Δ, is an inherent feature. For the j-th components

$$\Delta(\hat{\vec{r}}_i)_j \, \Delta(\hat{\vec{p}}_i)_j \geq \frac{\hbar}{2} \tag{2.3}$$

holds which is related to the property that the respective operators do not commute.

$$[\,(\hat{\vec{r}}_i)_j, (\hat{\vec{p}}_i)_j\,] = i\hbar \neq 0 \tag{2.4}$$

This uncertainty principle is named after Heisenberg.[35] It can be generalized that all non commuting operators of observables have complementary variables.

If the potential energy term in $\hat{H}$ is time-independent, the time evolution for an energy eigenstate $\psi(\vec{x})$ with energy E is a complex phase (2.6) and it is sufficient to solve the time-independent Schrödinger equation (2.5).

$$\hat{H}\psi(\vec{x}) = E\psi(\vec{x}) \tag{2.5}$$

$$\Psi(\vec{x}, t) = \psi(\vec{x})\, e^{-\frac{iEt}{\hbar}} \tag{2.6}$$

This concept can be applied in quantum chemistry, when a chemical system is not exposed to an external time-dependent electromagnetic field, since the potentials of electrons and nuclei within the system are not explicitly time-dependent. It is therefore possible to calculate the initial and excited states of a system and their time evolution in the absence of an electromagnetic field using (2.5) and (2.6), but the evolution of the wave function during the excitation has to be described with (2.1). In the following, special attention is given to isolated molecular systems in the gas phase, although some concepts can also be adopted to periodic systems.

Given such a molecular system with n electrons (indices i, j) and N nuclei (indices I, J), the corresponding Hamilton operator has, in atomic units, the form of (2.7) - (2.8), where r and R denote distances, M the nuclear masses and Z the nuclear charges. This operator includes kinetic energies $\hat{T}$ of the electrons (subscript e) and nuclei (subscript n) and Coulomb potentials $\hat{V}$ between the particles.

$$\hat{H} = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{T}_n + \hat{V}_{nn} \tag{2.7}$$

$$= -\sum_{i=1}^{n} \frac{1}{2}\Delta_i - \sum_{i=1}^{n}\sum_{I=1}^{N} \frac{Z_I}{r_{iI}} + \sum_{i<j}^{n} \frac{1}{r_{ij}} - \sum_{I=1}^{N} \frac{1}{2M_I}\Delta_I + \sum_{I<J}^{N} \frac{Z_I Z_J}{R_{IJ}} \tag{2.8}$$

In order to calculate the energy eigenstates, some common approximations are used. One of those addresses the number of variables. Nuclear and electronic coordinates are treated simultaneously as variables in (2.7) - (2.8). However, the electron has a much smaller mass m, and hence inertia, compared to the nuclei. Born and Oppenheimer could show that the total energy of a molecule is separable in electronic, vibronic and rotatory energy contributions up

to the order of magnitude m/M. Couplings of these contributions appear in higher orders of m/M.[36] The electronic energy is therefore obtained in a fixed configuration of the nuclei. This reduces the number of variables, since nuclear coordinates are treated parametrically. Consequently, the kinetic energy of the nuclei vanishes and the Coulomb potential between the nuclei is constant. It can be added after the calculation of the electronic energy. The electronic Hamilton operator $\hat{H}_{el}$ is given as

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} \tag{2.9}$$

Equivalently, the electronic wave function $\psi_{el}(\vec{x})$ depends only on the electron coordinates explicitly. The index "el" will be dropped since the Born-Oppenheimer approximation is used from here on. The coordinates of each electron are composed of three spatial and one discrete spin coordinate $\vec{x}_i = (\vec{r}_i, \sigma_i) \in \mathbb{R}^3 \times \mathbb{S}$. The wave function $\psi(\vec{x})$ is thus the mapping $(\mathbb{R}^3 \times \mathbb{S})^n \to \mathbb{C}$. It has to be square integrable $\psi(\vec{x}) \in \mathcal{L}^2((\mathbb{R}^3 \times \mathbb{S})^n)$ and normalized to give a reasonable probability density. Moreover, electrons are indistinguishable for which the probability density has to be invariant under arbitrary relabeling of indices.

The spin is an intrinsic property of quantum mechanical particles. As the angular momentum, the spin is also a quantized property. That the electron spin has two possible states could be seen in the Stern-Gerlach experiment, which originally intended to demonstrate the quantization of the angular momentum.[30] Silver atoms were sent through an inhomogeneous magnetic field. The beam of atoms splits into two discrete paths, which is due to a magnetic moment of the atom. Since the total orbital angular momentum of the silver atom is zero and the spins of the electrons cancel each other except for one electron, this specific splitting could be observed. This was not recognized at that time. Uhlenbeck and Goudsmit proposed later that the electron spin is $s = 1/2$ in order to explain the fine structure of atomic spectral lines.[37] Although spin does not describe a rotational motion as the angular momentum, the mathematical description is equivalent. Both operators are vectors with x, y and z components. It is not possible to measure two or more components simultaneously but one component and the length. Therefore, spin is also treated as an angular momentum. The corresponding operator relations for a prototype angular momentum $\hat{\vec{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)$ are given as

$$[\hat{J}_i, \hat{J}_j] = i\hbar \sum_{k=1}^{3} \epsilon_{ijk} \hat{J}_k \tag{2.10}$$

$$[\hat{J}^2, \hat{J}_i] = 0 \tag{2.11}$$

where the Levi-Civita symbol ϵ_{ijk} is used and $i, j, k \in \{x, y, z\}$. The spin and angular momentum of electrons interact in an atomic potential, which is known as spin-orbit coupling.[38] As this phenomenon is a relativistic effect, it is not present in a non-relativistic treatment.

Besides the similarities to the angular momentum, spin is an intrinsic property of a quantum particle. That is, each particle type has the same spin quantum number, whereas the angular momentum quantum number can change. Particles divide into two classes, namely such with half-integer and such with integer spin quantum number. According to relativistic quantum field theory, the wave function has to be anti-symmetric with respect to the interchange of particles with half-integer spin and symmetric for integer spin particles. This is the so called spin-statistics theorem.[39] It displays the connection between spin values and the occupation statistics of energy levels for two classes of particles, namely fermions and bosons. The spin-statistics theorem identifies half-integer spin particles as fermions. Due to the anti-symmetry of the wave-function, two particles can not have the same configuration. The probability amplitude vanishes. Energetically higher configurations have thus be occupied successively, known as the Fermi-Dirac statistics.[40, 41] Integer spin particles, instead are identified as bosons. Two ore more particles can occupy the energetically lowest configuration, which is referred to as Bose-Einstein statistics.[42]

Many electron systems, like molecules for instance, have to be described with an anti-symmetric wave function. Since the many body problem can generally not be solved in a closed form, this requirement must be forced strictly for potential and approximate many electron wave functions. A systematic technique is to construct a many particle function from one-particle functions $\phi(\vec{x}_k) \in \mathcal{L}^2(\mathbb{R}^3 \times \mathbb{S})$, which are called spinorbitals. A simple sum or product of these orbitals does not satisfy the anti-symmetry principle. Introducing the antisymmetrization operator $\hat{\mathcal{A}}$, an anti-symmetric wave function, called Slater determinant $\tilde{\psi}^{\text{SL}}$, is constructed in (2.13) from orthonormal ϕt.

$$\hat{\mathcal{A}}f(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n) = \sum_{\pi \in S_n} (-1)^{\pi} f(\vec{x}_{\pi(1)}, \vec{x}_{\pi(2)}, \cdots, \vec{x}_{\pi(n)}) \tag{2.12}$$

$$\tilde{\psi}^{\text{SL}}(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n) = \frac{1}{\sqrt{n!}} \hat{\mathcal{A}} \prod_i \phi_i(\vec{x}_i) \tag{2.13}$$

Where π is a permutation from the set of all permutations of n elements S_n and $(-1)^{\pi}$ the parity of π.

In practical quantum chemistry applications, the wave function is usually approximated as a linear combination of Slater determinants.

$$\tilde{\psi} = \sum_i C_i \, \tilde{\psi}_i^{\text{SL}} \tag{2.14}$$

This is in principle an exact approach, if the one-particle basis spans the one-particle Hilbert space $\mathcal{H}_1 = \mathcal{L}^2(\mathbb{R}^3 \times \mathbb{S})$. Given such an one-particle space, the functions in the n-particle Hilbert space $\mathcal{H}_n = \mathcal{L}^2((\mathbb{R}^3 \times \mathbb{S})^n)$ can be expanded in terms of products of the one-particle functions, from which also Slater determinants (2.13) are built. Although it is in principle possible to start with any basis, it is beneficial to provide a basis that approximates the desired state

with a small number of basis functions. In cases, where the state is predominately represented by a single Slater determinant, it is a common procedure to optimize a spinorbital basis for a single determinant wave function (section 2.1) first and then optimize the coefficients of a multi determinantal wave function subsequently built from these spinorbitals. If the desired state has more than one dominant determinant, the spinorbitals have to be optimized for a multi configurational wave function consisting of more than one determinant.

The theoretical backbone of the spinorbital optimization lies in the use of hermitian Hamilton operators. Due to the Hermiticity, the eigenfunctions ψ_i of the operator are an orthonormal basis of the Hilbert space $\mathcal{H}_n$. That is, any approximate wave function $\tilde{\psi}$ can be expanded in the basis of the eigenfunctions.

$$|\tilde{\psi}\rangle = \sum_i \langle \psi_i | \tilde{\psi} \rangle |\psi_i\rangle \tag{2.15}$$

In (2.15) the notation of the functions has been changed to the Dirac notation. The vectors of the Hilbert space are no longer represented in position space, but as abstract vectors $|\cdot\rangle$ and dual vectors $\langle\cdot|$. The scalar product is represented as $\langle\cdot|\cdot\rangle$. If E_0 is the lowest eigenvalue, it follows from (2.15) that the expectation value of the energy $\langle E \rangle_{|\tilde{\psi}\rangle}$ with the approximate wave function $|\tilde{\psi}\rangle$ is bound from below with E_0.

$$\langle E \rangle_{|\tilde{\psi}\rangle} = \langle \tilde{\psi} | \hat{H} | \tilde{\psi} \rangle \geq E_0 \tag{2.16}$$

The energy can thus be varied and minimized with respect to the spinorbitals which will appear in the expectation value expression since the determinants are built from them.

2.1 Hartree-Fock theory

The Hartree-Fock (HF) method [43, 44] uses a single Slater determinant ansatz $\tilde{\psi} = \tilde{\psi}^{\mathrm{SL}}$ for the wave function and minimizes the energy expectation value with respect to the spinorbitals ϕ_i according to the variational principle based on (2.16). A great simplification of the working equations and for subsequent applications can be achieved when orthonormality is forced on the orbitals. The energy functional then reads

$$\tilde{E}_{\mathrm{HF}}[\{\phi_i\}] = \langle \tilde{\psi} | \hat{H} | \tilde{\psi} \rangle = \sum_{i=1}^n \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_{i \neq j}^n \left[\langle ij|ij \rangle - \langle ij|ji \rangle \right] \tag{2.17}$$

where the functions ϕ_i, ϕ_j are abbreviated with the indices i, j for convenience and $\hat{h}$ is an one-particle operator and therefore depends on one variable, which can be named arbitrarily. The coordinates with index one will be used from here on in such cases. The operator contains

the kinetic energy of the electron and the potential energy of all nuclei.

$$\hat{h}(1) = -\frac{1}{2}\Delta_1 - \sum_{I=1}^{N} \frac{Z_I}{r_{1I}} \tag{2.18}$$

Two-electron integrals $\langle ij|kl \rangle$ are defined as

$$\langle ij|kl \rangle = \iint \phi_i^*(1)\phi_j^*(2)\frac{1}{r_{12}}\phi_k(1)\phi_l(2)\,\mathrm{d}\vec{x}_1\,\mathrm{d}\vec{x}_2 \tag{2.19}$$

In order to guarantee orthonormality among the orbitals during the variation of the energy, Lagrangian multipliers ϵ_{ij} are introduced and the Lagrangian functional $\mathfrak{L}$ is varied instead of the energy functional itself. The following equation holds for stationary solutions.

$$\frac{\delta \mathfrak{L}}{\delta \phi_i^*} = \frac{\delta}{\delta \phi_i^*}\left(\tilde{E} - \sum_{i,j} \epsilon_{ij}(\delta_{ij} - \langle i|j \rangle)\right) = 0 \tag{2.20}$$

It shall be noted at this point, that the spinorbitals are not uniquely defined. The wave function $\tilde{\psi}$ can be written as a determinant of a matrix $\mathbf{A}_\phi$ which contains the orbitals as entries. An unitary transformation among the $\{\phi_i\}$ yields a new basis $\{\phi_i'\}$. The transformation is represented as $\mathbf{A}_{\phi'} = \mathbf{A}_\phi \mathbf{U}$, where $\mathbf{U}$ is the unitary transformation matrix. Since $\det(\mathbf{U}) = e^{i\varphi}$ for any unitary matrix and $\det(\mathbf{A}_\phi \mathbf{U}) = \det(\mathbf{A}_\phi)\det(\mathbf{U})$, this will only effect a phase change in the wave function and leave the energy invariant.

Starting from (2.20), a set of eigenvalue equations (2.21), called canonical Fock equations, can be derived for the spinorbitals, which yield the lowest possible energy expectation value with a single determinant. During this derivation, an unitary transformation among the orbitals is chosen that diagonalizes ϵ.

$$\hat{f}(1)\phi_i(1) = \epsilon_{ii}\,\phi_i(1) \tag{2.21}$$

These equations are effective one-particle equations and the solutions are one-particle states. The definition of the Fock operator $\hat{f}$ in (2.22) reveals that it has to be constructed from the solutions of (2.21).

$$\hat{f}(1) = \hat{h}(1) + \hat{G}(1) \tag{2.22}$$

$$\hat{G}(1) = \sum_{j=1}^{n} \hat{G}_j(1) = \sum_{j=1}^{n}\left[\int \frac{1}{r_{12}}\phi_j^*(2)(1 - \hat{P}_{12})\phi_j(2)\,\mathrm{d}\vec{x}_2\right] \tag{2.23}$$

The second term on the right hand side of (2.25) represents a potential, where the operator $\hat{P}_{12}$ permutes the indices of the coordinates $\vec{x}_1$ and $\vec{x}_2$. It is generated by the one-particle states ϕ_j and represents a mean field potential due to the integration over second particle's coordinate. Note, that there is no self interaction of orbital ϕ_i, since the terms cancel for the summation index $j = i$. However, combining the potential from the one-particle states with

the nuclei electron potential in $\hat{h}$ (2.18) yields an effective one-particle potential

$$\nu_{\text{eff}}^{\text{HF}}(1) = -\sum_{I=1}^{N} \frac{Z_I}{r_{1I}} + \hat{G}(1) \tag{2.24}$$

and the Fock operator is alternatively given as

$$\hat{f}(1) = -\frac{1}{2}\Delta_1 + \nu_{\text{eff}}^{\text{HF}}(1) \tag{2.25}$$

A physical situation where the Hamilton operator is a sum over Fock operators

$$\hat{H}_0 = \sum_{i=1}^{n} \hat{f}(i) \tag{2.26}$$

describes a system where n electrons are in a field of nuclei and an average field of electrons. The exact eigenfunctions of this Hamilton operator are Slater determinants, which are used in Hartree-Fock theory. The Hartree-Fock wave function is therefore an exact solution to a mean field problem.

The Fock equation is solved when the potential due to the one-particle states reproduces the same one-particle states, which is denoted as self-consistency of the field. It is therefore not possible to solve the Fock equation in one step, an iterative procedure is necessary. That is, the orbitals are guessed initially and a corresponding Fock operator is built. The solutions of (2.21) with this operator are calculated and an updated operator is built with these orbitals until the solutions do not change within a predefined convergence criterion.

Roothaan-Hall equations

Many molecular systems with an even number of electrons have singlett ground states and a single determinant is a valid approximation. Latter is possible if the determinants, which can be built from the eigenfunctions of the Fock operator, are energetically not (nearly) degenerate. These systems are referred to as closed-shell systems. It is common to split the n molecular spinorbitals (MOs) into two sets of $n/2$ MOs with the same spatial part and different spin parts.

$$\{\phi_i(\vec{x})\} \qquad \longrightarrow \qquad \{\varphi_i(\vec{r})\alpha(\omega)\} \quad \text{and} \quad \{\varphi_i(\vec{r})\beta(\omega)\} \tag{2.27}$$

In the HF framework, this method is called restricted closed-shell Hartree-Fock (RHF). The aim is to find the $n/2$ spatial orbitals, for which eigenvalue equations can be derived by inserting the definition of the spinorbitals into (2.21). The spin integration over coordinate two is directly possible and for the spin integration over particle one, a projection on the spinstate of ϕ_i is necessary. Using the orthonormality of the spinstates, this yields the spinless Fock

equation (2.28), which is analogue to (2.21) but the coordinates refer to spatial coordinates only

$$\hat{f}(1)\varphi_i(1) = \epsilon_{ii}\,\varphi_i(1) \tag{2.28}$$

and $\hat{G}(1)$ from (2.23) becomes

$$\hat{G}(1) = \sum_{j=1}^{n/2} \hat{G}_j(1) = \sum_{j=1}^{n/2} \left[\int d\vec{r}_2 \, \frac{1}{r_{12}} \varphi_j^*(2)(2 - \hat{P}_{12})\varphi_j(2) \right] \tag{2.29}$$

Numerical solutions for the spatial orbitals in (2.28) are available for atomic systems. These numerical atomic orbitals (AOs) are fitted by a set of atom centered parameterized functions, called basis set. Given a set of AOs $\{\chi_\mu\}$, spatial molecular orbitals (MOs) are usually expressed as a linear combination of these AOs, which is known as the linear combination of atomic orbitals (LCAO) approach.

$$|\varphi_i\rangle = \sum_\mu |\chi_\mu\rangle\, C_{\mu i} \tag{2.30}$$

With this parametrization, only the expansion coefficients have to be found and thus represent the new target. The LCAO ansatz converts the spinless Fock equation (2.28) into an algebraic matrix eigenvalue problem (2.32), the Roothaan-Hall equation [45], if it is projected onto χ_ν for all ν. Greek letters shall denote atomic functions.

$$\langle\nu|\hat{f}|i\rangle = \langle\nu|i\rangle\,\epsilon_{ii} \qquad \forall\,(\nu, i) \tag{2.31}$$

$$\Rightarrow \quad \mathbf{FC} = \mathbf{SC}\epsilon \tag{2.32}$$

Where $\mathbf{S}_{\nu\mu} = \langle\nu|\mu\rangle$ are entries of the overlap matrix and the Fock operator representation $\mathbf{F}_{\nu\mu} = \langle\nu|\hat{f}|\mu\rangle$ consisting of an one-particle $\mathbf{h}_{\nu\mu} = \langle\nu|\hat{h}|\mu\rangle$ and a two-particle part $\mathbf{G}$.

$$\mathbf{G}_{\nu\mu} = \sum_{\lambda\sigma} \mathbf{D}_{\lambda\sigma} \left(\langle\nu\lambda|\mu\sigma\rangle - \frac{1}{2}\langle\nu\lambda|\sigma\mu\rangle \right) \tag{2.33}$$

$$\mathbf{D}_{\lambda\sigma} = 2 \sum_{i=1}^{n/2} \mathbf{C}_{\lambda i} \mathbf{C}_{\sigma i}^* \tag{2.34}$$

$\mathbf{D}$ is called density matrix. It fully covers the self-dependent part of the Fock matrix, which can be denoted by

$$\mathbf{F}[\mathbf{D}] = \mathbf{h} + \mathbf{G}[\mathbf{D}] \tag{2.35}$$

Therefore, $\mathbf{D}$ has to be updated after each iteration with the coefficients $\mathbf{C}$. For given coefficients, $\mathbf{F}$ is constructed and subsequently transformed into an orthogonal basis. Then (2.32) reduces to an ordinary matrix eigenvalue problem $\mathbf{F}'\mathbf{C}' = \mathbf{C}'\epsilon$ and can be solved for $\mathbf{C}'$ and back transformed to $\mathbf{C}$. If a predefined convergence threshold is reached in the coefficient

matrix, density matrix or in the energy

$$\tilde{E}_{\text{HF}} = \text{Tr}\,\mathbf{Dh} + \frac{1}{2}\,\text{Tr}\,\mathbf{DG}[\mathbf{D}] \tag{2.36}$$

the procedure stops and a self-consistent field (SCF) is obtained.

Although it is often possible to reach convergence within a sufficiently large atomic basis, the solution only displays the best in the given basis. However, the lowest possible energy within the HF theory requires a complete basis set. This energy is called Hartee-Fock limit E_{HF}. A complete basis set in terms of atomic basis functions is infinitely large and thus the number of solutions to (2.32). The $n/2$ spatial orbitals with the lowest orbital energies ϵ are used to construct the n spinorbitals and the HF wave function. These orbitals are called occupied orbitals and the remaining spinorbitals are called virtual orbitals.

Scaling and prescreening

The introduction of AOs leads to a generalized eigenvalue problem (2.32). The Fock matrix is $m \times m$ dimensional for m AOs. If the Fock matrix is known, standard techniques and implementations can be used to solve the equation which scale with $\mathcal{O}(m^3)$. The even more computationally demanding step is the construction of the Fock matrix, more precisely the calculation of the four center integrals (2.19) needed for the two-electron part in (2.33). There are m^4 integrals, but each of these only contribute significantly for small molecular systems. An increased system size corresponds directly to a larger fraction of integrals containing AOs with distant atomic centers and thus smaller values. The magnitude of the integrals can be estimated with the Schwarz inequality.

$$|\langle \mu\nu|\lambda\sigma \rangle| \leq \sqrt{\langle \mu\mu|\lambda\lambda \rangle}\,\sqrt{\langle \nu\nu|\sigma\sigma \rangle} \tag{2.37}$$

It is therefore possible to estimate the values of all four center integrals with screening integrals over pairs of AOs of the type $\langle \mu\mu|\nu\nu \rangle$. Since AOs are atom centered and decrease rapidly with respect to the distance from the center, the number of non-vanishing screening integrals grows linearly for large systems. This leads to an overall asymptotic scaling of $\mathcal{O}(m^2)$ for the construction of $\mathbf{F}$. Although this scaling is formally lower than the $\mathcal{O}(m^3)$ scaling for solving the eigensystem, former dominates the CPU timings due to the larger prefactor, even for large systems.[46] The $\mathcal{O}(m^3)$ scaling becomes more prominent for extremely large systems with very long distances between the atomic centers. In this case, the number of significant two-electron integrals will scale even lower due to the vanishing operator $1/r_{12}$ and the Fock matrix will be very sparse. This opens a gate for techniques that exploit the sparsity of the matrix and therefore have a lower scaling in this step.

So far, only the integrals in (2.33) have been discussed. However, these integrals are contracted with the density matrix and thus the magnitude of the density matrix entries, which are

contracted with a specific integral, has also to be considered in the prescreening process. If $D^{\max}$ is the largest density matrix entry to be contracted with the integral $\langle \mu\nu | \lambda\sigma \rangle$, the largest possible absolute contribution to the Fock matrix can be estimated with $|D^{\max} \langle \mu\nu | \lambda\sigma \rangle|$ and (2.37). It could be shown in this context, that a differential density approach during the SCF is prefarable as the density difference $\Delta\mathbf{D}$ from the $(i-1)$-th to the i-th iteration has vanishing entries for already converged elements of the density matrix $\mathbf{D}^{i-1}$. Therefore, an increasing fraction of integrals can be neglected in the calculation of the current $\Delta\mathbf{G}^i$.[47–49]

2.2 Density functional theory

In wave function theory (WFT), molecular states are described with wave functions, complicated functions which depend on all particle variables. Properties are then calculated with the wave function. The aim to describe physical properties with a simpler and lower dimensional quantity comes naturally. The density functional theory (DFT) method uses the one-particle electron density $\rho(1)$ which depends only on a single particle's coordinates. Before proceeding with a very brief introduction to the general concepts of DFT and the practical Kohn-Sham version, a definition of the one-particle density is provided. The starting point is the n-particle probability density, which reads in a time-independent framework as $|\Psi(\vec{x},t)|^2 = |\psi(\vec{x})|^2 = \rho_n(\vec{x})$. An integration of this term over $(n-p)$ particle coordinates and considering, that the particles are indistinguishable leads to a p-particle probability density

$$\rho_p(1,\ldots,p) = \binom{n}{p} \int \cdots \int \rho(1,\ldots,n)\, d(p+1)\cdots dn \tag{2.38}$$

The one-particle probability density $\rho(1)$ is a special case depending only on one-particle's coordinates. An integration over these coordinates yields the number of electrons n and the probability $1/n\, \rho(1)\, d1$ to find one particle in configuration $\vec{x}_1 = \{\vec{r}_1, \sigma_1\}$. Spinless density matrices are obtained when only integration over the spin variable is performed. The one-particle density is a useful quantity and a theory is build around it, which is presented below. However, in wave function theory, observables are represented as expectation values of operators evaluated with wave functions. For an arbitrary p-particle interrelating operator $\hat{O}_p$ the expectation value is $\langle \hat{O}_p \rangle = \langle \psi | \hat{O}_p | \psi \rangle$. The operator acts on p-particles in the ket vector $|\psi\rangle$. It is therefore only possible to express $\langle \hat{O}_p \rangle$ in terms of $\rho_p(1,\ldots,p)$ if $\hat{O}_p \psi = \psi \hat{O}_p$, which restores $\rho_n(1,\ldots,n)$ before integral evaluation. A more general quantity, namely the n-particle density matrix $\gamma_n(1',\ldots,n';1,\ldots,n) = \psi^*(1',\ldots,n')\psi(1,\ldots,n)$ is introduced to be able to express expectation values of arbitrary operators as

$$\langle \hat{O}_p \rangle = \int \cdots \int \left[\hat{O}_p(1,\ldots,p)\gamma_p(1',\ldots,p';1,\ldots,p) \right]_{\substack{i'=i \\ \forall i \leq p}} d1 \cdots dp \tag{2.39}$$

where the p-particle density matrix γ_p is defined analogously to ρ_p as

$$\gamma_p(1', \ldots, p'; 1, \ldots, p) = \binom{n}{p} \int \cdots \int \gamma_n(1', \ldots, n'; 1, \ldots, n) \, d(p+1) \cdots dn \qquad (2.40)$$

In the wave function picture, ρ_1, γ_1 and ρ_2 have to be known for the evaluation of the electronic energy of a system, which is the expectation value of a molecular Hamilton operator (2.9) and therefore interrelates at most two particles through the term $1/r_{ij}$.

In DFT, only ρ_1 is used for the same purpose. The theoretical foundation of DFT is given by the Hohenberg-Kohn theorems, which state that the energy $E[\rho]$ of a system with n electrons in an external field V, mostly nuclear potentials, is a functional of the one-particle density and it is also variational for the ground state.[50]

The explicit functional dependence $E[\rho]$ is unknown. Starting from a wave function perspective shows, that the one-particle density matrix and the two-particle density are required to describe the kinetic and electron-electron energies, respectively. However, the constrained energy functional variation with respect to densities normalized to n, results formally in an Euler-Lagrange equation with the Lagrange multiplier μ.

$$\mu = \frac{\delta E[\rho]}{\delta \rho} \qquad (2.41)$$

In order to obtain working equations to minimize the energy functional, a functional form of E has to be provided. Kohn and Sham introduced the use of orthonormal orbitals ϕ to build a density corresponding to a Slater determinant wave function of these orbitals.[51]

$$\rho = \sum_{i=1}^{n} \phi_i^* \phi_i \qquad (2.42)$$

The mayor benefit is, that an expression for the kinetic energy functional $T_s[\rho]$ can be found in terms of the orbitals

$$T_s[\rho] = -\frac{1}{2} \sum_{i}^{n} \langle \phi_i | \Delta | \phi_i \rangle \qquad (2.43)$$

This energy represents the kinetic energy of a non-interacting system and is therefore an approximation to the correct functional $T[\rho]$. It is also possible to approximate the energy functional for electron-electron interaction $E_{ee}[\rho]$ by a Coulomb term $J[\rho]$.

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(1)\rho(2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \qquad (2.44)$$

There is an overall remaining and unknown energy functional, which is referred to as the exchange-correlation functional $E_{xc}[\rho]$. A functional variation of the energy functional, with additional orthonormality constraints for the orbitals and Lagrange multipliers ϵ, yields the

Kohn-Sham equation for the orbitals

$$\hat{f}^{\mathrm{KS}}(1)\phi_i(1) = \left(-\frac{1}{2}\Delta_1 + \hat{v}_{\mathrm{eff}}^{\mathrm{KS}}(1)\right)\phi_i(1) = \epsilon_i\phi_i(1) \tag{2.45}$$

with

$$\hat{v}_{\mathrm{eff}}^{\mathrm{KS}}(1) = -\sum_{I=1}^{N}\frac{Z_I}{r_{1I}} + \hat{J}(1) + \hat{v}_{xc}(1) \tag{2.46}$$

$$\hat{J}(1) = \int\frac{\rho(2)}{r_{12}}d\vec{r}_2 \qquad\qquad \hat{v}_{xc}(1) = \frac{\delta E_{xc}[\rho]}{\delta\rho} \tag{2.47}$$

As there is no systematic way to choose or improve $E_{xc}[\rho]$, there has been tremendous effort to find exchange-correlation functionals that yield accurate results, compared with wave function methods. The interested reader is referred to review papers on this topic.[52] For a given exchange-correlation functional, (2.45) is solved in an AO basis self-consistently. In this basis, the Kohn-Sham Fock operator reads

$$\mathbf{F}[\mathbf{D}] = \mathbf{h} + \mathbf{J}[\mathbf{D}] + \mathbf{v}_{xc}[\mathbf{D}] \tag{2.48}$$

where $\mathbf{h}$ is the matrix representation of (2.18) and the energy reads

$$E[\mathbf{D}] = \mathrm{Tr}\,\mathbf{D}\mathbf{h} + J[\mathbf{D}] + E_{xc}[\mathbf{D}] \tag{2.49}$$

2.3 Orthonormal local orbital spaces

The canonical occupied and virtual Hartree-Fock molecular orbitals, obtained from (2.32), are extended over the whole system. They are not local in the sense that they have a limited spatial extent with respect to the system size. However, LMOs have applications in many fields of quantum chemistry. A major application of LMOs takes place in the context of dynamic electron correlation. The movement of interacting electrons is correlated. Since this is a local phenomenon, taking advantage of LMOs can lead to tremendous computational savings for extended systems, when locality can be fully exploited (section 2.6). Another application of LMOs is the comparison of different chemical systems, more precisely, structural units as bonds. It is possible to compare these structural units across different molecular systems if the localization scheme pushes the MOs of the different systems towards the same kind of locality by extremalizing a localization function (section 2.3.1). Given a set of LMOs, specific active parts of a system can be described in a fixed environment, which is represented by a set of frozen LMOs (section 2.4). The underlying assumption in such a model is that the physical process of interest takes mainly place in the active parts of the system. An example is a local electronic excitation.[53]

Having motivated the benefits of local orbital spaces, a selection of popular methods to

generate LMOs are presented in the following. The conventional strategy is based on unitary transformations of canonical HF orbitals, which is therefore described in more detail.

2.3.1 Unitary transformation of canonical orbitals

Unitary transformations among occupied or virtual HF orbitals leave the energy and the wave function invariant and the orthonormality of the orbitals is conserved. Since non-relativistic theories do not require complex orbitals, orthogonal transformations

$$\left|\varphi'_p\right\rangle = \sum_q \left|\varphi_q\right\rangle U_{qp} \tag{2.50}$$

with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ are used instead to achieve locality while extremalizing a localization function. The variety of methods stems from the fact that there is no unambiguous definition of the localization criterion. The overall transformations towards the optimized MOs can be realized as a sequence of consecutive orthogonal transformations, since any product of orthogonal transformations is an orthogonal transformation.

Edmiston-Ruedenberg localization scheme

Edmiston and Ruedenberg proposed a localization algorithm based on pair wise orbital rotations, also known as Jacobi sweeps, $(\left|\varphi_p\right\rangle, \left|\varphi_q\right\rangle) \mapsto (\left|\varphi'_p\right\rangle, \left|\varphi'_q\right\rangle)$, where γ is the rotation angle.[20]

$$\left|\varphi'_p\right\rangle = \quad \cos\gamma \left|\varphi_p\right\rangle + \sin\gamma \left|\varphi_q\right\rangle \tag{2.51}$$

$$\left|\varphi'_q\right\rangle = -\sin\gamma \left|\varphi_p\right\rangle + \cos\gamma \left|\varphi_q\right\rangle \tag{2.52}$$

In the original paper, the localization function $C^{\mathrm{ER}}\left[\left|\varphi\right\rangle\right] = \sum_i \left\langle ii|ii\right\rangle$ is maximized, which is the self-interaction energy. The resulting orbitals are thus called energy localized. Replacing the operator $1/r_{12}$ with an arbitrary operator $\hat{O}$ yields a more general localization function $C\left[\left|\varphi\right\rangle\right] = \sum_i \left\langle ii|\hat{O}|ii\right\rangle$. It is maximized by a maximization of the difference

$$\Delta C(\gamma) = C\left[\left|\varphi'\right\rangle\right] - C\left[\left|\varphi\right\rangle\right] = A_{ij} + \left(A_{ij}^2 + B_{ij}^2\right)^{1/2} \cos(4\gamma - 4\alpha) \tag{2.53}$$

where A_{ij}, B_{ij} and α are defined as

$$A_{ij} = \left\langle ij|\hat{O}|ij\right\rangle - \frac{1}{4}\left\langle ii - jj|\hat{O}|ii - jj\right\rangle \tag{2.54}$$

$$B_{ij} = \left\langle ij|\hat{O}|ii - jj\right\rangle \tag{2.55}$$

$$\tan 4\alpha = -\frac{A_{ij}}{B_{ij}} \tag{2.56}$$

ΔC can be maximized with $\gamma_{\max}$ satisfying $\cos(4\gamma_{\max} - 4\alpha) = 1$. In this case (2.53) reduces to

$$\Delta C(\gamma_{\max}) = A_{ij} + \left(A_{ij}^2 + B_{ij}^2\right)^{1/2} \qquad (2.57)$$

Each iteration step involves finding the pair of orbitals with the largest ΔC and the transformation according to (2.51) - (2.52) with the corresponding $\gamma_{\max}$. The computational scaling of the localization process with the system size is caused by the scaling of the integral evaluation needed in C. For C^{ER}, it scales with $\mathcal{O}(m^5)$ for two-electron integrals, where m is the number of AOs. For the scaling properties see section 2.5.4.

Pipek-Mezey localization functional

Pipek and Mezey minimized the spatial extent of each orbital in terms of the number of atoms it is extended over. This is done through the maximization of atomic charges q_I^i on atoms I.

$$C^{\mathrm{PM}}\big[\,|\varphi\rangle\,\big] = \sum_i \sum_I \left(q_I^i\right)^2 \qquad (2.58)$$

Originally, the gross atomic Mulliken populations

$$q_I^i = \langle i|\hat{P}_I|i\rangle \qquad (2.59)$$

where $\hat{P}_I$ is a projection operator onto the atomic orbitals χ_μ on Atom I

$$\hat{P}_I = \sum_{\mu \in I} \sum_\nu \frac{1}{2}\left(\left(\mathbf{S}^{-1}\right)_{\nu\mu}|\nu\rangle\langle\mu| + |\mu\rangle\langle\nu|\left(\mathbf{S}^{-1}\right)_{\mu\nu}\right) \qquad (2.60)$$

have been used.[21, 54] A benefit, compared to the Edmiston-Ruedenberg localization, is the lowered $\mathcal{O}(m^3)$ scaling as only one-electron integrals are needed. The biggest advantage of this method is that MOs with σ and π symmetries are not mixed. However, the atomic charges strongly depends on the AO basis. Consider the extreme case where the one-particle space is spanned with atomic functions on one center only. Then all MOs already have maximal charges on this atom, but are not local. This is a conceptual deficit which can be overcome with alternative definitions of q_I^i, which do not explicitly depend on the basis set choice.[55] Although atomic charges vary strongly between different definitions, the LMOs generated with the Pipek-Mezey localization scheme do not.[55]

Boys localization functional

There are different equivalent formulations of the Foster-Boys localization criterion.[22, 56] A convenient choice is the minimization of the sum of the second moments μ_2 and involves only

one-electron integrals. The localization function C^{FB} is given as

$$C^{\mathrm{FB}}\left[\,|\varphi\rangle\,\right] = \sum_i \langle i |\left(\hat{r} - \langle i|\hat{r}|i\rangle\right)^2 |i\rangle = \sum_i \mu_2^i \qquad (2.61)$$

The minimization can be performed in an analogous procedure as in the Edmiston and Rue-denberg case, where the transformation in (2.51) - (2.52) is applied to C^{FB}. The $\mathcal{O}(m^3)$ computational scaling of the Boys localization is equivalent to the scaling of Pipek-Mezey scheme. The LMOs produced with the Boys scheme yield the most local MOs in average, compared to the methods described above. This holds for both, the occupied space and the virtual space. Nevertheless, occupied MOs localize stronger than virtual MOs in all schemes using the Jacobi sweep algorithm, since the localization function has only strong and isolated minima for the occupied but not well separated local minima for the virtual orbitals.[57] The Jacobi sweep algorithm is not able to handle the latter case for large systems.[58]

Trust-region minimization method

As described above, the Jacobi sweep algorithm is only suited for the localization of occupied orbitals. In the context of local correlation, locality can be exploited most effectively if both, occupied and virtual orbitals, are well localized. The least localized orbital will thus represent the limiting factor. As the Boys localization function (2.61) minimizes the sum of the second moments, the obtained orbitals will be the most local in average, which is a reasonable choice. Nevertheless, there are individual orbitals which are more delocalized. Jansik *et al.* addressed both issues, finding a robust unitary transformation algorithm for the virtual space and reducing the maximum orbital spread. They used a trust-region minimization method [59], which yields significantly more local virtual orbitals in combination with the Boys localization function and could also reduce the orbital tails, compared to Boys orbitals, by using larger powers of the second or fourth moment μ_4^m.[60, 61]

$$C_m = \sum_p \left(\mu_2^p\right)^m \qquad \text{or} \qquad C_m = \sum_p \left(\mu_4^p\right)^m \qquad (2.62)$$

The driving force for more locality is the increasing penalty for orbitals with larger spatial spreads. The orthogonal transformation of the orbitals is represented in the second quantized form.

$$|\tilde{p}\rangle = e^{-\hat{\kappa}} a_{p\sigma}^\dagger \,|\rangle \qquad (2.63)$$

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} \hat{E}_{pq} \qquad (2.64)$$

$$\hat{E}_{pq} = a_{p\alpha}^\dagger a_{q\alpha} + a_{p\beta}^\dagger a_{q\beta} \qquad (2.65)$$

The localization function is therefore a function of κ, where the matrix κ has to be calculated. The key idea behind the trust-region approach is to find a trust-radius $||\kappa||$ first in which a second order Taylor expansion $\tilde{C}_m$ of C_m

$$\tilde{C}_m(\kappa) = \tilde{C}_m^{(0)} + \kappa^T \tilde{\mathbf{C}}_m^{(1)} + \frac{1}{2}\kappa^T \tilde{\mathbf{C}}_m^{(2)} \kappa \tag{2.66}$$

is a valid approximation. Afterwads, the approximate subproblem is solve in this trust-region. The trust-region optimization has also been successfully applied to the Edmiston-Ruedenberg and Pipek-Mezey localization functionals and to basis sets with diffuse functions.[62–64]

2.3.2 Cholesky decomposition and energy based localization

The one-electron density matrix (2.34) has been introduced in the context of the Hartree-Fock theory, where only occupied orbitals contribute. A pseudo-density matrix $\mathbf{D}^{\mathrm{virt}}$ can be defined analogously

$$\mathbf{D}_{\lambda\sigma}^{\mathrm{virt}} = 2 \sum_{a\in\mathrm{virt}} \mathbf{C}_{\lambda a}\mathbf{C}_{\sigma a}^* \tag{2.67}$$

using the virtual orbitals. Both, the density and pseudo-density matrices, are positive semidefinite and can be Cholesky decomposed with a lower triangular matrix $\mathbf{L}$.

$$\mathbf{D}_{\lambda\sigma}^{\mathrm{virt}} = \sum_p \mathbf{L}_{\lambda p}\mathbf{L}_{\sigma p}^* \tag{2.68}$$

The entries of $\mathbf{L}$ are obtained through the following formulas

$$\mathbf{L}_{jj} = \sqrt{\mathbf{D}_{jj} - \sum_{k=1}^{j-1} \mathbf{L}_{jk}\mathbf{L}_{jk}^*} \tag{2.69}$$

$$\mathbf{L}_{ij} = \frac{1}{\mathbf{L}_{jj}}\left(\mathbf{D}_{ij} - \sum_{k=1}^{j-1} \mathbf{L}_{ik}\mathbf{L}_{jk}^*\right) \qquad j > i \tag{2.70}$$

Note, that the values and size of the resulting matrix $\mathbf{L}$ depends on the order and the number of AOs in the basis choosen for a decomposition. A specific order of the density matrix entries is called the pivoting. Usually, the pivoting is according to the largest diagonal element in the current iteration. However, the Cholesky decomposition can also be used to generate orbitals on a specific part of a system by taking only the largest diagonal elements of the density matrix with AO indices belonging to the part of interest.[65] This procedure generates, due to the partial decomposition, a reduced number of MOs in the AO basis of the part of interest which is even more local because of the lower triangular shape of the new coefficient matrix $\mathbf{L}$. The partial decomposition produces a (pseudo)-density $\mathbf{D}^A$ for the part of interest and leaves a residual (pseudo)-densitiy $\mathbf{D}^B = \mathbf{D} - \mathbf{D}^A$. With this density splitting at hand, an

energy-based orbital localization in specific spatial regions and molecular fragments has been reported.[66, 67] Conceptually, the interaction energy of the active part A and the inactive part B is maximized in a multi-level HF framework. The decomposition of the density is displayed in the HF energy (2.36) [66] as

$$\tilde{E}_{\text{HF}} = \text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{h} + \frac{1}{2}\,\text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{G}[\mathbf{D}^{\text{A}}] + \text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{G}[\mathbf{D}^{\text{B}}] + \text{Tr}\,\mathbf{D}^{\text{B}}\mathbf{h} + \frac{1}{2}\,\text{Tr}\,\mathbf{D}^{\text{B}}\mathbf{G}[\mathbf{D}^{\text{B}}] \qquad (2.71)$$

The interaction energy between the parts is given as the difference of the energy of the joint system and the energies of the individual systems $\tilde{E}_{\text{HF}}^{\text{AB}} = \tilde{E}_{\text{HF}} - \tilde{E}_{\text{HF}}^{\text{A}} - \tilde{E}_{\text{HF}}^{\text{B}}$. Here, the individual energies $\tilde{E}_{\text{HF}}^{\text{P}}$ with P$\in$\{A,B\} are

$$\tilde{E}_{\text{HF}}^{\text{P}} = \text{Tr}\,\mathbf{D}^{\text{P}}\mathbf{h}^{\text{P}} + \frac{1}{2}\,\text{Tr}\,\mathbf{D}^{\text{P}}\mathbf{G}[\mathbf{D}^{\text{P}}] \qquad (2.72)$$

Therefore, a maximization of the interaction energy is equivalent to a minimization of the sum of the individual energies. In an iterative procedure, densities are calculated and Cholesky decomposed into the fragments and the interaction energy is minimized in the fragment MO basis.

2.4 Quantum embedding

Wave function based quantum mechanical methods are computationally demanding and the required computational resources scale drastically with system size, for many of these methods. However, there are cases in which a specific part of the system is considered as active and more important than the environment of this part. If the two parts are treated on different levels of theory, this is called an embedding. The strategies vary depending on the application. For extended systems, there are methods that combine quantum mechanics (QM) with molecular mechanics (MM) on different parts. Methods, where the active part is treated with QM the environment with MM are called QM/MM.[68–72] There is also a class of methods which uses more than two layers of QM and MM parts which are known as own N-layered integrated molecular orbital and molecular mechanics (ONIOM).[73, 74]

The following section focuses on fully QM embedding techniques. Even for this selection, a comprehensive overview is out of the scope of this thesis. The following sections are intended to review around and mainly focus on WFT embedding, especially WFT-in-WFT. The Huzinaga embedding [75] and the projector based embedding [76] are described in more detail as they are used in this thesis for the generation of occupied and virtual orbitals for electron correlation calculations (section 5.1). The projector based embedding originates from one-particle density functional embedding theory [77] which is therefore also introduced conceptionally. Greens function based embedding [78] is not covered in this section.

2.4.1 Wave function theory embedding

Huzinaga-Cantu embedding

Huzinaga and Cantu investigated the separability of many-electron systems into electronic group functions.[75] This was motivated by the observation, that the closed shell wave function of an atomic system could be split into group functions of the almost independent core and valence orbitals.[79] Huzinaga and Cantu expressed the approximate wave function of a molecular system with n electrons $\tilde{\psi}$ according to (2.73) as an antisymmetrized product, denoted by the partial antisymmetrizer $\hat{\mathcal{A}}_p$, of group functions $\tilde{\psi}_1, \tilde{\psi}_2, \ldots, \tilde{\psi}_N$. Each of those group functions obeys the Pauli principle and is normalized.

$$\tilde{\psi} = C\hat{\mathcal{A}}_p\left[\tilde{\psi}_1(1,\ldots,n_1)\,\tilde{\psi}_2(n_1+1,\ldots,n_1+n_2)\cdots\tilde{\psi}_N(n_{N-1}+1,\ldots,n)\right] \tag{2.73}$$

The total energy expectation value, calculated with $\tilde{\psi}$, is additive in the energies of the electronic groups $\tilde{E} = \sum_A \tilde{E}^A$, if the strong orthogonality condition (2.74) holds.[80]

$$\int \tilde{\psi}_A^*(1,i,j,\cdots)\,\tilde{\psi}_B(1,k,l,\cdots)\,\mathrm{d}\vec{x}_1 = \delta_{AB} \tag{2.74}$$

To see this, the molecular electronic Hamilton operator is partitioned to yield

$$\hat{H} = \sum_A \hat{H}^A + \frac{1}{2}\sum_A\sum_{A\neq B}\hat{V}_{ee}^{AB} \tag{2.75}$$

with

$$\hat{H}^A = \hat{T}_e^A + \hat{V}_{en}^A + \hat{V}_{ee}^{AA} \tag{2.76}$$

$$\hat{V}_{ee}^{AB} = \sum_{i\in A}\sum_{j\in B}\frac{1}{r_{ij}} \tag{2.77}$$

The evaluation of $\tilde{E} = \langle\tilde{\psi}|\hat{H}|\tilde{\psi}\rangle$ yields the electronic group energies

$$\tilde{E}^A = \langle\tilde{\psi}_A|\hat{H}^A|\tilde{\psi}_A\rangle + \frac{1}{2}\sum_{B\neq A}\left(\iint\frac{\gamma_A(1,1)\gamma_B(2,2)}{r_{12}}\,\mathrm{d}\vec{r}_1\,\mathrm{d}\vec{r}_2 - \frac{1}{2}\iint\frac{\gamma_A(2,1)\gamma_B(1,2)}{r_{12}}\,\mathrm{d}\vec{r}_1\,\mathrm{d}\vec{r}_2\right) \tag{2.78}$$

expressed in terms of the spinless one-particle density matrix $\gamma_K(1',1)$ of the group functions $\tilde{\psi}_K$, defined in (2.40). Consider a special case where each electronic group function is described by a single Slater determinant $\tilde{\psi}_K = \tilde{\psi}_{SL,K}$. The spinorbitals are assumed to be orthonormal within and across group functions.

$$\left\langle\varphi_i^A\middle|\varphi_j^B\right\rangle = \delta_{ij}\,\delta_{AB} \tag{2.79}$$

Fixing all electronic group functions except one, namely $\tilde{\psi}_A$, leaves $\tilde{E}^A$ and also the interaction energies of the fixed groups with group A variable in the total energy. Therefore, an effective group energy $\tilde{E}_{\mathrm{eff}}^A$ can be defined by counting the second and third terms in (2.78) doubly. The effective group energy for Slater determinant group functions reduces to ordinary terms

from the HF energy expression

$$\tilde{E}_{\text{eff}}^{\text{A}} = \text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{h} + \frac{1}{2}\sum_{\text{B}\neq\text{A}}\text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{G}[\mathbf{D}^{\text{B}}] + \frac{1}{2}\text{Tr}\,\mathbf{D}^{\text{A}}\mathbf{G}[\mathbf{D}^{\text{A}}] \tag{2.80}$$

and the total energy resembles the usual Hartree-Fock energy functional (2.36). Since the aim is to describe $\tilde{\psi}^{\text{A}}$ in the fixed environment, the effective energy is varied with respect to the orbitals belonging to $\tilde{\psi}^{\text{A}}$ while restricting the variation to fulfill the orthonormality condition (2.79). The resulting Fock equation, with ϵ being unitarily transformed to be diagonal, is given in as

$$\hat{f}\left|\varphi_i^{\text{A}}\right\rangle = \epsilon_i^{\text{A}}\left|\varphi_i^{\text{A}}\right\rangle + \sum_{\text{B}\neq\text{A}}\sum_{b\in\text{B}}\theta_{ib}\left|\varphi_b^{\text{A}}\right\rangle \tag{2.81}$$

where $\hat{f}$ is the same operator as in (2.29) with the difference that orbitals which do not belong to A are fixed. θ_{ib} are Lagrangian multipliers, which encode the orthonormality between orbitals on A and B. If orthonormality holds, then $\theta_{ib} = \left\langle\varphi_b^{\text{B}}|\hat{f}|\varphi_i^{\text{A}}\right\rangle$ and (2.81) can be reformulated as an eigenvalue equation in (2.82). Here, $\hat{\rho}^{\text{B}} = \sum_b |\varphi_b^{\text{B}}\rangle\langle\varphi_b^{\text{B}}|$ projects onto the orbital space of group B. However, $\hat{\rho}^{\text{B}}\hat{f}$ is not hermitian if the frozen orbitals are not eigenfunctions of the Fock operator. It has thus to be hermitized, which can be realized with the anticommutator $\{\hat{f},\hat{\rho}^{\text{B}}\}$. The conversion of (2.81) into (2.83) is an example of the more general coupling operator method [81], where equations of the type (2.81) can be converted into eigenvalue equations for group A, even if the Fock operators would differ between the groups.

$$\left(\hat{f} - \sum_{\text{B}\neq\text{A}}\hat{\rho}^{\text{B}}\hat{f}\right)\left|\varphi_i^{\text{A}}\right\rangle = \epsilon_i^{\text{A}}\left|\varphi_i^{\text{A}}\right\rangle \tag{2.82}$$

$$\xrightarrow{\text{hermitize}} \quad \hat{f}^{\text{HC}}\left|\varphi_i^{\text{A}}\right\rangle = \left(\hat{f} - \sum_{\text{B}\neq\text{A}}\{\hat{f},\hat{\rho}^{\text{B}}\}\right)\left|\varphi_i^{\text{A}}\right\rangle = \epsilon_i^{\text{A}}\left|\varphi_i^{\text{A}}\right\rangle \tag{2.83}$$

The modified Fock equation (2.83) of the Huzinaga-Cantu (HC) embedding reads in AO basis as

$$\left(\mathbf{F} - \sum_{\text{B}\neq\text{A}}\frac{1}{2}\left[\mathbf{S}\,\mathbf{D}^{\text{B}}\,\mathbf{F} + (\mathbf{S}\,\mathbf{D}^{\text{B}}\,\mathbf{F})^{\dagger}\right]\right)\mathbf{C}^{\text{A}} = \mathbf{S}\mathbf{C}^{\text{A}}\epsilon^{\text{A}} \tag{2.84}$$

From (2.83) one can verify that $\hat{f}^{\text{HC}}$ has an eigenvalue of $-\epsilon$, for any eigenfunction of $\hat{f}$ with eigenvalue ϵ that has been fixed in group B$\neq$A. Hence, freezing an occupied orbital of $\hat{f}$ turns it into a virtual orbital of $\hat{f}^{\text{HC}}$. The additional term in $\hat{f}^{\text{HC}}$ is therefore often referred to as the level-shift operator.

Besides the attempt of splitting core and valence subsystems, there has been an application of the Huzinaga-Cantu embedding where localized molecular orbitals of the total system are

divided into active and inactive orbitals and the inactive LMOs are frozen.[82] The method has been used in combination with HF and DFT orbitals and operators.

Effective core potentials

Huzinaga and Cantu originally used their embedding to optimize valence orbitals in the fixed environment of core orbitals.[75] Assuming that core orbitals do not change significantly when moving from an atomic to a molecular systems, approximations have been introduced for the core dependent parts of the operator $\hat{f}^{\mathrm{HC}}$. They have been replaced by simpler operators and potentials and adopted to different molecular systems.[83] Under the assumption, that the core orbitals of the atoms B are eigenfunctions of $\hat{f}$, the level-shift operator is given as

$$-\sum_{\mathrm{B}}\sum_{c}(2\epsilon_c^{\mathrm{B}})\,|\varphi_c^{\mathrm{B}}\rangle\langle\varphi_c^{\mathrm{B}}| \tag{2.85}$$

It is then possible to express $\hat{f}^{\mathrm{HC}}$ in terms of atomic model potentials (MPs) $\hat{V}_{\mathrm{B}}^{\mathrm{MP}}$

$$\hat{f}^{\mathrm{HC}} = -\frac{1}{2}\Delta_1 - \sum_{\mathrm{B}}\frac{Z_{\mathrm{B}}^{\mathrm{eff}}}{r_{1\mathrm{B}}} + \sum_{\mathrm{B}}\hat{V}_{\mathrm{B}}^{\mathrm{MP}}(1) + \sum_{v}\hat{G}_v(1) \tag{2.86}$$

$$\hat{V}_{\mathrm{B}}^{\mathrm{MP}} = -\frac{Z_{\mathrm{B}}^{\mathrm{core}}}{r_{1\mathrm{B}}} + \sum_{c}\hat{G}_c(1) + \sum_{c}(2\epsilon_c^{\mathrm{B}})\,|\varphi_c^{\mathrm{B}}\rangle\langle\varphi_c^{\mathrm{B}}| \tag{2.87}$$

$\hat{G}$ is defined according to (2.23) and the indices c and v implicitly refer to core and valence orbitals respectively. Furthermore, a nuclear charge $Z_{\mathrm{B}}^{\mathrm{core}}$ associated with the core of atom B and an effective charge $Z_{\mathrm{B}}^{\mathrm{eff}} = Z_{\mathrm{B}} - Z_{\mathrm{B}}^{\mathrm{core}}$ are introduced.

The aim is to find a proper and simplified representation of $\hat{V}_{\mathrm{B}}^{\mathrm{MP}}$ in atomic calculations, since $\hat{G}$ is an integral operator which has local and non-local parts, and adopt it to molecular systems. Two conceptionally different approaches have been reported in the literature. The MP method fits the parameters of the simplified potential while reproducing valence orbitals and energies.[84] The alternative is known as the *ab initio* model potenital (AIMP) method where the individual parts of the MP operator are approximated directly, only using core orbitals.[83] The aforementioned MP methods are one of two prominent realizations of effective core potenitals (ECPs).

The second ECP method, called pseudopotential (PP) method, is based on the Phillips-Kleinman equation [85]

$$\hat{f}^{\mathrm{PK}}\,|\tilde{\varphi}_v\rangle = \left(\hat{f} - \sum_{\mathrm{B}}\sum_{c}(\epsilon_v - \epsilon_c^{\mathrm{B}})\,|\varphi_c^{\mathrm{B}}\rangle\langle\varphi_c^{\mathrm{B}}|\right)|\tilde{\varphi}_v\rangle = \epsilon_v\,|\tilde{\varphi}_v\rangle \tag{2.88}$$

which is an eigenvalue problem for pseudoorbitals $|\tilde{\varphi}_v\rangle$

$$|\tilde{\varphi}_v\rangle = a_v |\varphi_v\rangle + \sum_B \sum_c |\varphi_c^B\rangle\langle\varphi_c^B|\tilde{\varphi}_v\rangle \qquad (2.89)$$

The core eigenfunctions $|\varphi_c\rangle$ and valence eigenfunctions $|\varphi_v\rangle$ of $\hat{f}$ are also eigenfunctions of the Phillips-Kleinman Fock operator $\hat{f}^{PK}$ with identical eigenvalues ϵ_v. Noting that $\hat{f}^{PK}$ and $\hat{f}^{HC}$ have an identical structure with a different level-shift operator, it is also possible to define so called atomic pseudopotentials $\hat{V}_B^{PP}$.

$$\hat{V}_B^{PP} = -\frac{Z_B^{core}}{r_{1B}} + \sum_c \hat{G}_c(1) + \sum_c (\epsilon_v - \epsilon_c^B) |\varphi_c^B\rangle\langle\varphi_c^B| \qquad (2.90)$$

The pseudopotential depends on the valence orbital energy, in contrast to the MPs. Two different concepts have gained great acceptance to fit PPs. The first is energy-consistency. A parameterized form of the PP is adjusted to resemble a set of all electron energies. The second is shape-consistency. Shape-consistent PPs are derived for specific one-electron states by the aim to reproduce the valence orbital energy and the shape of the valence orbital. Regardless of the optimization strategy, an analytical form of the PPs have to be provided. Using the spherical symmetry of atoms, each pseudoorbital belongs to a angular quantum number l. If the pseudoorbital is known, it is possible to invert (2.88) for an atomic calculation to obtain an expression for $\hat{V}_{B,l}^{PP}$.[86] The atomic PP can thus be written in a semilocal form

$$\hat{V}_B^{PP} = \hat{V}_{B,L}^{PP} + \sum_{l=0}^{L-1} \left(\hat{V}_{B,l}^{PP} - \hat{V}_{B,L}^{PP} \right) \sum_{m_l=-l}^{l} |l, m_l\rangle\langle l, m_l| \qquad (2.91)$$

where $|l, m_l\rangle$ are angular momentum eigenstates.

Shape-consistent PPs are generated by abandoning the constraint, that pseudoorbitals are linear combinations according to (2.89). Instead, pseudoorbitals are fitted to match all-electron orbitals at large distances.[87] The point wise inversion of (2.88) requires a division by the radial part of the pseudoorbital, which is therefore chosen to be nodeless. Note, that this procedure allows only one valence orbital with angular momentum l. The numerical $\hat{V}_{B,l}^{PP}$ are finally fitted to an analytical form to be transferred to molecular AO basis calculations. Contrary, the parameters of the analytical form of the PPs are directly fitted for energy-consistent PPs.[88] For this, several atomic all electron states are calculated and the difference between all electron energies and energies calculated with PPs is minimized over a given set of atomic states. One of the benefits is, that any choice of core valence separation is allowed. Heavy elements with relativistic effects are of special interest in this context, since the direct relativistic effects are concentrated in the inner core part of the atoms. Making use of this in form of ECPs, can implicitly incorporate relativistic effects on the valence at a non-relativistic level of theory.[89, 90]

Fragment molecular orbital method

Changing the perspective from small and medium sized molecular systems to very extended ones like proteins comes, on the QM level, inevitably with tremendous computational resource requirements for the treatment of the system as a whole. One of the methods to overcome this issue is the fragment molecular orbital (FMO) method.[91–93] The system is fragmented into N_f fragments by means of chemical functional groups without splitting electron bond pairs. The electronic energy of the system is expanded in a many-body expansion in terms of the fragments up to a maximum expansion order of three. The energy expansion (2.92) is expressed in the powerset $\mathcal{P}$ notation to keep the notation consistent with other chapters of the thesis.

$$E = \sum_{\substack{\mathbb{X} \in \mathcal{P}(\{1,...,N_f\}) \\ |\mathbb{X}| < 3}} \Delta E_{\mathbb{X}} \quad \text{with} \quad \Delta E_{\mathbb{X}} = E_{\mathbb{X}} - \sum_{\substack{\mathbb{Y} \in \mathcal{P}(\mathbb{X}) \\ |\mathbb{Y}| < |\mathbb{X}|}} \Delta E_{\mathbb{Y}} \tag{2.92}$$

The fragment energies and energies of fragment tuples are calculated in an electrostatic potential of the fragments which are not part of the respective tuples. The electrostatic potentials $\hat{V}_{\mathrm{P}}^{\mathrm{ES}}$ of the individual fragments P are given as

$$\hat{V}_{\mathrm{P}}^{\mathrm{ES}}(1) = \int \frac{\rho_{\mathrm{P}}(2)}{r_{12}} d\vec{r}_2 \tag{2.93}$$

The electron densities ρ_{P} are not known *a priori* and a calculation on the full system has to be avoided. The key idea of the FMO method is to calculate initial fragment densities, adopt these in $\hat{V}_{\mathrm{P}}^{\mathrm{ES}}$ for repeated fragment optimization calculations until a self-consistent field of all fragments is present. The converged fragment densities are than used in the pair fragment calculations etc. The Fock operator for any fragment tuple $\mathbb{X}$ in the field of all other fragments is given, at closed-shell HF level, as

$$\hat{f}_{\mathbb{X}}^{\mathrm{FMO}}(1) = \hat{f}_{\mathbb{X}}(1) + \sum_{\mathrm{P} \notin \mathbb{X}} \left(-\frac{Z_{\mathrm{P}}}{r_{1\mathrm{P}}} + \hat{V}_{\mathrm{P}}^{\mathrm{ES}}(1) \right) + \sum_{c} B_c |c\rangle\langle c| \tag{2.94}$$

where $\hat{f}_{\mathbb{X}}$ represents the HF operator only on $\mathbb{X}$. The Fock operator $\hat{f}_{\mathbb{X}}^{\mathrm{FMO}}$ is diagonalized in the AO basis of the atoms in $\mathbb{X}$. Additional basis functions of neighboring atoms are also included if $\mathbb{X}$ is connected to them by bond orbitals, which are assigned to $\mathbb{X}$. This is also the origin of the last term in (2.94). It is responsible for level-shifting core orbitals through atomic core basis functions $|c\rangle$ and sufficiently large B_c.

2.4.2 Density functional theory embedding: DFT-in-DFT, WFT-in-DFT and WFT-in-HF

Dividing a system in an active part A and an environment B is expressed in density embedding through the densities ρ_{A} and ρ_{B} associated with the two parts. The summed densities must be

equal to the density of the total system ρ. In the framework of Kohn-Sham DFT (section 2.2), each part of the energy functional can be expressed equivalently by substituting $\rho = \rho_A + \rho_B$, except the kinetic energy functional $T_s[\rho]$ (2.43). This functional is derived from a wave function of the total system and not directly from a density. Therefore, $T_s[\rho]$ will also have a non-additive part $T^{\mathrm{nadd}}[\rho_A, \rho_B]$, if the densities ρ_A and ρ_B are not built from mutually orthogonal orbitals.[77]

$$T_s[\rho] = T_s[\rho_A] + T_s[\rho_B] + T_s^{\mathrm{nadd}}[\rho_A, \rho_B] \tag{2.95}$$

Keeping ρ_B fixed and varying the so obtained energy functional with respect to ρ_A and therefore the Kohn-Sham orbitals ϕ^A associated with A, yields a modified version of the Kohn-Sham operator in (2.45) and contains $T^{\mathrm{nadd}}[\rho_A, \rho_B]$.

$$\hat{f}_A^{\mathrm{KS}}(1) = -\frac{1}{2}\Delta_1 + \hat{v}_{\mathrm{eff}}^{\mathrm{KS}}(1) + \frac{\delta T^{\mathrm{nadd}}[\rho_A, \rho_B]}{\delta \rho_A} \tag{2.96}$$

The authors of the original paper used the simple Thomas-Fermi model to approximate T^{nadd}.[77] Other approximate or exactly methods to calculate T^{nadd} have also been reported.[94, 95]

Projector based embedding

Another strategy is to prevent the usage of the non-additive kinetic energy term by forcing the orbitals of A and B to be orthogonal, since non-orthogonality is the origin of T^{nadd}. The energy of the system can, in case of orthogonality, be expressed according to (2.49) as

$$E[\mathbf{D}^A; \mathbf{D}^B] = \mathrm{Tr}\left(\mathbf{D}^A + \mathbf{D}^B\right)\mathbf{h} + J[\mathbf{D}^A + \mathbf{D}^B] + E_{xc}[\mathbf{D}^A + \mathbf{D}^B] \tag{2.97}$$

The orthogonality can be forced through the Fock operator for system A (2.96), where the non-additive kinetic energy is assumed to vanish by augmenting the operator with a level-shift operator

$$\hat{f}_A^{\mathrm{KS}}(1) = -\frac{1}{2}\Delta_1 + \hat{v}_{\mathrm{eff}}^{\mathrm{KS}}(1) + \mu \sum_b |\phi_b^B\rangle\langle\phi_b^B| \tag{2.98}$$

with a sufficiently large shifting constant μ.[76] A DFT-in-DFT embedding can in this framework be realized if $\hat{v}_{\mathrm{eff}}^{\mathrm{KS}}$ is chosen to be different for the calculation on A than on B. Since (2.98) is equivalent to a HF calculation on A if $\hat{v}_{\mathrm{eff}}^{\mathrm{KS}}$ is adjusted to yield $\hat{v}_{\mathrm{eff}}^{\mathrm{HF}}$, this can also be considered as a HF-in-DFT embedding. A HF-in-HF embedding is also possible if the frozen density is obtained from a HF calculation. Furthermore, any WFT-in-DFT embedding can be realized by incorporation of embedding terms into the core Hamilton operator of system A, such that the Fock operator (2.98) is obtained. The modified Hamilton operator is denoted as $\hat{H}^{\mathrm{A\ in\ B}}$. The total energy of the system, described by a wave function on subsystem A ψ^A

and the frozen density on subsystem B, is then [76]

$$E[\psi^{A}; D^{B}] = \left\langle \psi^{A} \middle| \hat{H}^{A \text{ in } B} \middle| \psi^{A} \right\rangle - \text{Tr } D^{A}(\mathbf{v}_{xc}[D^{A} + D^{B}]) - \mathbf{v}_{xc}[D^{A}])$$
$$+ E_{xc}[\mathbf{D}^{A} + \mathbf{D}^{B}] - E_{xc}[\mathbf{D}^{A}] + E[0; \mathbf{D}^{B}] \tag{2.99}$$

An interesting aspect of WFT-in-DFT or WFT-in-HF embedding is a truncation of the basis set on the active WFT subsystem, which has been investigated with respect to errors that can occur due to reduced orthogonality.[96] Within this analysis, the authors proposed an algorithm to lower the errors due to AO basis set truncation by using only orbitals in the vicinity of the subsystem A in the level-shift operator.

An alternative scheme for DFT-in-DFT and WFT-in-DFT uses the Huzinaga-Cantu embedding and the corresponding level-shift operator in the embedded Fock equation in order to avoid the arbitrary choice of the level-shift constant μ.[82] Both methods use localized orbitals for the partitioning of the system and to construct the frozen densities.

2.5 Wave function based electron correlation methods

In time-independent quantum chemistry, the time-independent Schrödinger equation (2.5) has to be solved, where antisymmetric wave functions with respect to particle interchange describe states. These functions depend on all particle variables, which makes an analytical solution impossible. The Hartee-Fock method (section 2.1) circumvents the many-particle problem by approximating the wave function as a single Slater determinant, fulfilling the antisymmetry requirement by construction, built from orthonormal one-particle functions. It turns out that energy minimizing one-particle functions can be obtained as eigenfunctions of an one-particle Fock operator $\hat{f}(1)$ (2.21), which contains an effective potential of the remaining electrons. The approximate Hartree-Fock wave function for the original problem is therefore equivalently an exact solution to the mean field problem with the Hamilton operator $\hat{H}_{0}$ (2.26).

Since the mean field character, which suggest a somehow uncorrelated system, the missing energy

$$E_{\text{corr}} = E - \tilde{E}_{\text{HF}} \tag{2.100}$$

has been termed electron correlation energy. However, the movement of the electrons is not uncorrelated in a strict mathematical sense. Correlation in a mathematical sense means that the joint probability of finding two particles in configurations $\vec{x}_1$ and $\vec{x}_2$ does not factorize to a product of individual probabilities. The joint probability is given by the two-particle density $\rho_2(\vec{x}_1, \vec{x}_2)$ and the individual probabilities by the one-particle density $\rho_1(\vec{x}_1)$ and $\rho_1(\vec{x}_2)$. For
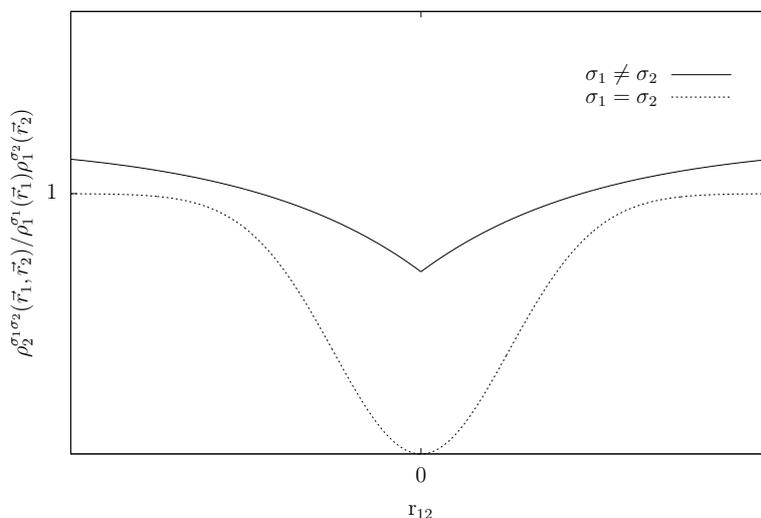
Figure 2.1: Schematic illustration of the Coulomb cusp ($\sigma_1 \neq \sigma_2$) and Fermi hole ($\sigma_1 = \sigma_2$).

the HF Slater determinant one obtains

$$\rho_1^{\mathrm{HF}}(\vec{x}_1) = \sum_i |\phi_i(\vec{x}_1)|^2 \tag{2.101}$$

$$\rho_2^{\mathrm{HF}}(\vec{x}_1, \vec{x}_2) = \sum_{ij} \left(|\phi_i(\vec{x}_1)|^2 |\phi_j(\vec{x}_2)|^2 - \phi_i^*(\vec{x}_1)\phi_j(\vec{x}_1)\phi_j^*(\vec{x}_2)\phi_i(\vec{x}_2) \right) \tag{2.102}$$

Considering that the two particles have σ_1 and σ_2 spins, that is the particle with σ_1 spin is associated with σ_1-spinorbitals and particle with σ_2 spin with σ_2-spinorbitals, spin integration yields the spinless two-particle density matrix blocks.

$$\rho_2^{\mathrm{HF},\sigma_1\sigma_2}(\vec{r}_1, \vec{r}_2) = \begin{cases} \rho_1^{\mathrm{HF},\sigma_1}(\vec{r}_1)\rho_1^{\mathrm{HF},\sigma_2}(\vec{r}_2) - \gamma_1^{\mathrm{HF},\sigma_1}(\vec{r}_1, \vec{r}_2)\gamma_1^{\mathrm{HF},\sigma_2}(\vec{r}_2, \vec{r}_1) & \sigma_1 = \sigma_2 \\ \rho_1^{\mathrm{HF},\sigma_1}(\vec{r}_1)\rho_1^{\mathrm{HF},\sigma_2}(\vec{r}_2) & \sigma_1 \neq \sigma_2 \end{cases} \tag{2.103}$$

Equation (2.103) shows that same-spin particles are mathematically correlated in HF theory, referred to as Fermi hole (fig. 2.1), but particles with different spins are not. Former is explicitly forced by construction to fulfill the antisymmetry requirement for the wave function. The exact wave function should also be correlated for two charged particles with different spin, which is known as the Coulomb cusp (fig. 2.1).

Many methods have been developed to access the correlation energy, which is about 1% of the total energy.[97] Energy differences in this order of magnitude can not fully resolve the needed chemical accuracy, or thermal energy, of about 1 kcal/mol. The methods rely on the fact that any many-particle wave function in the n-particle Hilbert space can be superposed from a product basis of one-particle Hilbert space basis functions. The HF orbitals are such a basis. Instead of using the product basis directly, linear combinations of Slater determinants

are used to ensure the antisymmetry of the wave function. There is a variety of methods tackling different weaknesses of HF theory. If a single Slater determinant is not an accurate description of the ground state, which often happens if several one-particle states are nearly degenerate, there is static correlation. The methods which specifically address this issue (MCSCF, CASSCF, RASSCF) are not discussed in the following. The main focus lies on methods (MBPT,CC), where HF is a reasonable reference and correlation effects can mainly be rooted to the Coulomb cusp, representing dynamical correlation.

2.5.1 Configuration interaction

The configuration interaction (CI) method requires a preceding HF calculation to provide an occupied space $\mathbb{O}$ and a virtual space $\mathbb{V}$. The indices $i, j, k, l, \ldots$ shall denote occupied and $a, b, c, d, \ldots$ virtual orbital indices. The HF reference determinant $|\tilde{\psi}_0\rangle$ consists only of occupied orbitals. A systematic exchange of one, two, etc. occupied orbitals $i_1, i_2, \ldots, i_n$ with virtual orbitals $a_1, a_2, \ldots, a_n$ leads to singly, doubly, $\ldots$, n-fold substituted determinants. A p-fold substituted determinant is expressed as $\left|\tilde{\psi}_{i_1 i_2 \ldots i_p}^{a_1 a_2 \ldots a_p}\right\rangle$. The CI wave function ansatz is a linear combination of substituted determinants. A special case is the full CI (FCI) method, where all possible determinants are used.

$$|\psi_{\text{FCI}}\rangle = |\tilde{\psi}_0\rangle + \sum_{\substack{i_1 \\ a_1}} C_{i_1}^{a_1} \left|\tilde{\psi}_{i_1}^{a_1}\right\rangle + \sum_{\substack{i_1 < i_2 \\ a_1 < a_2}} C_{i_1 i_2}^{a_1 a_2} \left|\tilde{\psi}_{i_1 i_2}^{a_1 a_2}\right\rangle + \cdots + \sum_{\substack{i_1 < i_2 < \cdots < i_n \\ a_1 < a_2 < \cdots < a_n}} C_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n} \left|\tilde{\psi}_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n}\right\rangle$$

(2.104)

This wave function is exact, since the Hilbert space $\mathcal{H}_n$ is spanned. Each determinant has an independent linear coefficient, which can be found variationally by minimizing the FCI energy $E^{\text{FCI}} = \langle\psi_{\text{FCI}}|\hat{H}|\psi_{\text{FCI}}\rangle$ with respect to the linear coefficients. This is equivalent to solving the matrix eigenvalue equation $\mathbf{HC} = \mathbf{EC}$, where $\mathbf{H}$ is the Hamilton operator represented in the basis of the substituted determinants. A classification of these determinants into singles, doubles, triples etc. symbolically written as $|S\rangle$, $|D\rangle$, $|T\rangle$ etc. respectively, allows to present the block structure (2.105) of the full matrix in a convenient way. The FCI matrix is $\binom{m}{n} \times \binom{m}{n}$ dimensional, where m is the number of spinorbitals. The FCI method is thus practically not feasible, except for very small systems.

$$\mathbf{H} = \begin{pmatrix} \langle\tilde{\psi}_0|\hat{H}|\tilde{\psi}_0\rangle & 0 & \langle\tilde{\psi}_0|\hat{H}|D\rangle & 0 & \cdots \\ & \langle S|\hat{H}|S\rangle & \langle S|\hat{H}|D\rangle & \langle S|\hat{H}|T\rangle & \cdots \\ & & \langle D|\hat{H}|D\rangle & \langle D|\hat{H}|T\rangle & \cdots \\ & & & \langle T|\hat{H}|T\rangle & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(2.105)

Note, that $\langle\tilde{\psi}_0|\hat{H}|S\rangle = \langle a|\hat{f}|i\rangle = 0$ for single substitutions $|S\rangle = |\tilde{\psi}_i^a\rangle$, since canonical HF orbitals have vanishing off-diagonal Fock matrix entries.[98] The structure of the matrix is

due to the two-particle interrelating Hermitian Hamilton operator and the orthonormality of the orbitals. The Slater-Condon rules apply under these conditions and any $\langle P|\hat{H}|Q\rangle$ entry vanishes when $|P\rangle$ and $|Q\rangle$ differ by more than two substitutions.[99, 100] An alternative, but equivalent, formulation of the FCI problem is a projective approach, where the Schrödinger equation

$$(\hat{H} - \tilde{E}_{\mathrm{HF}})\,|\psi_{\mathrm{FCI}}\rangle = E_{\mathrm{corr}}\,|\psi_{\mathrm{FCI}}\rangle \tag{2.106}$$

is projected onto each many-particle basis function in the expansion to obtain coupled equations for the coefficients and the energy. Using an intermediate normalization $\langle\tilde{\psi}_0|\psi_{\mathrm{FCI}}\rangle = 1$ and a projection onto $|\tilde{\psi}_0\rangle$ yields an equation for the correlation energy.[1, 2]

$$E_{\mathrm{corr}} = \left\langle\tilde{\psi}_0\right|(\hat{H} - \tilde{E}_{\mathrm{HF}})|\psi_{\mathrm{FCI}}\rangle = \sum_{\substack{i_1<i_2\\a_1<a_2}} C_{i_1 i_2}^{a_1 a_2}\left\langle\tilde{\psi}_0\Big|\hat{H}\Big|\tilde{\psi}_{i_1 i_2}^{a_1 a_2}\right\rangle \tag{2.107}$$

The exact correlation energy is calculated with the exact doubles coefficients, which can not be calculated independently of the other coefficients. Nevertheless, this opens a route for methods that focus on the accurate calculation of doubles coefficients.

As the FCI wave function expansion is not feasible, a truncation seems favorable for the sake of applicability. A truncation after single and double substitutions is referred to as CI singles doubles (CISD). Higher order truncations are possible as well. Truncating at the p-th substitution level corresponds to determinants with at most p-fold substitutions and the number of substitutions M is

$$M = \sum_{k=1}^{p} \binom{n}{k}\binom{m-k}{k} \tag{2.108}$$

Although the truncation is associated with less computational requirements, the fundamental properties of the exact wave function, size-extensivity and size-consistency (section 2.5.2), are lost along.

2.5.2 Coupled cluster theory

The coupled cluster (CC) theory [101] is distinguished from other methods by the wave function construction scheme. It is specially designed to provide the size-consistency and size-extensivity (2.5.2) features of the exact wave function at any truncation level, other than the truncated CI versions.

Size-extensivity and size-consistency

Any approximate wave function should provide a proper scaling of the correlation energy with respect to the system size. One criterion is called size-consistency. Assuming two well separated and non-interacting subsystems A and B, the correlation energy should be

additive.[102]

$$E_{\text{corr}}^{AB} = E_{\text{corr}}^A + E_{\text{corr}}^B \tag{2.109}$$

Size-consistency holds if the wave function of the total system factorizes with respect to the subsystem wave functions. Size-consistency is a natural and weak condition. A more sophisticated concept is the size-extensivity. A method is size-extensive if the correlation energy scales linearly with identical correlating units.

While the FCI wave function is both, size-consistent and size-extensive, a truncation causes the loss of both properties. The CISD wave functions for two subsystems ψ_A and ψ_B have at most doubly substituted determinants. Another CISD wave function for the combined system ψ_{AB} also allows for a maximum of two substitutions. It is therefore not possible to express the total wave function in a product separated form $\psi_{AB} \neq \psi_A \psi_B$, since this would require quadruple substitutions with double substitutions on A and B.

CC wave function ansatz

The CC wave function (2.112) is defined in second quantized form. That is, operators that act on a predefined set of many-body functions and convert these into each other. The many-body basis includes all possible Slater determinants from the eigenbasis of the Fock operator, which have already been presented in the context of CI. The substitution operator $\hat{t}_\mu$ substitutes a set of occupied orbitals by a set of virtual orbitals in the function it acts on. An application of an arbitrary $\hat{t}_\mu$ on the HF reference produces many-particle basis functions.

$$\hat{t}_\mu \left| \tilde{\psi}_0 \right\rangle = \hat{t}_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n} \left| \tilde{\psi}_0 \right\rangle = \left| \tilde{\psi}_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n} \right\rangle \tag{2.110}$$

The same orbital can not be doubly occupied in a determinant. Any manipulation that produces this case yields vanishing determinants. This fact, and also the antisymmetry of Slater determinants, are encoded in the algebra of elementary substitution operators which build $\hat{t}_\mu$. In particular, $(\hat{t}_\mu)^m = 0$ (nilpotency) and $[\hat{t}_\mu, \hat{t}_\nu] = 0$ (commutativity) holds for substitution operators. All possible determinants could be generated by application of the cluster operator

$$\hat{T} = \sum_{\substack{i_1 < i_2 < \cdots < i_n \\ a_1 < a_2 < \cdots < a_n}} t_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n} \hat{t}_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n} \tag{2.111}$$

to the HF reference. The expansion coefficients $t_{i_1 i_2 \ldots i_n}^{a_1 a_2 \ldots a_n}$ are called cluster amplitudes. However, this ansatz would reproduce the CI wave function and it is not desired for the sake of size-consistency and size-extensivity. The CC ansatz is a product form (2.112) and this ansatz is equivalent to a wave operator ansatz with the mentioned properties of the substitu-

tion operators

$$\left|\tilde{\Psi}_{\mathrm{CC}}\right\rangle = \prod_{\mu}\left(1 + t_\mu \hat{t}_\mu\right)\left|\tilde{\Psi}_0\right\rangle = \prod_{\mu} e^{t_\mu \hat{t}_\mu}\left|\tilde{\Psi}_0\right\rangle = e^{\sum_\mu t_\mu \hat{t}_\mu}\left|\tilde{\Psi}_0\right\rangle = e^{\hat{T}}\left|\tilde{\Psi}_0\right\rangle \qquad (2.112)$$

The term $(1 + t_\mu \hat{t}_\mu)$ can be augmented with higher order terms due to nilpotency and thus replaced by the full Taylor expansion of the exponential $e^{t_\mu \hat{t}_\mu}$. The third equality holds due to $[\hat{t}_\mu, \hat{t}_\nu] = 0$ for disjoint occupied and virtual spaces. In the obtained exponential form, the cluster operator can be truncated to an arbitrary substitution level k, that is $\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_k$. Nevertheless, the truncated CC wave function still includes all higher order substitutions represented as products of lower order substitutions. These higher order substitutions do not have independent expansion coefficients and are determined by lower order amplitudes. There are contributions of lower order substitution level amplitudes to higher once, even below the maximum substitution order k of the truncated $\hat{T}$. The operators $\hat{T}_1^2$ and $\hat{T}_2$ do for example both contribute to substitution order two. Although this behaviour is favorable with respect to wave function properties, it also complicates the variational solution tremendously. The CC energy functional $\tilde{E}_{\mathrm{CC}}[t_\mu]$ is non-linear in the amplitudes t_μ and furthermore involves the calculation of $\left\langle e^{\hat{T}}\tilde{\Psi}_0 \middle| \hat{H} \middle| e^{\hat{T}}\tilde{\Psi}_0 \right\rangle$ which will not truncate at any substitution order, even if $\hat{T}$ is truncated, since the bra as well as the ket vector contain all substitution orders and $\hat{H}$ can interrelate a maximum order difference of two. Variational CC is thus not feasible. However, a projection based approach exists, which at least circumvents the latter problem of variational CC.

Projected CC equations

Starting with the Schrödinger equation (2.113) one obtains the similarity transformed version by multiplication with $e^{-\hat{T}}$.

$$\hat{H}\left|\tilde{\Psi}_{\mathrm{CC}}\right\rangle = \hat{H}\, e^{\hat{T}}\left|\tilde{\Psi}_0\right\rangle = E_{\mathrm{CC}}\, e^{\hat{T}}\left|\tilde{\Psi}_0\right\rangle \qquad (2.113)$$

$$e^{-\hat{T}}\,\hat{H}\, e^{\hat{T}}\left|\tilde{\Psi}_0\right\rangle = E_{\mathrm{CC}}\left|\tilde{\Psi}_0\right\rangle \qquad (2.114)$$

The so obtained operator $e^{-\hat{T}}\,\hat{H}\, e^{\hat{T}}$ has the HF reference as an eigenfunction with the CC energy as an eigenvalue. It is however not Hermitian and can not be used in a variational treatment. Instead, projecting (2.114) onto the reference state yields an equation for the energy (2.115). Projections onto each many-particle basis function $\left|\tilde{\Psi}_\mu\right\rangle$ provide an equivalent number of equations and amplitudes (2.116).

$$\left\langle \tilde{\Psi}_0 \middle| e^{-\hat{T}}\,\hat{H}\, e^{\hat{T}}\,\tilde{\Psi}_0 \right\rangle = E_{\mathrm{CC}} \qquad (2.115)$$

$$\left\langle \tilde{\Psi}_\mu \middle| e^{-\hat{T}}\,\hat{H}\, e^{\hat{T}}\,\tilde{\Psi}_0 \right\rangle = 0 \qquad (2.116)$$

The amplitude equations are explicitly formulated through the use of the Baker-Campbell-Hausdorff (BCH) expansion [103–105]

$$e^{-\hat{T}}\,\hat{H}\,e^{\hat{T}} = \hat{H} + \left[\hat{H},\hat{T}\right] + \frac{1}{2!}\left[\left[\hat{H},\hat{T}\right],\hat{T}\right] + \frac{1}{3!}\left[\left[\left[\hat{H},\hat{T}\right],\hat{T}\right],\hat{T}\right] + \frac{1}{4!}\left[\left[\left[\left[\hat{H},\hat{T}\right],\hat{T}\right],\hat{T}\right]\hat{T}\right] \quad (2.117)$$

which truncates already at the fourth nested commutator order. These nonlinear and coupled equations have to be solved iteratively. The equation for the CC energy from (2.115) allows due to (2.117) a maximum of double substitutions and the energy therefore reads for all substitution orders larger or equal two as

$$E_{\mathrm{CC}} = \tilde{E}_{\mathrm{HF}} + \sum_{ia} f_{ia}t_i^a + \frac{1}{4}\sum_{ijab}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\left(\langle ij|ab\rangle - \langle ij|ba\rangle\right) \quad (2.118)$$

Truncated CC methods, as CCSD, have become very popular for medium sized systems and it has been shown that it can cope with most of the total correlation energy. The CCSD(T) [106] with a triple subsection correction on top of the CCSD energy is even viewed as the gold standard.[107] Nevertheless, the needed amplitude equations are expensive. This can be measured by the scaling of the tensor contractions (section 2.5.4) of the individual terms.

2.5.3 Many-body perturbation theory

The Rayleigh-Schrödinger perturbation theory (RSPT) is a formal theory in which the Hamilton operator $\hat{H}$, with unknown eigenfunctions, is divided into a unperturbed part $\hat{H}_0$, with known eigenstates, and a small perturbation $\hat{U}$.[108, 109] A linear expansion of the perturbed wave function $|\Psi\rangle$ and the energy E in terms of p-th order corrections $\left|\Psi^{(p)}\right\rangle$ and $E^{(p)}$ respectively is inserted into the Schrödinger equation. This results in equations relating the corrections of different orders.

$$\left(\hat{H}_0 - E^{(0)}\right)\left|\Psi^{(p)}\right\rangle = \sum_{q=1}^{p}\left(E^{(q)} - \hat{U}\delta_{q1}\right)\left|\Psi^{(p-q)}\right\rangle \qquad \forall\, p \in \mathbb{N} \quad (2.119)$$

A projection of this equation onto $\left\langle\Psi^{(0)}\right|$ gives an expression for the energy corrections

$$E^{(p)} = \left\langle\Psi^{(0)}\left|\hat{U}\right|\Psi^{(p-1)}\right\rangle \qquad \forall\, p \in \mathbb{N} \quad (2.120)$$

where p-th order energy corrections are calculated with $(p-1)$-th order wave function corrections. These energies can in turn be used to solve (2.119) for $\left|\Psi^{(p)}\right\rangle$. The correcting wave functions are expressed in the known basis of $\hat{H}_0$ for this purpose.
A specific choice the unperturbed Hamilton operator defines a perturbation theory. As the subject of interest is the correlation energy, it is possible to start from a mean field model as the unperturbed system, if it is a reasonable starting point. The Møller-Plesset pertur-

bation theory (MPPT) uses the mean field Hamiltonian $\hat{H}_0$ (2.26), a sum of Fock operators from HF theory. The zeroth order energy $\left\langle \Psi^{(0)} \middle| \hat{H}_0 \middle| \Psi^{(0)} \right\rangle$ is therefore not the Hartree-Fock energy although $|\psi_0\rangle = |\tilde{\psi}_{\mathrm{HF}}\rangle$. Instead, the HF energy is reproduced in the first order energy correction.

$$E_{\mathrm{MPPT}}^{(0)} + E_{\mathrm{MPPT}}^{(1)} = \left\langle \Psi^{(0)} \middle| \hat{H}_0 \middle| \Psi^{(0)} \right\rangle + \left\langle \Psi^{(0)} \middle| \hat{U}_{\mathrm{MPPT}} \middle| \Psi^{(0)} \right\rangle = \tilde{E}_{\mathrm{HF}} \tag{2.121}$$

The first correlation energy contributions are thus present in second order corrections. These can be calculated with first order wave function corrections, which are obtained from (2.119) by expanding $\left| \Psi^{(1)} \right\rangle$ in the eigenbasis of $\hat{H}_0$. That is, all possible Slater determinants, as in the FCI expansion (2.104). However, $\hat{U}_{\mathrm{MPPT}}$ is also a two particle interrelating operator. Therefore, $\langle \psi_0 | \hat{H} | X \rangle$ vanish for single and more than double substitutions X. The first order wave function correction and the second order energy are thus

$$\left| \Psi^{(1)} \right\rangle = \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab} \left| \tilde{\psi}_{ij}^{ab} \right\rangle \qquad \text{with} \qquad t_{ij}^{ab} = -\frac{\langle ab|ij \rangle - \langle ab|ji \rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{2.122}$$

$$E_{\mathrm{MPPT}}^{(2)} = \sum_{\substack{i<j \\ a<b}} \frac{|\langle ab|ij \rangle - \langle ab|ji \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{2.123}$$

The method where this perturbative expansion is stopped after second order is termed MP2.

2.5.4 Tensor contraction: two-particle integrals

The presented correlation methods are based on a previous Hartree-Fock calculations in which molecular one-particle eigenstates of the Fock operator are calculated approximately. The many-particle basis is then built upon this one-particle basis. Therefore, correlation methods do involve tensors with molecular orbital indices. This can be coefficients or amplitudes of Slater determinants from the wave function expansion or two-particle integrals of the type $I_{pqrs} = \langle pq|rs \rangle$. The MOs are expanded in terms of a fixed set of AOs (2.30). The calculation of the integrals in the MO basis can thus be viewed as a tensor contraction of AO integrals $I_{\mu\nu\lambda\sigma}$ with some coefficient tensor $C_{\mu\nu\lambda\sigma}^{pqrs}$

$$I_{pqrs} = \sum_{\mu\nu\lambda\sigma} C_{\mu\nu\lambda\sigma}^{pqrs} I_{\mu\nu\lambda\sigma} \tag{2.124}$$

If N_{AO} and N_{MO} denote the number of AOs and MOs, respectively, there are N_{MO}^4 values of I_{pqrs} and in order to calculate each of those a sum over N_{AO}^4 indices has to be carried out. Such an implementation would thus scale with $N_{AO}^4 N_{MO}^4$. Luckily, the tensor $C_{\mu\nu\lambda\sigma}^{pqrs}$ factorizes

into coefficient matrices (2-tensors)

$$I_{pqrs} = \sum_{\mu\nu\lambda\sigma} C_\mu^p C_\nu^q C_\lambda^r C_\sigma^s \, I_{\mu\nu\lambda\sigma} \tag{2.125}$$

and it is therefore possible to transform one AO index at a time. The first transformation would for example be

$$I_{\mu\nu\lambda s} = \sum_\sigma C_\sigma^s \, I_{\mu\nu\lambda\sigma} \tag{2.126}$$

That is, it has to be summed over N_{AO} indices for each of the $N_{AO}^3 N_{MO}$ values of $I_{\mu\nu\lambda s}$. This step therefore scales as $N_{MO} N_{AO}^4$. The scaling of the following steps can be obtained equivalently. As $N_{MO} \leq N_{AO}$ generally, the first transformation step is the most expensive one.

The scaling properties of other terms in the context of correlation methods can be investigated by analogous systematic analysis with respect to the indices and factorization of tensors.

2.6 Wave function based local correlation methods

Local correlation methods are based on the energy expression of the exact FCI wave function (2.107) which can be expressed in terms of contributions of occupied orbital pairs E_{ij} [1, 2]

$$E_{\text{corr}} = \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab} \left\langle \tilde{\psi}_0 \middle| \hat{H} \middle| \tilde{\psi}_{ij}^{ab} \right\rangle = \sum_{i<j} E_{ij} \tag{2.127}$$

where C_{ij}^{ab} are the exact doubles coefficients. From this starting point, Sinanoğlu and Nesbet separately developed the independent electron pair approximation (IEPA) in order to calculate these coefficients independently for each occupied orbital pair.[1, 2] Later, also couplings between pairs were treated in the coupled electron pair approximation (CEPA).[110] If both, the occupied and the virtual orbitals, are spatially localized, the needed occupied orbital pairs and the corresponding virtual orbitals could be restricted to reduce computational costs. This has been limited due to the issues regarding a proper localization of virtual orbitals (section 2.3.1). Alternative non-orthogonal virtual orbitals, projected atomic orbitals (PAOs) (section 2.6.1), have therefore been investigated to generate local virtual orbital spaces, called domains, which are spatially close and therefore can be assigned to specific localized occupied orbital pairs. This idea was first adopted by Pulay and Saebø to the MPPT, CISD and CEPA methods[111–115] and then extended to CCSD and CCSD(T) by Werner *et al.*[116, 117] Parallel to this field of research, Meyer *et al.* developed the CEPA method and combined this with the concept of PNOs (section 2.6.1).[110] The main difference is that PNOs are not selected due to their spatial properties. It has been shown that PNOs represent the orbital basis for a two electron system in which the FCI wave function converges most rapidly in

a specific order.[118] A truncation of the orbital basis therefore systematically reduces the domain size of a pair. The corresponding methods are denoted by a capital L in front of the non-local variant of the method, as for example LCCSD. The common ground of these methods (except IEPA) is that the whole system is treated in a single correlation calculation and an orbital invariant formulation of the underlying non-local methods is required. Section 2.6.2 contains a selection of methods and is not meant to be comprehensive. A different class of methods follows the strategy of fragmentation and some kind of many-body expansion (MBE) in this fragmentation, where the correlation energy is calculated piecewise. Methods as the CIM[18, 119–126] and DEC [16, 127–129] are shortly reviewed in section 2.6.3. The incremental scheme (IS) [13–15, 25, 130–159] is reviewed in more detail in the next chapter 3 as the work in this thesis has been performed in the incremental scheme framework.

2.6.1 Correlation domains

With projected atomic orbitals

The term domain is reserved in local correlation methods and refers to the virtual orbital space associated with a set of LMOs. There are different strategies to obtain domains. A very common approach involves the use of PAOs, which have also been used in the pioneering work by Pulay and Saebø on local correlation methods.[111] A specific PAO $|\varphi_\mu\rangle$ is constructed by projecting the closed shell occupied space with orbitals $|\varphi_i\rangle$ out of the atomic basis function $|\chi_\mu\rangle$.

$$|\varphi_\mu\rangle = \left(1 - \sum_i |\varphi_i\rangle\langle\varphi_i|\right) |\chi_\mu\rangle = \sum_\nu |\chi_\nu\rangle \, \mathbf{C}^{\mathrm{PAO}}_{\nu\mu} \tag{2.128}$$

The coefficients of the virtual orbitals are therefore given in the original AO basis as

$$\mathbf{C} = 1 - \frac{1}{2}\mathbf{D}\mathbf{S} \tag{2.129}$$

where $\mathbf{D}$ is the HF density matrix and $\mathbf{S}$ the overlap matrix in AO basis. According to (2.128), each AO corresponds directly to a virtual orbital. There are linear dependencies, since the number of virtual orbitals is exceeded by the number of occupied orbitals n_o. This is handled by diagonalization of the overlap matrix $\mathbf{S}^{\mathrm{PAO}} = \mathbf{C}^{\mathrm{PAO}\dagger}\mathbf{S}\mathbf{C}^{\mathrm{PAO}}$ and elimination of eigenfunctions with vanishing eigenvalues. The orthogonalization of the PAO basis is postponed and the PAOs which shall be treated together are selected first. For any localized occupied orbital φ_i, an orbital domain $[i]$ is defined by the set of PAOs which correspond to AOs with a large contribution to the LMO. The significance of AOs on an atom can for example be determined through the Mulliken orbital charges.[160] The orbital φ_i is represented in the truncated AO basis as

$$|\tilde\varphi_i\rangle = \sum_{\nu\in[i]} |\chi_\nu\rangle \, \tilde{\mathbf{C}}_{\nu i} \tag{2.130}$$

Whether the selected AOs and therefore $[i]$ is sufficient to describe φ_i is determined by a threshold on the residual norm.[161]

$$\min_{[i]} \langle \varphi_i - \tilde{\varphi}_i | \varphi_i - \tilde{\varphi}_i \rangle \leq t_{\text{res}} \tag{2.131}$$

The orbital domain $[i]$ is expanded until the threshold is hit. Orbital domains for LMO pairs or higher order tuples are formed by unification of the domains.

$$[i_1 i_2 \ldots i_p] = \bigcup_{q=1}^{p} [i_q] \tag{2.132}$$

With approximate natural orbitals

Natural orbitals are orbitals which diagonalize the one-particle density matrix (2.40) according to

$$\gamma_1(1';1) = \sum_p n_p \varphi_p^*(1') \varphi_p(1) \tag{2.133}$$

where n_p is the occupation number of natural orbital φ_p.[118] Equivalently, the diagonalization can be performed with the density matrix $\mathbf{D}$ in the MO basis

$$\mathbf{D}\mathbf{d}^{\text{NO}} = n\mathbf{d}^{\text{NO}} \tag{2.134}$$

and then transformed into the AO basis through $\mathbf{C}^{\text{NO}} = \mathbf{C}^{\text{MO}\dagger}\mathbf{d}^{\text{NO}}\mathbf{C}^{\text{MO}}$. It has been shown by Löwdin that these natural orbitals, sorted with respect to the occupation numbers, yield the most rapidly converging wave function expansion in terms of Slater determinants.[118] A truncation according to the occupation number would provide a solution to the slow convergence problem of CI. Note, that it is not the energy which converges most rapidly. However, the natural orbitals cannot be calculated without the exact wave function which corresponds to the density matrix. Different strategies evolved to approximate NOs. In a two electron system, natural orbitals are called natural orbitals of a pair. In the context of local correlation, it is desired to find approximate natural orbitals for the occupied orbital pairs to provide pair domains $[ij]$ also in more than two electron systems. These approximate natural orbitals are called PNOs, originally termed pseudonatural orbitals. They are determined through the density matrix of an approximate multideterminantal wave function which only contained substitutions from the i and j.[23] A self-consistent optimization of PNO and NOs of pairs has also been reported.[24]

Another option to produce approximate NOs is to use the MP2 method. It provides wave function corrections to HF in terms of many-body basis functions and is therefore suited for a calculation of an approximate density matrix. The virtual-virtual block of the density matrix in MO basis can either be diagonalized as a whole or the density matrix is splitted in contri-

butions of specific occupied orbitals and thereafter diagonalized in this subunit. MP2-NOs obtained in the former case are denoted as frozen natural orbitals (FNOs) [162] and related methods FNO methods. In the context of local correlation methods, as pointed out above, domains for occupied orbital pairs $[ij]$ are of special interest. Such domains can be obtained through a decomposition of the density matrix into pair density contributions $\mathbf{D}^{ij}$,

$$\mathbf{D}^{ij} = \frac{1}{1+\delta_{ij}} \left(\tilde{\mathbf{T}}^{ij\dagger}\mathbf{T}^{ij} + \tilde{\mathbf{T}}^{ij}\mathbf{T}^{ij\dagger} \right) \tag{2.135}$$

diagonalization and trucation of the orbital space with respect to the eigenvalues. It has been demonstrated, that this procedure can be combined with local correlation methods and yield accurate results.[3–12] The abbreviation $\tilde{\mathbf{T}}^{ij} = 2\mathbf{T}^{ij} - \mathbf{T}^{ij\dagger}$ in (2.135) has been used for the intermediates of the wave function coefficients $t_{ij}^{ab} = (\mathbf{T}^{ij})_{ab}$. If these coefficients are known, MP2-PNOs are obtained with the described procedure for $i \neq j$. This represents an orthonormal domain $[ij]$. MP2-PNOs to different occupied pairs are not orthogonal. This needs special attention in local PNO based methods. The density matrix, defined in (2.135), opens also a route for the calculation of OSVs when $i = j$ is investigated. This procedure yields orbital domains $[i]$.

For canonical occupied and virtual orbital spaces, the coefficients are obtained through (2.122) but locality can not be exploited. However, at least the occupied orbitals are localized and therefore not canonical in a local treatment. Pulay and Saebø reformulated the MP2 equations in an orbital invariant formulation for this purpose.[113] The wave function coefficients can then not be calculated directly, they have to be solved in an iterative procedure to yield vanishing residuals R_{ij}^{ab}. If the virtual orbitals diagonalize the Fock matrix, the equation reads

$$0 = R_{ij}^{ab} = \langle ab|ij \rangle + (\epsilon_a + \epsilon_b - f_{ii} - f_{jj})t_{ij}^{ab} - \sum_{k \neq i} f_{ik}t_{kj}^{ab} - \sum_{k \neq j} f_{jk}t_{kj}^{ba} \tag{2.136}$$

and (2.122) is obtained in a spacial case with canonical occupied orbitals. Solving (2.136), using the coefficients to diagonalize the full virtual-virtual block of the density matrix rather than a pair density, and using these MP2-NOs in a higher quality method is a common strategy and known as FNO method.[162] A non-iterative estimate

$$t_{ij}^{ab} = -\frac{\langle ab|ij \rangle}{\epsilon_a + \epsilon_b - f_{ii} - f_{jj}} \tag{2.137}$$

is obtained for the coefficients if the third and last term in (2.136) are neglected due do locality which is sometimes referred to as the semi-canonical approximation.[5]

The two-particle integral in (2.137) have to be calculated from the AO basis with m basis functions. For the canonical HF virtual orbitals, this transformation scales as $\mathcal{O}(m^5)$ and needs to be avoided, if possible. A workaround by Neese and Riplinger exploits the benefits of PAOs (previous section).[5] They used PAO pair domains $[ij]_{\mathrm{PAO}}$ to calculate the coefficients

in (2.137) and hence the density matrix in the domain basis after a recanonicalization, that is diagonalization of the Fock matrix. This yields so called domain based local pair natural orbitals (DLPNOs) and tremendous computational savings since both, AO basis and MO basis are truncated. Based on this strategy, Werner *et al.* designed a multi-step process for PNO generation. [10] First, PAO domains for each occupied orbital $[i]_{\mathrm{PAO}}$ are generated. They are then used to determine OSV domains $[i]_{\mathrm{OSV}}$ by diagonalization of $\mathbf{D}^{ii}$ in the PAO domain basis. Next, $[i]_{\mathrm{OSV}}$ and $[j]_{\mathrm{OSV}}$ are unified and canonicalized to form an OSV domain $[ij]_{\mathrm{OSV}}$ for the pair ij and finally the PNO domain is obtained by a diagonalization of the density matrix in domain $[ij]_{\mathrm{OSV}}$.

2.6.2 Direct local correlation methods

Direct local correlation methods have the common feature to describe the total system in a single correlation calculation while introducing locality related approximations. The common ground of all these methods is that the correlation energy of the system can be decomposed into and analysed with respect to occupied orbital pair contributions (2.127). For example, very distant occupied orbital pairs do not contribute to the correlation energy and can thus be discarded, which is known as the pair approximation. This can be rationalized from the energy expression (2.123) and the first order wave function coefficients (2.122) in MP2. Both involve integrals of the type $\langle ab|ij \rangle$, which is equivalently a Coulomb interaction of two charge distributions ρ_{ai} and ρ_{bj}. A multipole expansion reveals that the integral decays with R_{ij}^{-3} in magnitude and the energy thus with R_{ij}^{-6}, where R_{ij} is the distance of the centers of charges of the occupied orbitals i and j.[163] Therefore, a categorization of orbital pairs with respect to their expected contribution has been established. A common approach is to differentiate between strong pairs, weak pairs and distant pairs.[116] Distant pairs are neglected, weak pairs are treated approximately at a lower level of theory and strong pairs fully. In the PAO-LCCSD approach of Werner *at al.*, categorization has been done through a real space criterion for the distance of the orbitals in the two domains $[i]_{\mathrm{PAO}}$ and $[j]_{\mathrm{PAO}}$.[116] More recently, energy based selection of pairs is favored in different methods as in the DLPNO-CCSD(T) [6] of Neese *et al.* and in the new generation local methods PNO-LMP2 [10] and PNO-LCCSD(T) [12] of Werner *et al.* The energy screening is trimmed for computational efficiency. That is, OSV domains for each orbital $[i]_{\mathrm{OSV}}$ are formed in the PAO orbital domain basis $[i]_{\mathrm{PAO}}$. The pair domains $[ij]_{\mathrm{OSV}}$ are then constructed from the orbital domains and canonicalized in the virtual space. This allows for the semi-canonical approximation, explained above, and the MP2 pair energy reads as

$$E_{ij}^{\mathrm{SC\text{-}MP2}} = -\sum_{ab} \frac{2 \langle ab|ij \rangle \langle ab|ij \rangle}{\epsilon_a + \epsilon_b - f_{ii} - f_{jj}} + \sum_{ab} \frac{\langle ab|ij \rangle \langle ab|ji \rangle}{\epsilon_a + \epsilon_b - f_{ii} - f_{jj}} \qquad (2.138)$$

where $a, b \in [ij]_{\text{OSV}}$. Since the purpose of the pair screening is to identify distant pairs first, only this case is considered. Assuming that i and j are distant and a is close to i and b is close to j, the second term in (2.138) will fall off exponentially and is therefore neglected. An additional multipole expansion in the first term yields the dipole pair (DIP) energy

$$E_{ij}^{\text{SC-MP2-DIP}} = -\frac{8}{R_{ij}^6} \sum_{ab} \frac{(\langle i|\vec{r}|a\rangle \langle j|\vec{r}|b\rangle)^2}{\epsilon_a + \epsilon_b - f_{ii} - f_{jj}} \qquad (2.139)$$

which has been shown to yield accurate estimates for distant pairs.[5] Distant pairs are found with an energy threshold. The energies of non-distant pairs are recalculated with the more accurate energy estimate in (2.138) and another energy threshold identifies weak pairs. The obtained pair energies are kept for weak pairs. The remaining pairs are considered as strong. For these pairs, PNO domains are calculated and used in a higher level of theory.

A major difference between canonical orbital based correlation methods and local methods are the orthogonality relations of the orbitals. In canonical based formulations, all orbitals are orthogonal to each other. In direct local correlation methods, occupied orbitals are still orthogonal. However, orbital domains are orthogonal to their respective occupied orbitals but domains of different occupied orbital tuples are not. The non-vanishing overlap is denoted by $\langle a_{ij}|b_{kl}\rangle$, where the subscripts of the orbital indices a and b refer to the orbital domain they originate. That is, $a_{ij} \in [ij]$ and $b_{kl} \in [kl]$. These overlap contributions are introduced into the residual equations (2.136) by transforming the equations from the canonical basis into the used basis. The MP2 residual equations transform to the PNO-LMP2 [10] equations according to

$$R_{ij}^{a_{ij}b_{ij}} = \langle a_{ij}b_{ij}|ij\rangle + (\epsilon_{a_{ij}} + \epsilon_{b_{ij}} - f_{ii} - f_{jj})t_{ij}^{a_{ij}b_{ij}} - \sum_{k \neq i} f_{ik} \sum_{c_{kj}d_{kj} \in [kj]} \langle a_{ij}|c_{kj}\rangle t_{kj}^{c_{kj}d_{kj}} \langle d_{kj}|b_{ij}\rangle$$

$$- \sum_{k \neq j} f_{jk} \sum_{c_{ki}d_{ki} \in [ki]} \langle b_{ij}|c_{ki}\rangle t_{ki}^{c_{ki}d_{ki}} \langle d_{ki}|a_{ij}\rangle$$

$$(2.140)$$

In MP2, there are no singles coefficients. In CCSD however, there are also singles amplitudes. The choice of singles domains is not unambiguous but it has been common practice to use OSVs.[5, 8, 11] Local versions of MP2 and CCSD(T) have also been realized with the use of OSVs alone,[7–9] instead of the hybrid OSV-PNO or a PAO based methods mentioned above.

2.6.3 Fragmentation based local methods

The possibility to decompose the correlation energy of a system in terms of single occupied orbital or orbital pair contributions (2.127) suggests to treat subsystems separately to reproduce these contributions independently and then restore the correlation energy. The obvious benefits of such approaches are the embarrassingly parallel nature, the compatibility with existing

code of correlation methods and the reduced dimensionality of the subsystems. A variety
of methods have been reported in the literature which fall mainly into two different classes.
In one of those classes, the system is divided into fragments which are correlated within the
fragment space and corrections upon this single unit treatment are taken into account by si-
multaneous correlation of fragment tuples. This represents a many-body expansion. Related
methods are for example the IS [13–15, 25, 130–159], FMO-CC [164] and DEC [16, 127–129].
The other class of methods treats the individual fragments within an augmented description
and identifies the fragment contribution directly from a single calculation in the augmented
frame. Examples for methods falling into this class are the divide and conquer (DAC) [17,
165], CIM [18, 119–126] and local natural orbital (LNO) [19, 166] methods. Some of the
mentioned methods have only minor conceptional differences. The following sections cover a
small selection of these methods. The interested reader is referred to the original literature.

Cluster in molecule method

The correlation energy of the total system is expressed as a sum of the correlation energies
of local occupied MOs in CIM approaches [18, 119–126]

$$E_{\text{corr}} = \sum_i E_i \tag{2.141}$$

The contributions E_i are calculated within a so called cluster. A cluster includes an occupied
space which is close to the central orbital i by means of a real space distance threshold and
a virtual PAO domain for this central LMO according to Pulay and Boughtons prescription
[161] presented in section 2.6.1. For the notation of a cluster P, the previously used domain
descriptor $[i]$ is extended to differentiate between occupied spaces $[\underline{P}]$ and virtual spaces $[\overline{P}]$.
In this notation, energy contributions for single LMOs E_i are obtained from LMO pair energies
E_{ij} according to

$$E_i = \sum_{j \in [\underline{P}]} E_{ij} \qquad \text{with} \qquad E_{ij} = \sum_{ab \in [\overline{P}]} E_{ij}^{ab} \tag{2.142}$$

The original CIM method is limited in applicability as the virtual spaces can be very large,[125]
since the occupied space is not restricted to pairs. This issue has been addressed by either
reduction of occupied and virtual spaces of the cluster in the context of LNO-CC [19, 166] or
the combination of the CIM method with direct local correlation methods as CIM-DLPNO-
CCSD [126], which subsequently adjusts correlation domains to single LMOs or pairs of LMOs
within the cluster.
Another energy related improvement has been gained by consideration of distant pairs. [125]
The occupied space is selected through a real space threshold d in a cluster. The central
LMO is correlated together with each of those. That means only energy contributions of the
central LMO and the LMOs within a sphere of radius d are taken into account. There are

however LMOs inside the sphere with distances d_{jk} between $d < d_{jk} \leq 2d$. Their energies are not included from any cluster. An inclusion of the corresponding energies is called distant pair correction and given by

$$E_{\text{corr}} = \sum_i E_i + \sum_{\substack{jk \\ d < d_{jk} \leq 2d}} \tilde{E}_{jk} \qquad \text{with} \qquad \tilde{E}_{jk} = \frac{1}{M} \sum_P E_{jk}^{[P]} \qquad (2.143)$$

where $\tilde{E}_{jk}$ represents the averaged contribution of the pair jk, this pair can contribute through M different clusters $[P]$. The individual contributions for a specific cluster are denoted as $E_{jk}^{[P]}$.

Divide expand consolidate method

The divide expand and consolidate approach [16, 127–129] is formulated in the context of MP2 and CC theory, where the correlation energy is given according to (2.118). The DEC methods exploit the progress in localization of virtual orbitals [60–64] (section 2.3.1) by using these orthonormal MOs instead of alternative MOs as PAOs or NOs. With this choice of virtual MOs, the occupied-virtual blocks of the Fock matrix vanish. With a closed shell HF reference, the energy expression is then given as

$$E_{\text{corr}}^{\text{CC}} = \sum_{ijab} E_{ij}^{ab} = \sum_{ijab} \left(t_{ij}^{ab} + t_i^a t_j^b \right) \left(2 \langle ij|ab \rangle - \langle ij|ba \rangle \right) \qquad (2.144)$$

As in the CIM approach, described above, a partitioning of the energy is applied. In the CIM methods, individual LMO contributions are used. In contrast, the partitioning in DEC methods is with respect to atomic fragment E_P and fragment pair ΔE_{PQ} contributions

$$E_{\text{corr}}^{\text{CC}} = \sum_P \left(E_P + \sum_{Q < P} \Delta E_{PQ} \right) \qquad (2.145)$$

which are calculated separately. Before proceeding to the explicit expressions of the contributions, the notation shall be clarified. The occupied space of a fragment is denoted by $\underline{P}$. If this space is augmented with further LMOs close to $\underline{P}$, the notation $[\underline{P}]$ is used as shown for the CIM method above. As the DEC methods are local methods, there is also a domain approximation for the virtual space. An augmented virtual space for a fragment P shall be denoted as $[\overline{P}]$. In this notation, the energy contributions read

$$E_P = \sum_{\substack{ij \in \underline{P} \\ ab \in [\overline{P}]}} E_{ij}^{ab} \qquad (2.146)$$

$$\Delta E_{PQ} = \sum_{\substack{i \in \underline{P} \\ j \in \underline{Q}}} \sum_{ab \in [\overline{P}] \cup [\overline{Q}]} E_{ij}^{ab} \quad + \sum_{\substack{i \in \underline{Q} \\ j \in \underline{P}}} \sum_{ab \in [\overline{P}] \cup [\overline{Q}]} E_{ij}^{ab} \qquad (2.147)$$

Note, that E_{ij}^{ab} is only equal in all terms if the full system is treated simultaneously and the energies are calculated with the corresponding amplitudes. Instead of a full treatment, the occupied and virtual spaces are truncated for any fragment and fragment pair. A locality analysis of MP2 and CCSD amplitude equations [128] showed that the amplitudes, needed for the evaluation of E_P and ΔE_{PQ}, can be generated by augmenting the occupied and virtual spaces of the fragments with neighboring fragments and performing the correlation calculations in this extended orbital space, called amplitude orbital space (AOS). The energies are calculated in the subspace of the AOS, the energy orbital space (EOS), according to (2.146)-(2.147).

A crucial step in the DEC method is the identification of a sufficiently, but not unnecessary, large AOS since the correlation space size directly relates to the computational demands. The AOS of single fragments is selected in an iterative procedure using an energy threshold which is called fragment optimization threshold (FOT).[129] The basis of the iterative scheme is an ordered list of orbitals with decreasing importance measured with some metric such as the distance to the central fragment. The iterative procedure consists of two phases. In the first phase, the orbital space $[P]$ is expanded until the calculated energy with the actual and previous orbital spaces, $[P]_k$ and $[P]_{k-1}$ respectively, differ by less than the FOT. The orbital space which meets this condition first is chosen as a reference $[P]_{\text{ref}}$. In the second phase, the orbital spaces are reduced and a binary search is performed to find the smallest possible orbital space which yields an energy difference smaller than the FOT to the reference energy.

3 Incremental Scheme

The incremental scheme is an ansatz in which the energy of the system is reformulated in a many-body expansion in terms of domains consisting of occupied LMOs and a corresponding virtual space. In the original formulation of Stoll for non-atomic calculations, LMOs and canonical virtual orbitals of the total system have been used in all domains and the LMO partitioning of the system into domains has been provided by hand.[13] More recent implementations in the context of molecular calculations do also perform SCF calculations on the whole system. The occupied space undergoes a Boys localization [22] (section 2.3.1) and the LMOs are partitioned to yield disjoint occupied spaces.[14, 15, 147] Two alternative occupied LMO partitioning schemes have been implemented based on graph partitioning [14, 147] or K-means partitioning [15] in the incremental scheme framework. These schemes are presented in section 3.1.

As a result, N_{D_o} disjoint occupied spaces are obtained. The correlation energy is expanded in terms of the corresponding domains

$$E_{\text{corr}} = \sum_{\substack{\mathbb{I} \in \mathcal{P}(\{1,\ldots,N_{D_o}\}) \\ |\mathbb{I}| < I_o}} \Delta_o E_{\mathbb{I}} \qquad \text{with} \qquad \Delta_o E_{\mathbb{I}} = E_{\mathbb{I}} - \sum_{\substack{\mathbb{J} \in \mathcal{P}(\mathbb{I}) \\ |\mathbb{J}| < |\mathbb{I}|}} \Delta_o E_{\mathbb{J}} \qquad (3.1)$$

where the powerset $\mathcal{P}$ notation [25] is used and I_o denotes the order of the expansion. Each correlation energy $E_{\mathbb{I}}$ for an occupied domain tuple $\mathbb{I}$ in (3.1) is calculated independently in an embarrassingly parallel fashion.

For large molecules, a canonical virtual space is not feasible and does not fully exploit the locality of correlation effects. Therefore, a compressed local virtual space is necessary. This has been realized through PAOs [14, 147] or domain specific basis sets [15, 25] as presented in section 3.2.

Equipped with N_{D_o} domains, a correlation calculation on domain $[i]$ only yields an energy E_i associated with the respective domain. According to (3.1), these energies are corrected in the second order by two-site domain energies $\Delta E_{ij} = E_{ij} - E_i - E_j$. From the standard many-particle basis point of view with a HF reference, the energy calculated with the two-site domain $[ij]$ rises from individual substitutions within the domains $[i]$ and $[j]$ as well as simultaneous substitutions from the individual LMO spaces into the pair virtual space. Since the major part of the former contributions are included in E_i and E_j, the incremental contribution ΔE_{ij} consists mainly of the latter contributions. Higher order incremental correlation energy

corrections are defined recursively in (3.1) on the right hand side. That is, corrections with higher order domain tuples are calculated from corrections in lower orders. A more elaborate analysis of the incremental energy contributions is provided in section 3.3 within the context of the CCSD method.

An incremental energy calculation, as described above, requires independent calculations to obtain each $E_{\mathbb{I}}$. There are $n_k = \binom{N_{D_o}}{k}$ tuples and thus calculations at incremental order k. The number of domains, provided that their size is kept consistently, should scale linearly with system size in a method for large molecular systems. However, even the number of pairs $n_2 = N_{D_o}(N_{D_o} - 1)/2$ does not scale linearly with system size. This gets worse in higher incremental orders since the number of tuples scales polynomially with the number of domains $n_k \sim N_{D_o}^k$. A truncation on the tuple level to reduce the number would be beneficial with respect to computational resource requirements, even in the embarrassingly parallel implementations. Such a truncation is fortunately consistent with the ideas of local electron correlation treatment. As pointed out in section 2.6.2, dynamic electron correlation contributions of distant occupied LMOs fall asymptotically with R^{-6}, where R is the distance of the centres of charges of two LMOs. Therefore, distant pairs could be neglected which holds also true if two LMO spaces instead of individual orbitals are spatially separated. It should be noted that this statement is undermined by the results of a locality analysis of the MP2 and CCSD amplitude equations. The analysis showed that the amplitudes with occupied indices of spatially close LMOs do only couple strongly to amplitudes with occupied indices which are close to the respective LMOs.[128] The amplitudes of a two-site domain with sufficiently distant occupied LMOs will therefore reproduce mainly the amplitudes of the individual domains and the amplitudes with shared occupied indices will vanish. The strategy of neglecting domain tuples which contain distant domains from the energy expansion is therefore legit and has been extensively used in the literature.[14, 15, 25, 147, 155] A simple real space threshold d has been used for the distance of domains and accurate results were obtained at incremental order three.[14] As the incremental corrections decrease in general with increasing incremental order, an order dependent, and with increasing order sharpening, threshold

$$d_k = \frac{d}{(k-1)^2} \tag{3.2}$$

has been suggested and used successfully.[25, 147] An alternative tuple screening is presented in section 5.1 which is based on off-diagonal Fock matrix elements in the LMO basis.

In molecular calculations on closed-shell [14, 15, 141–147, 156] and open-shell [157–159] systems, accuracy and applicability of the incremental scheme have been demonstrated repeatedly. A major benefit of the incremental scheme is that it is independent of the electron correlation method, although at least size-consistency should be guaranteed. Existing quantum chemistry code can thus be reused. The scope of the incremental scheme are correlation energies with CCSD [14, 147], MP2, CCSD(T) [148], F12 methods [149] and DLPNO-CCSD(T)

[150]. Furthermore, incremental expansions have also been used for molecular properties.[151–154] Earlier work on periodic systems [130–140] shall be mentioned for the sake of completeness. Since this thesis is exclusively concerned with molecular systems, the interested reader is referred to the original literature for periodic systems.

3.1 Occupied space partitioning

The partitioning of the occupied LMO space is a crucial step. If strongly correlated orbital spaces are split, higher incremental orders are needed to correlate those orbital spaces simultaneously. A proper partitioning therefore enhances the convergence of the incremental series.

Two different occupied space partitioning schemes have been reported in the literature, a graph partitioning scheme [14] and the K-means partitioning [15]. These schemes are reviewed below. Furthermore, a new partitioning scheme is presented, which uses force fields for the generation of atom wise occupied domains.

Each of these partitioning methods starts from the set of LMOs of the system, which have to be grouped according to their spatial distribution. Instead of dealing with the whole information of the spatial distribution, a single three dimensional point is used as an identifier for each LMO. This point is the centre of charge of the orbital. For any LMO i, it is defined as $R_i(\langle i|x|i\rangle \,|\, \langle i|z|i\rangle \,|\, \langle i|z|i\rangle)$. Obviously, other information as the maximum spatial extent are suppressed in this simplified picture. However, the overall locality of the LMOs is handled by the localization method. Foster-Boys localization is preferred in this context, as it yields the most local MOs in average.

3.1.1 Graph partitioning

Distances between LMOs can be defined, if the set of LMOs is represented by a set of centres of charge. With a distance threshold, the set of LMOs can in turn be divided into connected and disconnected pairs. This situation is represented in graph theory by a graph $G(V, E)$ which is defined by the vertices V (centres of charge of the LMOs) and edges E (connections between LMOs). With this representation, the problem of LMO partitioning is equivalent to a graph partitioning. The METIS graph partitioning [167] has been applied in the context of the incremental scheme [14, 147] It tries to cut the graph into a predefined number of disjoint vertex subsets. This is done by recursive bisections of the graph. The algorithm of a single multilevel graph bisection shall only be briefly reviewed. The full graph $G = G_0$ is transformed iteratively into smaller graphs by merging adjacent vertices until a sufficiently small graph G_m is obtained (coarsening phase). G_m is partitioned into two equally sized sets of vertices with a minimum number of edges connecting the two sets (partitioning phase). The graph is successively uncoarsened and the partitions are updated intermediately (uncoarsening phase).

This algorithm requires a target number of partitions, which represents the number of domains N_{D_o}, and tries to distribute the vertices into N_{D_o} equally sized portions. A domain size of four has been used to estimate the number of domains for the graph partitioning.[14] However, it has been realized that in some cases close lying orbitals are separated into different domains. This problem could be eliminated by merging close lying LMOs into single vertices with higher weights in order to prevent cutting edges between merged vertices.[149]

3.1.2 K-means partitioning

The K-means clustering method [168, 169] aims to divide a set of vectors, into K subsets of vectors $D_1, \ldots, D_K$ which minimize the summed deviations from the centroids of the subsets

$$\underset{\{D_p\}}{\arg\min} \sum_{p=1}^{K} \sum_{\vec{R} \in D_p} |\vec{R} - \vec{C}_p|^2 \tag{3.3}$$

where the centroid $\vec{C}_p$ of a subset D_p is defined as

$$\vec{C}_p = \frac{1}{|D_p|} \sum_{\vec{R} \in D_p} \vec{R} \tag{3.4}$$

This method has been utilized for occupied space partitioning with $K = N_{D_o}$ and the set of centre of charge vectors of the LMOs as the set to be partitioned.[15] The algorithm [170] starts with an initial guess for the centroids (means). Then, the following two steps are executed repeatedly: (1) each center of charge is inserted into the set whose mean has the shortest distance to the centre of charge. (2) means are updated with the assigned centres of charge. The partitioning is converged, if the sets $D_1, \ldots, D_K$ do not change in the next iteration. This result represents a local minimum reached by the choice of the initial means. In order to find global minima, many random initial means have to be considered and the obtained local minima compared.

As in the graph partitioning scheme, the target number of domains has to be predefined. The choice is ambiguous. However, the number of domains can be estimated by chemical intuition and could be chosen identical with the number of functional groups.[15] Finding the number of domains is itself an optimization problem. The choice has to balance between the computational costs, which increases steeply with the domain size, and the accuracy at a given incremental order, which decreases for too small domains.

3.1.3 Atom wise partitioning with force fields

The partitioning methods presented in the last two sections require both a predefined number of domains. These methods either use a domain size parameter to estimate the number of domains or count the number of functional groups for this purpose. A new method is presented

in the following which does not force an arbitrary fixed domains size and does not require much chemical intuition. The method is motivated by the atoms in the molecule picture. An occupied space is associated with each non-hydrogen atom. A simple way to realize this would be to evaluate the distances of the LMO centres of charge and than assign each LMO to the atom with the shortest distance. However, this procedure does not try to find an efficient grouping with respect to their spatial distribution as the K-means method for instance. The incremental expansion would not be expected to converge fastest with this choice. Furthermore, there can be LMOs which have nearly the same distance to different atoms. A hard distance threshold seems in this context somehow inappropriate as it lead to the case that an atomic domain has no valence orbitals.

For the mentioned reasons, this simple strategy has been abandoned and an alternative atomic domain generation is proposed which is based on interaction potentials and force fields. A set of domain vectors $\{\vec{C}_p\}$ is used to describe the location of domains D_p with $|D_p|$ associated centres of charge. This quantity is related to the centroids in the K-means method. The

---

Algorithm 1 Optimization of domain centers.

> set threshold t_{CC}
> set threshold t_{DC}
> set convergence threshold t
> set $\vec{C}^{(0)} = (\vec{C}_1^{(0)}, \ldots, \vec{C}_{N_{D_o}}^{(0)})$ as atomic positions
> set iterations I
> $i = 0$
> **while** $i \leq I$ **do**
> > calculate $q^{(0)} = (q_1^{(0)}, \ldots, q_{N_{D_o}}^{(0)})$
> > calculate $\nabla V(\vec{C}^{(i)})$
> > **for** $p \in \{1, \ldots, N_{D_o}\}$ **do**
> > > $\lambda_p^{(i)} = \arg\min_{\lambda_p} V(\vec{C}^{(i)} - \lambda_p[\nabla V(\vec{C}^{(i)})]_p \, \vec{e}_p)$
> > > **if** $|\vec{C}_p^{(i)} - \lambda_p^{(i)}[\nabla V(\vec{C}^{(i)})]_p - \vec{C}_p^{(0)}| < t_{CC}$ **then**
> > > > $\vec{C}_p^{(i+1)} = \vec{C}_p^{(i)} - \lambda_p^{(i)}[\nabla V(\vec{C}^{(i)})]_p$
> > > **else**
> > > > $\vec{C}_p^{(i+1)} = \vec{C}_p^{(i)}$
> > > **end if**
> > > **if** $|V(\vec{C}^{(i+1)}) - V(\vec{C}^{(i)})| < t$ **then**
> > > > **break**
> > > **end if**
> > > $i = i + 1$
> > **end for**
> **end while**

---

basic idea behind the new method is to start initially with domains located at non-hydrogen atomic positions and let these rearrange in the vicinity of the atoms, such that the domains are still associated with the initial atoms at any stage of the algorithm. The centres of charge which are closest to the final domain position are then included into the final atomic domain. The details of the procedure are described below. The initial domain positions are chosen to be non-hydrogen atomic positions. The domain positions $\vec{C}_p$ are allowed to change but not more than a predefined threshold t_{CC}, that is $|\vec{C}_p - \vec{C}_p^{(0)}| < t_{CC}$. The domain positions are embedded in a fixed set of LMO centres of charge R with vectors $\vec{R}_q \in R$. For any domain p, the associated centres of charge Q_p are defined as the centres of charge in a sphere with a predefined radius t_{DC}.

$$Q_p = \{\vec{R}_q \in R \ : \ |\vec{R}_q - \vec{C}_p| < t_{DC}\} \tag{3.5}$$

The domain positions are moved due to pseudo-interactions with centres of charge through a Mie type potential [171] V_{DC} which is attractive for large distances and repulsive for short distances. Furthermore, different domains are assumed to have repulsive pseudo-interactions V_{DD} proportional to the number of surrounding centres of charge, which is considered through pseudo-charges $q_p = |Q_p|$. Each centre of charge is associated with a negative unit pseudo-charge $q_C = -1$. With this, the pseudo-interaction potential is defined as

$$V = \ V_{DC} + V_{DD} \tag{3.6}$$

$$V_{DC} = \ \sum_{pq} q_C q_q \left[\left(\frac{\sigma_1^{DC}}{|\vec{R}_p - \vec{C}_q|} \right)^{m_1^{DC}} - \left(\frac{\sigma_2^{DC}}{|\vec{R}_p - \vec{C}_q|} \right)^{m_2^{DC}} \right] \tag{3.7}$$

$$V_{DD} = \ \frac{1}{2} \sum_{p \neq q} q_p q_q \left(\frac{\sigma_1^{DD}}{|\vec{C}_p - \vec{C}_q|} \right)^{m^{DD}} \tag{3.8}$$

where in the summation over p in V_{DC} centres of charge of core orbitals are not included due to the singularity when domain centres and core orbital centres coincide. With these potentials, the initial domain positions will transform according to the associated force. The slightly modified steepest descent method [172] is utilized to find repeated updates of $\vec{C}_p$. The outline of the algorithm is given in algorithm 1. It contains a repeated update of pseudo-charges and updates of $\vec{C}_p$ along the gradient $\nabla V = \left(\frac{\partial V}{\partial \vec{C}_1}, \ldots, \frac{\partial V}{\partial \vec{C}_{N_{D_o}}} \right)$. The components of the gradient are given as

$$\frac{\partial V}{\partial \vec{C}_q} = \ \frac{\partial V_{DC}}{\partial \vec{C}_q} + \frac{\partial V_{DD}}{\partial \vec{C}_q} \tag{3.9}$$

$$\frac{\partial V_{DC}}{\partial \vec{C}_q} = \ -\sum_{pq} q_C q_q \left(\vec{R}_p - \vec{C}_q \right) \left[m_1^{DC} \left(\frac{\sigma_1^{DC}}{|\vec{R}_p - \vec{C}_q|} \right)^{2+m_1^{DC}} - m_2^{DC} \left(\frac{\sigma_2^{DC}}{|\vec{R}_p - \vec{C}_q|} \right)^{2+m_2^{DC}} \right] \tag{3.10}$$

$$\frac{\partial V_{DD}}{\partial \vec{C}_q} = \frac{m^{DD}}{2} \sum_{p \neq q} q_p q_q \left(\vec{C}_p - \vec{C}_q \right) \left(\frac{\sigma_1^{DD}}{|\vec{C}_p - \vec{C}_q|} \right)^{2+m^{DD}} \tag{3.11}$$

A global minimization would suggest to find a single prefactor λ in each iteration, that minimizes V. Here, a local instead of a global minimization is performed. Therefore, individual constrained prefactors λ_p have been used for each component of the gradient. The process is converged, if the difference in V is less than a threshold t.

After convergence, the set of centres of charge is divided into disjoint subsets $D_1, \ldots, D_{N_{D_o}}$, where an individual centre of charge is added to the set D_p to which it has the smallest distance by means of the converged domain centers.

$$D_p = \{ \vec{R}_q \in R \; : \; p = \arg\min_r |\vec{R}_q - \vec{C}_r| \} \tag{3.12}$$

The presented algorithm reliably generates atomic domains over a variety of tested systems and does not compete with the computation time scales of the SCF or correlation calculations. In order to obtain meaningful partitions with the K-means method, many initial configurations have to be taken into account. If N_{D_o} is chosen to coincide with the number of non-hydrogen atoms, the configuration space of the domain centers increases steeply and the number of cycles has to be adjusted, such that it becomes in issue for very large systems. One possibility to overcome this problem could be to restrict the initial guesses to small spheres around the atomic positions.

In figure 3.1, two examples are presented. It is observed that the domain centers for atoms with surrounding LMOs do not change significantly and the domain centre stays in the vicinity of the closer LMOs. Other domains centres for atoms which are not coordinated by LMOs
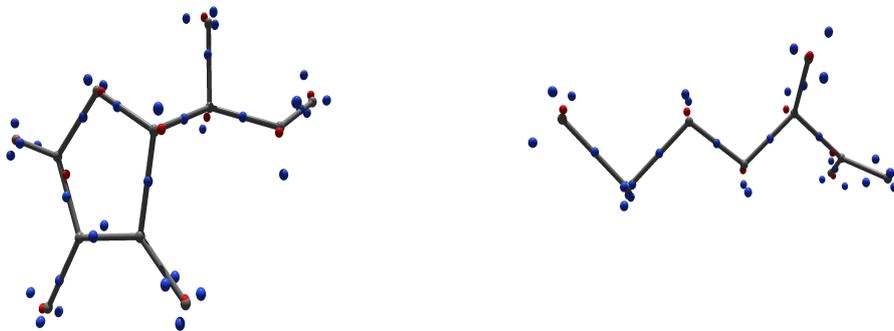
Figure 3.1: Optimized domain centers (red) starting from non-hydrogen atomic positions (gray). LMO centres of charge (blue) presented as well.

reach out to the closest LMOs and avoid LMOs wich are already close to other domain centres.

The presented force field based method for generation of atomic LMO domains is combined with the ordinary incremental scheme and also with a new incremental scheme (section 5.1), which includes an additional expansion of the virtual space. The atomic domains, obtained from the LMOs of the whole system, are used in this context for the generation of occupied and virtual spaces local to specific atomic domain tuples. In an SCF embedding calculation, the environment of an atomic domain tuple is frozen and the occupied and virtual spaces are recalculated for the active domain tuple. The special choice of atomic domains is beneficial with this respect, since the partitioning is performed by means of atomic sites. The treatment of the active part with a basis set of the respective atoms seems to be a reasonable choice.

3.2 Virtual space truncation

A general strategy of local correlation methods is to compress the virtual space for localized occupied orbitals. This is often referred to as the domain approximation. There are several possibilities to generate local virtual spaces. Some of them are summarized in section 2.6.1. This section is exclusively concerned with the utilized ideas within the incremental scheme. The traditional PAO virtual spaces (section 2.6.1), which have been used in many local correlation methods either directly or as intermediate virtual spaces, have also been used within the incremental scheme.[14] An alternative route for virtual space compression employs domain specific basis sets.[15, 147] The involved ideas are presented in the following section.

3.2.1 Domain specific basis sets

Without any restrictions, the selected basis set would be applied to the whole system and the system would hence be treated at the same level of theory. However, in some applications a more refined treatment of a specific part is desired. The dual basis set approach [173] addresses this by using different basis sets on different atoms, where the part of interest is described with a larger basis set. Since the number of virtuals increases linearly with the number of basis functions, such a strategy reduces the number of virtual orbitals significantly and focuses on a proper description of the virtual space on the part of interest.

The concept of domain specific basis sets follows these ideas in the environment of the incremental scheme.[15, 147] For each domain tuple, a larger basis set is used in the vicinity of the respective occupied domain tuple. Two different versions have been reported in this context. In the first approach [147, 174, 175], the whole system is treated with a universal basis set $\mathcal{B}_U$ at SCF level, the canonical MOs are localized and partitioned. For each domain tuple, individual SCF calculations are performed in a domain specific basis set $\mathcal{B}_D$, which is formed as follows: a larger basis set is used for all atoms which lie in a sphere with a predefined

radius from any of the domains in the tuple. The basis set of the remaining part is chosen to be very small. In order to provide consistency among occupied domains, a mapping between the LMOs from $\mathcal{B}_U$ and from $\mathcal{B}_D$ has to be done. Initially, the centres of charge were mapped [14] but this was ambiguous for some cases.[15] This mapping has later been replaced by a localization based mapping. [174, 175] The latest version maximizes the sum of squared overlaps between a set of fixed LMOs calculated with $\mathcal{B}_U$ and MOs calculated with $\mathcal{B}_D$. The localization function

$$C^{\mathrm{TL}}\left[\left|\varphi^{\mathcal{B}_D}\right\rangle\right] = \sum_i \left\langle i^{\mathcal{B}_U}|i^{\mathcal{B}_D}\right\rangle^2 \tag{3.13}$$

is maximized with the Jacobi sweep algorithm used by Edmiston and Ruedenberg (section 2.3.1).[175] This procedure is called template localization.

A second and alternative approach is termed zero-buffer (B0) approximation and differs in the selection of the atoms for which the larger basis set is used.[15] In the previously described method, the same atom can be considered to have a larger basis for different domains, if it is close to those domains. This is excluded within the B0 approximation by distributing the atoms of the molecule uniquely among the domains. There is therefore no buffer region with additional atoms for a domain. The allocation of the atoms is accomplished with the K-means clustering [168, 169], presented in section 3.1. However, the situation here is much simpler. Each atom is just associated with the closest domain centre, which is considered as the fixed centroid in the K-means method. This automatically satisfies the condition (3.3). Within the B0 approximation, also a different strategy has been proposed for the generation of the orbital spaces of the domain tuples. The HF calculation of the system is treated in a basis $\mathcal{B}_U$ with a maximum angular momentum quantum number and the basis of the domain tuples $\mathcal{B}_D$ extends this basis set with higher angular momentum quantum numbers on the selected atoms.[15] By construction, $\mathcal{B}_U \subset \mathcal{B}_D$ which preserves the occupied space of the tuple and enables the calculation of the associated virtual spaces without an additional HF calculation in the basis $\mathcal{B}_D$.[15]

3.2.2 Scaling properties

The domain approximation leads to computational savings in general and this holds also within the incremental scheme. In order to provide a systematic description, first the sources of the computational demands have to be clarified. The incremental scheme requires an HF calculation of the total system in an AO basis $\mathcal{B}^{\mathrm{AO}}$. As discussed in section 2.1, the four index integral $\langle \mu\nu|\lambda\sigma\rangle$ evaluation scales formally with $\mathcal{O}(|\mathcal{B}^{\mathrm{AO}}|^4)$, but reduces asymptotically to $\mathcal{O}(|\mathcal{B}^{\mathrm{AO}}|^2)$ when a prescreening is applied. The solution of the Roothaan-Hall eigensystem scales with $\mathcal{O}(|\mathcal{B}^{\mathrm{AO}}|^3)$ and a smaller prefactor. As long as the HF calculation of the total system can not be abandoned, the calculation of the total energy will asymptotically not reduce in scaling lower than $\mathcal{O}(|\mathcal{B}^{\mathrm{AO}}|^3)$.

Another source for computational demands is the calculation of the correlation energy. Many correlation methods are formulated in the MO basis $\mathcal{B}^{\mathrm{MO}}$, which consists of n_o occupied orbitals and $n_v = |\mathcal{B}^{\mathrm{AO}}| - n_o$ virtual orbitals. Without any restrictions, the number of MO four index integrals $\langle pq|rs \rangle$ is equal to the number of AO integrals. These integrals have to be calculated and then used in the correlation methods. The evaluation requires a transformation from the AO basis integrals to the MO integrals (section 2.5.4), which scales formally with $\mathcal{O}(|\mathcal{B}^{\mathrm{MO}}||\mathcal{B}^{\mathrm{AO}}|^4)$. The transformation process can be performed with integrals on storage, which further requires equivalent disk space. However, it is also possible to recalculate needed AO integrals during the correlation calculation and use them directly without storing, which eliminates tremendous disk requirements and the input-output bottleneck at the cost of repeatedly calculating the same integrals.[176, 177]

The calculation of the MP2 energy in a canonical basis does only require the four index MO integrals and therefore scales as the transformation does. There are no higher order scaling terms, even for the non-canonical case. The projective CCSD method has higher order scaling terms in the residual equations, which are due to tensor contractions of amplitudes and one- and two-particle integrals. Although n_o and n_v increase both linearly with the system size for a fixed basis set, n_v is usually significantly larger and depends on the basis set. It is therefore appropriate to distinguish between these quantities. The highest scaling terms in CCSD scales with $\mathcal{O}(n_o^2 n_v^4)$. The scaling is even higher in CCSD(T) with $\mathcal{O}(n_o^3 n_v^4)$ and in CCSDT with $\mathcal{O}(n_o^3 n_v^5)$.

Up to this point, no approximations have been considered. Suppose, that an incremental scheme is used. The scaling properties for an incremental calculation can be estimated with the most expensive incremental calculation and the number of incremental calculations. The latter is determined by the number of domains, the incremental order and the tuple screening. The former depends on the choice of the AO and MO bases.

The incremental scheme in combination with a domain specific basis set $\mathcal{B}^{\mathrm{AO}}_{\mathrm{DS}}$ shows slow basis set increase with the system size for the most expensive calculation, since a large basis is only used on selected atoms. However, the AO basis increases with molecular size and hence the computational costs for AO-MO integral transformations and the memory requirements for MO integrals as well. The scaling of the coupled cluster calculation itself is reduced. For the most expensive calculation, the number of virtual orbitals n_v^{inc} still increases with system size in the incremental scheme but the maximum number of occupied orbitals n_o^{inc} is limited due to the truncated expansion order. Therefore, n_o^{inc} does not count as an overall scaling factor and a single CCSD calculation scales as $\mathcal{O}((n_v^{\mathrm{inc}})^4)$ with respect to molecular size. Analogous arguments hold for the other cases.

The scaling properties of the most expensive calculation change fundamentally when PAO domains are used. In this case, the AO basis as well as the MO basis are truncated. For small and medium sized molecules this might cause that the full AO basis has to be used in the

most expensive incremental calculation. However, for large systems there is an asymptotic limit for the basis set and the number of virtual orbitals. Beyond this limit, there is no additional computational cost for an incremental calculation with respect to system size and the computational cost is constant. The overall scaling is then determined by the number of incremental calculations. Any asymptotically linear scaling incremental method should provide constant scaling in the most expensive calculation and a linear scaling in tuples.

3.3 CCSD energy and amplitudes in the incremental scheme framework

In this section, the relationship between the incremental scheme and the CCSD energy is investigated. The starting point is the closed-shell CCSD energy (2.144) with a slightly more compact notation

$$E_{\text{corr}}^{\text{CC}} = \sum_{ijab} \tau_{ij}^{ab} \, G_{iajb} \tag{3.14}$$

where $\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b$ and $G_{iajb} = 2 \langle ij|ab \rangle - \langle ij|ba \rangle$ are used. Further, the virtual spaces in all tuple domains are assumed to be the canonical virtual orbitals of the total system and shall be denoted by $[\overline{S}]$. The energy of an single domain $[I]$ is then expressed as

$$E_I = \sum_{\substack{i \in \underline{I} \\ j \in \underline{I}}} \sum_{ab \in [\overline{S}]} \left(\tau_{\underline{I}}^{[\overline{S}]} \right)_{ij}^{ab} G_{iajb} \tag{3.15}$$

where the subscripts and superscripts on τ denote the occupied $\underline{I}$ and virtual $[\overline{S}]$ spaces used in the correlation calculation. For any pair domain, the energy correction is given as $\Delta E_{IJ} = E_{IJ} - E_I - E_J$, where E_{IJ} is calculated in the unified occupied space $\underline{IJ}$. Therefore, the summation over the occupied indices in the CCSD energy will decompose according to

$$\sum_{\substack{i \in \underline{IJ} \\ j \in \underline{IJ}}} = \sum_{\substack{i \in \underline{I} \\ j \in \underline{I}}} + \sum_{\substack{i \in \underline{J} \\ j \in \underline{J}}} + \sum_{\substack{i \in \underline{I} \\ j \in \underline{J}}} + \sum_{\substack{i \in \underline{J} \\ j \in \underline{I}}} \tag{3.16}$$

into all possible pairs of domain indices. With this decomposition, ΔE_{IJ} is given by

$$
\begin{aligned}
\Delta E_{IJ} = &\sum_{\substack{i \in \underline{I} \\ j \in \underline{I}}} \sum_{ab \in [\overline{S}]} \left[\left(\tau_{\underline{IJ}}^{[\overline{S}]} \right)_{ij}^{ab} - \left(\tau_{\underline{I}}^{[\overline{S}]} \right)_{ij}^{ab} \right] G_{iajb} \quad + \quad \sum_{\substack{i \in \underline{J} \\ j \in \underline{J}}} \sum_{ab \in [\overline{S}]} \left[\left(\tau_{\underline{IJ}}^{[\overline{S}]} \right)_{ij}^{ab} - \left(\tau_{\underline{J}}^{[\overline{S}]} \right)_{ij}^{ab} \right] G_{iajb} \\
&+ \sum_{\substack{i \in \underline{I} \\ j \in \underline{J}}} \sum_{ab \in [\overline{S}]} \left(\tau_{\underline{IJ}}^{[\overline{S}]} \right)_{ij}^{ab} G_{iajb} \quad + \quad \sum_{\substack{i \in \underline{J} \\ j \in \underline{I}}} \sum_{ab \in [\overline{S}]} \left(\tau_{\underline{IJ}}^{[\overline{S}]} \right)_{ij}^{ab} G_{iajb}
\end{aligned}
\tag{3.17}
$$

In order to categorize and extend to higher orders, the four index amplitude tensor τ_{ij}^{ab} is recognized as a matrix with occupied indices for each virtual pair ab. That is $\tau_{ij}^{ab} = \left(\tau^{ab} \right)_{ij}$.

3 Incremental Scheme

The matrix τ^{ab} splits into domain blocks τ^{ab}_{IJ}

$$\tau^{ab} = \begin{pmatrix} \tau^{ab}_{11} & \tau^{ab}_{12} & \cdots & \tau^{ab}_{1N_{D_o}} \\ \tau^{ab}_{21} & \tau^{ab}_{22} & \cdots & \tau^{ab}_{2N_{D_o}} \\ \cdots & \vdots & \cdots & \vdots \\ \tau^{ab}_{N_{D_o}1} & \tau^{ab}_{N_{D_o}2} & \cdots & \tau^{ab}_{N_{D_o}N_{D_o}} \end{pmatrix} \tag{3.18}$$

One can recognize that E_I and ΔE_{IJ} both contribute to τ^{ab}_I and the latter also contributes to τ^{ab}_{IJ} and τ^{ab}_{JI}. For higher order energy corrections, similar arguments are used as above. Let $\mathbb{I}$ be any domain tuple with order $|\mathbb{I}|$. The calculation of $\Delta E_{\mathbb{I}}$ requires the calculation of $E_{\mathbb{I}}$ and $E_{\mathbb{J}}$, where $\mathbb{J}$ is a proper subset of $\mathbb{I}$. For $E_{\mathbb{I}}$, the summation over the occupied indices in the energy expression will decompose into sums over all pairs ($i \in \underline{I}$, $j \in \underline{J}$) with $I, J \in \mathbb{I}$. Correspondingly, the $E_{\mathbb{J}}$ will only contribute to the summations with ($i \in \underline{I}$, $j \in \underline{J}$) and $I, J \in \mathbb{J}$. Grouping all terms with respect to the sums they go in, yields $N^2_{D_o}$ sums of the type

$$\sum_{\substack{i \in \underline{I} \\ j \in \underline{J}}} \sum_{ab \in [\overline{S}]} \left(\tau^{ab}_{IJ} \right)_{ij} G_{iajb} \tag{3.19}$$

where the blocks τ^{ab}_{IJ} of the matrix τ^{ab} are build incrementally and given as

$$\tau^{ab}_{IJ} = \sum_{\substack{\mathbb{I} \in \mathcal{P}(\{1,\ldots,N_{D_o}\}) \\ IJ \in \mathbb{I} \wedge |\mathbb{I}| < I_o}} \left(\Delta\tau^{[\overline{S}]}_{\underline{\mathbb{I}}} \right)^{ab} \tag{3.20}$$

$$\left(\Delta\tau^{[\overline{S}]}_{\underline{\mathbb{I}}} \right)^{ab} = \left(\tau^{[\overline{S}]}_{\underline{\mathbb{I}}} \right)^{ab} - \sum_{\substack{\mathbb{J} \in \mathcal{P}(\mathbb{I}\backslash\{I,J\}) \\ |\mathbb{J}| < |\mathbb{I}\backslash\{I,J\}|}} \left(\Delta\tau^{[\overline{S}]}_{\underline{\mathbb{J}}} \right)^{ab} \tag{3.21}$$

The total CCSD energy of the incremental scheme is obtained as the sum of all $N^2_{D_o}$ contributions according to (3.19). Equivalently, the overall CCSD energy can be expressed simply with the matrix τ^{ab}

$$E^{\text{inc-CC}}_{\text{corr}} = \sum_{ij \in \underline{S}} \sum_{ab \in [\overline{S}]} (\tau^{ab})_{ij} G_{iajb} \tag{3.22}$$

This result demonstrates that an incremental expansion of the energy in the context of CCSD displays an incremental improvement of the CCSD amplitude quantity τ^{ab}_{ij} mediated through an incremental expansion of the occupied spaces. It seems that incremental corrections on the amplitudes with respect to occupied spaces work well, as the incremental scheme has been proven to yield accurate results in combination with CCSD. [14, 147]

There is an interesting connection to the DEC method, reviewed in section 2.6.3. The DEC method follows a different strategy than the incremental scheme. Instead of an incremental calculation of the CCSD energy, fragment and fragment pair energy contributions are obtained

from single calculations. The main difference is, that the occupied spaces of the fragment and fragment pair domains are augmented with proximate fragment occupied spaces. This is based on locality analysis of MP2 and CCSD amplitude equations that demonstrate the coupling of amplitudes with indices of proximate occupied LMOs.[128] An augmentation of the occupied space for a fragment I represents therefore the desire to improve the quality of the amplitudes and thus τ_{II}^{ab} on this fragment. The τ_{II}^{ab} block would be calculated in the occupied space $[\underline{I}]$ and virtual space $[\overline{I}]$. For the the sake of comparison, $[\overline{I}] = [\overline{S}]$ is assumed. Then, $\tau_{II}^{ab} = \left(\tau_{[\underline{I}]}^{[\overline{S}]}\right)^{ab}$ and similarly $\tau_{IJ}^{ab} = \left(\tau_{[\underline{IJ}]}^{[\overline{S}]}\right)^{ab}$. Comparing these τ^{ab} blocks with the corresponding blocks from the incremental scheme (3.20) - (3.21), leads to the following observation: the amplitude blocks τ_{IJ}^{ab} can be generated either by including all relevant neighboring occupied spaces at once in a correlation calculation or incrementally by correlating subsets of the relevant occupied spaces.

4 Incremental Scheme with expansions for occupied and virtual spaces

The incremental evaluation of energies has a long history and started on atomic calculations by Nesbet [2] who picked this idea up from the work of Bethe and Goldstone [178] and therefore named the incremental energy expansion after those authors. As presented in the previous chapter, there have been many applications in quantum chemistry for molecules [14, 15, 25, 141–159] and periodic systems [13, 130–140]. All of these methods use an incremental expansion with respect to occupied orbitals. This may be historically explained with the fact that the localization of virtual orbitals has not been equivalently well established as the localization of occupied orbitals. Methods for proper localization of the virtual space of molecular systems [57, 60–63] have started to evolve about forty years past the development of the Edmiston-Ruedenberg [20] or Foster-Boys [22] algorithms which worked only well for occupied spaces. The incremental series is not expected to converge rapidly for any of the expanded orbital spaces without a proper localization on the respective space. Furthermore, a screening for tuple selection would be difficult. There is only one exception in the literature in which an incremental expansion of the virtual space is used as an ansatz to access the FCI limit.[179] The authors used the full occupied space and the subsets of canonical virtual orbitals in the incremental calculations. An energy based algorithm has been proposed for the selection of virtual orbital tuples, since the use of locality was no option with canonical orbitals. Virtual orbital tuples of next higher orders have been generated by appending additional orbitals to the previous tuples. The additional orbitals are then only used in the next tuples generation cycle when their incremental energy contribution is larger then a given threshold.

Such an incremental expansion does not exploit the locality of electron correlation and therefore alternatives are investigated in this thesis. The next two sections address some key ideas regarding the attempt to formulate combined occupied and virtual space incremental expansions. Furthermore, a novel method which falls into this category is presented.

4.1 Incremental error propagation

The incremental expansion of the correlation energy in terms of either occupied or virtual orbitals that span the corresponding space of the target system is formally exact if the expan-

sion is not truncated. Energies of occupied and virtual tuples shall be denote by $E_\mathbb{I}$ and $E^\mathbb{A}$, respectively. The identifiers $\mathbb{I}$ and $\mathbb{A}$ for occupied and virtual fragment tuples are chosen for convenience. A further expansion of individual energies $E_\mathbb{I}$ or $E^\mathbb{A}$ in the complementary space without truncation also reproduces the energies exactly. Without any guidance regarding the tuple selection and truncation, there is no motivation for such an approach. The number of virtual tuples grows polynomially with incremental order as the occupied tuples which was pointed out at the beginning of this chapter. Obviously, no truncation or a truncation at a very high order (4.4) would represent a cumbersome redefinition of the original energy. However, even with a systematic approach, there are associated errors to the incremental expansion. These are formulated in the next two sections.

4.1.1 Expansion in one orbital space

In the ordinary incremental scheme, only the occupied space is expanded. This case is therefore explicitly shown in the following. The arguments hold equivalently for the expansion in the virtual space only. Suppose that a set of virtual orbitals spans the virtual space of the system. In this case, an error with respect to the occupied space $\Delta_\mathbb{I}$ is associated with each occupied tuple energy $E_\mathbb{I}$, since this energy is obtained by using exclusively the respective occupied space instead of the full occupied space, which would yield the total energy E. The exact contribution $E\big|_\mathbb{I}$ of the occupied tuple $\mathbb{I}$ to E could be obtained by an isolation of the contributions of $\mathbb{I}$ from E for which

$$E\big|_\mathbb{I} = E_\mathbb{I} + \Delta_\mathbb{I} \tag{4.1}$$

holds. A substitution of $E_\mathbb{I}$ into (3.1) yields a sum of two incremental expansions: one in terms of $E\big|_\mathbb{I}$ which displays the incremental expansion with the exact energies and another (4.2) in terms of $\Delta_\mathbb{I}$ which represents the propagation of the errors.

$$\Delta_{\text{corr}} = \sum_{\substack{\mathbb{I} \in \mathcal{P}(\{1,\dots,N_{D_o}\}) \\ |\mathbb{I}| < I_o}} \Delta\Delta_\mathbb{I} \qquad \text{with} \qquad \Delta\Delta_\mathbb{I} = \Delta_\mathbb{I} - \sum_{\substack{\mathbb{J} \in \mathcal{P}(\mathbb{I}) \\ |\mathbb{J}| < |\mathbb{I}|}} \Delta\Delta_\mathbb{J} \tag{4.2}$$

Lower order incremental errors contribute to higher order incremental errors several times. Here, the objective is the error propagation caused by the incremental expansion itself and this should be distinguished from other error sources in practical calculations. An example of additional errors is the convergence threshold for the energy in correlation calculations. However, such errors propagate equivalently to the errors of the method itself. A numerical analysis of the error propagation with respect to energy convergence has already been reported and the arguments can be adopted.[180] To be precise, the sum over all incremental errors $\Delta_\mathbb{I}$ at a given order $p = |\mathbb{I}|$ contributes explicitly n_{pq} times to the incremental energies in order q

with

$$n_{pq} = \binom{N_{D_o} - q}{p - q} \tag{4.3}$$

and also contributes indirectly to order q through recursive direct contributions in orders r with $p < r < q$.[180] The numerical study showed, that the errors due to convergence thresholds can be kept below chemical accuracy (up to fifth order and with $N_{D_o} \leq 20$) within the incremental scheme, if the errors of individual correlation energies are randomly distributed in the same order of magnitude $\sim 10^{-10}$. This holds even if the errors increase by a factor of 10 with each incremental order.

In the context of errors related to the incremental expansion itself, it could be expected that the individual errors $\Delta_{\mathbb{I}}$ will decrease with the incremental order, since more of the correlation space is included in the calculations at higher orders. Therefore, the single domain errors should represent a upper absolute limit for the individual errors. Indeed, also the domain size has to be considered with this respect. The smallest domain with size $D^{\min}$ will be described worst and the ratio $\frac{D^{\min}}{n_o}$ is an indicator for the accuracy. It has to be tested to which extent this ratio could in principle be reduced for arbitrary orbital spaces to maintain the validity of the method.

However, from the perspective of local correlation with local occupied orbitals, only a local subspace $\mathbb{L}_o(\mathbb{I})$ of the total occupied space will significantly correlate with the occupied tuples $\mathbb{I}$. This is an idea which is followed by methods like CIM [18, 119–126], LNO-CC [19, 166], DEC [16, 127–129] and divide-and-conquer [17, 165] (section 2.6.3). The target fragment correlation energies are obtained in these methods by augmenting the occupied space of the fragment with further relevant LMOs and later identify the energy contributions of the target fragment. This procedure was also underpinned by locality analyses of the MP2 and CCSD amplitude equations.[128] How the amplitudes calculated with augmented occupied spaces relate to the incremental scheme is presented in section 3.3. In the context of locality, the ratio $\frac{D^{\min}}{n_o}$ has to be replaced by $\frac{D^{\min}}{\mathbb{L}_o^{\min}}$ which does not vanish asymptotically with increasing molecular sizes.

No approximations on the virtual space have been considered in the preceding discussion. As in other local correlation methods, it is assumed that the virtual correlation space is also limited for an occupied tuple analogous to the occupied space. How this approximation has been implemented in the incremental scheme is presented in section 3.2. Such a restriction of the virtual space causes an error $\Delta_{\mathbb{I}}^{\mathbb{L}_v(\mathbb{I})}$ for each tuple $\mathbb{I}$ due to the smaller virtual space $\mathbb{L}_v(\mathbb{I})$. It is pointed out in the manuscript in section 5.1 how this error propagates and that it is desirable to minimize each $\Delta_{\mathbb{I}}^{\mathbb{L}_v(\mathbb{I})}$ individually.

4.1.2 Expansion in both orbital spaces

For an occupied tuple energy $E_{\mathbb{I}}$, the expansion in the virtual space would be given in analogy to (3.1) as

$$E_{\mathbb{I}} = \sum_{\substack{\mathbb{A} \in \mathcal{P}(\{1,\ldots,N_{D_v}\}) \\ |\mathbb{A}| < I_v}} \Delta^v E_{\mathbb{I}}^{\mathbb{A}} \qquad \text{with} \qquad \Delta E_{\mathbb{I}}^{\mathbb{A}} = E_{\mathbb{I}}^{\mathbb{A}} - \sum_{\substack{\mathbb{B} \in \mathcal{P}(\mathbb{A}) \\ |\mathbb{B}| < |\mathbb{A}|}} \Delta^v E_{\mathbb{I}}^{\mathbb{B}} \qquad (4.4)$$

where N_{D_v} and I_v denote the number of virtual fragments and the virtual expansion order, respectively, which do not have to coincide with those of the occupied space in general. The incremental expansion for the virtual space introduces an error $\Delta_{\mathbb{I}}^{\mathbb{A}}$ for each calculated energy $E_{\mathbb{I}}^{\mathbb{A}}$, since subsets of the full orbital spaces are used. These incremental errors propagate equivalently to $\Delta_{\mathbb{I}}$ as shown in (4.2) and yield a total incremental error for an occupied tuple

$$\Delta_{\mathbb{I}} = \sum_{\substack{\mathbb{A} \in \mathcal{P}(\{1,\ldots,N_{D_v}\}) \\ |\mathbb{A}| < I_v}} \Delta^v \Delta_{\mathbb{I}}^{\mathbb{A}} \qquad (4.5)$$

which in turn propagates through (4.2) to the total error of the incremental scheme. Note, that it has been assumed that the virtual space can be divided into N_{D_v} subsets. This is always possible by an arbitrary disaggregation of any appropriate virtual space. However, by doing so there is no hope for fast convergence of the virtual space expansion. The presented incremental expansion would only be meaningful in the local correlation sense, if the virtual space could be decomposed into local virtual subspaces. One possibility would be to use well localized unitary transformed virtual orbitals which are currently available [57, 60–63] but not used in the context of the incremental scheme yet. With such a local fragmentation of the virtual space, it is possible to reuse the argumentation in the last section regarding the largest errors: the calculation with the largest error will be the one with the smallest occupied and virtual space and the ratios $\frac{D_o^{\min}}{\mathbb{L}_o^{\min}}$ and $\frac{D_v^{\min}}{\mathbb{L}_v^{\min}}$ represent a measure for accuracy. The peculiarities of an expansion in the virtual space with alternative virtual spaces than virtual LMOs are discussed in the next section.

4.2 Paradigms in local incremental expansions of the virtual space

The attempt to combine an incremental scheme for occupied as well as virtual spaces may seem to be an obvious extension to the incremental expansion of the virtual space. That this is not the case with traditional approaches of local methods will be clear from the following discussion. A straightforward extension to the incremental expansion of the occupied space with the additional expansion of the virtual space is presented in section 4.1.2. However, with disjoint subsets of the virtual space, it is equivalently possible to define the inverse incremental expansion

$$E_{\text{corr}} = \sum_{\substack{\mathbb{A} \in \mathcal{P}(\{1,\dots,N_{D_v}\}) \\ |\mathbb{A}| < I_v}} \Delta^v E^{\mathbb{A}} \qquad \text{with} \qquad \Delta^v E^{\mathbb{A}} = E^{\mathbb{A}} - \sum_{\substack{\mathbb{B} \in \mathcal{P}(\mathbb{A}) \\ |\mathbb{B}| < |\mathbb{A}|}} \Delta^v E^{\mathbb{B}} \qquad (4.6)$$

where the occupied space is expanded for each virtual tuple $\mathbb{A}$

$$E^{\mathbb{A}} = \sum_{\substack{\mathbb{I} \in \mathcal{P}(\{1,\dots,N_{D_o}\}) \\ |\mathbb{I}| < I_o}} \Delta_o E_{\mathbb{I}}^{\mathbb{A}} \qquad \text{with} \qquad \Delta_o E_{\mathbb{I}}^{\mathbb{A}} = E_{\mathbb{I}}^{\mathbb{A}} - \sum_{\substack{\mathbb{J} \in \mathcal{P}(\mathbb{I}) \\ |\mathbb{J}| < |\mathbb{I}|}} \Delta_o E_{\mathbb{J}}^{\mathbb{A}} \qquad (4.7)$$

and the energy is then calculated with the increments of the virtual space. Virtual LMOs are candidates for this as they have the same properties as occupied LMOs and could therefore be treated on equal footing. This choice may not seem obvious at the first glance for other virtual spaces. However, some elements of this expansion will be beneficial when dealing with the pitfalls of traditional virtual spaces for a local ansatz. PAOs and OSVs are for example not orthogonal between virtual domains but orthogonal to the occupied LMOs. They have to be orthogonalized for each virtual tuple and therefore differ from the virtual orbitals of the individual virtual domains. Consider a pair of virtual domains AB for simplicity. A pair increment for any occupied tuple $\mathbb{I}$ would then be given as

$$\Delta_o E_{\mathbb{I}}^{AB} = E_{\mathbb{I}}^{AB} - E_{\mathbb{I}}^{A} - E_{\mathbb{I}}^{B} \qquad (4.8)$$

where $E_{\mathbb{I}}^{AB}$ is calculated in the virtual space $[\overline{AB}]$ and $E_{\mathbb{I}}^{A}$ and $E_{\mathbb{I}}^{B}$ in the spaces $[\overline{A}]$ and $[\overline{B}]$, respectively. The individual orbital sets are not proper subsets of the pair orbitals. For example $[\overline{A}] \not\subset [\overline{AB}]$ although $[\overline{A}] \cup [\overline{B}]$ and $[\overline{AB}]$ span the same space. This causes inconsistencies in the incremental energy expansion of the virtual space for a specific occupied tuple. In contrast, occupied orbital domains are proper subsets of the full LMO space for any tuple. This validates the calculation of specific virtual tuple contributions $\Delta_o E_{\mathbb{I}}^{\mathbb{A}}$ according to the right hand side of (4.7). This energy contribution does already take care of the incremental expansion of the occupied space. As $\mathbb{A}$ may not be needed for all $\mathbb{I}$ and *vice versa* the two versions of nested incremental schemes may be abandoned in the local picture. Instead of seeking $E_{\mathbb{I}}$ through a virtual space expansion and than calculating $\Delta_o E_{\mathbb{I}}$, it seems reasonable to use $\Delta_o E_{\mathbb{I}}^{\mathbb{A}}$ contributions to access $\Delta_o E_{\mathbb{I}}$ directly through important virtual tuples. More precisely, if two virtual tuples are chosen to contribute to $\mathbb{I}$, a possible expansion could be

$$\Delta_o E_{\mathbb{I}} = \Delta_o E_{\mathbb{I}}^{\mathbb{A}_1} + \Delta_o E_{\mathbb{I}}^{\mathbb{A}_2} - \Delta_o E_{\mathbb{I}}^{\mathbb{A}_1 \cap \mathbb{A}_2} \qquad (4.9)$$

where the last term is necessary to avoid double counting of domain contributions which lie in the intersection of the two virtual tuples. This kind of expansion does not fully circumvent the problem, related to the non-proper subsets issue with local virtual tuples, but it only requires that the virtual space of the intersection $[\overline{\mathbb{A}_1 \cap \mathbb{A}_2}]$ is a proper subset of $[\overline{\mathbb{A}_1}]$ and $[\overline{\mathbb{A}_2}]$.

This is not as strict as the requirement that any subset of any tuple has to be a proper subset. Nevertheless, a generalization of (4.9) to more than two virtual tuples

$$\Delta_o E_{\mathbb{I}} = \sum_{p=1}^{n_{\mathbb{A}}} \left(\Delta_o E_{\mathbb{I}}^{\mathbb{A}_p} - \sum_{q<p} \Delta_o E_{\mathbb{I}}^{\mathbb{A}_p \cap \mathbb{A}_q} \right) \tag{4.10}$$

where $\cup_p \mathbb{A}_p = \mathbb{L}_v(\mathbb{I})$ demonstrates that many tuples with proper subsets could be needed. The next sections are concerned with explicit formulations of the presented ideas under consideration that the properties of the virtual tuple spaces have peculiarities in the local correlation framework. It seems therefore beneficial to classify the different orbital spaces according to their properties first and than proceed with the design of the expansion.

A list of virtual spaces is presented in table 4.1, where important properties as "orthogonal tuples", "tuples have proper subtuples" and "local tuples" are compared and the virtual spaces classified according to these properties. Class IA contains virtual LMOs which fulfill all requirements and can be used in an expansion shown in (4.7) and the inverse expansion where the occupied space is expanded first, equivalently. Cholesky decomposed virtual spaces fall into class IB. These orbitals are orthonormal and do not require further orthogonalization. They therefore do not entail the proper subtuple problem. However, the locality of the orbitals depends on the pivoting order. In this context, a graph theoretical tool, namely the Cuthill-Mckee ordering [181] could help to maximize the locality. For this, the connectivity of the atoms could be represented in a sparse connectivity matrix which is then transformed to yield a minimal bandwidth. The ordering of the so obtained atomic basis can subsequently be used for the pivoting.

Table 4.1: Classification of virtual spaces according to their properties with regard to an application within an incremental virtual space expansion. CDO: Cholesky decomposition orbitals; SO: Schmidt orthogonalization; LO: Löwdin orthogonalization; ✓* : requires special construction; (✓): holds approximately.

virtual space	orthogonal tuples	tuples have proper subtuples	local tuples	class
LMO	✓	✓	✓	IA
CDO	✓	✓	✓*	IB
SO-PAO	✓	✓*	✓	IIA
SO-OSV	✓	✓*	✓	IIA
EGV	✓	(✓)	✓	IIB
CMO	✓	✓	✗	III
LO-PAO	✓	✗	✓	IV
LO-OSV	✓	✗	✓	IV
PAO	✗	✓	✓	V
OSV	✗	✓	✓	V

The second class contains virtual spaces which have orthogonal virtuals within a tuple and are local on the tuple. Schmidt orthogonalized local virtual domains as PAOs and OSVs fall into class IIA with a special preparation. A random Schmidt orthogonalization of the local domains would maintain their locality on the tuple but there would be no lower order tuples, also orthogonalized randomly, which are proper subsets of the generated tuple. The same argument holds for Löwdin orthogonalized local domains, although Löwdins orthogonalization preserves the original orbitals to maximal extent in a least squares sense.[182] It is however possible to orthogonalize the local domains in a manner that proper subsets are obtained for the incremental expansion. This procedure and the corresponding expansion are shown in section 4.2.2. The subclass IIB contains domain tuples which approximately have proper subsets. Embedding generated virtuals (EGVs) are an example for this class. In which sense they provide approximate subtuples and how this can be used within the framework of virtual space expansion is demonstrated in section 4.2.1. The corresponding method has also been applied successfully in molecular calculations, which is presented in the manuscript in section 5.1.

Class III spaces are obviously not suited for local approaches. The virtual spaces in the classes IV and V are traditional virtual spaces for local methods but are less desirable within the context of a local virtual space expansion with standard quantum chemistry codes.

4.2.1 Virtual expansion in terms of domain tuples with proper subtuples

In this thesis, a virtual space expansion is realized with embedding generated virtuals (EGVs) which is presented in the manuscript in section 5.1. The EGVs are obtained through embedding calculations, where the orbitals on the tuple $\mathbb{A}$ are calculated in a frozen environment of the remaining occupied domains. These orbitals are local on $\mathbb{A}$, since the basis set is restricted to this specific tuple. Atomic domains are used in this context as presented in section 3.1.3. A proper description of the virtual space of a tuple is not guaranteed with EGVs, if a tuple has domain "holes". That is, if connecting atomic domains are missing in the active part of the embedded system (fig. 4.1). This leads to "holes" in the AO basis and thus to an inconsistent description of the virtual space. Contrary, if a terminal atomic domain A is removed from $\mathbb{A}$

connected   disconnected
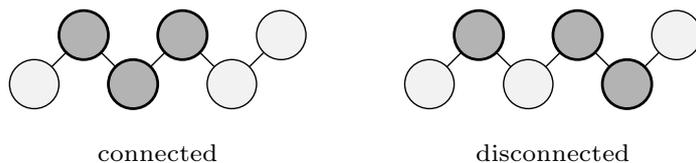
Figure 4.1: Connected and disconnected tuples
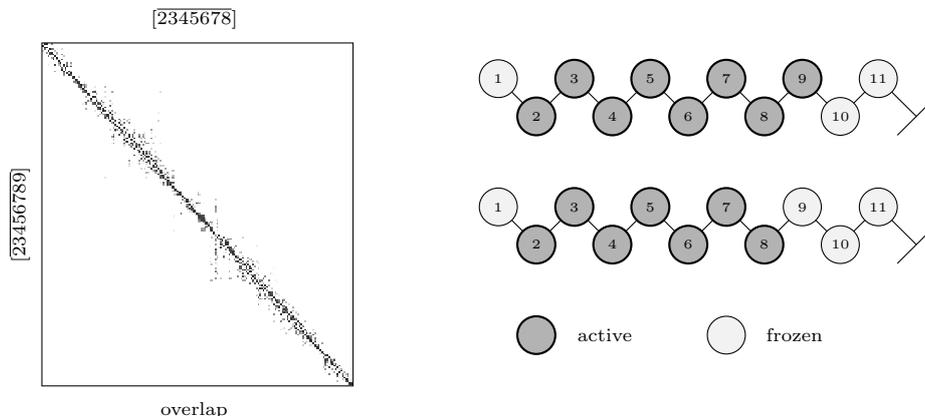
Figure 4.2: Overlap of virtual orbitals between embedding generated virtuals for the tuples (23456789) and (2345678) in the molecule $C_{18}H_{38}$ where each carbon atom is associated with a domain according to the sketch on the right hand side. Basis set: CC-PVDZ.

and added to the environment, the virtual orbitals on $\mathbb{B} = \mathbb{A} \setminus A$ will not differ extremely from the original ones. This is demonstrated exemplarily in figure 4.2 where the overlap between virtual orbitals on $\mathbb{B}$ and $\mathbb{A}$ are shown. In comparison, symmetrically orthogonalized PAOs on $\mathbb{B}$ and $\mathbb{A}$ differ much more. EGVs are therefore considered to have approximately proper subsets of tuples with similar tuple sizes where only terminal domains are removed. A removal of central domains causes more distortion on the virtual space.

From the observations above, one could attempt to formulate an expression for the incremental energies $\Delta_o E_{\mathbb{I}}$ in terms of $\Delta_o E_{\mathbb{I}}^{\mathbb{A}}$. Such an expansion is presented in (4.9) but needs a redefinition to be used more efficiently with EGVs. As stated above, double counting has to be prevented by subtracting terms as $\Delta_o E_{\mathbb{I}}^{\mathbb{A}_1 \cap \mathbb{A}_2}$. With EGVs it is therefore additionally important to select the tuples such that there are no disconnected intersections $\mathbb{A}_1 \cap \mathbb{A}_2$. The overall constraints on the tuples can be summarized as follows:

(1) only connected tuples are used

(2) only tuples with connected intersections are used

(3) connected intersections are not too small

The last requirement is due to conservation of the "tuples have approximate proper subsets" condition. These requirements have been met in the method presented in the manuscript in section 5.1 by using a connected base tuple and additional connected tuples which consist of domains from the base and extra domains which are different in all additional tuples.
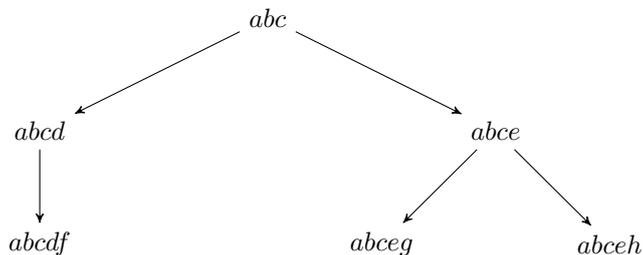
Figure 4.3: Rooted tree visualizing intersections of the tuples $(abcdf), (abceg)$ and $(abceh)$.

4.2.2 Virtual expansion with Schmidt orthogonalized local domains

PAOs are non-specific orbitals but if they are used to form virtual spaces for a set of LMOs as shown in section 2.6.1, they represent local virtual spaces for the individual occupied space fragments. These virtual spaces $[\bar{p}]$ are not orthogonal between each other and are also linearly dependent. The former holds also for OSV virtual spaces for occupied LMOs. An incremental expansion of the virtual space requires the use of virtual spaces for tuples $[\bar{\mathbb{A}}]$. In order to run correlation calculations with standard quantum chemistry codes, the virtual spaces should be orthogonalized. A symmetric orthogonalization (LO) of a tuple virtual space yields virtual spaces for which no subtuple space is a proper subset. That is

$$\mathbb{B} \subset \mathbb{A} \; \Rightarrow \; [\bar{\mathbb{B}}]_{\mathrm{LO}} \not\subset [\bar{\mathbb{A}}]_{\mathrm{LO}} \qquad \forall \, \mathbb{B} \tag{4.11}$$

This would cause inconsistencies in an incremental expansion. A special use of the Schmidt orthogonalization (SO) poses a workaround. The Schmidt orthogonalization of virtual domains in a predefined order $a_1 \ldots a_p$ provides virtual spaces $[\overline{a_1 \ldots a_p}]_{\mathrm{SO}}$ for which all subsets with the same ordering are proper subsets

$$[\overline{a_1 \ldots a_q}]_{\mathrm{SO}} \subset [\overline{a_1 \ldots a_p}]_{\mathrm{SO}} \qquad \forall \, q < p \tag{4.12}$$

since subsets are formed earlier in the orthogonalization process and not changed by further orthogonalization to other domains. This is a distinguishing property of Schmidt orthogonalization. Although it is not possible to realize a virtual space expansion according to (4.10), an alternative expansion is possible which exploits (4.12). Assuming that a set of virtual tuples $\mathbb{A}_p$ contains all relevant domains in $\mathbb{L}_v(\mathbb{I})$ for the correlation spaces of $\mathbb{I}$, the smallest common subset of these tuples can be identified as the root subset $\mathbb{A}^{(0)} = \cap_p \mathbb{A}_p$. All other tuples and subsets can be formed by successfully adding domains to the root tuple. This is illustrated in figure 4.3 by a rooted tree. The leafs of the graph represent the virtual tuples which were originally considered for the expansion. Any shared parent vertex represents also a common

tuple subset. This representation is of special use with respect to a virtual space expansion with Schmidt orthogonalized domains. If the extra domain in a child vertex is Schmidt orthogonalized to the tuple domain space in the parent vertex, the virtual tuple space of the parent vertex becomes a proper subspace of the child's virtual space. The same argument also holds for all other ascendants of a vertex. In figure 4.3 the virtual space $\overline{[abc]}_{\text{SO}}$ would be a proper subspace of $\overline{[abcd]}_{\text{SO}}$ and $\overline{[abce]}_{\text{SO}}$ for instance. Besides, $\overline{[abc]}_{\text{SO}}$ and $\overline{[abce]}_{\text{SO}}$ are proper subsets of $\overline{[abceg]}_{\text{SO}}$ and $\overline{[abceh]}_{\text{SO}}$.

One can therefore realize a virtual space expansion which uses PAOs by starting with the energy contribution of the root tuple $\Delta_o E_{\mathbb{I}}^{\mathbb{A}^{(0)}}$ and adding incremental energy contributions for tuples along the rooted graph. The grouping of vertices with respect to the distance to the root vertex is reasonable in this context. Let $S^{(d)}$ denote the set of tuples with the distance d. For any tuple $\mathbb{A}^{(d)}$ at distance d, the contribution $\Delta_o E_{\mathbb{I}}^{\mathbb{A}^{(d)}} - \Delta_o E_{\mathbb{I}}^{\text{P}(\mathbb{A}^{(d)})}$ has to be added where $\text{P}(\mathbb{A}^{(d)})$ is the parent of $\mathbb{A}^{(d)}$. The expansion for the virtual space is therefore formulated as

$$\Delta_o E_{\mathbb{I}} = \Delta_o E_{\mathbb{I}}^{\mathbb{A}^{(0)}} + \sum_{d=1}^{d^{\max}} \sum_{\mathbb{A}^{(d)} \in S^{(d)}} \left(\Delta_o E_{\mathbb{I}}^{\mathbb{A}^{(d)}} - \Delta_o E_{\mathbb{I}}^{\text{P}(\mathbb{A}^{(d)})} \right) \tag{4.13}$$

5 Incremental expansion with embedding generated virtuals

A new incremental scheme which follows a combined occupied and virtual space expansion ansatz is presented in a manuscript which is included in this chapter. All presented results, figures and tables in the manuscript are exclusively obtained and prepared by the author of this thesis, Ilyas Türkmen. The supervisor and co-author Prof. Dr. M. Dolg suggested corrections for the final written and did not contribute with original work.

The following sections provide additional remarks on the graph theoretical concepts that are used in the manuscript and a follow up idea on the inclusion of approximately natural orbitals in the framework of the incremental scheme.

5.1 Manuscript

<div align="center">

Manuscript

Linear Scaling Incremental Scheme for Correlation Energies with Embedding Generated Virtuals

</div>

<div align="center">

I. Türkmen and M. Dolg

J. Chem. Theory Comput. 2024, 20, 8, 3154-3168

</div>

A linear scaling Incremental Scheme for correlation energies with embedding generated virtuals

Ilyas Türkmen[1] and Michael Dolg[1]

Institute for Theoretical Chemistry, University of Cologne, Greinstr. 4, D-50939 Cologne, Germany

A novel incremental scheme is presented including an incremental expansion of the virtual space for the calculation of electron correlation energies, which is compatible with any size-extensive correlation method and scales asymptotically linear for large molecules. The performance is studied for organic molecules, water clusters and a La(III) water complex, where also the compatibility with pseudopotentials is examined. The computational requirements are already reduced tremendously for medium sized water clusters and hydrocarbons with respect to the canonical CCSD as well as the ordinary incremental scheme references. Correlation energies within chemical accuracy have been observed for all studied systems. The novelty of method is that relatively small virtual spaces are used in combination with tuples of localized occupied spaces. The corresponding orthonormal occupied and virtual orbitals are obtained from QM/QM embedding calculations and can thus be used with standard quantum chemistry codes for correlation calculations. It is presented how relevant virtual spaces are selected and the correlation energies are linked in the new virtual space expansion.

Keywords: electron correlation methods; local correlation; incremental scheme; linear scaling; embedding; ab initio

I. INTRODUCTION

Wave function-based local correlation methods are in continuous progress and have been pioneered by the work of Sinanoglu[76] and Nesbet[57]. All methods try to exploit the locality of dynamic electron correlation by introducing approximations in order to combine savings in computation time with an acceptable loss of accuracy. The strategy is to incorporate only interactions of spatially adjacent parts of the N_o occupied and N_v virtual orbitals. As reliable standard ab initio methods such as coupled cluster singles and doubles (CCSD) and CCSD with perturbative triples (CCSD(T)) scale with $N_o^2 N_v^4$ and $N_o^3 N_v^4$, respectively, local approaches become more important with increasing system size.

Many local correlation methods have evolved to manifested the concepts of local electron correlation. These methods can conceptually by divided into two groups which fundamentally differ in their approaches. The direct methods treat the whole system in a single correlation calculation and adjust specific local virtual spaces to the localized occupied molecular orbitals (LMOs). The methods in this group are distinguished by the choice of the virtual spaces. Localized canonical virtual orbitals have been no option for a for long time, as they could not be obtained with a Jacobi sweep algorithm, other than occupied LMOs.[82] Although there has been tremendous progress in this context[29–31,35,82], alternative non-orthogonal virtual spaces have been used traditionally in the framework of local correlation for historical reasons. This started with the pioneering work of Pulay and Saebo through the introduction of projected atomic orbitals[64] (PAOs). They are used in direct local correlation methods as PAO-LMP2[65], PAO-LCCSD[25] and PAO-LCCSD(T)[74] to form virtual spaces for all LMOs. Another subgroup of direct methods uses pair natural orbitals[1,8,56] (PNOs) and orbital specific virtuals[87] (OSVs) which are virtual spaces for pairs or single LMOs obtained from diagonalization of LMO pair densities. With these virtual spaces, local MP2[85,87], CCSD[55,66,73,88] and CCSD(T)[48,67,75] have been implemented. Additionally, also explicitly correlated F12 versions available.[46,47].

The alternative to a direct calculation of the full system is fragmentation. That is, the full system is partitioned into smaller parts and the energy of the full system is reformulated in terms of fragment energies. Fragment energies can be accessed either directly from a calculation within the fragment spaces only or indirectly by extending the correlation space of the fragment and identifying the contributions of the fragment afterwards. The fragment molecular orbital (FMO) CC method[13] and the incremental scheme method[17–20,79,89,90] fall into the former class. Examples for the latter are the divide-and-conquer[38,39], divide-expand-consolidate[12,34,40,93] (DEC), cluster-in-molecule[24,41–45,59,60] (CIM) and local natural orbital[68,69] (LNO) methods. In this work, we would like to extend the ideas of the original incremental scheme by Stoll[79], which has also previously been used in the context of atomic calculations by Nesbet.[58] The ordinary incremental scheme represents a many-body expansion of the correlation energy in terms of sets of occupied LMOs. The correlation energies are obtained by correlating each tuple of occupied LMO sets with a single virtual space and calculating incremental contributions with the many-body expansion. This method has been successfully applied for the calculation of correlation energies of molecular systems[19–22,37,50,51,61,62,77,89,91,92] as well as periodic systems[4,5,52–54,70–72,78–81]. A truncation

of the series at a maximum tuple size, called incremental order I_o, of three is usually sufficient to obtain chemical accuracy. Another advantage is the compatibility with existing implementations of correlation methods. It has for example been combined with non-local CCSD(T) methods[18] as well as with the efficient local correlation method DLPNO-CCSD(T).[15] Originally, the total set of canonical virtual orbitals was used for each occupied tuple, which is not feasible for large systems.[58] A compression of the virtual space was achieved by the use of PAO spaces[20] or domain specific basis sets[17,89,90] and ensured the applicability for larger systems. An exclusive expansion of the virtual space was investigated, but not in the context of local correlation.[11] There is also no incremental scheme which combines an expansion of occupied and virtual spaces. To our knowledge, there is also no other local correlation method, that realized this approach. Particularly in chemical systems in which several atoms are correlated in a small area, such as at metal centers of complexes, the virtual space to be considered at once can become very large. An expansion of the virtual space in smaller virtual spaces would for example be advantageous in order to be able to calculate accurate binding energies of complexes with high coordination numbers.

In this paper, we investigated a combined occupied and virtual space expansion within the framework of the incremental scheme in order to further reduce the computational resources for the most expensive calculation in the incremental scheme. We show for the system $(H_2O)_{21}$ that the computation time and memory requirements for the most expensive correlation calculation are overall reduced by more than two orders of magnitude in the new method compared to the ordinary incremental scheme.

The traditional choice for virtual spaces in the local correlation framework would be non-specific orthogonalized PAOs, occupied orbital specific (OSV), or orbital pair specific ones (PNOs). These spaces are not suited for the virtual space expansion as will be described below. Alternatively, occupied and virtual orbitals from QM/QM embedding calculations are used which are referred to as embedding generated virtuals (EGVs). The Huzinaga embedding[33] and projector-based embedding[23,49] methods are employed for this purpose. EGVs are SCF orbitals and thus orthonormal. They can be used in standard CCSD codes. These two embedding approaches have previously been used in the framework of density functional theory (DFT) and wave function theory (WFT) embedding.[27,49] Within these embedding methods, the system is partitioned into two parts, where one of them is embedded into the other and the total system is calculated by using different levels of theory on the different parts. In our approach, we do not treat different parts on different levels of theory. We partition the system as in the ordinary incremental scheme into domains of occupied spaces. The embedding calculations are only used as a tool to generate specific virtual spaces to be correlated with the tuples of occupied domains. The calculated incremental energy contributions are subsequently used to calculate the correlation energy of the full system.

II. THEORY

Notation

We will use a somehow new methodology in this paper. The virtual space expansion involves different virtual spaces for the same occupied space. There is no one to one correspondence as in other local correlation methods where a single identifier is sufficient. We shall therefore extent the conventional notation and restore as much as possible. Occupied orbital and virtual orbital indices are denoted by $i, j, k, \ldots$ and $a, b, c, \ldots$, respectively as usual. Groups of occupied orbitals, also denoted as occupied domains, will be referenced by capital letters $I, J, K, \ldots$ and the associated virtual orbital counterparts as $A, B, C, \ldots$. We will also provide expressions which contain tuples of domain indices. These are denoted by doublestruck letters $\mathbb{I}, \mathbb{J}, \mathbb{K}, \ldots$ and $\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots$, respectively. These quantities are only used to reference the domains and domain tuples. The corresponding occupied orbital spaces and virtual orbital spaces are denoted by $[\mathbb{I}]$ and $[\mathbb{A}]$, respectively.

A. Incremental scheme for occupied space only

Occupied orbitals of the full system are localized and occupied domains are formed. We chose a new approach for this step which is presented in section II A 1. Occupied spaces $[\mathbb{I}]$ of any domain tuple $\mathbb{I}$ are constructed as a unification of the LMO spaces of the individual domains

$$[\mathbb{I}] = \bigcup_{I \in \mathbb{I}} [I] \tag{1}$$

Each occupied tuple $\mathbb{I}$ has a single associated virtual space $[\overline{\mathbb{L}(\mathbb{I})}]$, which shall span the virtual correlation space of the occupied tuple and where $\mathbb{L}(\mathbb{I})$ indicates that this could possibly be a local virtual space. The chosen virtual spaces vary in the literature. Canonical orbitals[58], orthogonalized PAO spaces[20], virtual orbitals in the context of domain specific basis sets[17,89] and frozen natural orbitals[83] have been reported. We will also present some results obtained with embedding generated virtual orbitals in a single virtual space treatment, additional to the virtual space expansion.

The correlation energy of the full system $\tilde{E}_{\text{corr}}$ is expanded incrementally over the occupied tuples $\mathbb{I}$ which is given in the established power set notation[20] as

$$\tilde{E}_{\text{corr}} = \sum_{\mathbb{I} \in \mathcal{P}(\{1, \ldots, N_D\}) \wedge |\mathbb{I}| \leq I_o} \Delta \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} \tag{2}$$

where $\mathcal{P}(\{1,\ldots,N_D\})$ is the power set of the domain indices for the N_D domains and

$$\Delta\epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} = \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} - \sum_{\mathbb{J}\in\mathcal{P}(\mathbb{I})\,\wedge\,|\mathbb{J}|<|\mathbb{I}|} \Delta\epsilon_{\mathbb{J}}^{\mathbb{L}(\mathbb{J})} \qquad (3)$$

The subscripts and superscripts of ϵ denote the occupied and virtual spaces used in the correlation calculation, respectively. The energy $\epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ is thus calculated in the occupied space $[\mathbb{I}]$ and virtual space $[\overline{\mathbb{L}(\mathbb{I})}]$.

1. Occupied space partitioning of the full system

We chose to use atomic occupied domains, where each non-hydrogen atom is considered as a domain center. This is advantageous with the perspective that such occupied domains are used in embedding calculations (section II B), in which we can treat basis set restrictions and domain restrictions for the active part on equal footing.

In the following, we present a method for the generation of non-hydrogen atomic domains, which differs from the previously used graph partitioning[20] and K-means partitioning[89] schemes, which did not aim to generate atomic domains. We shall note at this point that the partitioning is a crucial step, since the convergence of the incremental expansion will be faster with a more appropriate partitioning. Our partitioning scheme is briefly summarized as follows.

The canonical occupied orbitals of a HF calculation on the full system are localized with the Foster-Boys localization[3] in combination with the Edmiston and Ruedenberg algorithm[9]. The centers of charges $\vec{R}_i$ of occupied LMOs φ_i^{LMO} are then calculated according to

$$\vec{R}_i = \langle \varphi_i^{\mathrm{LMO}}|\vec{r}|\varphi_i^{\mathrm{LMO}}\rangle \qquad (4)$$

We than proceed with the identification of LMOs which correspond to the non-hydrogen atomic domains. First, the centers of domains is set equal to the non-hydrogen atomic positions of the molecule and the number of LMOs in a sphere is obtained. A pseudo-charge is assigned to the domain center according to this number. The domain centers are then allowed to move in a predefined sphere, such that they can be associated with the atomic position at any instance. The positions of the domain centers are successively updated according to a pseudo-force due to pseudo-potentials between the domains and between domains and center of charges, which have a negative unit pseudo-charge assigned. Domain centers which already have close lying center of charges will migrate closer towards the center of those and free therefore more loose center of charges, which in turn can be accessed by other domains. This algorithm is repeated until the pseudo-potential difference becomes smaller in magnitude than a predefined threshold. The actual disjoint occupied orbital spaces are then formed by assigning each LMO center of charge to the closest optimized domain center.

2. Error propagation for occupied space

Each occupied domain tuple space $[\mathbb{I}]$ has an individual virtual space $[\overline{\mathbb{L}(\mathbb{I})}]$ assigned. This introduces an error $\Delta^{\mathbb{L}(\mathbb{I})}$ if this spaces does not span the total virtual space of the system $[\overline{\mathbb{S}}]$ which can be expressed as

$$\epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} = \epsilon_{\mathbb{I}}^{\mathbb{S}} - \Delta^{\mathbb{L}(\mathbb{I})} \qquad (5)$$

If this is substituted into (2), it is observed that the total error $\Delta\tilde{E}_{\mathrm{corr}}$ has an equivalent incremental expansion as the energy $\tilde{E}_{\mathrm{corr}}$ itself, where each $\epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ is replaced by $\Delta^{\mathbb{L}(\mathbb{I})}$. In order to minimize the error systematically, each $\Delta\Delta^{\mathbb{L}(\mathbb{X})}$, arising from (3), should be attempted to be minimized. This would be the case if the error $\Delta^{\mathbb{L}(\mathbb{I})}$ is equal to the sum of the incremental errors of the subtuples $\mathbb{J}\in\mathcal{P}(\mathbb{I})$. The same effect is also achieved if each $\Delta^{\mathbb{L}(\mathbb{X})}$ is minimized individually. That is, the virtual space $[\overline{\mathbb{L}(\mathbb{I})}]$ should be sufficiently large to reproduce the energies obtained with the total virtual space. Note, that the discussed error does not take the error of the incremental scheme itself into account, which still remains even if the full virtual space is used.

3. Approximations due to locality of correlation

Dynamic correlation is short ranged and falls asymptotically as R^{-6}, which is the dispersion interaction. Hence, if the distance of two occupied domains I and J is large, the correlation energy of the pair IJ will be approximately equal to the sum of the correlation energies of the individual domains.

$$\epsilon_{IJ} \approx \epsilon_I + \epsilon_J \qquad (6)$$

Moreover, any higher order occupied domain tuple containing I and J does not contribute significantly to the correlation energy. For example, adding a third domain K would mainly reconstruct the correlation in IJ and JK. Only the simultaneous correlation of K with I and J is added, which is expected to be a small contribution with respect to the R scaling. These arguments can be propagated through all incremental orders. Incremental energies associated with any occupied domain tuple $IJ\in\mathbb{I}$ can therefore be neglected, since $\Delta\epsilon_{\mathbb{I}}\approx 0$.

We apply an occupied domain tuple selection criterion which differs form the real space distance threshold $d_{\mathrm{thresh}}^{\mathrm{s}}$ for pairs of LMO center of charges used in other incremental schemes.[19,20,89] The selection scheme utilizes a Fock matrix analysis. For this, the diagonal occupied-occupied block of the Fock matrix $\mathbf{F}$ of the total system is transformed for this purpose into the basis of the LMOs. This transformation requires the transformation matrix from the canonical to the LMO basis $\mathbf{C}^{\mathrm{LMO}}$.

$$\mathbf{F}^{\mathrm{LMO}} = \mathbf{C}^{\mathrm{LMO}}\mathbf{F}\mathbf{C}^{\mathrm{LMO\,T}} \qquad (7)$$

The occupied-occupied block of this matrix is is diagonally dominant if sorted according to occupied domains. The magnitude of the off-diagonal elements represents a measure for the effect of the corresponding orbitals at the correlation level.[40] For the purpose of occupied domain tuple selection, the Fock matrix in the LMO basis $\mathbf{F}^{\mathrm{LMO}}$ is mapped to a smaller matrix $\mathbf{F}^{\mathrm{max}}$ containing the maximum absolute Fock matrix values of each pair of domains

$$\mathbf{F}_{IJ}^{\mathrm{max}} = \max_{i \in [\underline{I}], j \in [\underline{J}]} |\mathbf{F}_{ij}^{\mathrm{LMO}}| \tag{8}$$

which represents the strongest links between LMOs. Occupied domain tuples in which all pairs of domains have $\mathbf{F}^{\mathrm{max}}$ values above a threshold $F_{\mathrm{thresh}}^{\mathrm{o}}$ are considered in the incremental expansion of the occupied space. That is

$$\mathbb{I} \in \mathcal{P}(1, \ldots, N_D) \; : \; \mathbf{F}_{IJ}^{\mathrm{max}} > F_{\mathrm{thresh}}^{\mathrm{o}} \quad \forall \, I, J \in \mathbb{I} \tag{9}$$

B. QM/QM embedding and augmentation

In QM/QM embedding, a system $\mathbb{S}$ is divided into two parts, a part $\mathbb{X}$ containing the sites of interest and the remaining environment $\mathbb{E} = \mathbb{S} \backslash \mathbb{A}$. The occupied orbital space of the environment $[\underline{\mathbb{E}}]$ is frozen at the SCF level in an embedding calculation and the active part is optimized in the fixed environment. This enables the calculation of the orbital spaces $[\underline{\mathbb{X}}]$ and $[\overline{\mathbb{X}}]$ of the parts of interest. The so obtained orbitals diagonalize an embedding Fock operator and thus provide orbitals for a subsequent correlation calculation with a standard correlation method code.

An embedding calculation requires a predefined set of frozen orbitals. These orbitals can be obtained from a preceding Hartree-Fock calculation of the full system. We will freeze a selection of occupied domain spaces obtained according to the procedure in section II A 1.

1. Huzinaga equation based embedding

Within the Huzinaga embedding, the wavefunction of the total system is set up as an antisymmetrized product of closed-shell electronic group functions of which one is the active $\mathbb{X}$ and the others $\mathbb{E}_i$ are frozen parts of the system.[33] Assume that the system is divided in two parts for simplicity. Following this ansatz a modified Fock equation is derived which only varies the orbitals on $\mathbb{X}$.

$$\left(\hat{\mathcal{F}} - \{ \hat{\mathcal{F}}, \hat{\rho}_{\mathbb{E}} \} \right) |\varphi_i^{\mathbb{X}}\rangle = \epsilon_{ii} |\varphi_i^{\mathbb{X}}\rangle \tag{10}$$

It contains the Fock operator of the total system $\hat{\mathcal{F}}$ in which the occupied orbitals on $\mathbb{E}$ frozen. The additional anticommutator in (10) originates from the orthogonality condition between the active and the frozen parts

and the requirement, that the Fock operator should be Hermitian.[32] This extra term causes upward orbital energy shifts for frozen orbitals, where

$$\hat{\rho}_{\mathbb{E}} = \sum_{j \in [\mathbb{E}]} |\varphi_j^{\mathbb{E}}\rangle \langle \varphi_j^{\mathbb{E}}| \tag{11}$$

is the projector onto the frozen part $\mathbb{E}$.

The Huzinaga equation is given in matrix form as

$$\left(\mathbf{F}^{\mathbb{X}} - \mathbf{P}_{\mathbb{E}}^{\mathrm{H}} \right) \mathbf{C}^{\mathbb{X}} = \mathbf{S}^{\mathbb{X}} \mathbf{C}^{\mathbb{X}} \epsilon \tag{12}$$

with

$$\mathbf{P}_{\mathbb{E}}^{\mathrm{H}} = \frac{1}{2} \left[\mathbf{S}^{\mathbb{XE}} \mathbf{D}^{\mathbb{E}} \mathbf{F}^{\mathbb{EX}} + (\mathbf{S}^{\mathbb{XE}} \mathbf{D}^{\mathbb{E}} \mathbf{F}^{\mathbb{EX}})^{\dagger} \right] \tag{13}$$

where the superscripts of the matrices label the atomic basis sets, in which the operators are represented. A superscript $\mathbb{X}$ refers for instance to the AO basis in which the orbitals on $\mathbb{X}$ are represented. We will not use the full AO basis on the active part. Instead, a restricted basis set is used with basis functions of the atoms in the active part.

The frozen one particle density matrix $\mathbf{D}^{\mathbb{E}}$ contributes to (10) in several terms either directly or indirectly. The latter is due to the Fock matrix $\mathbf{F}$, which depends on the overall one particle density matrix $\mathbf{D} = \mathbf{D}^{\mathbb{X}} + \mathbf{D}^{\mathbb{E}}$ and thus on $\mathbf{D}^{\mathbb{E}}$. Since $\mathbf{D}^{\mathbb{E}}$ is frozen, the associated Coulomb J and exchange K contributions can be viewed as a modelpotential $\mathbf{V}^{\mathrm{MP}}$, which has to be calculated once.

$$\mathbf{V}^{\mathrm{MP}} = G[\mathbf{D}^{\mathbb{E}}] = J[\mathbf{D}^{\mathbb{E}}] - K[\mathbf{D}^{\mathbb{E}}] \tag{14}$$

The self-consistently obtained orbitals for $\mathbb{X}$ minimize the effective energy on $\mathbb{X}$ and are orthonormal in this modelpotential.[33]

2. Projector based embedding

The projector based embedding method[23,49] realizes the two concepts of minimizing the energy of the active part and forcing the orthogonality between active and frozen parts directly in a modified Fock equation.

$$\left(\hat{\mathcal{F}} + \mu \hat{\rho}_{\mathbb{E}} \right) |\varphi_i^{\mathbb{X}}\rangle = \epsilon_{ii} |\varphi_i^{\mathbb{X}}\rangle \tag{15}$$

Here, μ is an adjustable parameter and $\hat{\rho}_{\mathbb{E}}$ is given in (11). The additional term shifts the orbital energies of occupied LMOs from $\mathbb{E}$ by μ and drives the orthogonality. Equation (15) reads in matrix form as

$$\left(\mathbf{F}^{\mathbb{E}} + \mathbf{P}_{\mathbb{E}}^{\mathrm{P}} \right) \mathbf{C}^{\mathbb{E}} = \mathbf{S}^{\mathbb{E}} \mathbf{C}^{\mathbb{E}} \epsilon \tag{16}$$

with

$$\mathbf{P}_{\mathbb{E}}^{\mathrm{P}} = \mathbf{S}^{\mathbb{XE}} \mathbf{D}^{\mathbb{E}} \mathbf{S}^{\mathbb{XE} \dagger} \tag{17}$$

and thus contains the same modelpotential contributions (14) as the Huzinaga embedding. The two methods differ

in the level shift matrix. It has to be calculated only once in the simpler projection based method whereas it has to be calculated in each iteration of the SCF in the Huzinaga method since the shift depends on the Fock matrix. However, the shift operator in the projection based embedding does not guarantee orthogonality between the active and frozen parts for truncated basis sets. This causes errors in the energy in the context of HF-in-HF or WFT-in-HF embedding which can be minimized by restricting the projector orbitals close to the active site.[2] We will not be bothered by this fact, since we do not attempt to calculate the energy of the full system in an embedded picture. Instead, the correlation energy of the total system is calculated incrementally and the embedding is only used as a tool to access occupied and virtual spaces. There is thus no persistent active part.

3. Effect of basis set restriction

LMOs on $\mathbb{E}$ are kept without modification and used in (12) or (16). These orbitals are therefore orthonormal to each other. The orthonormality of active orbitals is also guaranteed by the Lagrange multipliers. The orthogonality between the embedded part $\mathbb{X}$ to the frozen part $\mathbb{E}$, however, could be restored when the full basis set is also used on the active part. Indeed, this orthogonality only requires that the basis set on $\mathbb{X}$ contains the basis set on $\mathbb{E}$. However, employing the AO basis of the total system for the active part would not reduce the computational requirements in the subsequent correlation calculation, since the cardinalities of the virtual spaces of $\mathbb{X}$ and the total system $\mathbb{S}$ would be equal $\left|\overline{[\mathbb{X}]}\right| = \left|\overline{[\mathbb{S}]}\right|$ in this case. The AO basis should thus be truncated for convenience. As mentioned before, we do not attempt to calculate the embedded system energies in which the loss of orthogonality would cause issues.[2]
We restrict the basis set on $\mathbb{X}$ to the atoms on $\mathbb{X}$. That is, the non-hydrogen atoms corresponding to the domains in $\mathbb{X}$ and the hydrogen atoms which have the shortest distance to these atoms. This restriction has a tremendous effect on the computational requirements of this method. The largest tuple $\mathbb{X}^{\max}$ in our incremental scheme will asymptotically be independent of system size, which is provided by the selection mechanism of these tuples presented in section II C. Consequently, the AO basis of a single embedding calculation becomes asymptotically independent of molecular size and in turn the size of the virtual spaces as well. Note, that the number of occupied orbitals used in a single correlation calculation is independent of molecular size in any incremental scheme with many-body expansion for the occupied space with a fixed incremental order. Therefore, the transformation of two particle integrals for the subsequent correlation calculation which scales as $N_{\mathrm{MO}} N_{\mathrm{AO}^4}$ will be reduced to a constant. The scaling of the correlation calculation itself also changes. Considering the scaling $N_o^2 N_v^4$ of the CCSD and $N_o^3 N_v^4$ of the CCSD(T) methods, these scal-

ings are both asymptotically constant with the presented basis set truncation.

4. Virtual space augmentation

The incremental expansion of the virtual space requires the possibility to correlate different virtual tuple spaces $\overline{[\mathbb{A}]}$ with occupied tuple spaces $[\mathbb{I}]$. We realize the generation of such occupied and virtual space pairs through embedding calculations with a concept which we call virtual space augmentation presented in figure 1. For this, the desired virtual tuple $\mathbb{A}$ has to be provided. An embedding calculation is then performed where the occupied space $[\underline{\mathbb{S} \setminus \mathbb{A}}]$ is frozen and the part $\mathbb{A}$ is actively treated. The so obtained virtual space $\overline{[\mathbb{A}]}$ is already the target space and does thus not need any further manipulation. The generated occupied space $[\underline{\mathbb{A}}]$ however has to be localized in order to restore the occupied tuple space $[\mathbb{I}]$. We use the so called template localization, which has been used in the incremental scheme with domain specific basis sets in a slightly different context.[14,16] The occupied orbitals in $[\underline{\mathbb{A}}]$ are unitarily transformed with a Jacobi sweep algorithm to maximize the localization function

$$\mathcal{L} = \sum_i \langle \varphi_i^{\mathbb{S},\mathrm{LMO}} | \varphi_i^{\mathbb{A}} \rangle^2 \qquad (18)$$

where $\varphi^{\mathbb{S},\mathrm{LMO}}$ denotes occupied LMOs of the full system calculation. The so obtained localized orbitals $\varphi^{\mathbb{A},\mathrm{LMO}}$ can than be mapped to the original occupied domains and thus the occupied space of interest $[\mathbb{I}]$ be isolated form $[\underline{\mathbb{A}}]$.
The outlined procedure overall yields orthonormal occupied LMO spaces for the tuple $\mathbb{I}$ and a larger virtual space for the tuple $\mathbb{A}$. The tuple $\mathbb{A}$ will be called augmentor of occupied tuple $\mathbb{I}$. The virtual space expansion involves several augmentors for a specific occupied tuple.
There are obviously alternatives to the outlined virtual space construction. Indeed, virtual space augmentation and also occupied space augmentation are built into the DEC,[12,34,40,93] CIM[24,41–45,59,60] and LNO[68,69] methods where the augmented occupied and virtual spaces are not obtained through embedding calculations. However, the decisive difference between our method and the above mentioned methods is that not a single augmented correlation space but several augmented spaces are used in individual correlation calculations and one has to cope with the peculiarities of the energy expansion with respect to these virtual spaces.

C. Incremental scheme for occupied and virtual space

In an incremental scheme with a single virtual space for each occupied domain tuple $\mathbb{I}$, the virtual space has to be sufficiently large in order to avoid errors and error propagation (section II A 2). As electron correlation
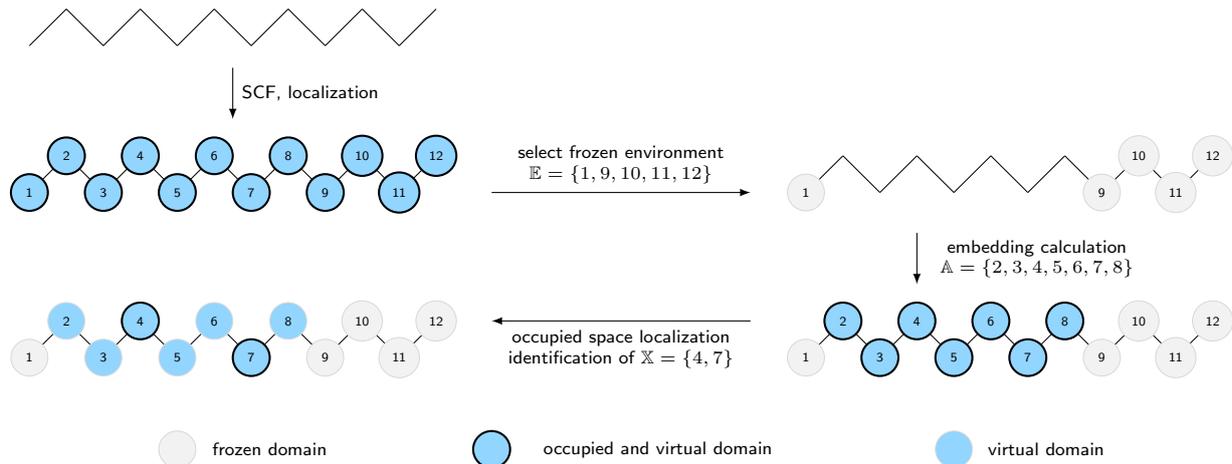
FIG. 1: General working procedure for embedding driven virtual space augmentation.

methods scale most expensively through the size of the virtual space, this represents the bottleneck for the incremental scheme with an expansion of the occupied space only. Therefore, the additional incremental expansion of the virtual space is attempted for each $\mathbb{I}$.

An analogous power set expansion as in (2)-(3) is not studied. It would require disjoint sets of local virtual orbitals which are orthogonal to each other, analogous to the occupied LMOs. An option would be localized virtual orbitals, which can be accessed in the meantime through special localization schemes[29–31,35,82] but are not considered in this work. However, even with virtual LMOs one would expect that the incremental expansion would converge at higher orders than for the occupied space. This assumption is based on a locality analysis of MP2 and CCSD amplitude equations which demonstrated that the virtual correlation space of a local occupied domain is generally more extended.[40] For this reason, we take care in our virtual space expansion that the virtual spaces are not too small in a single calculation. Other commonly used virtual spaces as orthogonalized PAO or OSV spaces lead to inconsistent virtual space expansions. This can be demonstrated in a simple example. Let $[\mathbb{A}_1]_{\text{OPAO}}$ and $[\mathbb{A}_1]_{\text{OPAO}}$ be two OPAO spaces which shall be considered for an occupied tuple $\mathbb{I}$. The virtual space expansion could be formulated as

$$\epsilon_{\mathbb{I}} = \epsilon_{\mathbb{I}}^{\mathbb{A}_1} + \epsilon_{\mathbb{I}}^{\mathbb{A}_2} - \epsilon_{\mathbb{I}}^{\mathbb{A}_1 \cap \mathbb{A}_2} \qquad (19)$$

where the doubly counted contributions are subtracted with the last term. Two problems become evident already within this simple case: (1) the intersection may be small, (2) the virtual space $[\mathbb{A}_1 \cap \mathbb{A}_2]_{\text{OPAO}}$ is not a proper subspace of the other spaces and thus inconsistent due to the orthogonalization.

From the above discussion, there are two main guidelines for the virtual space expansion

(1) minimum virtual space size

(2) avoid inconsistencies in virtual space

We have adopted these guidelines for the design and will proceed with the concepts of the obtained method. The virtual space expansion is not formulated for the individual correlation energies $\epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ which are used in (3) to calculate incremental energies contributions $\Delta \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ with respect to the occupied space. Instead we attempt to calculate $\Delta \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ directly by applying the occupied space expansion to each augmentor

$$\Delta_{\text{o}} \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} \quad \leftarrow \quad \Delta_{\text{o}} \epsilon_{\mathbb{I}}^{\mathbb{A}} = \epsilon_{\mathbb{I}}^{\mathbb{A}} - \sum_{\mathbb{J} \in \mathcal{P}(\mathbb{I}) \, \wedge \, |\mathbb{J}| < |\mathbb{I}|} \Delta \epsilon_{\mathbb{J}}^{\mathbb{A}} \qquad (20)$$

This avoids inconsistencies due the ambiguity of the augmentors for different occupied tuples. Conversely, the occupied space expansion in (20) is well defined as the occupied spaces of any $\mathbb{J}$ are proper subsets of $\mathbb{I}$. From this starting point, we expand the correlation energy contributions $\Delta \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})}$ in terms of a base augmentor $\mathbb{A}(\mathbb{I})$ and additional augmentors $\mathbb{A}(\mathbb{I} \cup \mathbb{Y})$ according to

$$\Delta_{\text{o}} \epsilon_{\mathbb{I}}^{\mathbb{L}(\mathbb{I})} = \Delta_{\text{o}} \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I})} + \sum_{\mathbb{Y}} \Delta_{\text{o}}^{\text{v}} \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I} \cup \mathbb{Y})} \qquad (21)$$

The indices o and v are given to emphasize which type of incremental expansion is involved in the individual term. Each augmentor is chosen to have at least a predefined minimum size L^{min}. The base augmentor contains the most relevant domains for the correlation with $\mathbb{I}$. The relevance is measured on the basis of the Fock matrix as presented in section II C 2. Relevant domains $Y_1, Y_2, \ldots$ which are not in the base augmentor are considered incrementally. At first order, only a single domain Y_k is used to form an augmentor $\mathbb{A}(\mathbb{X} \cup Y_k)$ and is correlated in a calculation. At second order, pairs of these domains ($|\mathbb{Y}| = 2$) are correlated simultaneously within the virtual spaces of the corresponding augmentors $\mathbb{A}(\mathbb{I} \cup \mathbb{Y})$ and the same logic applies to higher orders. However, the correlation of a domain tuple $\mathbb{Y}$ with more than one domain contains all simultaneous correlations of the domains contained in the power set $\mathcal{P}(\mathbb{Y})$. Thus, in order to

avoid multiple counting due to contributions from lower orders, all lower level contributions have to be subtracted which reads as

$$\Delta_o^v \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup\mathbb{Y})} = \Delta_o \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup\mathbb{Y})} - \sum_{\mathbb{W}\in\mathcal{P}(\mathbb{Y})\wedge|\mathbb{W}|>0} \Delta_o \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup\mathbb{Y})\setminus\mathbb{W}} \quad (22)$$

For the first order, that is $|\mathbb{Y}| = 1$ the expansion reads much simpler

$$\Delta_o^v \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup Y)} = \Delta_o \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup Y)} - \Delta_o \epsilon_{\mathbb{I}}^{\mathbb{A}(\mathbb{I}\cup Y)\setminus Y} \quad (23)$$

and is may favored to to prevent from too large augmentors since if $|\mathbb{A}(\mathbb{I} \cup \mathbb{Y}) \setminus \mathbb{Y}| = L^{\min}$ then $|\mathbb{A}(\mathbb{I}\cup\mathbb{Y})| = L^{\min} + |\mathbb{Y}|$.
We used embedding generated occupied and virtual spaces for the individual calculations. Although $[\mathbb{A}(\mathbb{I} \cup Y) \setminus Y]$ is not exactly a proper subsets of $[\mathbb{A}(\mathbb{I}\cup Y)]$, it can be approximately. There is obviously an effect due to the removal of basis functions on Y but the embedding calculations mainly restores the virtual space. This effect is smallest when we force Y by construction to be a terminal domain in the augmentor $\mathbb{A}(\mathbb{I} \cup Y)$. That is, no central atoms are removed in the active part of the embedding calculation. The construction of augmentors is presented in II C 3.

1. Reachable domains

The incremental expansion of the virtual space in (21) and (22) involves the correlation of each occupied tuple $\mathbb{I}$ with a set of virtual tuples, called augmentors. The aim is to find a set of important virtual domains, we called them reachables $\mathbb{R}(\mathbb{I})$, for the virtual correlation space and calculate the energy contribution of these without using all at once.
We propose to use the same Fock matrix $\mathbf{F}^{max}$ from (9) as it is used for the selection of the occupied tuples but with a weaker threshold on the virtual space. This is justified due to the fact that our embedding generated virtual spaces are local to the occupied space of the tuple due to the basis set restriction and we therefore can apply the locality argument for the occupied tuple selection equivalently for the virtual space.
The reachable domains for a single occupied domain I are defined as

$$\mathbb{R}(I) = \{ A : \mathbf{F}_{AI}^{\max} > F_{\text{thresh}}^R \} \quad (24)$$

where F_{thresh}^R is a predefined threshold. The reachables for an occupied domain tuple $\mathbb{I}$ are then taken as the intersection of the reachables of the individual domains.

$$\mathbb{R}(\mathbb{I}) = \bigcap_{I\in\mathbb{I}} \mathbb{R}(I) \quad (25)$$

The reason for an intersection rather than a unification is to keep the augmentors as small as possible without

compressing the virtual spaces too much. A virtual domain in the intersection is simultaneously important for all occupied domains in the tuple and therefore has to be considered.

2. Base augmentor

The base augmentor $\mathbb{A}(\mathbb{I}) \subset \mathbb{R}(\mathbb{I})$ contains the most relevant domains for an occupied tuple $\mathbb{I}$ up to a minimum augmentation size $L^{\min}$. The construction of a base augmentor is illustrated in figure 2. The first step is the formation of a minimum connection $\mathbb{C}(\mathbb{I})$ between the domains in $\mathbb{I}$. This can be motivated with a locality analysis of the MP2 and CCSD amplitude equations. It has been observed that not only the virtual space local to a specific occupied space couple the amplitudes. Instead, the virtual space which is proximate to the surrounding of an occupied space also contributes in the amplitude equations.[40] The correlation between the occupied spaces of a disconnected tuple $\mathbb{I}$ is therefore mediated likewise through its connecting virtual space. As we are using QM/QM embedding to generate the augmentors virtual spaces, we assume that with disconnected augmentors the virtual space would be described poorly around the "holes" in the active part and therefore exclude this option. The connection is obviously chosen to be minimal since a control over the virtual space size is desired.
Note, that the size of the minimum connection can be larger than $L^{\min}$. No further augmentation is done in this case and the largest size of augmentors at incremental order three is called $L^{\max}$. If the size of the minimum connection is smaller than $L^{\min}$, the augmentor is further augmented with reachable domains. The reachable domains A of the tuple $\mathbb{I}$ are sorted for this purpose according to a geometrical mean $\bar{F}^{\max}(A)$ of $\mathbf{F}^{\max}$ values and added to the minimum connection in descending order. The geometrical mean

$$\bar{F}^{\max}(A) = \left(\prod_{I\in\mathbb{I}} \mathbf{F}_{AI}^{\max} \right)^{1/|\mathbb{I}|} \quad (26)$$

provides a measure for the relevance of A to $\mathbb{I}$. It will only be large if A is important for all domains in $\mathbb{I}$ other than an arithmetic mean for instance.
The correlation of the occupied orbital space $[\mathbb{I}]$ with the base augmentor virtual space $[\mathbb{A}(\mathbb{I})]$ will represent a main part of the correlation with respect to the total correlation space, if the base is large enough. Since the aim is to reduce the virtual space of the most expensive calculation, the base can not be chosen too large. However, the correlated domains will also be described poorly if the base is small since further domains local to them are expected to improve the quality of the amplitudes in the CC calculation.[40] The choice of $L^{\min}$ could in principle be guided by a simple energy screening first incremental order in occupied space where $L^{\min}$ is successively enlarged and only a single (base) augmentor is used. $L^{\min}$
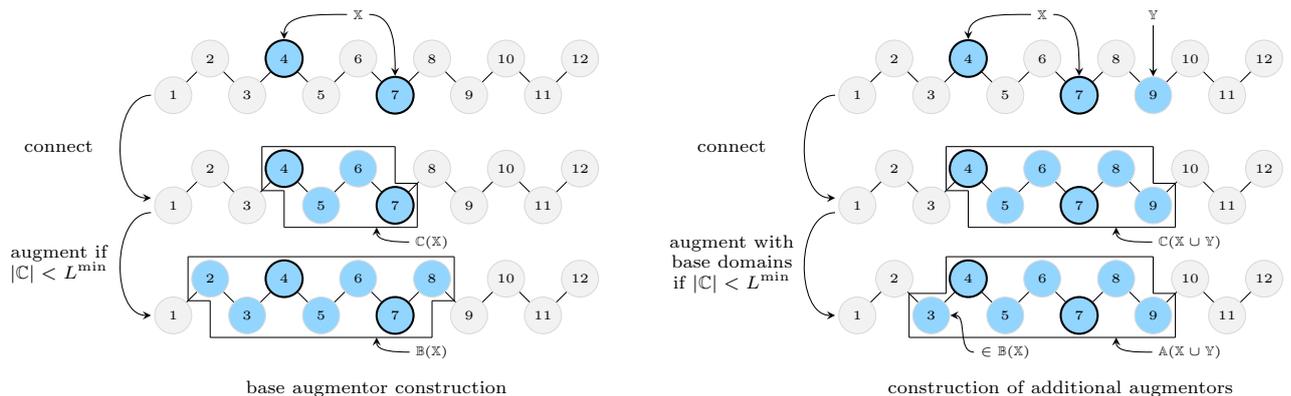
FIG. 2: A schematic picture of constructing augmentors for a given active domain tuple.

3. Additional augmentors

Assume that the virtual space of the reachables $\mathbb{R}(\mathbb{I})$ span the virtual space for an occupied tuple $\mathbb{I}$. This means that a correlation calculation with the orbital spaces $[\mathbb{I}]$ and $[\mathbb{R}(\mathbb{I})]$ would reproduce the energy calculated with the total virtual space. In contrast, a correlation with the base augmentor space $[\mathbb{A}(\mathbb{I})]$ does not include the needed interactions with the virtual domains $Y \in \mathbb{R}(\mathbb{I}) \setminus \mathbb{A}(\mathbb{I})$. Although these virtual domains are less relevant compared to the base, their contributions can not be neglected. The associated contributions are included in the virtual space expansion by assigning augmentors $\mathbb{A}(\mathbb{I} \cup Y)$ to each Y and correlating these augmentor spaces with $\mathbb{I}$. Higher order corrections can also be included for simultaneous correlations of $\mathbb{I}$ with pairs (Y_p, Y_q) or higher order tuples $\mathbb{Y}$ by assigning individual augmentors.

We focus in the present work on first-order corrections but would like to provide a possible and strict selection criterion for higher order tuples $\mathbb{Y}$. That is, each pair of reachable domain in $\mathbb{Y}$ should additionally be relevant for each other.

$$\mathbf{F}_{YY'}^{\max} > F_{\text{thresh}}^{Y} \quad \forall Y, Y' \in \mathbb{Y} \tag{27}$$

The construction scheme for any augmentor $\mathbb{A}(\mathbb{I} \cup \mathbb{Y})$ is similar to the base construction. It is also presented in figure 2. The first step is the formation of a minimum connection of $\mathbb{I} \cup \mathbb{Y}$. This set is further augmented with domains if it is smaller than $L^{\min}$. The augmentation is intentionally made with domains from the base. The reason for augmentation from the base is to avoid double counting. Suppose that the single contributions of the domains Y and Y' through additional augmentors

$\mathbb{A}(\mathbb{I} \cup Y)$ and $\mathbb{A}(\mathbb{I} \cup Y')$. There will be a double counting of the pair contributions (Y, Y') if $Y' \in \mathbb{A}(\mathbb{I} \cup Y)$ and $Y \in \mathbb{A}(\mathbb{I} \cup Y')$ which can be avoided with an augmentation from the base.

Another criterion for $\mathbb{Y}$ is that each domain has to be terminal. This can be seen from (22). The augmentors $\mathbb{A}(\mathbb{I} \cup \mathbb{Y}) \setminus \mathbb{W}$ are used in the expansion, where $\mathbb{W}$ is an element of the power set of $\mathbb{Y}$. There will be disconnected $\mathbb{A}(\mathbb{I} \cup \mathbb{Y}) \setminus \mathbb{W}$ if any Y is not terminal. The construction of connected augmentors with specific terminal domains is implemented with standard techniques of graph theory, where the non-hydrogen atomic domain centers are treated as nodes and bonds between these domains are represented as edges.

4. Ring subsystems

We discussed above that augmentors should be connected for a proper description of the corresponding virtual spaces generated by embedding methods. In general, the virtual space for a domain should not differ qualitatively from one augmentor to another to guarantee consistency among the augmentors and the virtual spaces should also not differ qualitatively with respect to their contribution to the full system. Although this requirement holds for all domains, ring systems represent a special case. Ring structures as benzene have large correlation effects. Assuming that only a subset of the ring domains is included in an augmentor, the QM/QM embedding will not reproduce the correct correlation space for the corresponding subspace of the ring. Obviously, this effect is expected for any subsystem of the total system. Nevertheless, ring systems lead to more inconsistencies and errors, as other subunits. The domains belonging to a ring are therefore treated as a unit. In the construction of the augmentors, either the full ring is used for the correlation space or none of the domains in the ring.

Ring subsystems are detected through the connectivity of the domains. We shall note here that a virtual space expansion according to (21) and (22) is not possible for fused ring systems since either double counting or inconsistencies would occur. A single augmentor approach for occupied domain tuples on the fused ring parts of the system would be appropriate.

III. APPLICATIONS

A. Computational details

The new incremental scheme is implemented in the *Quantum Objects Library*, an in-house quantum chemistry software library. The existing code generated integral evaluation and density contraction[28] has been used for the implementation of the embedding schemes. The basis set cc-pVDZ[7,86] is used through this paper. Additionally, the pseudopotential ECP46MWB and the corresponding basis ECP46MWB-II[6] is used for La(III). All correlation calculations are CCSD calculations, which are performed with the in-house coupled cluster program.[10,26]

B. Fock matrix based selection of occupied tuples

In order to exploit the local nature of electron correlation in the framework of the incremental scheme, a selection on the occupied domain tuples is required. It is also mandatory for any linear scaling incremental scheme since the number of tuples at incremental order k scales as N_D^k. A widely used selection criterion is the spatial domain distance and it has been shown that this approach yields accurate results.[17–20,89,90] In this work, we are focusing on the incremental expansion of the occupied and virtual spaces using off-diagonal Fock matrix elements in the LMO basis as presented in II A 3 and II C 1. These Fock matrix elements represent a measure for the importance of occupied domains in the correlation space.[40] The selection scheme is tested first within the ordinary incremental scheme in which only the occupied space is expanded before testing the selection criterion for the virtual space. In order to avoid any other error sources, other than the error of the incremental expansion itself, we used the full canonical virtual space for the correlation calculations with the occupied domain tuples. These occupied tuples are selected fully automatically for the studied chemical systems according to (9). This procedure requires the calculation of the $\mathbf{F}^{\max}$ values in (8) and the provision of the threshold $F_{\text{thresh}}^{\text{o}}$. In figure 3 we present the information about all tuple energy contributions according to (3) of six different systems plotted against $\mathbf{F}^{\max}$ values of the occupied tuples. It is demonstrated that the presented selection according to Fock matrix elements is legitimate in the framework

of the incremental scheme, since the tuple energy contributions decrease with the $\mathbf{F}^{\max}$ values for all studied systems. More precisely, the energy contributions of occupied domain pairs (I, J) decrease exponentially with decreasing maximum off-diagonal Fock matrix elements $\mathbf{F}_{IJ}^{\max}$ which are defined according to (9). Whereas $\mathbf{F}^{\max} \geq 10^{-1}\ E_h$ corresponds to absolute energy contributions larger than $10^{-4}\ E_h$, the energy contributions are significantly smaller for $F^{\max} \leq 10^{-3}\ E_h$ and below $10^{-7}\ E_h$. Analogously, there is an exponential decay for 3-tuples. Since there are $\mathbf{F}_{IJ}^{\max}$ values for each pair in a 3-tuple, we chose to plot the energy contributions against the smallest of the $\mathbf{F}_{IJ}^{\max}$ values, which is denoted by $\min|\mathbf{F}^{\max}|$ and represents the weakest pair in the tuple. This is consistent with the tuple selection criterion as only tuples are included in which all pairs have larger $\mathbf{F}^{\max}$ values than a predefined threshold. The energy contributions of 3-tuples are smaller than the contributions from pairs at the same Fock matrix value in any of the investigated molecular systems. In contrast, there is a small overlap region of pair and 3-tuple energy contributions (figure 3), if different systems are compared, since the magnitude of off-diagonal Fock matrix elements depends on the atoms in the molecule and the geometry.[63] Nevertheless, the largest energy contributions for 3-tuples with $\min|\mathbf{F}^{\max}| \leq 0.01\ E_h$ is smaller than the pair contributions with $\mathbf{F}^{\max} \leq 0.005\ E_h$. Therefore, the 3-tuple value for the threshold can be chosen smaller than the corresponding value for the pairs.

Although the minimum of $\mathbf{F}^{\max}$ is a reasonable choice, there is still potential to improve this criterion to narrow the range of energy contributions at a fixed Fock-matrix selection value. Furthermore, instead of a universal Fock matrix threshold, it could be adapted to the atoms and geometries involved, which is beyond the scope of this work.
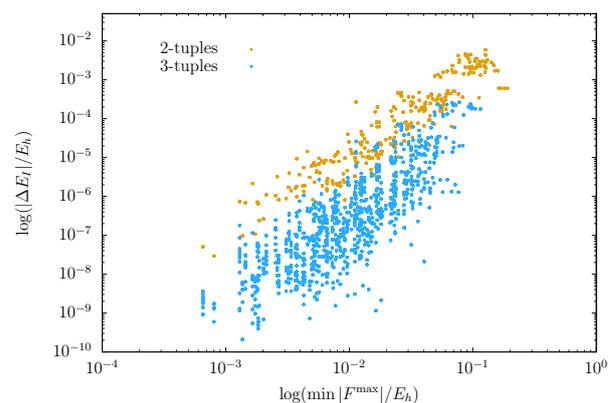
FIG. 3: Correlation between occupied space incremental energy contribution $\Delta E_{\mathbb{I}}$ in second and third order and the minimum value of $\mathbf{F}^{\max}$ in the corresponding tuple $\mathbb{I}$. Systems: $(H_2O)_{11}$, $[La(H_2O)_8]^{3+}$, $C_{12}H_{26}$, Met, $(Gly)_3$, $AscH^-$. Basis: cc-pVDZ.

TABLE I: Correlation energy errors (kJ/mol) of the screened occupied space incremental scheme with respect to the canonical CCSD $\Delta \tilde{E}_{\text{corr}}$ with three different threshold $(F_{\text{thresh}}^{\text{o},2}, F_{\text{thresh}}^{\text{o},3})$ combinations: (A) (0 E_h,0 E_h) , (B) (0.005 E_h,0.005 E_h), (C) (0.005 E_h,0.01 E_h). Incremental order: 3.

Molecule	$\Delta \tilde{E}_{\text{corr}}^{(A)}$	$\Delta \tilde{E}_{\text{corr}}^{(B)}$	$\Delta \tilde{E}_{\text{corr}}^{(C)}$
$(H_2O)_{11}$	-0.03	-1.16	-1.41
$[La(H_2O)_8]^{3+}$	0.16	0.16	0.16
$C_{12}H_{26}$	0.04	-0.11	-0.16
Met	-0.38	-0.69	-0.12
$(Gly)_3$	0.53	0.21	0.30
$AscH^-$	-2.29	-2.29	-2.16

Nevertheless, the presented Fock-matrix based selection on the occupied space yields accurate results in combination with the ordinary incremental scheme. Exemplary correlation energy errors with thresholds $F_{\text{thresh}}^{\text{o},2} = 0.005$ E_h at second order, $F_{\text{thresh}}^{\text{o},3} = 0.005$ E_h and $F_{\text{thresh}}^{\text{o},3} = 0.01$ E_h at third order are shown in table I. The absolute errors are smaller than 2.3 kJ/mol in all cases with respect to the canonical CCSD valence correlation energies. Furthermore, the truncation with the selected thresholds does not introduce significant additional errors compared to the full incremental expansion. A more refined truncation at incrementally order three is also validated. The selected thresholds are motivated by the analysis of figure 3 and the values $F_{\text{thresh}}^{\text{o},2} = 0.005$ E_h and $F_{\text{thresh}}^{\text{o},3} = 0.01$ E_h seem reasonable. Note, that the truncation at incremental order three is more important for saving computational resources. Therefore, $F_{\text{thresh}}^{\text{o},2}$ could be chosen more generously although this leads to pairs with larger distances and thus larger augmentor spaces, which is not desired.

C. Incremental expansion in occupied space: single augmentor and F^R threshold

A single augmentor approach is closely related to the ordinary incremental schemes with domain specific basis sets.[19,89] Only a single augmentor is assigned to each occupied tuple in this version without an incremental expansion for the virtual space. As in the ordinary incremental schemes, the individual correlation calculations are independent of each other and can therefore be performed in an embarrassingly parallel manner. The single augmentor strategy will not be followed as our main goal is to demonstrate that a virtual space expansion is possible. However, we investigate the virtual space sizes which would be needed in such a single augmentor approach to include all reachable domains according to (25) and provide thresholds F_{thresh}^R for the virtual space truncation. We observe that for $F_{\text{thresh}}^R = 0.001$ E_h most of the occupied space tuples require the full virtual space of the

molecule for the test systems in table I. This holds even if the virtual space thresholds are set equal to the occupied space thresholds, which is the upper limit. That is, $F_{\text{thresh}}^{R,2} = 0.005$ E_h and $F_{\text{thresh}}^{R,3} = 0.01$ E_h. Nevertheless, for larger systems as for example $C_{18}H_{38}$, the virtual spaces of each tuple become smaller than the full space. The distribution of the virtual domain sizes for the chain system $C_{18}H_{38}$ are presented in figure 4. The largest augmentor size is still 13 for single occupied domains and 12 in the higher incremental orders with these sharp thresholds. For $F_{\text{thresh}}^R = 0.001$ E_h the histogram is shifted significantly to the right and the full virtual space would be needed for four domains, six pairs and two 3-tuples. However, the CCSD correlation energy obtained with the incremental scheme using the former thresholds deviates from the full canonical CCSD calculation by only 0.84 kJ/mol. In the following discussion for the virtual space expansion, we assume that a generous second order threshold $F_{\text{thresh}}^{R,2} = 0.001$ will be sufficient and that a reduced threshold of $F_{\text{thresh}}^{R,3} = 0.01$ is a convenient compromise between accuracy and feasibility.

D. Incremental expansion in occupied and virtual space: energies and base size dependence

The new incremental expansion uses a base augmentor and additional augmentors through which missing virtual space contributions are included. The minimum size L^{min} for the base augmentor is not known $a\ priori$. It should obviously be smaller than the number of all reachables. In figure 5, the correlation energies are presented which are obtained with either a single base augmentor with size L^{min} or with the additional incremental expansion on top of this. The two methods are denoted by inc-occ-base and inc-occ-virt, respectively. If L^{min} is systematically increased, inc-occ-base errors monotonically converge towards the ordinary incremental scheme CCSD
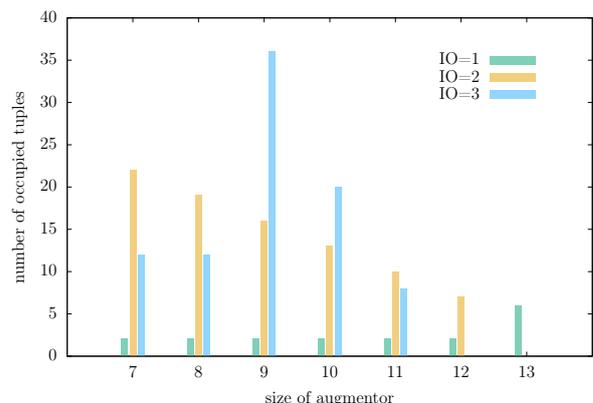
FIG. 4: Distribution of single augmentor sizes for $C_{18}H_{38}$ with virtual space thresholds $F_{\text{thresh}}^{R,2} = 0.005$ E_h and $F_{\text{thresh}}^{R,3} = 0.01$ E_h. Basis: cc-pVDZ.

energies. This represents a confirmation for the chosen order (II C 2) in which reachables are added to an augmentor. However, the errors can be very large for smaller values of $L^{\min}$ and convergence quite slow depending on how strongly the system is correlated. AscH$^-$ is a prominent example, since it contains a ring structure. The virtual space expansion also fails for this system with small $L^{\min}$ values but converges faster for larger values. That the ring system is treated as a unit causes the presence of very small augmentors and thus inconsistencies in the expansion. This originates from (22). If the ring is eliminated from an augmentor, the remainder will be small if $L^{\min}$ is also small. Systems with rings require therefore larger $L^{\min}$ values in general.

However, the error of the new incremental scheme drops much faster than the respective single augmentor errors and the error is less than 5 kJ/mol with $L^{\min} \geq 7$ for most of the systems. Furthermore, the errors of the inc-occ-virt method are significantly smaller than those of the inc-occ-base. These results constitute evidence for the proposed virtual space expansion. That is, virtual space contributions can be taken into account through additional augmentors in combination with EGVs. Nevertheless, a direct comparison of the energies would not be fair since the additional augmentors can be larger in size than $L^{\min}$ if the minimum connection to a reachable domain is longer than $L^{\min}$. The energies shall therefore be compared with this discrepancy in mind. For the water cluster and the La^{3+} complex, $L^{\min} = L^{\max}$ holds and the energies can be compared directly. Whereas the inc-occ-virt approach yields accurate results for $L^{\min} = 6$, the errors of the single augmentor with the same size are significantly higher. A single augmentor has to include almost the full virtual space in order to yield similar results. For other systems, additional augmentors have $L^{\max} \geq 8$ and do not represent significant reductions. This is due to the quite generously chosen truncation parameters $F_{\text{thresh}}^{R,2} = 0.001 \ E_h$ and $F_{\text{thresh}}^{R,3} = 0.005 \ E_h$ and the fact that these systems are still not large enough to see the full potential of the method. Larger systems are discussed below. Before moving on to those discussions, it is shown how $L^{\min}$ can be selected systematically. As figure 5 shows, the base size can be too small in the inc-occ-virt method in some cases like Gly$_3$ or AscH$^-$. However, a proper minimum size can be found by systematically increasing $L^{\min}$ in the inc-occ-base method and selecting $L^{\min}$ as the smallest value L for which the energy difference is smaller than a given threshold ϵ_{thresh}

$$E(L^{\min}) - E(L^{\min} - 1) < \epsilon_{\text{thresh}} \qquad (28)$$

A threshold can be applied to the incremental energy at any order since the energy converges at each individual order with increasing $L^{\min}$. A reasonable choice is a threshold for the first incremental order to save computational resources in this preparation step. At third order for example a threshold of 10 kJ/mol provides that all inc-occ-virt energies are within chemical accuracy for the systems in figure 5. Exemplary results obtained with
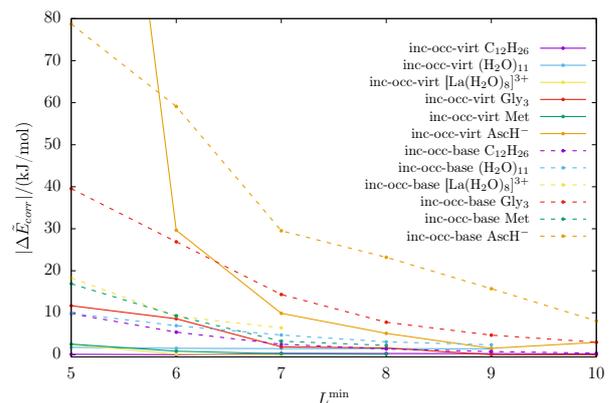
FIG. 5: Absolute correlation energy errors $\Delta \tilde{E}_{\text{corr}}$ with respect to canonical CCSD of a single base augmentor approach (inc-occ-base) and a virtual space expansion approach (inc-occ-virt). Thresholds: $F_{\text{thresh}}^{\text{o},2} = 0.005$ E_h, $F_{\text{thresh}}^{\text{o},3} = 0.01 \ E_h$, $F_{\text{thresh}}^{R,2} = 0.001 \ E_h$ and $F_{\text{thresh}}^{R,3} = 0.005 \ E_h$. Basis: cc-pVDZ.

this energy based $L^{\min}$ selection and more loosened F^R thresholds are shown in table II which also contains the $L^{\max}$ information.

The correlation energy errors of the new inc-occ-virt approach are compared with a single augmentor approach in which the base size coincides with the $L^{\max}$ of the correspond inc-occ-virt calculation in order to provide a fair comparison. The $L^{\max}$ values are in general larger than $L^{\min}$ except for cluster like geometries where the distance threshold for connectivity is not hit. As table II shows, the errors are in general significantly smaller with the inc-occ-virt approach. That is, a virtual space expansion with the same maximum cost in the most expensive calculation as in a single virtual space calculation yields more accurate results. Even for these small systems it

TABLE II: Correlation energy errors of the base contribution $\Delta \tilde{E}_{\text{corr}}^{\text{base}}$ and total errors including additional contributions $\Delta \tilde{E}_{\text{corr}}$ in kJ/mol with respect to canonical CCSD. The number of domains (N_D), minimum base size $L^{\min}$ and largest augmentor sizes $L^{\max}$ are given. Basis: cc-pVDZ; projective embedding; Thresholds: $F_{\text{thresh}}^{\text{o},2} = 0.005 \ E_h$, $F_{\text{thresh}}^{\text{o},3} = 0.01 \ E_h$, $F_{\text{thresh}}^{R,2} = 0.005 \ E_h$ and $F_{\text{thresh}}^{R,2} = 0.01 \ E_h$.

Molecule	N_D	$L^{\min}$	$L^{\max}$	$\Delta \tilde{E}_{\text{corr}}^{\text{base}}$	$\Delta \tilde{E}_{\text{corr}}$
Met	9	6	8	-2.24	-1.09
[La(H$_2$O)$_8$]$^{3+}$	9	6	6	-9.20	-0.17
(H$_2$O)$_{11}$	11	5	5	-9.89	-1.79
AscH$^-$	12	10	11	-4.42	-2.83
(Gly)$_3$	13	8	11	-1.30	-2.44
C$_{12}$H$_{26}$	12	5	7	-2.51	-0.29
C$_{18}$H$_{38}$	18	5	7	-5.99	-1.29

(a) $(H_2O)_{11}$ summed additional contributions

(b) $(H_2O)_{11}$ mean additional contribution

(c) $C_{12}H_{26}$ summed additional contributions

(d) $C_{12}H_{26}$ mean additional contribution
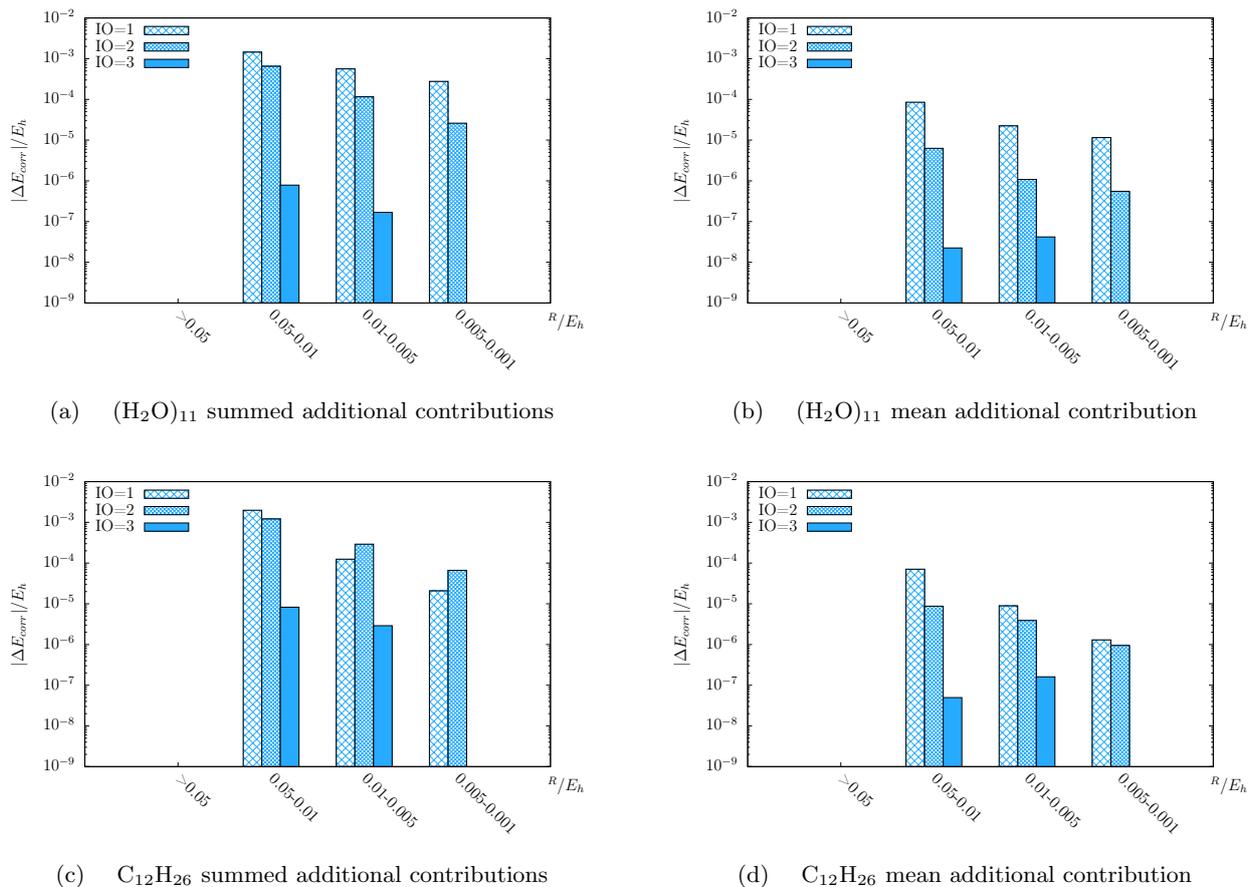
FIG. 6: Magnitude of incremental correlation energy contributions for the virtual space in addition to base contribution in relation to the off-diagonal Fock matrix value F^R of the contributing domains.

is possible to save computational effort without giving up chemical accuracy. The errors of the single augmentor approach are at least twice as for the inc-occ-virt method.

E. Incremental expansion in occupied and virtual space: F^R dependence of energy contributions

The incremental expansion of the virtual space, as presented in section II C, yields accurate results for the threshold $F_{\text{thresh}}^{R,2} = 0.001\ E_h$ and $F_{\text{thresh}}^{R,3} = 0.005\ E_h$, which is demonstrated in the previous sections. In order to be useful as a threshold, the absolute energy contributions have to shrink in magnitude with more narrow thresholds. This is also one of the underlying assumptions of the incremental virtual space expansion.
As can be seen from figure 6 that this assumption holds true in the way it is implemented here. The summed additional contributions, corresponding to the sum over $\mathbb{Y}$ in (21), are provided as well as the mean contribution per additional augmentor. Both systematically decrease with smaller off-diagonal Fock-matrix values F^R for the

two shown systems $(H_2O)_{11}$ and $C_{12}H_{26}$. There is an exception in this trend at incremental order three in the mean contributions, where $0.001\ E_h \leq F^R \leq 0.05\ E_h$ yields slightly larger contributions than $0.01\ E_h \leq F^R \leq 0.05\ E_h$. However, these contributions are at least one order in magnitude smaller than those at incremental order one and two. In general, the mean energy contributions are smaller in higher incremental orders for a given F^R range. This legitimates a sharper truncation at higher orders. Indeed, for the presented systems, no additional augmentors would be needed at incremental order three since the summed contributions are smaller than 10^{-5} E_h. Contrary, additional augmentors are mandatory at lower incremental orders. The contributions lie in the order of magnitude $10^{-3}\ E_h$ for $0.01\ E_h \leq F^R \leq 0.05\ E_h$ and still about $10^{-4}\ E_h$ for $0.001\ E_h \leq F^R \leq 0.005\ E_h$. Note, that smaller F^R values usually correspond to more spatially distant domains. Augmentors will also tend to be larger and lead to large correlation spaces in the individual calculations. Such calculations at incremental order three are the most expensive ones. Larger thresholds should therefore be used at incremental order three, since the most expensive calculations fortunately yield

negligible contributions. Another aspect which underpins a sharper truncation at higher orders are numerical issues which can occur in an expansion with very small contributions. We have used a CCSD energy threshold of 10^{-10} E_h for the presented calculations.

F. Incremental expansion in occupied and virtual space: recommended parameters

The investigations of the previous sections demonstrate that the off-diagonal Fock matrix elements are suited for a selection of the occupied domain tuples and the corresponding domains for the virtual correlation space. The results regarding the minimum base size and thresholds can be summarized as follows.

- the thresholds for the occupied space $F_{\text{thresh}}^{\text{o},2} = 0.005$ E_h and $F_{\text{thresh}}^{\text{o},3} = 0.01$ E_h at incremental orders two and three, respectively, yield accurate results.

- the thresholds for the virtual space $F_{\text{thresh}}^{R,2} = 0.001$ E_h and $F_{\text{thresh}}^{R,3} = 0.01$ E_h at incremental orders two and three, respectively, yield accurate results.

- the minimum base size $L^{\min}$ can be determined systematically with an energy criterion applied to a single augmentor incremental scheme.

Since all steps, such as occupied space partitioning and occupied and virtual space selection, are automated, the presented incremental scheme method can in principle be used as a black-box application with the recommended settings for the type of systems presented in this work. The applicability to other types of systems has to be tested in advance. It is also possible to implement an automated adjustment of the virtual space parameters F^R, and hence the correlation contributions to be included, based on the magnitude of the previously calculated correlation energy contributions with higher F^R values. It should be noted at this point, that we have exclusively used the cc-pVDZ basis throughout this paper and that the Fock matrix thresholds may be basis set depended. However, initial tests with smaller and larger basis sets indicate that the recommended parameters seem to yield reasonable results. The localization of the orbitals may contribute to the transferability.

G. Incremental expansion in occupied and virtual space: scaling with system size

The number of occupied tuples scales, without any truncation, with N_D^2 and N_D^3 for incremental orders two and three, respectively. Furthermore, the virtual space of the total system becomes larger with increasing system size and a CCSD calculation for example scales as N_v^4 for an incremental calculation since N_o is independent of
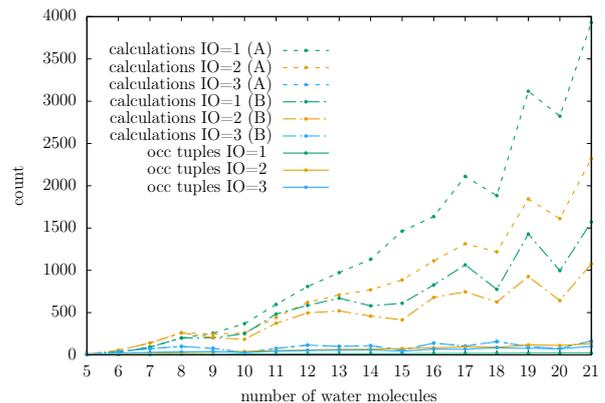
FIG. 7: Scaling of the number of occupied tuples up to incremental order three and the number of calculations with augmentors of size five for increasing water cluster $(H_2O)_x$ sizes x. Thresholds: $F_{\text{thresh}}^{\text{o},2} = 0.005$ E_h and $F_{\text{thresh}}^{\text{o},3} = 0.01$ E_h, (A) $F_{\text{thresh}}^{R,2} = 0.001$ E_h, (B) $F_{\text{thresh}}^{R,2} = 0.005$ E_h and $F_{\text{thresh}}^{\text{o},3} = 0.01$ E_h. Basis: cc-pVDZ

molecular size due to the fixed incremental order. Both of these issues have to be addressed by the new incremental scheme approach in order to achieve an overall linear scaling. That is:

(a) the scaling of the number of tuples has to be reduced to linear in N_D.

(b) the computational demand for the correlation with the virtual space should be independent of the system size for any occupied tuple beyond a critical system size.

The former criterion is fulfilled due to the implementation of the threshold $F_{\text{thresh}}^{\text{o}}$ as described in section II A 3. A more strict threshold on incremental order three, compared to second order, leads to a similar linear scaling in both orders as demonstrated in figure 7 for three dimensional TIP4P[36,84] optimized water cluster geometries. The absolute number of tuples obviously depends on the topology of the molecular system. However, for a given threshold $F_{\text{thresh}}^{\text{o},2}$, the possible pairs for an exemplary domain lie in a "pair shell" around this domain. The linear scaling is guaranteed for the number of pairs since the size of this shell is asymptotically independent of system size for large systems. This holds also true for the third incremental order since only 3-tuples are formed with domains which are in the same pair shell, even if $F_{\text{thresh}}^{\text{o},3} = F_{\text{thresh}}^{\text{o},2}$.

Criterion (b) is fulfilled due to the use of the threshold F_{thresh}^R as described in section II C 1. As discussed above, the threshold defines a shell around each domain. The sizes of these shells are independent of system size although they can be significantly larger for smaller thresholds. However, using a single augmentor for each tuple which contains all reachable domains in a single calculation would yield an asymptotically linear

scaling method. Although, the number of reachables is lowered using the intersection of the reachables of the individual domains instead of the unification in (25), the single augmentor can be very large. The size depends on the topology of the system and the strength of the correlation between the sites of the system. A large virtual space pushes the limits of computational resources rapidly due to the high order scaling of correlation methods. The incremental expansion of the virtual space comes in at this point. The advantages are, that the computational demand of the most expensive calculation remains independent of system size and it is overall significantly lower than the single augmentor virtual space. This is demonstrated in an example for the system $(H_2O)_x$. For the largest system $(H_2O)_{21}$, there are still tuples which have all domains as reachables with $F_{\text{thresh}}^{R,2} = 0.001\ E_h$ and $F_{\text{thresh}}^{R,3} = 0.01\ E_h$ and the maximum number of reachables is 19 with $F_{\text{thresh}}^{R,2} = 0.005\ E_h$. In our incremental calculation we set $L^{\min} = 5$ which is in this case also equal to $L^{\max}$. That is, the most expensive calculation for the system $(H_2O)_{21}$ is equivalent to a calculation with the occupied space of three water molecules and a virtual space of five water molecules. The computational demand of the most expensive CCSD calculation is reduced to about $(5/21)^4 \approx 0.3\ \%$ of the single augmentor calculation and the transformation from the AO basis to the MO is even more reduced due to the fifth power scaling.

In order to demonstrate the overall linear scaling of the method, the increase in the number of needed calculations with the virtual space of five water molecules at a time is shown in figure 7. The linear behaviour starts earlier with the more loosened threshold on the second incremental order but it is present for both threshold combinations at all incremental orders. Much less calculations are needed at the more expensive incremental order three. That the number at the lower incremental orders is much higher is rooted to the fact that an occupied space expansion is made for each augmentor according to (20).

As described above, the number of incremental calculations scales linearly. Since no communication is required between the individual calculations and all virtual spaces are generated independently for each occupied tuple, it is possible to perform each of these correlation calculations independently as in the ordinary incremental schemes. The new approach is therefore also embarrassingly parallel. Assuming that each calculation runs on a single computational node and there are at least as many nodes as there are calculations, then it would be possible to run all calculations in parallel and the calculation would take as long as the slowest individual calculation. For the example $(H_2O)_{21}$ in figure 7 at least 3000 nodes would be required to calculate all augmentors with size five at once.

IV. SUMMARY AND CONCLUSIONS

We have presented a linear scaling incremental scheme method to calculate electron correlation energies. The new concept is based on the idea that correlation energy contributions of the virtual space can be calculated incrementally for a given local occupied space, which was shown to hold true. The incremental expansion includes a base contribution with the most relevant domains for the correlation space and additional contributions upon this base contribution. Large correlation spaces are expanded in terms of smaller correlation spaces and reduce the maximal computational requirements significantly. We have shown that a single, relatively small, correlation space for each tuple (base augmentor) is not accurate enough to reconstruct the total correlation energy. However, taking additional contributions into account with an equivalent computational demand yields accurate results of chemical accuracy or even better. In order to achieve similar correlation energies, a single calculation with a significantly larger correlation space has to be performed.

The selection of the occupied and virtual domains is based on off-diagonal Fock matrix elements in the LMO basis. We could demonstrate that this procedure is suited on both spaces, since the correlation energy contributions for the correlation of occupied with occupied domains and occupied with virtual domains decreases in magnitude with decreasing off-diagonal Fock matrix elements. The presented selection of occupied tuples and relevant domains for the virtual correlation space, based on thresholds, leads to a linear scaling of the number of occupied tuples and a constant computational demand for the correlation calculations of these with respect to the system size. Thus, an overall asymptotic linear scaling is guaranteed for large systems.

We provided a set of thresholds based on the investigated systems, which yield accurate results in all cases. The minimum base size can be obtained through small calculations with single augmentors until an auxiliary predefined energy threshold is reached. It is also possible to adjust the virtual space thresholds on the fly by an evaluation of the magnitudes of the previously calculated contributions.

This flexibility is provided due to the independence of the individual calculations. The presented method is embarrassingly parallel since each correlation calculation is performed individually and without any exchange of information. It is hence possible to calculate the correlation energy of a large system with the wall time of the largest individual calculation if a sufficiently large number of computational nodes is accessible.

[1] AHLRICHS, R., DRIESSLER, F., LISCHKA, H., STAEMMLER, V., AND KUTZELNIGG, W. Pno-ci (pair natural orbital configuration interaction) and cepa-pno (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. ii. the molecules beh2, bh, bh3, ch4, ch3–, nh3 (planar and pyrami-

dal), h2o, oh3+, hf and the ne atom. *J. Chem. Phys. 62* (1975), 1235.

[2] BARNES, T., GOODPASTER, J., MANBY, F., AND MILLER, T. Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys. 139* (2013), 024103.

[3] BOYS, S. F., AND FOSTER, J. M. Canonical configurational interaction procedure. *Revs. Modern. Phys. 32* (1960), 300.

[4] DE LARA-CASTELLS, M. P., BARTOLOMEI, M., MITRUSHCHENKOV, A. O., AND STOLL, H. Transferability and accuracy by combining dispersionless density functional and incremental post-hartree-fock theories: Noble gases adsorption on coronene/graphene/graphite surfaces. *J. Chem. Phys. 143* (2015), 194701.

[5] DE LARA-CASTELLS, M. P., STOLL, H., AND MITRUSHCHENKOV, A. O. Assessing the performance of dispersionless and dispersion-accounting methods: Helium interaction with cluster models of the tio2(110) surface. *J. Phys. Chem. A 118* (2014), 6367.

[6] DOLG, M., STOLL, H., SAVIN, A., AND PREUSS, H. Energy-adjusted pseudopotentials for the rare earth elements. *Theor. Chim. Acta 75* (1989), 173.

[7] DUNNING, THOM H., J. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys. 90*, 2 (1989), 1007–1023.

[8] EDMISTON, C., AND KRAUSS, M. Configuration-interaction calculation of h3 and h2. *J. Chem. Phys. 42* (1965), 1119.

[9] EDMISTON, C., AND RUEDENBERG, K. Localized atomic and molecular orbitals. *Rev. Mod. Phys. 35* (1963), 457.

[10] ENGELS-PUTZKA, A., AND HANRATH, M. A fully simultaneously optimizing genetic approach to the highly excited coupled-cluster factorization problem. *J. Chem. Phys. 134* (2011), 124106.

[11] ERIKSEN, J. J., LIPPARINI, F., AND GAUSS, J. Virtual orbital many-body expansions: A possible route towards the full configuration interaction limit. *J. Phys. Chem. Letters 8* (2017), 4633–4639.

[12] ETTENHUBER, P., BAUDIN, P., KJÆRGAARD, T., JØRGENSEN, P., AND KRISTENSEN, K. Orbital spaces in the divide-expand-consolidate coupled cluster method. *J. Chem. Phys. 144* (2016), 164116.

[13] FEDOROV, D. G., AND KITAURA, K. Coupled-cluster theory based upon the fragment molecular-orbital method. *J. Chem. Phys. 123* (2005), 134103.

[14] FIEDLER, B., HIMMEL, D., KROSSING, I., AND FRIEDRICH, J. More stable template localization for an incremental focal-point approach—implementation and application to the intramolecular decomposition of tris-perfluoro-tert-butoxyalane. *J. Chem. Theory Comput. 14* (2018), 557–571.

[15] FIEDLER, B., SCHMITZ, G., HAETTIG, C., AND FRIEDRICH, J. Combining accuracy and efficiency: An incremental focal-point method based on pair natural orbitals. *J. Chem. Theory Comput. 13* (2017), 6023–6042.

[16] FRIEDRICH, J. Localized orbitals for incremental evaluations of the correlation energy within the domain-specific basis set approach. *J. Chem. Theory Comput. 6* (2010), 1834–1842.

[17] FRIEDRICH, J., AND DOLG, M. Implementation and performance of a domain-specific basis set incremental approach for correlation energies: Applications to hydrocarbons and a glycine oligomer. *J. Chem. Phys. 129* (2008), 244105.

[18] FRIEDRICH, J., AND DOLG, M. Fully automated incremental evaluation of mp2 and ccsd(t) energies: Application to water clusters. *J. Chem. Theory Comput. 5* (2009), 287–294.

[19] FRIEDRICH, J., HANRATH, M., AND DOLG, M. Energy screening for the incremental scheme: Application to intermolecular interactions. *J. Phys. Chem. A 111* (2007), 9830.

[20] FRIEDRICH, J., HANRATH, M., AND DOLG, M. Fully automated implementation of the incremental scheme: Application to ccsd energies for hydrocarbons and transition metal compounds. *J. Chem. Phys. 126* (2007), 154110.

[21] FRIEDRICH, J., HANRATH, M., AND DOLG, M. Evaluation of incremental correlation energies for open-shell systems: Application to the intermediates of the 4-exo cyclization, arduengo carbenes

[22] FRIEDRICH, J., PERLT, E., ROATSCH, M., SPICKERMANN, C., AND KIRCHNER, B. Coupled cluster in condensed phase. part i: Static quantum chemical calculations of hydrogen fluoride clusters. *J. Chem. Theory Comput. 7* (2011), 843–851.

[23] GOODPASTER, J. D., BARNES, T. A., MANBY, F. R., AND MILLER, III, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys. 140* (MAY 14 2014).

[24] GUO, Y., LI, W., AND LI, S. Improved cluster-in-molecule local correlation approach for electron correlation calculation of large systems. *J. Phys. Chem. A 118* (2014), 8996–9004.

[25] HAMPEL, C., AND WERNER, H.-J. Local treatment of electron correlation in coupled-cluster theory. *J. Chem. Phys. 104* (1996), 6286.

[26] HANRATH, M., AND ENGELS-PUTZKA, A. An efficient matrix-matrix multiplication based antisymmetric tensor contraction engine for general order coupled cluster. *J. Chem. Phys. 133* (2010), 064108.

[27] HÉGELY, B., NAGY, P., FERENCZY, G., AND KÁLLAY, M. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys. 145* (2016), 064107.

[28] HELD, J., HANRATH, M., AND DOLG, M. An efficient hartree–fock implementation based on the contraction of integrals in the primitive basis. *J. Chem. Theory Comput. 14* (2018), 6197.

[29] HØYVIK, I., JANSÍK, B., AND JØRGENSEN, P. Orbital localization using fourth central moment minimization. *J. Chem. Phys. 137* (2012), 224114.

[30] HØYVIK, I., JANSÍK, B., AND JØRGENSEN, P. Pipek–mezey localization of occupied and virtual orbitals. *J. Comput. Chem. 34* (2013), 1456–1462.

[31] HØYVIK, I., AND JØRGENSEN, P. Localized orbitals from basis sets augmented with diffuse functions. *J. Chem. Phys. 138* (2013), 204104.

[32] HUZINAGA, A. Coupling operator method in the scf theory. *J. Chem. Phys. 51* (1969), 3971 –3975.

[33] HUZINAGA, S., AND CANTU, A. A. Theory of separability of many-electron systems. *J. Chem. Phys. 55* (1971), 5543.

[34] HØYVIK, I.-M., KRISTENSEN, K., JANSÍK, B., AND JØRGENSEN, P. The divide-expand-consolidate family of coupled cluster methods: numerical illustrations using second order møller-plesset perturbation theory. *J. Chem. Phys. 136* (2012), 014105.

[35] JANSÍK, B., HØST, S., KRISTENSEN, K., AND JØRGENSEN, P. Local orbitals by minimizing powers of the orbital variance. *J. Chem. Phys. 134* (2011), 194104.

[36] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79* (1983), 926–935.

[37] KALVODA, S., PAULUS, B., DOLG, M., STOLL, H., AND WERNER, H.-J. Electron correlation effects on structural and cohesive properties of closo-hydroborate dianions $(b_n h_n)$2- (n = 5–12) and $b_4 h_4$. *Phys. Chem. Chem. Phys. 3* (2001), 514–522.

[38] KOBAYASHI, M., AND NAKAI, H. Divide-and-conquer-based linear-scaling approach for traditional and renormalized coupled cluster methods with single, double, and noniterative triple excitations. *J. Chem. Phys. 131* (2009), 114108.

[39] KOBAYASHI, M., AND NAKAI, H. How does it become possible to treat delocalized and/or open-shell systems in fragmentation-based linear-scaling electronic structure calculations? the case of the divide-and-conquer method. *Phys. Chem. Chem. Phys. 14* (2012), 7629.

[40] KRISTENSEN, K., ZIÓŁKOWSKI, M., JANSÍK, B., KJÆRGAARD, T., AND JØRGENSEN, P. A locality analysis of the divide–expand–consolidate coupled cluster amplitude equations. *J. Chem. Theory Comput. 7* (2011), 1677–1694.

[41] LI, S., SHEN, J., LI, W., AND JIANG, Y. An efficient implementation of the cluster-in-molecule approach for local electron correlation calculations. *J. Chem. Phys. 125* (2006), 074109.

[42]LI, W., GUO, Y., AND LI, S. A refined cluster-in-molecule local correlation approach for predicting the relative energies of large systems. *Phys. Chem. Chem. Phys. 14* (2012), 7854–7862.

[43]LI, W., AND PIECUCH, P. Improved design of orbital domains within the cluster-in-molecule local correlation framework: Single-environment cluster-in-molecule ansatz and its application to local coupled-cluster approach with singles and doubles. *J. Phys. Chem. A 114* (2010), 8644–8657.

[44]LI, W., AND PIECUCH, P. Multilevel extension of the cluster-in-molecule local correlation methodology: Merging coupled-cluster and møllerplesset perturbation theories. *J. Phys. Chem. A 114* (2010), 6721–6727.

[45]LI, W., PIECUCH, P., GOUR, J. R., AND LI, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys. 131* (2009), 114109.

[46]MA, Q., SCHWILK, M., KÖPPL, C., AND WERNER, H.-J. Scalable electron correlation methods. 4. parallel explicitly correlated local coupled cluster with pair natural orbitals (pno-lccsd-f12). *J. Chem. Theory Comput. 13* (2017), 4871–4896.

[47]MA, Q., AND WERNER, H.-J. Scalable electron correlation methods. 2. parallel pno-lmp2-f12 with near linear scaling in the molecular size. *J. Chem. Theory Comput. 11* (2015), 5291–5304.

[48]MA, Q., AND WERNER, H.-J. Scalable electron correlation methods. 5. parallel perturbative triples correction for explicitly correlated local coupled cluster with pair natural orbitals. *J. Chem. Theory Comput. 14* (2018), 198–215.

[49]MANBY, F. R., STELLA, M., GOODPASTER, J. D., AND MILLER, T. F. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput. 8* (2012), 2564–2568.

[50]MATA, R. A., AND STOLL, H. An incremental correlation approach to excited state energies based on natural transition/localized orbitals. *J. Chem. Phys. 134* (2011), 034122.

[51]MÖDLL, M., DOLG, M., FULDE, P., AND STOLL, H. Quantum chemical ab initio calculations of the magnetic interaction in alkalithioferrates(iii). *J. Chem. Phys. 106* (1997), 1836.

[52]MÜLLER, C., HERMANSSON, K., AND PAULUS, B. Electron correlation contribution to the n2o/ceria(111) interaction. *Chem. Phys. 362* (2009), 91.

[53]MÜLLER, C., HERSCHEND, B., HERMANSSON, K., AND PAULUS, B. Application of the method of increments to the adsorption of co on the ceo2 (110) surface. *J. Chem. Phys. 128* (2008), 214701.

[54]MÜLLER, C., USVYAT, D., AND STOLL, H. Local correlation methods for solids: Comparison of incremental and periodic correlation calculations for the argon fcc crystal. *Phys. Rev. B: Condens. Matter Mater. 83* (2011), 245136.

[55]NEESE, F., HANSEN, A., AND LIAKOS, D. G. Efficient and accurate approximations of the local coupled-cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys. 131* (2009), 064103.

[56]NEESE, F., WENNMOHS, F., AND HANSEN, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys. 130* (2009), 114108.

[57]NESBET, R. K. Electronic correlation in atoms and molecules. *Adv. Chem. Phys. 9* (1965), 321–363.

[58]NESBET, R. K. Atomic bethe-goldstone equations. i. the be atom. *Phys. Rev. 155* (1967), 51–55.

[59]NI, Z., GUO, Y., NEESE, F., LI, W., AND LI, S. Cluster-in-molecule local correlation method with an accurate distant pair correction for large systems. *J. Chem. Theory Comput. 17* (2021), 756–766.

[60]NI, Z., LI, W., AND LI, S. Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems. *J. Comput. Chem. 40* (2019), 1130–1140.

[61]PAULUS, B. Wave-function-based ab initio correlation treatment for the buckminsterfullerene c60. *Int. J. Quantum Chem. 100* (2004), 1026–1032.

[62]PERLT, E., FRIEDRICH, J., VON DOMAROS, M., AND KIRCHNER, B. Importance of structural motifs in liquid hydrogen fluoride. *ChemPhysChem 12* (2011), 3474–3482.

[63]PIPEK, J. Long-range behavior of the off-diagonal fock matrix elements of localized molecular orbitals. *Chem. Phys. Lett. 143* (1988), 293–298.

[64]PULAY, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett. 100* (1983), 151.

[65]PULAY, P., AND SAEBØ, S. Orbital-invariant formulation and second-order gradient evaluation in møller–plesset perturbation theory. *Theor. Chim. Acta 69* (1986), 357.

[66]RIPLINGER, C., AND NEESE, F. An efficient and near linear scaling pair natural orbital based local coupled-cluster method. *J. Chem. Phys. 138* (2013), 034106.

[67]RIPLINGER, C., SANDHOEFER, B., HANSEN, A., AND NEESE, F. Natural triple excitations in local coupled-cluster calculations with pair natural orbitals. *J. Chem. Phys. 139* (2013), 134101.

[68]ROLIK, Z., AND KÁLLAY, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *J. Chem. Phys. 135* (2011), 104111.

[69]ROLIK, Z., SZEGEDY, L., LADJÁNSZKI, I., LADÓCZKI, B., AND KÁLLAY, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys. 139* (2013), 094105.

[70]ROSCISZEWSKI, K., PAULUS, B., FULDE, P., AND STOLL, H. Ab initio calculation of ground-state properties of rare-gas crystals. *Phys. Rev. B: Condens. Matter Mater. Phys. 60* (1999), 7905.

[71]ROSCISZEWSKI, K., PAULUS, B., FULDE, P., AND STOLL, H. Ab initio coupled-cluster calculations for the fcc and hcp structures of rare-gas solids. *Phys. Rev. B: Condens. Matter Mater. Phys. 62* (2000), 5482.

[72]SCHMITT, I., FINK, K., AND STAEMMLER, V. The method of local increments for the calculation of adsorption energies of atoms and small molecules on solid surfaces part i. a single cu atom on the polar surfaces of zno. *Phys. Chem. Chem. Phys. 11* (2009), 11196.

[73]SCHWILK, M., MA, Q., KOPPL, C., AND WERNER, H.-J. Scalable electron correlation methods. 3. efficient and accurate parallel local coupled cluster with pair natural orbitals (pno-lccsd). *J. Chem. Theory Comput. 13* (2017), 3650–3675.

[74]SCHÜTZ, M., AND WERNER, H.-J. Local perturbative triples correction (t) with linear cost scaling. *Chem. Phys. Lett. 318* (2000), 370.

[75]SCHÜTZ, M., YANG, J., CHAN, G. K.-L., MANBY, F. R., AND WERNER, H.-J. The orbital-specific virtual local triples correction: Osv-l (t). *J. Chem. Phys. 138* (2013), 054109.

[76]SINANOGLU, O. Many-electron theory of atoms, molecules and their interactions. *Adv. Chem. Phys. 6* (1964), 315–412.

[77]SPICKERMANN, C., PERLT, E., VON DOMAROS, M., ROATSCH, M., FRIEDRICH, J., AND KIRCHNER, B. Coupled cluster in condensed phase. part ii: Liquid hydrogen fluoride from quantum cluster equilibrium theory. *J. Chem. Theory Comput. 7* (2011), 868–875.

[78]STAEMMLER, V. Method of local increments for the calculation of adsorption energies of atoms and small molecules on solid surfaces. 2. co/mgo(001). *J. Phys. Chem. A 115* (2011), 7153.

[79]STOLL, H. Correlation energy of diamond. *Phys. Rev. B 46* (1992), 67006704.

[80]STOLL, H., AND DOLL, K. Approaching the bulk limit with finite cluster calculations using local increments: The case of lih. *J. Chem. Phys. 136* (2012), 074106.

[81]STOLL, H., PAULUS, B., AND FULDE, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys. 123* (2005), 144108.

[82]SUBOTNIK, J., DUTOI, A., AND HEAD-GORDON, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *J. Chem. Phys. 123* (2005), 114108.

[83]VERMA, P., HUNTINGTON, L., COONS, M. P., KAWASHIMA, Y., YAMAZAKI, T., AND ZARIBAFIYAN, A. Scaling up electronic structure calculations on quantum computers: The frozen natural orbital based method of increments. *J. Chem. Phys. 155* (2021), 034110.

[84]WALES, D. J., AND HODGES, M. P. Global minima of water

clusters (h2o) n, n 21, described by an empirical potential. *Chem. Phys. Lett. 286* (1998), 65–72.

[85] WERNER, H.-J., KNIZIA, G., KRAUSE, C., SCHWILK, M., AND DORNBACH, M. Scalable electron correlation methods i.: Pno-lmp2 with linear scaling in the molecular size and near-inverse-linear scaling in the number of processors. *J. Chem. Theory Comput. 11* (2015), 484–507.

[86] WOON, D. E., AND DUNNING, THOM H., J. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of Chemical Physics 98* (1993), 1358–1371.

[87] YANG, J., KURASHIGE, Y., MANBY, F. R., AND CHAN, G. K. L. Tensor factorizations of local second-order møller–plesset theory. *J. Chem. Phys. 134* (2011), 044123.

[88] YANG, J., MANBY, F. R., CHAN, G. K. L., SCHÜTZ, M., AND WERNER, H.-J. The orbital-specific virtual local coupled-cluster singles and doubles method. *J. Chem. Phys. 136* (2012), 144105.

[89] ZHANG, J., AND DOLG, M. Third-order incremental dual-basis set zero-buffer approach: An accurate and efficient way to obtain ccsd and ccsd(t) energies. *J. Chem. Theory Comput. 9* (2013), 2992–3003.

[90] ZHANG, J., AND DOLG, M. Approaching the complete basis set limit of ccsd(t) for large systems by the third-order incremental dual-basis set zero-buffer f12 method. *J. Chem. Phys. 140* (2014), 044114.

[91] ZHANG, J., AND DOLG, M. Third-order incremental dual-basis set zero-buffer approach for large high-spin open-shell systems. *J. Chem. Theory Comput. 11* (2015), 962.

[92] ZHANG, J., HEINZ, N., AND DOLG, M. Understanding lanthanoid(iii) hydration structure and kinetics by insights from energies and wave functions. *Inorganic Chemistry 53* (2014), 7700–7708.

[93] ZIÓŁKOWSKI, M., JANSÍK, B., KJÆRGAARD, T., AND JØRGENSEN, P. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys. 133* (2010), 014107.

TABLE OF CONTENTS
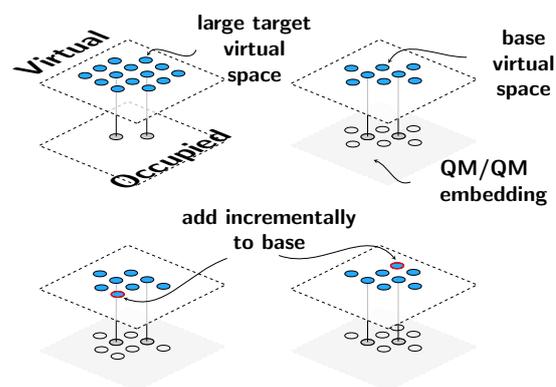
5.1.1 Implementation

The new method which is presented in the manuscript above uses EGVs. The Huzinaga-Cantu [75] and projective embedding [76] methods which have been used for this purpose are presented in the manuscript and also in section 2.4 and will not be discussed further. Instead, some of the used graph theoretical concepts are presented here. As shown in the manuscript, only connected augmentors are used within the new incremental scheme with an additional virtual space expansion. The construction of such augmentors has been shown in the manuscript schematically and shall be described in more detail below.

The basis of the construction scheme is a representation of the domain centres as vertices in an undirected graph. Two domains are considered to be connected if the distance between the centres is smaller than a threshold of for example 3.5 Bohrs. Each connection is represented in a graph as an edge. The graph of the n-hexane molecule is for example presented in figure 5.1. Such a graph can be analysed by various graph theoretical methods. In this work, the boost graph library (BGL) implementations [183] are used and adopted for special needs. One of the used methods is the breadth-first search (BFS). Starting from a root vertex, the graph is examined first across all vertices which are directly connected to the root vertex and afterwards across the vertices which are connected to the previously examined ones. Choosing vertex 3 as a root vertex in the n-hexane example, one obtains the rooted graph in figure 5.1. It is possible to track distances of the root vertex to all other vertices during the BFS. Furthermore, for each vertex p as destination starting from a root vertex r the next parent vertex, denoted by $n(r,p)$, in the direction of p can be stored. If two vertices s and t are connected, s is already a parent of t and therefore $n(s,t) = s$. In the n-hexane example, we could for instance try to connect 3 with 5. The first vertex after 3 should than be $n(3,5) = 4$. Obviously, any vertex can be chosen as the root vertex and the information can be stored for each of the vertices. We can therefore update the root vertex to 4 and find the next vertex in the direction of 5. That is $n(4,5) = 4$ and the connection is completed with the vertices
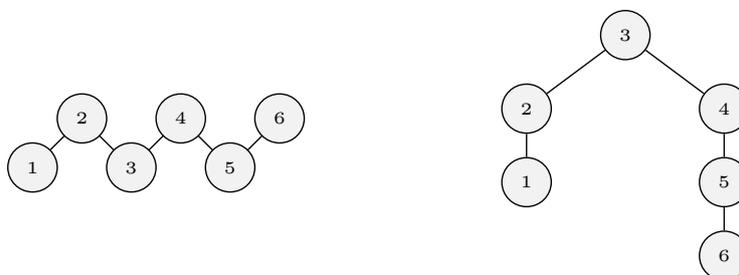
Figure 5.1: Graph representation of the domains in n-hexane (left) and rooted breath-first search graph with root vertex 3 (right).

C(3,5)=(3,4,5). In general, the connection of any pair $C(p,q) = (c_0, c_1, \ldots, c_{|C(p,q)|})$ is defined recursively by

$$c_0 = p\,; \quad c_{|C(p,q)|} = q\,; \quad c_i = n\left(c_{i-1}, c_{|C(p,q)|}\right) \quad 1 \leq i \leq |C(p,q)| - 1 \tag{5.1}$$

There are also other possibilities to find shortest paths between two vertices.[184] However, the distances between the vertices is not known *a priori* and the above described strategy to find connections is used for convenience when applied to more than two vertices. The construction of a base augmentor at incremental order three for instance requires a connection of three domains. In order to find the shortest connection of a tuple of vertices $\mathbb{X}$, an optimal sequence of shortest pair connections has to be found. This can be realized by creating all permutations π of $(x_1, x_2, \ldots, x_{|\mathbb{X}|})$

$$\left(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(|\mathbb{X}|)}\right) \quad \text{with} \quad \pi \in S_{|\mathbb{X}|} \tag{5.2}$$

and comparing the lengths L_π of the connections in these sequences by summing the pair distances

$$L_\pi = \sum_{p=1}^{|\mathbb{X}|-1} |C\left(x_\pi(p), x_\pi(p+1)\right)| \tag{5.3}$$

One of the shortest connections is chosen randomly. Unlike the base augmentor, additional augmentors must take into account that certain vertices have to be terminal. That is, the remainder is still connected if these vertices are removed. Suppose that $y \in \mathbb{Y}$ is one of the domains that shall be terminal. The terminality check is implemented by searching for y as a connecting parent. That is $n(p,q) = y$. The domain y is terminal, if this does not hold for any $p, q \in C(\mathbb{X} \cup \mathbb{Y})$.

5.2 Additional virtual space truncation with approximate natural orbitals

Approximate natural orbitals are a widely used tool in local correlation methods (section 2.6.1). Their main scope of application is to compress virtual orbital spaces in order to lower the computational cost for a correlation calculation. The involved steps are summarized conceptually as follows: the virtual-virtual block D_{ab} of the MP2 density matrix is (approximately) calculated and it is either fully diagonalized or parts of it. In a canonical spin-orbital basis, the MP2 density block D_{ab} reads [185]

$$D_{ab} = \frac{1}{2} \sum_{ijc} t_{ij}^{ac} t_{ij}^{bc} \tag{5.4}$$

where the amplitudes t are defined in (2.122). If the full matrix is diagonalized, FNOs [162] are obtained. These orbitals can be used to replace the full canonical virtual space, which has also been applied within the incremental scheme method.[186] The MP2 density matrix is used in other methods to access virtual orbitals for specific occupied orbitals (OSVs) or pairs of occupied orbitals (PNOs) [3–12] In these methods, a fraction of the density matrix in (5.4) is diagonalized instead of the full matrix, summing over the virtual orbital index c for a fixed pair ij. This procedure yields small virtual spaces when a local occupied MO basis is used.[5, 7]

These orbital spaces are well suited for direct correlation methods since the correlation spaces for each occupied orbital and occupied orbital pairs can be restricted to the respective OSV or PNO space. Within the incremental scheme, LMOs are grouped and these groups are correlated individually and in higher ´orders also together with other groups. A compressed virtual space that is tailored to specific tuples of occupied orbitals would undoubtedly be an advantage in this context. The use of approximate natural orbitals specific for occupied orbital tuples has not yet been investigated. These approximate orbitals could be called tuple natural orbitals (TNOs) in analogy to PNOs. They can be generated by diagonalizing the density matrix block $D_{ab}^{\mathbb{I}}$ for a tuple of occupied orbitals with $i \in \underline{\mathbb{I}}$. This block is evaluated by running the sum in (5.4) over all virtual orbitals and occupied orbitals in $\underline{\mathbb{I}}$. In the closed-shell case, the density matrix block is given explicitly by

$$D_{ab}^{\mathbb{I}} = \sum_{i,j\in\underline{\mathbb{I}}} \sum_{c} \frac{4\langle ij|ac\rangle \langle ij|bc\rangle - 2\langle ij|ac\rangle \langle ij|cb\rangle}{\epsilon_{ij}^{ac}\epsilon_{ij}^{bc}} \tag{5.5}$$

where $\epsilon_{ij}^{ab} = f_{aa} + f_{bb} - f_{ii} - f_{jj}$. As usual, the occupied space is localized and therefore the semi-canonical approximation is applied [4] by keeping expression (5.5) which exactly holds true for canonical orbitals. Other common techniques as using a local pseudo-canonicalized PAO space for the calculation of the density [5] are equivalently legit for TNOs as they are for OSVs or PNOs.

In section 5.1, a new incremental scheme was investigated which uses EGVs. The beneficial properties of an EGV space are that it is smaller than the full virtual space and local due to basis set restrictions and that it diagonalizes the embedding Fock matrix. They can therefore be used directly for the calculation of the density matrix block in (5.5). A diagonalization of this density matrix yields TNOs. The dimensions of the EGV space and the TNO space are equal but the TNO space can be further truncated according to the density matrix eigenvalues.

6 Summary and Outlook

In this thesis, general aspects of an incremental and local virtual space expansion have been discussed. A possible PAO based approach was proposed and another strategy based on EGVs has been implemented. The implementation uses a new atomic domain construction algorithm, graph theoretical methods, Fock matrix based truncation parameters for the selection of occupied and virtual spaces and the QM/QM embedding approach to access relevant virtual spaces. This leads overall to a method which scales asymptotically linear with the size of the system and can in principle be combined with any size-consistent electron correlation method such as the CC method.

The new incremental scheme can be used as a black box application. It was demonstrated that correlation energies are accurate and restore chemical accuracy for a variety of molecular systems. The reduction of computational requirements becomes evident for large systems but is already present for smaller ones. For the system $(H_2O)_{21}$, the computational requirements are already reduced by a factor of more than 100 in the most expensive calculation. The method is thus suited to treat even larger systems. There is a limitation of applicability if the largest calculation can not be performed in the desired basis set. For such a situation, an additional virtual space truncation would be beneficial. It was proposed that the concepts of approximate natural orbitals could be adopted to the incremental scheme framework by a calculation of domain tuple natural orbitals. A truncation according to the eigenvalues of the approximate natural orbitals is expected to yield the desired truncation.

Another investigated topic is an analysis of the total CCSD correlation energy calculated with the ordinary incremental scheme. It could be shown that not only the energy is incrementally calculated when the incremental scheme is applied. The amplitudes of the full system are also incrementally calculated from various incremental amplitude contributions. This fact represents a bridge between the DEC method and the incremental scheme. A possible further investigation could be to apply combined occupied and virtual space expansions in the framework of other fragmentation based methods.

The developed method was used for single point calculations on closed shell systems. The next step to analyse the behaviour of the method could be the application to potential energy surfaces in order to check whether the virtual space expansion causes discontinuous potential energy surfaces.

Alternatively one could target a combined incremental expansion with LMOs in both orbital spaces, since virtual LMOs are nowadays accessible with more refined localization techniques.

Bibliography

[1]   O. Sinanoğlu, *Adv. Chem. Phys.* **1964**, *6*, 315–412.

[2]   R. K. Nesbet, *Adv. Chem. Phys.* **1965**, *9*, 321–363.

[3]   F. Neese, F. Wennmohs, A. Hansen, *J. Chem. Phys.* **2009**, *130*, 114108.

[4]   F. Neese, A. Hansen, D. G. Liakos, *J. Chem. Phys.* **2009**, *131*, 064103.

[5]   C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106.

[6]   C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.

[7]   J. Yang, Y. Kurashige, F. R. Manby, G. K. L. Chan, *J. Chem. Phys.* **2011**, *134*, 044123.

[8]   J. Yang, F. R. Manby, G. K. L. Chan, M. Schütz, H.-J. Werner, *J. Chem. Phys.* **2012**, *136*, 144105.

[9]   M. Schütz, J. Yang, G. K.-L. Chan, F. R. Manby, H.-J. Werner, *J. Chem. Phys.* **2013**, *138*, 054109.

[10] H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, M. Dornbach, *J. Chem. Theory Comput.* **2015**, *11*, 484–507.

[11] M. Schwilk, Q. Ma, C. Köppl, H.-J. Werner, *J. Chem. Theory Comput.* **2017**, *13*, 3650–3675.

[12] Q. Ma, H.-J. Werner, *J. Chem. Theory Comput.* **2018**, *14*, 198–215.

[13] H. Stoll, *Phys. Rev. B* **1992**, *46*, 6700−6704.

[14] J. Friedrich, M. Hanrath, M. Dolg, *J. Chem. Phys.* **2007**, *126*, 154110.

[15] J. Zhang, M. Dolg, *J. Chem. Theory Comput.* **2013**, *9*, 2992–3003.

[16] M. Ziółkowski, B. Jansík, T. Kjærgaard, P. Jørgensen, *J. Chem. Phys.* **2010**, *133*, 014107.

[17] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **2009**, *131*, 114108.

[18] S. Li, J. Shen, W. Li, Y. Jiang, *J. Chem. Phys.* **2006**, *125*, 074109.

[19] Z. Rolik, M. Kállay, *J. Chem. Phys.* **2011**, *135*, 104111.

[20] C. Edmiston, K. Ruedenberg, *Revs. Modern Phys.* **1963**, *35*, 457–465.

[21] J. Pipek, *Int. J. Quantum Chem.* **1989**, *36*, 487–501.

Bibliography

[22] S. Boys, J. Foster, *Revs. Modern. Phys.* **1969**, *32*, 300–302.

[23] C. Edmiston, M. Krauss, *J. Chem. Phys.* **1965**, *42*, 1119.

[24] R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, W. Kutzelnigg, *J. Chem. Phys.* **1975**, *62*, 1235.

[25] J. Friedrich, M. Dolg, *J. Chem. Phys.* **2008**, *129*, 244105.

[26] P. Ehrenfest, *Ann. Phys.* **1911**, *341*, 91–118.

[27] H. Hertz, *Ann. Phys.* **1887**, *267*, 983–1000.

[28] A. H. Compton, *Phys. Rev.* **1923**, *21*, 483–502.

[29] J. J. Balmer, *Ann. Phys.* **1885**, *261*, 80–87.

[30] W. Gerlach, O. Stern, *Z. Physik* **1922**, *9*, 349–352.

[31] L. De Broglie, *Nature* **1923**, *112*, 540.

[32] R. Thomson, G., *Nature* **1927**, *119*, 890.

[33] M. Born, *Z. Physik* **1926**, *37*, 863–867.

[34] E. Schrödinger, *Phys. Rev.* **1926**, *28*, 1049–1070.

[35] W. Heisenberg, *Z. Physik* **1927**, *43*, 172–198.

[36] M. Born, R. Oppenheimer, *Ann. Phys.* **1927**, *389*, 457–484.

[37] G. Uhlenbeck, S. Goudsmit, *Sci. Nat.* **1925**, *13*, 953–954.

[38] M. Blume, R. Watson, *Proc. R. Soc. Lond. A* **1962**, *270*, 127–143.

[39] W. Pauli, *Phys. Rev.* **1940**, *58*, 716–730.

[40] E. Fermi, *Z. Physik* **1926**, *36*, 902–912.

[41] P. Dirac, *Proc. R. Soc. Lond. A* **1926**, *112*, 661–677.

[42] W. Heisenberg, *Z. Physik* **1926**, *38*, 411–426.

[43] D. Hartree, *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 111–132.

[44] V. Fock, *Z. Physik* **1930**, *61*, 126–148.

[45] C. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[46] D. Strout, G. Scuseria, *J. Chem. Phys.* **1995**, *102*, 8448–8452.

[47] J. Almlöf, K. Faegri Jr., K. Korsell, *J. Comput. Chem.* **1982**, *3*, 385–399.

[48] M. Häser, R. Ahlrichs, *J. Comput. Chem.* **1989**, *10*, 104–111.

[49] D. Cremer, J. Gauss, *J. Comput. Chem.* **1986**, *7*, 274–282.

[50] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

[51] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.

[52] J. Toulouse in *Density-Functional Theory*, (Eds.: E. Cancès, G. Friesecke), Springer, **2022**.

[53] A. Severo Pereira Gomes, C. R. Jacob, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222–277.

[54] P. Mezey, J. Pipek, *J. Chem. Phys.* **1989**, *90*, 4916–4926.

[55] S. Lehtola, H. Jonsson, *J. Chem. Theory Comput.* **2014**, *10*, 642–649.

[56] S. Boys, *Quantum Theory of Atoms, molecules, and the Solid State*, ACADEMIC PRESS, New York, **1966**, pp. 253–262.

[57] J. Subotnik, A. Dutoi, M. Head-Gordon, *J. Chem. Phys.* **2005**, *123*, 114108.

[58] I.-M. Høyvik, P. Jørgensen, *Chem. Rev.* **2016**, *116*, 3306–3327.

[59] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons , Ltd, **2000**.

[60] B. Jansik, S. Høst, K. Kristensen, P. Jørgensen, *J. Chem. Phys.* **2011**, *134*, 194104.

[61] I. Høyvik, B. Jansik, P. Jørgensen, *J. Chem. Phys.* **2012**, *137*, 224114.

[62] I. Høyvik, B. Jansik, P. Jørgensen, *J. Comput. Chem.* **2013**, *34*, 1456–1462.

[63] I. Høyvik, P. Jørgensen, *J. Chem. Phys.* **2013**, *138*, 204104.

[64] S. Folkestad, R. Matveeva, I. Høyvik, H. Koch, *J. Chem. Theory Comput.* **2022**, *18*, 4733–4744.

[65] A. Sánchez de Merás, C. Koch, H., I., L. Boman, *J. Chem. Phys.* **2010**, *132*, 204105.

[66] T. Giovannini, H. Koch, *J. Chem. Theory Comput.* **2021**, *17*, 139–150.

[67] T. Giovannini, H. Koch, *J. Chem. Theory Comput.* **2022**, *18*, 4806–4813.

[68] X. Assfeld, J.-L. Rivail, *Chem. Phys. Lett.* **1996**, *263*, 100–106.

[69] D. M. Philipp, R. A. Friesner, *J. Comput. Chem.* **1999**, *20*, 1468–1494.

[70] J. Pu, J. Gao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 632–650.

[71] Q. Sun, G. K.-L. Chan, *J. Chem. Theory Comput.* **2014**, *10*, 3784–3790.

[72] G. G. Ferenczy, *J. Comput. Chem.* **2013**, *34*.

[73] M. Svensson, S. Humbel, R. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* **1996**, *100*, 19357–19363.

[74] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, K. Morokuma, *Chem. Rev.* **2015**, *115*, 5678–5796.

[75] S. Huzinaga, A. A. Cantu, *J. Chem. Phys.* **1971**, *55*, 5543–5548.

[76] F. Manby, M. Stella, J. D. Goodpaster, T. F. Miller, *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

Bibliography

[77] T. Wesolowski, A. Warshel, *J. Phys. Chem.* **1993**, *97*, 8050–8053.

[78] J. Inglesfield, *J. Phys. C: Solid State Phys.* **1981**, *14*, 3795.

[79] V. Fock, M. Veselov, M. Petrashen, *Zh. Eksp. Teor. Fiz* **1940**, *10*, 723–739.

[80] R. McWeeny, *Proc. Roy. Soc.* **1959**, *253*, 242–259.

[81] A. Huzinaga, *J. Chem. Phys.* **1969**, *51*, 3971–3975.

[82] B. Hégely, P. Nagy, G. Ferenczy, M. Kállay, *J. Chem. Phys.* **2016**, *145*, 064107.

[83] S. Huzinaga, L. Seijo, Z. Barandiarán, M. Klobukowski, *J. Chem. Phys.* **1987**, *86*, 2132–2145.

[84] S. Huzinaga, M. Klobukowski, Y. Sakai, *J. Phys. Chem.* **1984**, *88*, 4880–4886.

[85] J. Phillips, L. Kleinman, *Phys. Rev.* **1959**, *116*, 287–294.

[86] L. Kahn, P. Baybutt, D. Truhlar, *J. Chem. Phys.* **2008**, *65*, 3826–3853.

[87] P. A. Christiansen, Y. S. Lee, K. S. Pitzer, *J. Chem. Phys.* **1979**, *71*, 4445–4450.

[88] M. Dolg, U. Wedig, H. Stoll, H. Preuss, *J. Chem. Phys.* **1987**, *86*, 866–872.

[89] M. Dolg in *Theoretical and computational chemistry, Vol. 11*, Elsevier, **2002**, pp. 793–862.

[90] M. Dolg, X. Cao, *Chem. Rev.* **2012**, *112*, 403–480.

[91] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701–706.

[92] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2000**, *318*, 614–618.

[93] D. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832–6840.

[94] A. Götz, S. Beyhan, L. Visscher, *J Chem Theory Comput.* **2009**, *12*, 3161–3174.

[95] J. Goodpaster, N. Ananth, F. Manby, T. Miller, *J. Chem. Phys.* **2010**, *133*.

[96] T. Barnes, J. Goodpaster, F. Manby, T. Miller, *J. Chem. Phys.* **2013**, *139*, 024103.

[97] F. Jensen, *Introduction to computational chemistry*, John wiley & sons, **2017**.

[98] L. Brillouin, *Actual. Sci. Ind.* **1933**, *71*.

[99] J. C. Slater, *Phys. Rev.* **1929**, *34*, 1293–1322.

[100] E. U. Condon, *Phys. Rev.* **1930**, *36*, 1121–1133.

[101] F. Coester, H. Kümmel, *Nuclear Physics* **1958**, *9*, 225–236.

[102] R. J. Bartlett, *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.

[103] H. Baker, *Proc. London Math. Soc.* **1905**, *24*, 24.

[104] J. E. Campbell, *Proc. London Math. Soc.* **1897**, *28*, 381.

[105] F. Hausdorff, *Leipz. Ber.* **1906**, *58*, 19.

[106] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **1989**, *157*, 479–483.

[107] J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.

[108] E. Schrödinger, *Ann. Phys.* **1926**, *385*, 437–490.

[109] J. Rayleigh, Macmillan, **1894**.

[110] W. Meyer, *Int. J. Quantum Chem.* **1971**, *5*, 341–348.

[111] P. Pulay, *Chem. Phys. Lett.* **1983**, *100*, 151.

[112] S. Saebø, P. Pulay, *Chem. Phys. Lett.* **1985**, *113*, 13.

[113] P. Pulay, S. Saebø, *Theor. Chim. Acta* **1986**, *69*, 357.

[114] S. Saebø, P. Pulay, *J. Chem. Phys.* **1987**, *86*, 914.

[115] S. Saebø, P. Pulay, *J. Chem. Phys.* **1988**, *88*, 1884.

[116] C. Hampel, H.-J. Werner, *J. Chem. Phys.* **1996**, *104*, 6286.

[117] M. Schütz, H.-J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370.

[118] P.-O. Löwdin, *Phys. Rev.* **1955**, *97*, 1474–1489.

[119] W. Li, P. Piecuch, J. R. Gour, S. Li, *J. Chem. Phys.* **2009**, *131*, 114109.

[120] W. Li, P. Piecuch, *J. Phys. Chem. A* **2010**, *114*, 6721–6727.

[121] W. Li, P. Piecuch, *J. Phys. Chem. A* **2010**, *114*, 8644–8657.

[122] W. Li, Y. Guo, S. Li, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7854–7862.

[123] Y. Guo, W. Li, S. Li, *J. Phys. Chem. A* **2014**, *118*, 8996–9004.

[124] Z. Ni, W. Li, S. Li, *J. Comput. Chem.* **2019**, *40*, 1130–1140.

[125] Z. Ni, Y. Guo, F. Neese, W. Li, S. Li, *J. Chem. Theory Comput.* **2021**, *17*, 756–766.

[126] Y. Guo, U. Becker, F. Neese, *J. Chem. Phys.* **2018**, *148*, 124117.

[127] I.-M. Høyvik, K. Kristensen, B. Jansík, P. Jørgensen, *J. Chem. Phys.* **2012**, *136*, 014105.

[128] K. Kristensen, M. Ziółkowski, B. Jansík, T. Kjærgaard, P. Jørgensen, *J. Chem. Theory Comput.* **2011**, *7*, 1677–1694.

[129] P. Ettenhuber, P. Baudin, T. Kjærgaard, P. Jørgensen, K. Kristensen, *J. Chem. Phys.* **2016**, *144*, 164116.

[130] K. Rosciszewski, B. Paulus, P. Fulde, H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *60*, 7905.

[131] K. Rosciszewski, B. Paulus, P. Fulde, H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *62*, 5482.

[132] H. Stoll, B. Paulus, P. Fulde, *J. Chem. Phys.* **2005**, *123*, 144108.

[133] C. Müller, B. Herschend, K. Hermansson, B. Paulus, *J. Chem. Phys.* **2008**, *128*, 214701.

[134] I. Schmitt, K. Fink, V. Staemmler, *Phys. Chem. Chem. Phys.* **2009**, *11*, 11196.

[135] C. Müller, K. Hermansson, B. Paulus, *Chem. Phys.* **2009**, *362*, 91.

[136] V. Staemmler, *J. Phys. Chem. A* **2011**, *115*, 7153.

[137] C. Müller, D. Usvyat, H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *83*, 245136.

[138] H. Stoll, K. Doll, *J. Chem. Phys.* **2012**, *136*, 074106.

[139] M. P. de Lara-Castells, H. Stoll, A. O. Mitrushchenkov, *J. Phys. Chem. A* **2014**, *118*, 6367.

[140] M. P. de Lara-Castells, M. Bartolomei, A. O. Mitrushchenkov, H. Stoll, *J. Chem. Phys.* **2015**, *143*, 194701.

[141] S. Kalvoda, B. Paulus, M. Dolg, H. Stoll, H.-J. Werner, *Phys. Chem. Chem. Phys.* **2001**, *3*, 514–522.

[142] B. Paulus, *Int. J. Quantum Chem.* **2004**, *100*, 1026–1032.

[143] J. Friedrich, E. Perlt, M. Roatsch, C. Spickermann, B. Kirchner, *J. Chem. Theory Comput.* **2011**, *7*, 843–851.

[144] C. Spickermann, E. Perlt, M. von Domaros, M. Roatsch, J. Friedrich, B. Kirchner, *J. Chem. Theory Comput.* **2011**, *7*, 868–875.

[145] E. Perlt, J. Friedrich, M. von Domaros, B. Kirchner, *Chem. Phys. Chem.* **2011**, *12*, 3474–3482.

[146] R. A. Mata, H. Stoll, *J. Chem. Phys.* **2011**, *134*, 034122.

[147] J. Friedrich, M. Hanrath, M. Dolg, *J. Phys. Chem. A* **2007**, *111*, 9830.

[148] J. Friedrich, M. Dolg, *J. Chem. Theory Comput.* **2009**, *5*, 287–294.

[149] J. Friedrich, J. Hänchen, *J. Chem. Theory Comput.* **2013**, *9*, 5381–5394.

[150] B. Fiedler, G. Schmitz, C. Haettig, J. Friedrich, *J. Chem. Theory Comput.* **2017**, *13*, 6023–6042.

[151] B. Fiedler, S. Coriani, J. Friedrich, *J. Chem. Theory Comput.* **2016**, *12*, 3040.

[152] J. Friedrich, S. Coriani, T. Helgaker, M. Dolg, *J. Chem. Phys.* **2009**, *131*, 154102.

[153] J. Yang, M. Dolg, *J. Chem. Phys.* **2007**, *127*, 084108.

[154] J. Friedrich, H. McAlexander, A. Kumar, T. D. Crawford, *Phys. Chem. Chem. Phys.* **2015**, *17*, 14284.

[155] J. Zhang, M. Dolg, *J. Chem. Phys.* **2014**, *140*, 044114.

[156] J. Zhang, N. Heinz, M. Dolg, *Inorg. Chem.* **2014**, *53*, 7700–7708.

[157] M. Mödl, M. Dolg, P. Fulde, H. Stoll, *J. Chem. Phys.* **1997**, *106*, 1836.

[158] J. Friedrich, M. Hanrath, M. Dolg, *J. Phys. Chem. A* **2008**, *112*, 8762.

[159] J. Zhang, M. Dolg, *J. Chem. Theory Comput.* **2015**, *11*, 962.

[160] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833–1840.

[161] J. W. Boughton, P. Pulay, *J. Comput. Chem.* **1993**, *14*, 736.

[162] A. G. Taube, R. J. Bartlett, *Collect. Czechoslov. Chem. Commun.* **2005**, *70*, 837–850.

[163] G. Hetzer, P. Pulay, H.-J. Werner, *Chem. Phys. Lett.* **1998**, *290*, 143.

[164] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2005**, *123*, 134103.

[165] M. Kobayashi, H. Nakai, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7629.

[166] Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, M. Kállay, *J. Chem. Phys.* **2013**, *139*, 094105.

[167] G. Karypis, V. Kumar, *SIAM J. Sci. Comput.* **1998**, *20*, 359–392.

[168] E. Forgy, *Biometrics* **1965**, *21*, 768–769.

[169] J. B. MacQueen in Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, *Vol. 1*, (Eds.: L. M. L. Cam, J. Neyman), University of California Press, **1967**, pp. 281–297.

[170] S. Lloyd, *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137.

[171] G. Mie, *Ann. Phys.* **1903**, *316*, 657–697.

[172] P. Debye, *Math. Ann.* **1909**, *67*, 535–558.

[173] R. Jurgens-Lutovsky, J. Almlöf, *Chem. Phys. Lett.* **1991**, *178*, 451–454.

[174] J. Friedrich, *J. Chem. Theory Comput.* **2010**, *6*, 1834–1842.

[175] B. Fiedler, D. Himmel, I. Krossing, J. Friedrich, *J. Chem. Theory Comput.* **2018**, *14*, 557–571.

[176] M. Schütz, R. Lindh, H.-J. Werner, *Mol. Phys.* **1999**, *96*, 719–733.

[177] M. Schütz, G. Hetzer, H.-J. Werner, *J. Chem. Phys.* **1999**, *111*, 5691–5705.

[178] H. A. Bethe, J. Goldstone, N. F. Mott, *Proc. R. Soc. A.* **1957**, *238*, 551–567.

[179] J. J. Eriksen, F. Lipparini, J. Gauss, *J. Phys. Chem. Letters* **2017**, *8*, 4633–4639.

[180] J. Friedrich, M. Hanrath, M. Dolg, *Chem. Phys.* **2007**, *338*, 33–43.

Bibliography

[181] E. Cuthill, J. McKee in Proceedings of the 1969 24th National Conference, Association for Computing Machinery, **1969**, pp. 157–172.

[182] P. Löwdin, *J. Chem. Phys.* **1950**, *18*, 365–375.

[183] J. Siek, L.-Q. Lee, A. Lumsdaine in **2002**.

[184] E. Dijkstra, *Numer. Math.* **1959**, *1*, 269–271.

[185] G. W. Trucks, E. Salter, C. Sosa, R. J. Bartlett, *Chem. Phys. Lett.* **1988**, *147*, 359–366.

[186] P. Verma, L. Huntington, M. P. Coons, Y. Kawashima, T. Yamazaki, A. Zaribafiyan, *J. Chem. Phys.* **2021**, *155*, 034110.

List of Figures

List of Tables

Appendix

Acronyms and abbreviations

MBE	many-body expansion
FMO	fragment molecular orbital
WFT	wave function theory
DFT	density functional theory
CI	configuration interaction
FCI	full CI
CC	coupled cluster
IEPA	independent electron pair approximation
CEPA	coupled electron pair approximation
SCF	self-consistent field
AO	atomic orbital
MO	molecular orbital
LMO	localized molecular orbital
LCAO	linear combination of atomic orbitals
PAO	projected atomic orbital
PNO	pair natural orbital
TNO	tuple natural orbital
EGV	embedding generated virtual
FNO	frozen natural orbital
OSV	orbital specific virtual
IS	incremental scheme
LNO	local natural orbital
DLPNO	domain based local pair natural orbital
CIM	cluster-in-molecule
DAC	divide and conquer
DEC	divide expand consolidate
RSPT	Rayleigh-Schrödinger perturbation theory

List of Tables

MPPT	Møller-Plesset perturbation theory
HF	Hartree-Fock
QM	quantum mechanics
HC	Huzinaga-Cantu
MM	molecular mechanics
ONIOM	own N-layered integrated molecular orbital and molecular mechanics
MP	model potential
AIMP	*ab initio* model potenital
ECP	effective core potenital
PP	pseudopotential
EOS	energy orbital space
AOS	amplitude orbital space
FOT	fragment optimization threshold
BGL	boost graph library
BFS	breadth-first search

Acknowledgements

Eidesstaatliche Erkärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen:

- I. Türkmen, M. Dolg, J. Chem. Theory Comput. 2024, 20, 8, 3154-3168.

_____   _____   _____

(Datum)   (Name)   (Unterschrift)

Curriculum Vitae

Personal Information

Last name	Türkmen
First name	Ilyas
Date of birth	8th of April 1992

Education

since 07/2019	PhD studies
	Institute for Theoretical Chemistry, University of Cologne
10/2014 - 06/2019	M.Sc. studies in Chemistry
	University of Cologne
10/2011 - 12/2014	B.Sc. studies in Chemistry
	University of Cologne
09/2002 - 07/2011	Abitur
	Martin-Luther-King-Gesamtschule Dortmund

Work experience

since 07/2019	Predoctoral fellow
	Institute for Theoretical Chemistry, University of Cologne
02/2015 - 06/2019	Research assistant
	University of Cologne
05/2014 - 01/2015	Student assistant
	University of Cologne