

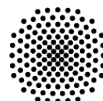
Development of the Transcorrelated Full Configuration Interaction Quantum Monte Carlo Method

Von der Fakultät Chemie der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Jacobus Philip Haupt
aus George, Südafrika

Hauptberichter: Prof. Dr. Ali Alavi
Mitberichter: Prof. Dr. Andreas Köhn
Prüfungsvorsitzender: Prof. Dr. Joris van Slageren
Tag der mündlichen Prüfung:

MAX-PLANCK-INSTITUT
FÜR FESTKÖRPERFORSCHUNG



Universität Stuttgart

Universität Stuttgart

Max-Planck-Institut für Festkörperforschung, Stuttgart

2024

It is important to realize that in physics today, we have no knowledge of what energy is.

— Richard Feynman, *Feynman Lectures on Physics, Volume 1, Chapter 4*

CONTENTS

Erklärung über die Eigenständigkeit der Dissertation	ix
Acknowledgements	x
Acronyms	xi
Zusammenfassung	xiii
Abstract	xiv
1. Introduction	1
1.1. Overview of the Thesis	2
1.2. Principal Approximations	2
1.2.1. The Born-Oppenheimer Approximation	3
1.2.2. Core Orbitals	4
1.2.3. Model Hamiltonians	5
1.3. The Variational Principle	6
1.4. The Schrödinger Equation as a Matrix Problem	6
1.4.1. Orbitals	8
1.4.2. Density Matrices	8
1.4.3. Electron Correlation	10
1.5. The Hartree-Fock Method	11
1.5.1. Correlation Energy	12
1.5.2. The Roothaan-Hall Equations	12
1.6. Post-Hartree-Fock Methods	13
1.6.1. Configuration Interaction	13
1.6.2. Multi-Configurational Self-Consistent Field	15
1.6.3. Perturbation Theory	16
1.6.4. Coupled Cluster Theory	18
2. Explicitly Correlated Methods	21
2.1. The Cusp Conditions	21
2.2. Hylleraas Methods	22
2.3. Explicitly Correlated Gaussians	24
2.4. F12 Methods	25
2.5. The Transcorrelated Method	25
2.5.1. The Method of Boys and Handy	26

2.5.2. Modern Resurgence	26
2.5.3. Comparison to F12/R12	26
3. Monte Carlo Methods	27
3.1. Classical Monte Carlo Methods	27
3.1.1. A Very Bad Game of Darts	27
3.1.2. A More Mathematical Description	29
3.1.3. Metropolis-Hastings Algorithm	31
3.2. Variational (Quantum) Monte Carlo	32
3.2.1. Trial Wave Functions	33
3.3. Diffusion Monte Carlo	34
3.4. Full Configuration Interaction Quantum Monte Carlo	36
3.4.1. Main Concepts	37
3.4.2. Basic Algorithm	38
3.4.3. Energy Estimators	40
3.4.4. Annihilation Plateaus	41
3.4.5. The Initiator Approximation	41
3.4.6. Reduced Density Matrix Sampling	41
3.4.7. Combining TC with Modern Electronic Structure	44
4. Optimizing Jastrow Factors for the Transcorrelated Method	46
4.1. Introduction	46
4.2. Jastrow Factor	46
4.3. Optimisation Strategy	46
4.3.1. Choosing an Appropriate Sample Size	46
4.4. Results	46
4.4.1. Neglecting Three-Body Excitations	46
4.4.2. Basis Set Convergence	46
4.5. Conclusion and Outlook	46
4.5.1. The xTC Approximation	46
5. The Transcorrelated Method for Multi-Reference Problems	47
5.1. Introduction	47
5.2. Size Consistency	47
5.2.1. Jastrow Factor Forms	47
5.3. CASSCF Ansatz	47
5.4. Transcorrelated Trial Wavefunction	47
5.5. Results	47
5.5.1. Binding Curves	47
5.5.2. Excitation Energies	47
5.6. Conclusion and Outlook	48

6. The PyTCHInt Library	49
6.1. Introduction	49
6.2. Matrix Element Evaluation	49
6.3. Deterministic Optimisation	49
6.4. Performance	49
6.5. Interface	49
6.6. Conclusion and Outlook	49
7. Summary and Outlook	50
List of Figures	51
List of Tables	53
Bibliography	54
A. Appendix (todo)	68
B. Curriculum vitae	69

Erklärung über die Eigenständigkeit der Dissertation

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit

*„Development of the Transcorrelated Full Configuration Interaction Quantum
Monte Carlo Method“*

ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt.

Datum, Ort

Unterschrift

Acknowledgements

TODO: thank people

Acronyms

1RDM	1-electron reduced density matrix	8, 44
2RDM	2-electron reduced density matrix	10, 44
AO	atomic orbital	12
BOA	Born-Oppenheimer approximation	3, 4, 14, 51
CAS	complete active space	5
CASSCF	complete active space self-consistent field	16
CBS	complete basis set	7, 8, 13–15, 21, 51
CC	coupled cluster	18, 19, 45
cc-pVXZ	correlation-consistent polarised valence with X -zeta quality ($X=D, T, Q, 5, 6, \dots$)	21
CCSD	coupled cluster with single and double excitations	19, 20
CCSD(T)	coupled cluster with single and double excitations and perturbative triples	20
CCSDT	coupled cluster with single, double and triple excitations	19
CI	configuration interaction	13, 15, 16, 18, 19, 24, 38, 41, 44
CSF	configuration state function	37
DC	distinguishable cluster	20
DCSD	distinguishable cluster with single and double excitations	20
DMC	diffusion Monte Carlo	34–37
DMRG	density matrix renormalization group	45
ECG	Explicitly Correlated Gaussian	24
EOM	equation of motion	20
FCI	full configuration interaction	13, 15, 16, 19, 36–38

FCIQMC	full configuration interaction quantum Monte Carlo	2, 14, 15, 36–38, 41, 43–45, 51, 52
GTG	Gaussian-type Geminal	24
GTO	Gaussian-type orbital	8, 12
HF	Hartree-Fock	11–13, 15–19, 25, 34
MC	Monte Carlo	27–30, 32, 33
MCSCF	multiconfigurational self-consistent field	15, 16, 36
MO	molecular orbital	12
MP2	second-order Møller-Plesset perturbation theory	18
QMC	quantum Monte Carlo	2, 32, 36, 41
RHF	restricted Hartree-Fock	12, 15, 51
ROHF	restricted open-shell Hartree-Fock	12
SCF	self consistent field	12
SD	Slater determinant	7, 11, 13, 15, 16, 18, 25, 37, 39, 44
STO	Slater-type orbital	8, 9, 22, 51
TC	transcorrelation	2, 44, 45
UHF	unrestricted Hartree-Fock	12
VMC	variational Monte Carlo	25, 32, 33, 36

Zusammenfassung

TODO: Hier werde ich eine Zusammenfassung schreiben.

Abstract

TODO: abstract

Introduction

According to modern quantum theory, to fully describe a (nonrelativistic) system of N_A nuclei and N_e electrons, we must solve the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{x}, t) = \hat{H} \Psi(\mathbf{x}, t) \quad (1.1)$$

where $\mathbf{x} = (\mathbf{r}, \sigma)$, $\mathbf{r} \in \mathcal{R}^{3(N_e+N_A)}$ gives the spatial coordinates, and σ the corresponding spins. For nuclei with atomic numbers Z_I (positive charge Ze), the Hamiltonian operator \hat{H} may be written

$$\begin{aligned} \hat{H} = & - \sum_{I=1}^{N_A} \frac{\hbar^2}{2m_I} \nabla_I^2 + \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{I,J=1; I \neq J}^{N_A} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \\ & - \sum_{i=1}^{N_e} \frac{\hbar^2}{2m_e} \nabla_i^2 + \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{i,j=1; i \neq j}^{N_e} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i=1}^{N_e} \sum_{I=1}^{N_A} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} \end{aligned} \quad (1.2)$$

where N_A and N_e are the number of nuclei and electrons, respectively, and we have represented the nuclear coordinates by \mathbf{R}_I and the electron coordinates by \mathbf{r}_i .

Since quantum chemistry and condensed matter sciences are in general concerned with nonrelativistic processes involving electrons and nuclei, this might boldly be called the *theory of everything*.¹ Hence, we may be tempted to conclude this dissertation early, but in practice the evaluation of the Hamiltonian in equation 1.2 is impossible. First, there is no closed-form solution of equation 1.1 with this Hamiltonian. Second, numerical evaluation of \hat{H} is far from trivial.

Let's consider a simple example, the Argon atom. Say we want to solve this partial differential equation on a grid. Let's choose a very coarse $10 \times 10 \times 10$ grid. Then *at each time step* we need to store $10^{3 \times (18+18)} = 10^{108}$ values, corresponding to all the particle positions and the grid. Considering there are "only" $\sim 10^{80}$ atoms in the known universe,² this is completely unreasonable.

This system has only 18 electrons and 18 nuclei, a far cry from the 10^{23} or higher number of electrons in a typical condensed matter system, for example. Moreover, we have not taken into account floating point precision, or that we would need to calculate

this for possibly many time steps. Clearly, drastic approximations and more sophisticated methods are required.

1.1. Overview of the Thesis

This dissertation fits into the field of nonrelativistic electronic structure theory, the branch of quantum chemistry concerned with the description of electrons and their correlation inside molecules and materials. More specifically, this dissertation focuses on high-accuracy (but often high-cost) *ab initio* methodologies, especially full configuration interaction quantum Monte Carlo (FCIQMC) and transcorrelation (TC). As such, we will only be discussing small systems consisting of only a few atoms, as they are tractable with a full, all-electron treatment with these methods. In principle, these methods should be able to be embedded^{3,4} into more large-scale calculations using multiscale techniques, but this is outside the scope of my work. Nevertheless, the work herein is focused on methodologies, and not on particular physical systems.

The outline of the dissertation is as thus:

- Chapter 1 (this chapter) provides a basic overview of electronic structure theory methods and some of its principal concepts. Sections 1.5 and 1.6 in particular are largely based on the appropriate chapters of Ref 5.
- Chapter 2 reviews the current works in so-called “explicitly correlated” methods, notably the well-established R12/F12 and the recently-reinvigorated TC.
- Chapter 3 provides a basic introduction to quantum Monte Carlo (QMC) and how it relates to FCIQMC.
- Chapter 4 discusses optimization strategies of Jastrow factors in the context of TC.
- Chapter 5 discusses an extension of the methods in the previous chapters to ensure size consistency and success when targeting strongly multi-reference problems.
- Chapter 6 provides an overview of the software `pytchint` developed in the group for evaluation of TC integrals.
- Finally, chapter 7 provides a review and an outlook for the field.

1.2. Principal Approximations

As discussed, to make any progress in electronic structure theory, we must make use of approximations. Of course, we must always be cautious and suspicious when using these,

and make sure they are valid for the systems in question. Thankfully, over the course of the last century, a cornucopia of different approximations has been developed to tackle the Schrödinger question, some of which will be discussed in this section.

1.2.1. The Born-Oppenheimer Approximation

The most important approximation used in electronic structure theory is the Born-Oppenheimer approximation (BOA).⁶ The BOA relies on the fact that the nuclei are much heavier than the electrons, with the mass of a single proton being almost 2000 times the mass of an electron. As an intuitive picture, we may think of the nuclei as moving much slower than the electrons, which can adapt themselves to the instantaneous positions of the nuclei. In mathematical terms, this means we can take the total wave function to be a product of its nuclear and electronic components,

$$\Psi_{\text{total}} = \Psi_{\text{nuc}} \Psi_{\text{elec}}. \quad (1.3)$$

Notice that the first two terms of equation 1.2 are independent of the electronic coordinates and, ipso facto, have no effect on Ψ_{elec} . This leads to the *electronic Hamiltonian* under the BOA, which can be written ~~in atomic units~~ as

$$\hat{H}_{\text{elec}} = - \sum_i \frac{1}{2} \nabla_i^2 - \sum_{i,I} \frac{Z_I}{r_{iI}} + \sum_{i>j} \frac{1}{r_{ij}}, \quad (1.4)$$

where we have simplified notation by using miniscule roman letters for the electrons and capital roman letters for the nuclei, and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $r_{iI} = |\mathbf{r}_i - \mathbf{R}_I|$, as well as by using atomic units.

In the language of second quantisation, equation 1.4 can be written as

$$\hat{H}_{\text{elec}} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} V_{pqrs} a_p^\dagger a_r^\dagger a_s a_q, \quad (1.5)$$

where p, q, r, s are general spin-orbital indices, and a_p (a_p^\dagger) is the annihilation (creation) operator for an electron in spin-orbital p . These must obey the anti-commutation relations

$$\{a_p, a_q^\dagger\} := a_p a_q^\dagger + a_p^\dagger a_q = \delta_{pq} \quad (1.6)$$

$$\{a_p, a_q\} = \{a_p^\dagger, a_q^\dagger\} = 0 \quad (1.7)$$

so that the electrons satisfy the Pauli exclusion principle.

In equation 1.5, h_{pq} is a one-body integral,

$$h_{pq} = \int d^3r \phi_p^*(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 - \sum_I \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}|} \right) \phi_q(\mathbf{r}) \quad (1.8)$$

and V_{pqrs} is a two-body integral,

$$V_{pqrs} = \int d^3r \int d^3r' \phi_p^*(\mathbf{r}) \phi_r^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_q(\mathbf{r}) \phi_s(\mathbf{r}'), \quad (1.9)$$

with a spatial-orbital basis $\{\phi_i(\mathbf{r})\}$.

For our purposes, we neglect the first term of equation 1.2, and treat the second term as approximately constant, so $\hat{H}_{\text{nuc}} = \sum_{IJ} Z_I Z_J r_{IJ}^{-1}$. Thus, we are left with the problem of solving the electronic structure problem, which is the subject of this dissertation. Solving the electronic problem for different nuclear positions gives us the potential energy surface.

Note that while chemists and physicists typically talk about the positions of nuclei in a molecule or solid as if they are fixed in place, as will be done in this dissertation, this is really a colloquialism. If the nuclei had an exact position and zero kinetic energy, the BOA would, ~~ipso facto~~, be in direct contradiction of the Heisenberg uncertainty principle. Instead, on the timescale of the electrons, due to the much higher mass of nuclei, in the BOA we treat the nuclei as approximately localised in a state in which their motion is much slower than that of the electrons (but, importantly, not zero). This keeps the approximation from being in conflict with the fundamental postulates of quantum theory.

The BOA is an immensely practical tool as it substantially simplifies our equations, and in many applications it is an excellent approximation. It will be a fundamental assumption throughout the rest of this dissertation, though it need not always be valid in all of quantum chemistry.

While we have done a lot to drastically reduce the complexity of equation 1.2, equation 1.5 is still intractable for large system sizes, scaling combinatorially with the size of the Hilbert space, as a function of the system size N_e (henceforth N), and the basis set size M . Hence, in addition to using a smaller basis, we still need extra approximations and sophisticated methodologies.

1.2.2. Core Orbitals

In the ground state of a single atom, electrons first occupy the lowest orbitals. The first occupied shells are typically where the electrons are the most tightly bound.*

Since they tend to be further from the nucleus in an atom, the valence (non-core) orbitals are typically the most affected by the introduction of additional atoms in the system. It is therefore often the valence orbitals that are most important in chemical systems. For this reason, we sometimes “freeze” the core orbitals, meaning they remain doubly occupied by electrons, and only correlate the valence orbitals. It is also common to refer to core electrons and valence electrons, meaning those electrons occupying the core and valence orbital space, respectively (though this is technically imprecise language, as electrons are indistinguishable).

*This neglects symmetries: energy may be higher due to the angular momentum while staying spatially closer to the nucleus.

Similarly, we may also delete virtual orbitals, typically those of high energy, to further reduce the size of the problem. The remaining space is known as an active space, and is the basic idea of the complete active space (CAS) methods.

1.2.3. Model Hamiltonians

In 1929, the surrealist artist René Magritte displayed a now-famous painting of a pipe with the caption *Ceci n'est pas une pipe* (French for “This is not a pipe”). It was meant to depict the idea that the painting itself is in a way treacherous: it may appear to be a pipe, but you cannot stuff it or smoke from it, as it is a representation of a pipe.

In a similar way, physicists and chemists often use *model Hamiltonians*, which do away with aspects of equation 1.2 that are not expected to be relevant to the problem at hand, resulting in new Hamiltonians that may be considered a representation of equation 1.2, much like Magritte’s painting.* Compared to *ab initio* Hamiltonians like equation 1.5, model Hamiltonians are generally much simpler, but depend on parameters whose values we may not necessarily know a priori.

The most famous model Hamiltonian is the Hubbard model,⁷ most typically used to describe electrons in a periodic lattice, and can be written as

$$\hat{H}_{\text{Hubb}} = - \sum_{\langle m,n \rangle} \sum_{\sigma} t_{mn} a_{n\sigma}^{\dagger} a_{m\sigma} + \sum_m U_m \hat{n}_{m\uparrow} \hat{n}_{m\downarrow}, \quad (1.10)$$

where $\langle m, n \rangle$ denotes nearest neighbour sites, t_{mn} (often taken to be constant, t) is a parameter called the hopping amplitude, and U_m (often taken to be constant, U) is a parameter called the on-site repulsion. Here we have used a spatial orbital basis labels $\{m, n\}$ and spin labels $\sigma \in \{\uparrow, \downarrow\}$. $\hat{n}_{m\sigma} := a_{m\sigma}^{\dagger} a_{m\sigma}$ is the number operator.

Despite its apparent simplicity, the Hubbard model is a rich model, and has been used to describe a variety of phenomena, from metal-insulator transitions to high-temperature superconductivity. Moreover, it is not always so simple to solve, and has been the subject of much research.^{8,9}

Let’s consider an especially simple version of the model, with only two sites, and open boundary conditions:

$$\hat{H}_{\text{Hubb}} = -t \sum_{\sigma} (a_{1\sigma}^{\dagger} a_{2\sigma} + a_{2\sigma}^{\dagger} a_{1\sigma}) + U(\hat{n}_{1\uparrow} \hat{n}_{1\downarrow} + \hat{n}_{2\uparrow} \hat{n}_{2\downarrow}). \quad (1.11)$$

We could use this as an approximate description of the H_2 molecule, where the electrons are restricted to only the lowest 1s orbitals. Compared to equation 1.2, this is a tremendously easy problem now.

*We might further argue that all of science can be described this way, as scientific models are always “mere” representations of Nature, and are not Nature itself.

Of course, when solving a problem using a model Hamiltonian, we must be careful in choosing the correct one. In particular, the exact solution to the Hubbard model in one dimension⁸ does not appear to be in agreement with the *ab initio* solution of a one-dimensional chain of hydrogen atoms.^{10,11}

1.3. The Variational Principle

One way of finding approximations to the lowest energy eigenstate of a quantum problem like that of equation 1.5 is the variational method. Several methods discussed in this dissertation build on the variational method, and the variational method itself is based on the variational principle, which states¹² for any “trial wave function” $|\tilde{\Psi}\rangle$ and Hermitian operator (such as the electronic Hamiltonian) \hat{H} ,

$$\frac{\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle}{\langle \tilde{\Psi} | \tilde{\Psi} \rangle} \geq E_0 := \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (1.12)$$

where E_0 is the lowest eigenvalue (e.g. the ground state energy). That is, the energy of $|\tilde{\Psi}\rangle$ is an upper bound of the exact energy of the system.

To prove this, consider an expansion of $|\tilde{\Psi}\rangle$ terms of the exact eigenkets $|k\rangle$ of \hat{H} , so that $\hat{H}|k\rangle = E_k|k\rangle$ and $|\tilde{\Psi}\rangle = \sum_k c_k |k\rangle$. Then

$$\begin{aligned} \frac{\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle}{\langle \tilde{\Psi} | \tilde{\Psi} \rangle} &= \frac{\sum_{kk'} c_k^* c_{k'} \langle k | \hat{H} | k' \rangle}{\sum_{kk'} \langle k | k' \rangle c_k^* c_{k'}} \\ &= \frac{\sum_k |c_k|^2 (E_k - E_0)}{\sum_k |c_k|^2} + E_0 \\ &\geq E_0 \end{aligned} \quad (1.13)$$

where, in going from the first to the second line, we used the orthonormality of the eigenbasis $\{|k\rangle\}$ and that the left- and right-eigenvectors are the same. That is, we used the fact the \hat{H} is a Hermitian operator. This is an important observation for future discussions, such as in section 2.5.

1.4. The Schrödinger Equation as a Matrix Problem

Implicit in the operator algebra of the fermionic creation and annihilation operators a^\dagger and a in equation 1.5 is the Pauli exclusion principle: we cannot have two or more fermions occupying identical states. In the context of electronic structure theory and this dissertation, the fermions and states in question are electrons and orbitals, respectively.

This handles the antisymmetry requirement of the fermionic wave function, $\hat{P}_{ij}^\sigma \Psi = -\Psi$, where \hat{P}_{ij}^σ permutes the i th and j th electrons with spin σ .

Our wave function can hence be expanded by a linear combination of Slater determinants (SDs),

$$\Psi = \sum_i c_i |D_i\rangle, \quad \text{where } |D_i\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \cdots & \chi_1(\mathbf{x}_N) \\ \chi_2(\mathbf{x}_1) & \chi_2(\mathbf{x}_2) & \cdots & \chi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(\mathbf{x}_1) & \chi_N(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \quad (1.14)$$

for a spin-orbital basis $\{\chi_p(\mathbf{x}_q)\}$. A spin-orbital is typically written as a product of a spatial part and a spin part, e.g. $\chi(\mathbf{x}) = \phi(\mathbf{r})\omega(\sigma)$.

Given such a finite spin-orbital basis of size $2M$ (M spatial orbitals with 2 different spins each),* the SD expansion is also finite. Therefore, we may write \hat{H} as a square matrix and equation 1.5 as an eigenvalue equation,

$$\hat{H}\Psi = E\Psi. \quad (1.15)$$

Note, however, that in order to make this a finite-dimensional matrix eigenvalue problem, we needed to use a finite spin-orbital basis. Thus, E in equation 1.15 is not the true ground state energy of the system, but indeed only an approximation within that basis set. This is the case even if we were to solve equation 1.15 exactly. In order to get the true ground state energy, we need to reach the complete basis set (CBS), i.e. $M \rightarrow \infty$.

With equation 1.15, we have transformed the hopelessly intractable partial differential equation 1.2 into a finite algebraic eigenvalue problem, which is much better suited for solving on a computer.

If we have N_\uparrow spin-up electrons and N_\downarrow spin-down electrons, we then have

$$|\mathcal{H}| = \binom{M}{N_\uparrow} \binom{M}{N_\downarrow} \quad (1.16)$$

where $|\mathcal{H}|$ denotes the size of the Hilbert space, and we have assumed that the number of spin-up orbitals M is equal to the number of spin-down orbitals. Here, $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ is a binomial coefficient.

The combinatorial scaling is still unfavourable. For a closed-shell system with 20 electrons in 20 orbitals, $|\mathcal{H}| = 3.41 \times 10^{10}$. Assuming we wish to have double precision,[†] to store the whole matrix we would need about 10^{23} bytes, or 100 zettabytes. So, while

*We assume here the same number of spin- \uparrow and spin- \downarrow orbitals for simplicity, though it need not always be the case.

[†]A double precision floating-point number has 1 bit for the sign, 11 bits for the exponent, and 52 bits for digits in the number (known as the mantissa or significand).¹³ This resolves to 64 bits or 8 bytes.

we've substantially reduced the complexity of the problem, it is still extremely demanding. More approximations are therefore needed.

1.4.1. Orbitals

Orbitals (and spin-orbitals), a term already used in this text, describe the spatial distribution (and in the case of spin-orbitals, also the spin) of a single electron. We assume that spatial orbitals form an orthonormal set, and that they form a complete space for an arbitrary wave function. However, in theory to have a CBS would mean needing an infinite number of basis functions. In practice, we must of course have a finite basis, and so this naturally leads to a hierarchy of basis sets, where the typically larger basis sets are both more expensive and more accurate.

Many different basis set families exist, but among the most popular are those due to Dunning and coworkers.¹⁴ Additionally, many different basis set extrapolation techniques have also been developed.¹⁵⁻²²

Strictly speaking, any kind of ^{set of} functions can be used as a basis set, as long as they are a complete set. However, in practice we may broadly categorise them into two groups: Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs).

The exact solution for the Hydrogen atom is the Slater-type function $\frac{1}{\sqrt{\pi}}e^{-r}$, and similarly we know that molecular orbitals also decay as $\sim e^{-\zeta r}$.

However, these STOs are difficult to work with, as we ~~need to evaluate~~ integrals of the form

$$\int d^3r_1 \int d^3r_2 \phi_p^*(\mathbf{r}_1) \phi_r^*(\mathbf{r}_2) \frac{1}{r_{12}} \phi_q(\mathbf{r}_1) \phi_s(\mathbf{r}_2) \quad (1.17)$$

are routinely needed.

Therefore, GTOs have been developed,²³ which have the form $e^{-\alpha r^2}$. GTOs are much easier to handle, as the product of a gaussian is also a gaussian, and the integral of a gaussian is exactly known, so equations like 1.17 have known solutions. With this comes a tradeoff, since we know STOs more faithfully capture the form of molecular orbitals, but GTOs are much better suited for computations. For this, the most common basis sets typically use a linear combination (or “contraction”) of gaussians fitted to a STO. An illustration is shown in figure 1.1.

It must be noted, however, that STOs and GTOs have different asymptotic behaviour for $r \rightarrow \infty$ and $r \rightarrow 0$. The latter in particular will be discussed in section 2.1.

1.4.2. Density Matrices

The 1-electron reduced density matrix (1RDM) is defined by its matrix elements

$$\gamma_{pq} = \langle \Psi | a_q^\dagger a_p | \Psi \rangle, \quad (1.18)$$

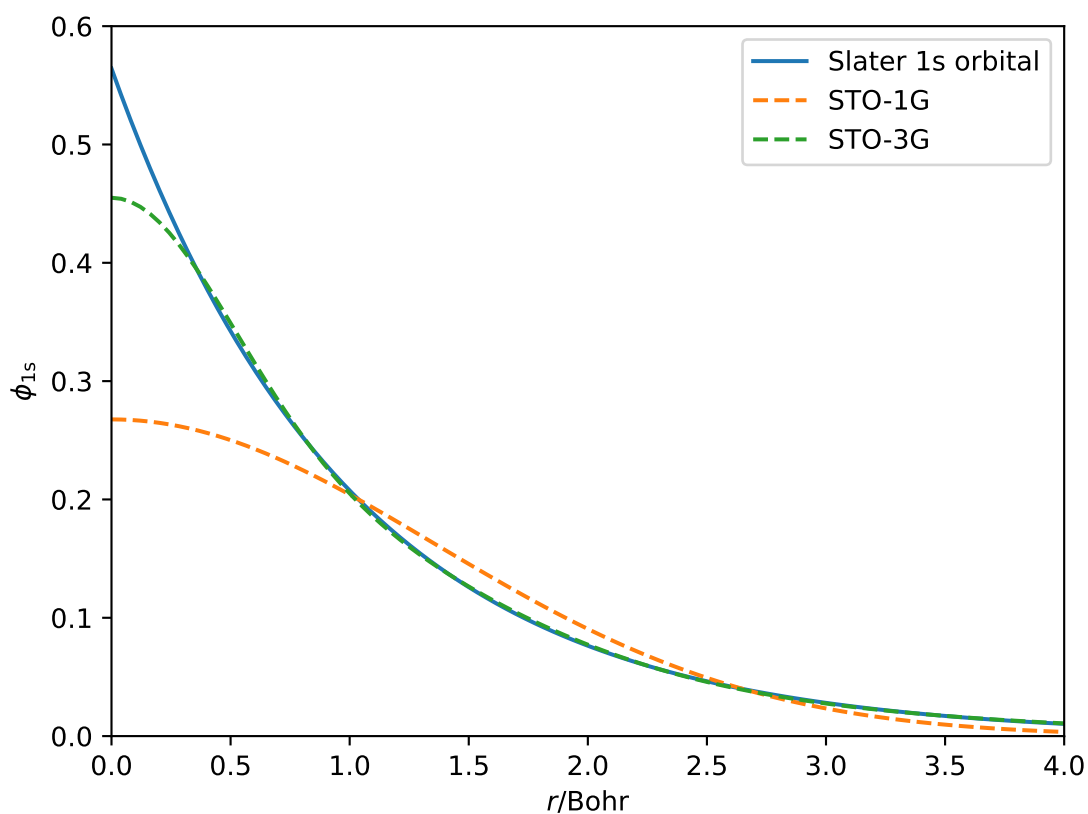


Figure 1.1. Illustration of an s -type orbital and the fitted gaussians (STO- x G means x gaussians are used to approximate the STO). Note that for much of the curve, there is good agreement, but not for the short- and long-range. The short-range behaviour in particular leads to complications, and is the focus of the methods discussed in chapter 2. The long-range behaviour is typically less of a problem, since the integrals we are interested in typically decay with r . Fitting parameters for the gaussians were taken from reference 24.

or written in a spin-orbital basis:

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \sum_{pq} \phi_q(\mathbf{x}_2)^* \phi_p(\mathbf{x}_1) \gamma_{pq} \quad (1.19)$$

Similarly, the 2-electron reduced density matrix (2RDM) is defined as

$$\Gamma_{pqrs} = \langle \Psi | a_s^\dagger a_r^\dagger a_q a_p | \Psi \rangle. \quad (1.20)$$

Any one- and two-electron Hermitian operator Ω can be written (in a spin-orbital basis) as

$$\langle \Psi | \Omega | \Psi \rangle = \sum_{pq} \gamma_{pq} \Omega_{pq} + \sum_{pqrs} \Gamma_{pqrs} \Omega_{rpsq}. \quad (1.21)$$

1.4.3. Electron Correlation

Two variables are independent (or uncorrelated) if the joint probability distribution is the product of their expected values, i.e.

$$P(\mathbf{x}_1, \mathbf{x}_2) = P(\mathbf{x}_1)P(\mathbf{x}_2). \quad (1.22)$$

By Bayes' theorem,^{25,26} this can be rewritten in terms of the conditional probability,

$$P(\mathbf{x}_1 | \mathbf{x}_2) = \frac{P(\mathbf{x}_1, \mathbf{x}_2)}{P(\mathbf{x}_2)} = \frac{P(\mathbf{x}_1)P(\mathbf{x}_2)}{P(\mathbf{x}_2)} = P(\mathbf{x}_1). \quad (1.23)$$

The variables are said to be correlated if the above is not true.

In the case of electronic structure, the variables in question are $\mathbf{x} := (\mathbf{r}, \sigma)$, the spatial coordinates and spin of the electrons. Therefore, when we speak of electron correlation, we are referring to these relations. Furthermore, as electrons are indistinguishable, for every pair $(\mathbf{x}_1, \mathbf{x}_2)$,

$$P(\mathbf{x}_1) = P(\mathbf{x}_2) = \frac{1}{N} \rho(\mathbf{x}) \quad (1.24)$$

$$P(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{N(N-1)} \rho(\mathbf{x}_1, \mathbf{x}_2) \quad (1.25)$$

where $\rho(\mathbf{x})$ is the electron density and $\rho(\mathbf{x}_1, \mathbf{x}_2)$ is the probability of finding electrons at $(\mathbf{x}_1, \mathbf{x}_2)$ simultaneously, known as the pair density,

$$\rho(\mathbf{x}) = N \int d^3x_2 \cdots d^3x_N \Psi^*(\mathbf{x}, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N) \Psi(\mathbf{x}, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N) \quad (1.26)$$

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = N(N-1) \int d^3x_2 \cdots d^3x_N \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N). \quad (1.27)$$

There are two key sources of electron correlation: Fermi correlation, which arises from the fact that electrons obey Fermi statistics, i.e. the wave function is antisymmetric with respect to exchange of \mathbf{x}_1 and \mathbf{x}_2 , and Coulomb correlation, which arises from the fact that electrons repel each other.

1.5. The Hartree-Fock Method

The typical starting point to most *ab initio* electronic structure methods is the Hartree-Fock (HF) method.²⁷⁻²⁹ The key approximation in HF is that we treat the exact N -body wave function solution as approximately a single SD.* Then, we variationally (see section 1.3) optimise the energy of the system with respect to parameters pertaining to the spin-orbitals. We will refer to this SD as $|\kappa\rangle$.

Then, for some reference configuration $|0\rangle$,

$$|\Phi_0\rangle = e^{-\kappa} |0\rangle. \quad (1.28)$$

Here,

$$\kappa = \sum_{pq} \kappa_{pq} a_p^\dagger a_q \quad (1.29)$$

is an anti-Hermitian operator (i.e. $\kappa_{pq}^* = -\kappa_{qp}$) whose matrix elements κ_{pq} are the orbital rotation coefficients and will be the variational parameters. The anti-Hermiticity of κ ensures that the orbital rotation $e^{-\kappa}$ is unitary.

With $\langle\Phi_0|\Phi_0\rangle = 1$, we want

$$E_{\text{HF}} = \min_{\kappa_{pq}} \langle\Phi_0| \hat{H} |\Phi_0\rangle. \quad (1.30)$$

This is a nonlinear equation, and hence the parameters κ_{pq} must be determined iteratively. Since $e^{-\kappa}$ is unitary, the orthonormality of the spin-orbitals is preserved, $\langle\phi_p|\phi_q\rangle = \delta_{pq}$.

The HF method can also be understood as a mean-field theory. That is, we treat the N -body problem as N one-body problems, where a single electron is in an effective “averaged” potential from the other $N - 1$ electrons. This effective one-body Hamiltonian is known as the Fock operator,

$$f = \sum_{pq} f_{pq} a_p^\dagger a_q. \quad (1.31)$$

The Fock operator f may be written as

$$f = h + V_{\text{eff}} \quad (1.32)$$

where the Fock potential’s matrix elements are $V_{\text{eff},pq} = \sum_{i \in \text{occ}} (V_{pqii} - V_{p i i q})$.


*More precisely, the HF approximation is to treat the wave function as a single configuration. In the context of this dissertation, that configuration will always be a SD.

The Hartree-Fock orbitals, which ultimately dictate the parameters for the minimisation procedure, are determined by diagonalising the Fock matrix,

$$f_{pq} = \epsilon_p \delta_{pq} \quad (1.33)$$

where the eigenvalues ϵ_p are known as the orbital energies.

Since f itself depends on the orbitals, the solution must be determined self-consistently. Hence HF is often referred to as a self consistent field (SCF) method.

Often, the orbitals for the \uparrow and \downarrow electrons are restricted to be the same. This is known as the restricted Hartree-Fock (RHF) method. Otherwise, we might instead use the unrestricted Hartree-Fock (UHF) method, where they are treated independently. However, since this breaks spin symmetry, we might have spin contamination. ~~R~~estricted open-shell Hartree-Fock (ROHF) is a variant that can treat open-shell systems which RHF otherwise would not be able to treat while remaining an eigenstate of the S^2 (total spin) operator. 

1.5.1. Correlation Energy

The correlation energy (for a given basis set) is defined as the difference between the exact energy and the HF energy,

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}. \quad (1.34)$$

It should be noted, however, that despite this name the HF method captures Pauli exchange, and hence does actually correlate electrons. Nevertheless, we typically neglect this when we speak of “correlation energy”.

Similarly, the Coulomb hole is defined as the remaining part of the wave function that is not captured by the HF method,

$$\Psi_{\text{hole}} = \Psi_{\text{exact}} - \Psi_{\text{HF}}. \quad (1.35)$$

1.5.2. The Roothaan-Hall Equations

The Roothaan-Hall equations are a matrix representation of the HF approximation.^{30,31} Given a basis set, typically GTOs, the Roothaan-Hall equations are given by the generalised eigenvalue problem,

$$FC = SC\epsilon \quad (1.36)$$

where C is a matrix of coefficients, F is the Fock matrix (which depends on C), S is the overlap matrix (which reduces to the identity matrix for orthonormalised bases), and ϵ is a diagonal matrix of orbital energies. The Roothaan-Hall equations are obtained by expanding the unknown molecular orbitals (MOs) in a basis of known functions, typically atomic orbitals (AOs).

Since this representation is in matrix form, instead of in terms of derivatives and integrals, it is more amenable to conventional computational techniques.

1.6. Post-Hartree-Fock Methods

While the HF method is convenient and can be used efficiently to solve for many electrons, it is oftentimes not sufficient to describe the electronic structure of a molecule. To account for the remaining correlation, numerous post-Hartree-Fock methods (that is, methods run after HF) have been formulated, some of which are described in this section.

Correlation energy is typically categorised as either dynamical or static, although their effects are not mutually exclusive and sometimes their ~~difference~~ ^{distinction} can be hazy.

Dynamical correlation is remaining correlation that arises due to the instantaneous repulsion from the motion (dynamics) of the electrons.

Static correlation, on the other hand, is related to the degeneracy or near-degeneracy of configurations. In general, to accurately describe a system with strong static correlation, a so-called multi-reference method is needed. As will be seen in chapter 5, static correlation plays a big role for molecules at dissociation.

We therefore have two “axes” on which to approach the exact solution to equation 1.5:

- The basis set chosen, where going to a larger basis set (in a systematic way) improves the result, e.g. cc-pVDZ to cc-pVTZ to cc-pVQZ, until eventually reaching the CBS limit.
- The “hierarchy of theories” where the more accurate the method, the closer to the full configuration interaction (FCI) limit (discussed below). Many post-Hartree-Fock methods can be systematically improved, for example by considering additional excitations in a coupled cluster method.

1.6.1. Configuration Interaction

Configuration interaction (CI) is perhaps the most conceptually straightforward post-Hartree-Fock method. Instead of using a single SD, we approximate the wave function as a linear combination of SDs. That is,

$$|\Phi_{\text{CI}}\rangle = \sum_p c_p |D_p\rangle \quad (1.37)$$

where $|D_p\rangle$ is a SD. This may alternatively be written*

$$|\Phi_{\text{CI}}\rangle = (\mathbb{1} + \sum_{ia} c_i^a \hat{a}_a^\dagger \hat{a}_i + \sum_{ijab} c_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j + \dots) |\Phi_0\rangle \quad (1.38)$$

*We adopt the common notation convention where i, j, k, \dots denote occupied orbitals, a, b, c, \dots virtual (or unoccupied) orbitals, and p, q, r, \dots unspecified.

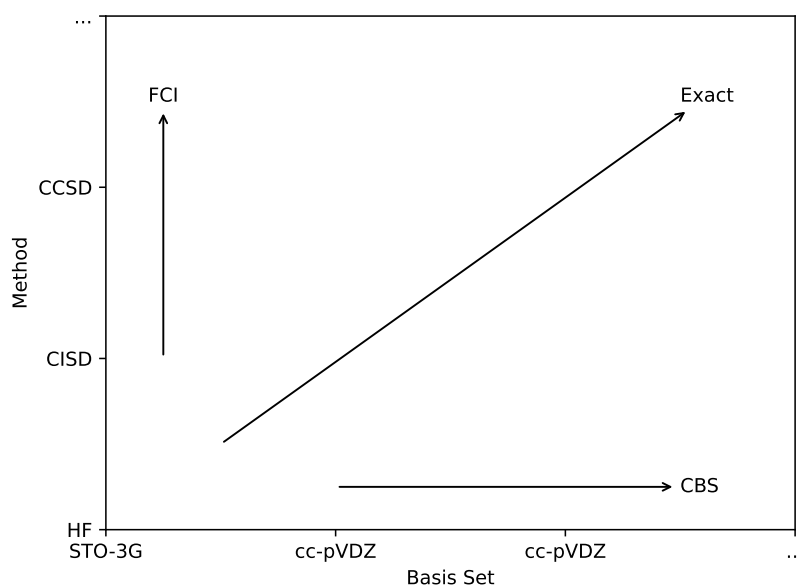


Figure 1.2. A sketch for the hierarchy of theories and basis sets. There are two axes on which to systematically improve until reaching the exact solution to the time-independent Schrödinger equation under the BOA: increasing basis set size (x axis) and increasing the level of theory (y axis). For most of this dissertation, we assume to be at or close to the highest level of theory (the FCI limit), as we mostly employ FCIQMC. We therefore focus on methods of rapidly converging to the CBS limit without the inherently expensive process of adding more basis functions.

where the number of terms in the sum depends on how many excitations we consider, i.e. how far we truncate the CI expansion. The energy and wave function are then found as the lowest eigenvalue and corresponding eigenvector, respectively, of the Hamiltonian written in the basis of these SDs. This can be done, for example, by exact diagonalisation. If all possible excitations are considered, then we have reached the CBS limit, and the solution is exact within that basis set.

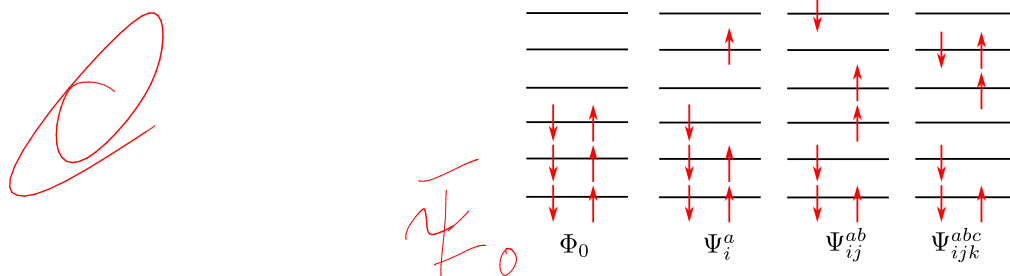


Figure 1.3. Illustration of a RHF reference and examples of single-, double-, and triple-excitations. The method CISDT, for example, would consider all determinants of the forms illustrated.

There are some important limitations to this method, however. Perhaps most notable is that the CI method is not size consistent. For two noninteracting systems A and B , we expect the composite system $A + B$ to have energy equal to the sum of that for A and B , i.e. $E_{A+B} = E_A + E_B$. However, within a truncated CI method, this is not the case*, due to the additive ansatz 1.38.

Additionally, if we wish to go to the FCI ~~method~~ ^{limit} and ensure size consistency and accurate results, we once again have bad scaling, as discussed in section 1.4. However, there exist sparse diagonalisation routines that need not store the entire matrix H , such as the Davidson,³² Lanczos,³³ and Arnoldi³⁴ (for non-Hermitian H) methods, which target the ground state and need to store only a fraction of the matrix. Still, these scale with the size of the eigenvector, which is still prohibitively expensive for large systems. A more practical approach to reaching FCI accuracy will be discussed later in section 3.4, where we discuss the FCIQMC algorithm.

1.6.2. Multi-Configurational Self-Consistent Field

As a generalisation of the HF method, multiconfigurational self-consistent field (MCSCF) is particularly well-suited for problems with static correlation.^{5,35,36} As in HF, we minimise the electronic energy with respect to variational parameters; however, now we also simultaneously minimise with respect to expansion coefficients for a set of configurations

*Of course, in FCI we necessarily have size consistency, since we have the exact solution for the basis set.

(e.g. SDs). i.e. consider the MCSCF wave function

$$|\kappa, \mathbf{c}\rangle = e^{-\kappa} \sum_i c_i |D_i\rangle, \quad (1.39)$$

where κ is familiar from HF, and the c_i coefficients familiar from CI. In MCSCF, we optimise

$$E_{\text{MCSCF}} = \frac{\langle \kappa, \mathbf{c} | H | \kappa, \mathbf{c} \rangle}{\langle \kappa, \mathbf{c} | \kappa, \mathbf{c} \rangle}. \quad (1.40)$$

Like in HF, the variational parameters appear nonlinearly and must be optimised iteratively. Determining an appropriate set $\{|D_i\rangle\}$ tends to be challenging, and even for small systems generating an MCSCF wave function can prove intractable.

Complete Active Space Self-Consistent Field

One “flavour” of MCSCF that has proven particularly successful is complete active space self-consistent field (CASSCF).^{37–40} In CASSCF, instead of inspecting individual configurations, we consider a set of configurations that satisfy a set of criteria. In particular, we partition the orbitals into three sets: the core, active, and virtual (unoccupied) regions. The core orbitals, as alluded to in section 1.2.2, are approximated to be doubly occupied. The virtual orbitals are approximated to always be unoccupied. The active orbitals are the remaining orbitals, which can have occupations between 0 and 2.

The MCSCF expansion is found by considering all possible excitations of the active electrons in the active space. Notice that in the limit of an empty active space, we recover the HF method, and in the limit of an active space containing all orbitals, we recover the FCI method.

1.6.3. Perturbation Theory

Perturbation theory is a set of approximate mathematical methods for solving problems involving small disturbances (perturbations) to a problem with a known solution (the unperturbed problem). If these perturbations are not too large, then the solution of the perturbed problem is close to that of the unperturbed problem, and can be expressed as the solution of the unperturbed problem plus some corrections. Perturbation theory fails, however, if the perturbation is large.

Rayleigh-Schrödinger

Perturbation theory applied to time-independent problems is sometimes referred to as Rayleigh-Schrödinger perturbation theory.^{12,41,42} Consider a Hamiltonian,

$$H_{\text{PT}} = H_0 + \lambda H', \quad (1.41)$$

where H_0 , referred to as the unperturbed Hamiltonian, λ is an arbitrary (real) parameter controlling the strength of the perturbation, and H' is the perturbation. We assume that we know the exact solution to H_0 , such that we have all eigenstates $\{|\Psi_n^{(0)}\rangle\}$ and their corresponding eigenvalues $\{E_n^{(0)}\}$.

To obtain the solution to the true Hamiltonian H , we expand in terms of the perturbation λ ,

$$|\Psi_n\rangle = \sum_{k=0}^{\infty} \lambda^k |\Psi_n^{(k)}\rangle \quad (1.42)$$

and

$$E_n = \sum_{k=0}^{\infty} \lambda^k E_n^{(k)}. \quad (1.43)$$

Since in this work we are generally interested in the ground state, we are typically targeting $n = 0$.

By inserting these expansions into the time-independent Schrödinger equation, we equate terms of the same order in λ , which leads to the corrections to the energy and wave function. We may then truncate the expansion to some order. This allows for a systematically improvable result.

Møller-Plesset

Møller-Plesset perturbation theory⁴³ is a special case of Rayleigh-Schrödinger perturbation theory, and is the variant most commonly seen in quantum chemistry.

In Møller-Plesset perturbation theory, the unperturbed Hamiltonian is chosen to be the Fock operator 1.31. The perturbed Hamiltonian is known as the fluctuation potential and is the difference between the true Coulomb interaction and effective one-electron potential discussed in section 1.5. *the*

That is,

$$H_0 = f, \quad H' = \sum_{i < j} r_{ij}^{-1} - V_{\text{eff}} = H - H_0. \quad (1.44)$$

If we apply Rayleigh-Schrödinger perturbation theory, we find the zeroth order wave function to be the HF wave function,

$$|\Psi_0^{(0)}\rangle = |\Phi_0\rangle, \quad f|\Phi_0\rangle = \sum_{\substack{i \\ E_0^{(0)}}} \epsilon_i |\Phi_0\rangle \quad (1.45)$$

where we have also identified the zeroth-order energy as the sum of orbital energies $E^{(0)} = \sum_i \epsilon_i$. With a bit more busywork, we get

$$E^{(1)} = \left\langle \Psi_0^{(0)} \left| H' \right| \Psi_0^{(0)} \right\rangle = \langle \Phi_0 | H' | \Phi_0 \rangle \quad (1.46)$$

$$E^{(2)} = \sum_{n>0} \frac{|\langle \Psi_n^{(0)} | H' | \Psi_0^{(0)} \rangle|^2}{E_0^{(0)} - E_n^{(0)}} = - \sum_{i>j, a>b} \frac{|\langle \Phi_{ij}^{ab} | H' | \Phi_0 \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \quad (1.47)$$

where $|\Phi_{ij}^{ab}\rangle = a_a^\dagger a_b^\dagger a_i a_j |\Phi_0\rangle$ is a doubly-excited state with respect to the HF wave function, and $\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j = E_0^{(0)} - E_n^{(0)}$ is the energy difference between two eigenstates of the Fock operator.

Some features worth noting about Møller-Plesset perturbation theory:

- The sum of zeroth- and first-order energies is the HF energy: $E_{\text{HF}} = E_0^{(0)} + E_0^{(1)}$.
- By the variational principle, $E_0^{(0)} < E_n^{(0)}$ for all $n > 0$ (except for degenerate ground states), so $E_0^{(2)} < 0$, i.e. the energy always decreases.
- If the ground state is degenerate, the term diverges. It might be possible to lift the degeneracy by a change of basis, however.
- If the HF solution is already a good approximation for the system, then Møller-Plesset perturbation theory can provide surprisingly good (and size consistent) results. Hence, this method is typically only applicable to single-reference problems (i.e. those without strong static correlation).

While higher-order approaches exist and see use, the most popular is second-order Møller-Plesset perturbation theory (MP2), due to its excellent compromise between cost and accuracy. That is, for only a little extra work after a successful HF, applying MP2 can improve our results considerably.

1.6.4. Coupled Cluster Theory

The lack of size consistency in CI theory arises from the linear ansatz in section 1.6.1. By using an exponential ansatz, we arrive at one of the most successful theories in electronic theory, coupled cluster (CC).^{44–47} While the methods presented here are inherently single-reference and build on a single SD, multi-reference generalisations to CC theory exist and are the study of active research.^{48–53} Therefore, the presented “flavours” of CC are generally expected to fail for systems with strong static correlation.

instead

Standard Coupled Cluster

To ensure a multiplicatively-separable wave function, we use a multiplicative ansatz,

$$|\Psi_{CC}\rangle = e^T |\Phi_0\rangle, \quad (1.48)$$

where $|\Phi_0\rangle$ is the reference (typically HF) wave function, and T is the cluster operator. The cluster operator is defined $T = T_1 + T_2 + T_3 + \dots$, where T_n is the n -body cluster operator, made up of all possible n -body excitations, for example, $T_1 = \sum_{i < j} t_{ij} a_i^\dagger a_j^\dagger a_j a_i$ for the one-body cluster operator, and $T_2 = \sum_{i < j, a < b} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i$ for the two-body cluster operator.

It is worth noting that CC and CI are identical, differing only in their parametrisation, when neither are truncated. They both provide FCI-level accuracy.

By inserting equation 1.48 into the time-independent Schrödinger equation and pre-multiplying by e^{-T} , we obtain

$$\langle \Phi_0 | e^{-T} H e^T | \Phi_0 \rangle = E \quad (1.49)$$

$$\langle \Phi_{ij\dots}^{ab\dots} | e^{-T} H e^T | \Phi_0 \rangle = 0 \quad (1.50)$$

where $|\Phi_{ij\dots}^{ab\dots}\rangle$ is a n -body excitation with respect to the reference wave function.

Note, however, that these equations do not rely on the Variational Principle. Instead of minimising a functional, we solve equations 1.50 for the amplitudes, which give the energy. While we still cannot reasonably always include all excitations in the cluster operator T , due to the multiplicative ansatz, even if we truncate the cluster operator, we get size-consistent results. In particular, any truncated CC wave function ansatz will contain contributions from all determinants in Fock space. The simplest CC truncation commonly used is the coupled cluster with single and double excitations (CCSD) method. In a sense, CCSD and any other truncated CC method is an approximation for FCI where the coefficients are not approximated to be zero (like in CI) but instead the higher-order excitations terms are generated from the lower-order excitation terms in a size-consistent fashion.

CC theory owes much of its success to its high accuracy at a reasonable cost. CCSD and coupled cluster with single, double and triple excitations (CCSDT) in particular scale as $\mathcal{O}(M^6)$ and $\mathcal{O}(M^8)$ respectively, with M the number of spatial orbitals. Every additional set of excitations incurs an additional factor of $\mathcal{O}(M^2)$ scaling.

T is wrong

Distinguishable Cluster Theory

The more recently-developed distinguishable cluster with single and double excitations (DCSD) method's central philosophy is removing irrelevant or nonphysical terms while maintaining many of CC's desirable properties, such as size consistency.⁵⁴

This method was shown to give quantitatively better results than standard CC for a variety of systems, and qualitatively correct results for systems where CC can fail spectacularly, such as the N_2 binding curve. Since its first publication, distinguishable cluster (DC) has been extended to explicitly-correlated F12 theory,⁵⁵ equation of motion (EOM),⁵⁶ tailored coupled cluster,⁵⁷ and more.

Combining Coupled Cluster with Perturbation Theory

It is possible to avoid the computationally-expensive scaling of CC somewhat, by treating some terms perturbatively. The most famous of these is the coupled cluster with single and double excitations and perturbative triples (CCSD(T)) method,⁵⁸ in which the single- and double-excitations are treated fully, whereas the triple-excitations are treated as perturbations, as in section 1.6.3.

CCSD(T) scales as $\mathcal{O}(M^7)$ and for the price provides extremely accurate results. The perturbative correction is a step that is applied only once at the end of a CCSD calculation. This is why sometimes this method is referred to as the “gold standard” of quantum chemistry.

Explicitly Correlated Methods

With the advent of modern computers combined with a vast array of sophisticated algorithms from which to choose, *ab initio* quantum chemistry has become a tremendously powerful tool, going beyond the study of small atoms, to molecules and solids, and are among the most effective and systematically improvable techniques to date. Nevertheless, convergence to the CBS limit is notoriously slow.

In particular, consider the popular basis set family developed by Dunning and coworkers, correlation-consistent polarised valence with X -zeta quality ($X=D, T, Q, 5, 6, \dots$) (cc-pVXZ).^{14,21,59–61} The size of these basis sets scale as $M \in \mathcal{O}(X^3)$, and since for standard post-Hartree-Fock discussed in chapter 1.6 we require four-index integrals, our computation time will scale at least as $t \in \mathcal{O}(X^{12})$.⁶²

Meanwhile, the CBS correlation error scales as $\epsilon \in \mathcal{O}(X^{-3})$ ^{16,18} or $\epsilon \in \mathcal{O}(M^{-1})$,⁶³ resulting in $t \in \mathcal{O}(M^{-\frac{1}{4}})$. Thus, the methods discussed so far come with the painful cost of requiring very large basis sets to approach high-accuracy results.

Explicitly correlated methods are a class of electronic structure methods specifically designed to address this unfavourable scaling by explicitly including the interelectronic distance r_{12} , and is the subject of this chapter. As the R12/F12 family of methods is the most mature of the explicitly correlated electronic structure methods, many reviews focusing on this topic already exist. This chapter in particular is in large part based on three reviews: references 62, 64 and 65.

2.1. The Cusp Conditions

Consider two charged point particles in a system described by the Hamiltonian of equation (1.5). By the Schrödinger equation, the local energy

$$E_L := \frac{H\Psi}{\Psi} \quad (2.1)$$

must be constant in the exact solution. However, when these two particles coalesce, i.e. $r_{12} \rightarrow 0$, the Coulomb potential, r^{-1} , diverges. Thus, for the local energy to be constant, we must have that near coalescence points, the kinetic energy exactly cancels the Coulomb energy. A more formal treatment of this argument leads to the electron-electron Kato

cuspid condition,⁶⁶

$$\left. \widetilde{\frac{\partial \Psi}{\partial r_{12}}} \right|_{r_{12} \rightarrow 0} = \frac{1}{2} \Psi(r_{12} = 0) \quad (2.2)$$

where the tilde represents spherical averaging.

This cusp condition was also later generalised.^{67,68} Early literature on the subject suggested that the success of explicitly correlated methods were due to the superior description of short-range correlation effects, and in particular in their much more faithful capturing of the cusp conditions like equation 2.1.^{69–72} However, further study found that the correlation error from a bad description of the wave function around a small sphere centred on the cusp is actually negligible.^{62,73–75} Instead, the success of explicitly correlated methods is actually due to the superior description of the overall shape and size of the Coulomb hole, which has a radius on the order of the atomic radius.

To understand why gaussian-type basis sets fail so spectacularly at capturing the cusp behaviour, it is instructive to consider a simpler example, like that of approximating $|x|$ by its Fourier decomposition. Such an illustration is found in figure 2.1. It is also worth noting that STOs do not suffer as badly from this limitation,⁷⁶ but as discussed in chapter 1.4.1, they are unsuccessful due to their lack of practicality.

2.2. Hylleraas Methods

Almost 30 years prior to Kato’s landmark paper describing cusps in the analytical form for the wave function, there was already work being done to understand the significance of including r_{12} terms in the wave function. Instead of being motivated by rigorous mathematics like Kato’s, Slater was motivated by studies in the He atom. In particular, he tried to construct a wave function which faithfully represents both the core region as well as the Rydberg limit (i.e. a highly excited atom where the electron is very far from the nucleus).^{76–79} This led him to suggest multiplying the wave function by a factor

$$e^{-2(r_1+r_2)+r_{12}/2}, \quad (2.3)$$

which can easily be shown to satisfy the cusp equation (2.2).

However, the first successful explicitly correlated electronic structure calculation is typically attributed to Hylleraas,⁶⁵ where he aimed to improve convergence of orbital expansions for helium.^{80,81} In this method, the coordinates $s := r_1 + r_2$, $t := r_1 - r_2$ and $u := r_{12}$ are used to construct the wave function

$$\Psi_N(s, t, u) = e^{-\alpha s} \sum_k^N c_k s^{l_k} t^{2m_k} u^{n_k}. \quad (2.4)$$

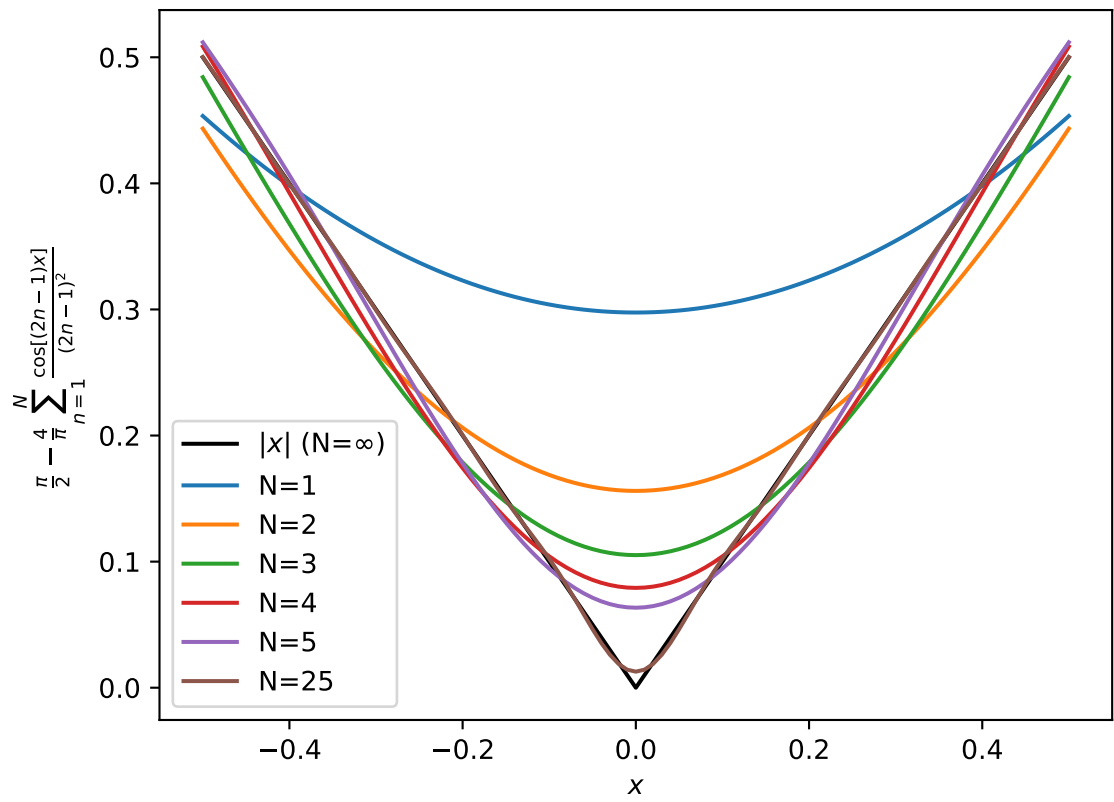


Figure 2.1. A toy example of the Coulomb cusp. Here, the Fourier expansion $x \approx \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=1}^N \frac{\cos[(2n-1)x]}{(2n-1)^2}$ is plotted for a few values of N , including the exact solution. As can be seen, even for many terms, the Fourier expansion is a poor descriptor in the cusp region. Indeed, the only way to describe it exactly is with an infinite number of terms.

In particular, using only three terms ($N = 3$), and variationally optimising for the parameters c_1, c_2, c_3 , Hylleraas was able to reach within 1.3 millihartree from the exact result.

Since then, there was rapid development on this approach and combining it with CI (which came to be known as the CI-Hyl methods).^{82–90} In CI-Hyl methods, the wave function is expanded as in CI,

$$\Psi = \sum_k c_k \Phi_k \quad (2.5)$$

where

$$\Phi_k = \mathcal{A} r_{ij}^{\nu_k} \prod_i \chi_{k_i}(\mathbf{x}_i) \quad (2.6)$$

where χ_k is a spin-orbital basis and \mathcal{A} is the antisymmetriser operator.

However, CI-Hyl methods were to eventually fall out of favour. This is because the expansions involve exceedingly difficult integrals involving many electrons and over products of correlation factors. This significantly restricts the tractability and scalability of the method, and it has since largely gone unused.

2.3. Explicitly Correlated Gaussians

Boys⁹¹ and Singer⁹² independently introduced gaussian basis functions with explicit correlation for calculations on molecules.⁹³ These methods are referred to as Explicitly Correlated Gaussians (ECGs), or in the case of functions of two electrons, Gaussian-type Geminals (GTGs). A spherical GTG may be written (with nuclear coordinates \mathbf{R}_1 and \mathbf{R}_2) as

$$g(\mathbf{r}_1, \mathbf{r}_2) = \exp(-\zeta_1 |\mathbf{r}_1 - \mathbf{R}_1|^2 - \zeta_2 |\mathbf{r}_2 - \mathbf{R}_2|^2 - \gamma |\mathbf{r}_1 - \mathbf{r}_2|^2). \quad (2.7)$$

This can be interpreted as two s -type orbitals and an interelectronic correlation factor $e^{-\gamma r_{12}}$. While they do not have the correct cusp behaviour, much like how gaussian basis functions approximately capture the electron-nuclear cusp, a linear combination of GTGs approximately capture the electron-electron cusp.

One major strength of this approach is that all integrals have closed-form algebraic expressions,⁹⁴ and avoids nonlinear optimisation.^{95,96} ECG methods have been extended to post-Hartree-Fock methods, such as MP2^{97,98}, and methods to avoid its many difficult integrals have also been developed.^{99–103}

While not as popular as F12 methods (see section 2.4), ECGs have been used for highly accurate variational calculations,¹⁰⁴ as well as for applications outside of standard electronic structure theory, such as bosons,¹⁰⁵ positronium (a bound state of an electron and a positron),¹⁰⁶ and non-Born-Oppenheimer systems.¹⁰⁷

2.4. F12 Methods

The most influential class of explicitly correlated methods to date are the F12 methods. The core principle of these methods is to augment the wave function from a conventional (typically SD) basis with an explicitly-correlated correction, called the F12 (or R12) correction. The original formulation **TODO: cite Kutzelnigg** parametrised a two-electron system (such as helium) wave function as

$$|\Psi\rangle = (1 + tQ_{12}r_{12})|\Psi_0\rangle + \sum_p t_p |\Psi_p\rangle \quad (2.8)$$

where Ψ_0 is a reference wave function (such as HF), $|\Psi_p\rangle$ are excited-state SDs, t, t_p are parameters (“amplitudes”) to be optimised, and

$$Q_{12} := \sum_{\alpha\beta} |\alpha\beta\rangle \langle\alpha\beta| \quad (2.9)$$

is referred to as a “strong orthogonality projector”, which ensures the r_{12} term is orthogonal to the reference and singly-excited determinants. Here α, β refer to virtual orbitals in the formally complete basis. The fact that, due to Q_{12} , the r_{12} term commutes with the standard excitation operators aids in including this explicit correlation into conventional electron correlation methods.

TODO: ... see review papers haettig pg 30 for basic intro then 35 - 50 then excited state papers

TODO: revisit brief mention of how it builds on gaussian-type geminals

TODO: make sure to include some excited state information

TODO: mention but maybe don't go into too much detail on MR-F12 methods

2.5. The Transcorrelated Method

TODO: brief recap of Boys-Handy method, some extensions by other people, and how it is used in our group

Hirschfelder first introduced a similarity transform method¹⁰⁸

then was worked on by Boys and Handy **TODO: cite**, to be discussed here

uses ideas from variational Monte Carlo (VMC), discussed more in section 3.2.

recent renewal of interest (lots and lots of citations) with various applications, including DMRG, quantum computing, etc.

will talk in more detail about the Jastrow factor in section 3.2.1 since it is also used in vmc

TODO: also mention biorthogonal orbitals etc.

TODO: short proof of isospectrality

TODO: reference section 1.3 and how it is not trivial to make TC variational since the proof of the variational principle uses the Hermiticity of the operator

2.5.1. The Method of Boys and Handy

2.5.2. Modern Resurgence

TODO: Also cite Werner and quantum computing people

2.5.3. Comparison to F12/R12

discuss additive vs multiplicative ansatz

have some performance comparison plots, maybe can ask permission to reproduce from Evelin or someone else

also discuss many-body integrals for f12 vs at most 3-body for tc

TODO: Maybe have a picture of basis set convergence? Maybe even just HF, compare F12 vs TC vs Hyl(?) vs standard

Monte Carlo Methods

Monte Carlo (MC) methods are a class of numerical methods that use random sampling to numerically solve problems. It has found applications in an impressive range of fields, from physics to finance.^{109–114} It is particularly useful for problems with high dimensionality, where deterministic methods are often impractical. In quantum chemistry and physics, since a ‘dimension’ can refer to any degree of freedom, high-dimensional problems are commonplace, and so MC methods are a natural choice.

While the name *Monte Carlo* was coined by Stanislaw Ulam, after the famous casino in Monaco,^{115,116} the foundational concept was already developed in the 18th century by the French mathematician Georges-Louis Leclerc, Comte de Buffon. As one of the earliest example applications, in the Buffon needle problem, one can randomly toss needles onto a lined sheet of paper and determine π .^{117–119}

Monte Carlo methods is a broad term, and as such it is not possible to give a comprehensive overview in a single chapter, and there exist many reviews and textbooks on MC and related topics.^{120–125} Here, we will focus on only a few concepts particularly relevant for this dissertation, largely following reference 113 and the relevant chapters of reference 114.

3.1. Classical Monte Carlo Methods

We start our discussion with classical MC methods. While there are numerous possible applications, notably in molecular dynamics,¹²⁰ here we restrict ourselves to the topic of Monte Carlo integration. In particular, we consider the classical textbook problem of calculating the value of π , then we provide a more rigorous framework.

3.1.1. A Very Bad Game of Darts

If we imagine throwing darts at a dartboard randomly, we can approximate π . If the radius of the circle is r , then its area is πr^2 . The area of the square circumscribing the circle is $4r^2$. Therefore, the ratio of the area of the circle to the area of the square is $\pi/4$. If we randomly sample a point in the square (“throw a dart”), the probability that the point is inside the circle is proportional to its area. Since we sample inside the square,



the probability of landing inside the circle is

$$P(\text{inside circle}) = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}. \quad (3.1)$$

Therefore, if we sample a large number of points, the ratio of the number of darts that land inside the unit circle to the total number of darts, we can approximate the probability distribution P and thus get an estimate for π as

$$\pi \approx 4 \frac{N_{\text{in}}}{N_{\text{out}}}. \quad (3.2)$$

This is illustrated in figure 3.1, and captures the core philosophy of MC methods.

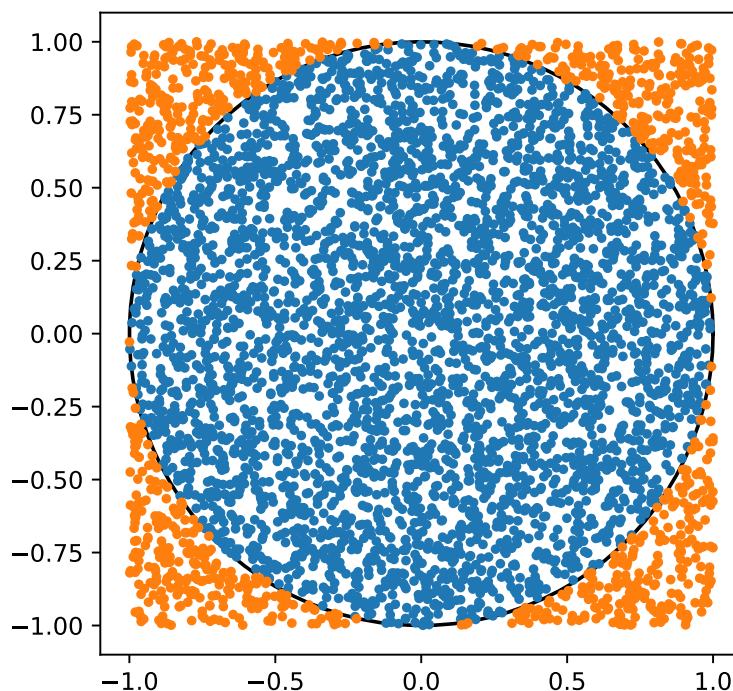


Figure 3.1. Our “game of darts”. Points inside the unit circle are coloured blue whereas points outside are orange. Using stochastic sampling, this naive approach uses 5000 randomly generated numbers between -1 and 1 to approximate $\pi \approx 4N_{\text{in}}/N_{\text{out}} \approx 3.1464$.
coordinates $(x, y) \in [-1, 1] \times [-1, 1]$

~~As~~ controlling the stochastic error is critical in MC methods. ^{To} estimate the reliability of our estimate, we can determine the standard deviation of the estimate. Our sampling scheme is a binomial distribution, with $p = \pi/4$. Hence,

$$\sigma_{\pi}^2 = \text{Var}(4P(\text{inside circle})) = \frac{16p(1-p)}{N} \quad (3.3)$$

or

$$\sigma_\pi = 4\sqrt{\frac{\frac{\pi}{4}(1 - \frac{\pi}{4})}{N}} \propto \frac{1}{\sqrt{N}}. \quad (3.4)$$

3.1.2. A More Mathematical Description

As we have expressed the problem of the previous section in terms of areas, it is clear that it can also be formulated in terms of integrals. For this particular problem, we have:

$$\pi = \int_{-1}^1 dx \int_{-1}^1 dy \Theta(1 - x^2 - y^2), \quad (3.5)$$

where Θ is the Heaviside step function. More generally, consider the integral of some smooth function f over $[a, b] \subseteq \mathbb{R}$,*

$$I = \int_a^b dx f(x). \quad (3.6)$$

Standard finite element methods for solving integrals of this type typically involving dividing the integration domain into N subintervals of length h and determining the weights w_i from e.g. a polynomial approximation. That is,

$$I \approx \sum_{i=1}^N w_i f(x_i). \quad (3.7)$$

The error σ in these sorts of methods is typically $\sigma \propto h^{-k} \propto N^{-k}$, where $k \in \mathbb{Z}_{>0}$. For a multi-dimensional integral, $\sigma \propto N^{-k/d}$, where d is the number of dimensions.¹³

In MC methods, $\forall i$ take $w_i = 1$ and $x_i \in [a, b]$ randomly sampled. That is,

$$I \approx \sum_{i=1}^N f(x_i). \quad (3.8)$$

If, for example, we choose N to be randomly sampled, then the variance is ~~then~~

$$\sigma^2 = \left\langle \left(\frac{b-a}{N} \sum_{i=1}^N f_i \right)^2 \right\rangle - \left(\left\langle \frac{b-a}{N} \sum_{i=1}^N f_i \right\rangle \right)^2 \quad (3.9)$$

$$= \frac{(b-a)^2}{N} (\bar{f}^2 - \bar{f}^2) \quad (3.10)$$

where $f_i := f(x_i)$ and x_i is a random number drawn, the angular brackets denote an average over all possible realisations, and the overbar represents an average of the function over the domain ($[a, b]$ in this discussion). i.e. the error in this method is proportional to the variance of f . Perhaps more interestingly, $\sigma \propto N^{-1/2}$, in line with the central limit

*We present the one-dimensional case for simplicity. The generalization to higher dimensions is straightforward.

theorem.¹²⁶ Comparing this error with standard quadrature, we see that MC integration is more efficient than an order- k algorithm when $d > 2k$. That is, although this particular MC example is naive, using simply a uniform distribution, it is still more efficient than standard methodologies for very large dimensions.

There exist several methods to reduce errors in MC methods.¹²⁷ Among the most important ones is importance sampling.¹²⁸ In the previous example, it is clear that if significant contributions to the integral come from a small region of the domain, only a few points would be sampled by the MC algorithm there when using a uniform distribution. This would lead to large statistical errors. Mathematically, this is from the large variance of the function. In importance sampling, we sample from a distribution p which has roughly the same shape as f such that f/p is roughly constant over the integration domain. Of course, being a probability distribution, we require

$$p(x) > 0 \quad \forall x \quad (3.11)$$

and

$$\int dx \, p(x) = 1. \quad (3.12)$$

The integral is

$$I = \int dx \, f(x) = \int dx \, \frac{f(x)}{p(x)} p(x). \quad (3.13)$$

Then, if we sample points according to p , we have

$$I \approx \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)}, \quad (3.14)$$

where the naive MC method is recovered when $p(x) = 1/(b-a)$, i.e. the uniform distribution.

In this case, the variance in the result is

$$\sigma^2 = \left\langle \left(\frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)} \right)^2 \right\rangle - \left(\left\langle \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)} \right\rangle \right)^2. \quad (3.15)$$

From this, we can see that if f/p is constant, the error vanishes, so it is important to choose a good p . If p is chosen poorly, the variance can be worsened, so this is a delicate problem.

In practice it is often difficult to estimate f , and a choice for p is problem-specific. However, when we have some approximation for its overall shape, importance sampling can be a powerful tool. Alternatively, another method known as adaptive Monte Carlo¹²⁹ seeks the most significant regions of f by random sampling, so that no a priori knowledge of the functional form is required.

3.1.3. Metropolis-Hastings Algorithm

A particularly successful method for generating random samples x_i from a probability distribution $\pi(x)$, where direct sampling of π is difficult, is the Metropolis-Hastings algorithm.^{116,130} Often, $\pi(x)$ is known only up to a normalisation constant, $\pi(x) = \pi(x)/C$ where $C = \int dx \pi(x)$ is intractable.

The Metropolis-Hastings algorithm is based on Markov chains,¹³¹ and we construct such a Markov chain so that its stationary distribution is $\pi(x)$, which is to say if the Markov chain starts at $\pi(x)$ for step t , it is still $\pi(x)$ for step $t + 1$.

The probability of having a sequence of states x_1, x_2, \dots, x_N is

$$p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2|x_1)p(x_3|x_2) \cdots p(x_N|x_{N-1}). \quad (3.16)$$

where $p(x|x')$ is the transition probability from $x' \rightarrow x$. Then, the probability at step t to be in state x , $\pi_t(x)$ is given by

$$\pi_{t+1}(x) = \sum_{x' \in \Omega} \pi_t(x')p(x|x') \quad (3.17)$$

where Ω is the set of all possible states. This is called the master equation. In the stationary state, $\pi_t(x) = \pi(x)$, so

$$\pi(x) = \sum_{x' \in \Omega} \pi(x')p(x|x'). \quad (3.18)$$

Finding the general solution to this problem is nontrivial. However, a sufficient (but not necessary) condition is called detailed balance:

$$\pi(x)p(x|x') = \pi(x')p(x'|x). \quad (3.19)$$

This ensures that the probability of going from x to x' is the same as the probability of going from x' to x , which implies that the probability is stationary.

In order to actually construct the algorithm, we must introduce the trial step probability $\omega(x|x')$, and the acceptance probability $A(x|x')$. Then,

$$p(x|x') = \omega(x|x')A(x|x'). \quad (3.20)$$

$\omega(x|x'), A(x|x') \in [0, 1]$ for each pair x, x' , and $\sum_{x'} \omega(x|x') = 1$. Furthermore, the original formulation of the algorithm required $\omega(x|x') = \omega(x'|x)$, which leads to (plugging into the master equation)

$$\frac{A(x|x')}{A(x'|x)} = \frac{\pi(x')}{\pi(x)}. \quad (3.21)$$

The algorithm proceeds in two stages: we propose a step $x' \rightarrow x$, and then accept or reject it. The probability of accepting is

$$A(x|x') = \min \left(1, \frac{\pi(x')}{\pi(x)} \right). \quad (3.22)$$

If we accept, we set the new state to x , otherwise we stay at x' .

In practice, on a computer, accepting is done by generating a random number $r \in [0, 1)$ and accepting if $r < \frac{\pi(x')}{\pi(x)}$, and otherwise rejecting. Furthermore, we don't just have a single Markov chain, but a collection of so-called "walkers", that each perform their own Markov chain (for parallelisation). The integrand is then sampled at each position where the walkers reach.

One final note, is that since the current state of a Markov chain is dependent on the previous state, the Markov chain is not independent of itself. This is referred to as autocorrelation. One method to reduce this autocorrelation and give essentially independent samples is known as blocking.¹³²

3.2. Variational (Quantum) Monte Carlo

As our first foray into QMC, we consider VMC. In section 1.3 we introduced the variational principle by parametrising the wave function and then finding the minimum of the expectation value of the energy occurring in the parametrisation space. If we have a large number of electrons and/or a large number of parameters, then the integrals involved in the evaluation of the energy will necessarily be high-dimensional. This is where MC comes in. For basic trial wave functions and small atoms like Hydrogen or Helium, direct integration may be possible, but as discussed in the previous section, this quickly becomes impossible for larger systems.

For a trial wave function Ψ (we omit the tilde from section 1.3), the expectation value of the energy is

$$\langle E \rangle = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (3.23)$$

$$= \frac{\int d^3N r \Psi^* H \Psi}{\int d^3N r |\Psi|^2} \quad (3.24)$$

$$= \frac{\int d^3N r \Psi^* \Psi \frac{H\Psi}{\Psi}}{\int d^3N r |\Psi|^2} \quad (3.25)$$

$$= \frac{\int d^3N r |\Psi|^2 E_L(\mathbf{r})}{\int d^3N r |\Psi|^2}, \quad (3.26)$$

where we have defined the local energy as

$$E_L(\mathbf{r}) := \frac{H\Psi}{\Psi}. \quad (3.27)$$

This rewriting of the integral is particularly suitable for *evaluation* using MC. Then we may vary the parameters and stop according to some minimisation algorithm. Notice that the more strongly Ψ resembles the exact wave function, the less strongly E_L varies with \mathbf{r} . In particular, if Ψ is equal to an exact eigenstate, then E_L is constant. Therefore, an alternative objective function to the energy expectation of equation 3.23 is the variance.^{133–135}

Defining

$$p(\mathbf{r}) = \frac{|\Psi|^2}{\int d^{3N}r |\Psi|^2}, \quad (3.28)$$

we may write equation 3.23 as

$$\langle E \rangle = \int d^{3N}r p(\mathbf{r}) E_L(\mathbf{r}). \quad (3.29)$$

This form of the energy is amenable to the Metropolis-Hastings approach outlined in section 3.1.3. Recall also that p need not be normalised when using the Metropolis-Hastings algorithm for sampling.

In principle, the algorithm is doable with just a single walker, but in practice we reduce the statistical error by using many. The algorithm then becomes

```

Initialise  $N$  walkers in random positions.
until convergence criterion met*
  for each walker at position  $r$ 
    sample the local energy  $E_L$  at  $r$ 
    propose a new position  $r'$  with probability  $p = |\Psi(r')/\Psi(r)|^2$ 
    if  $r'$  is accepted
      set  $r$  to  $r'$ 

```

The energy is then calculated as the expectation value of the local energy, averaged over the samples generated in this procedure. Steps at the beginning (before equilibrium) are discarded in a process called equilibration,¹³⁶ and a blocking procedure should be employed. ~~Decision~~ *The decision* to stop is generally based on compute time and/or precision required.

3.2.1. Trial Wave Functions

While VMC is a powerful tool, the quality of the solution is constrained by the quality of the trial wave function. Moreover, the evaluation of the trial wave function is expensive, and we therefore want a form that is easy to evaluate.

The form of the trial wave function is typically chosen based on a Jastrow factor, as already introduced in section 2.5. Generally,¹³⁷

$$\Psi_{\text{trial}} = e^J \Phi. \quad (3.30)$$

*The convergence criterion in VMC is typically just to sample a predefined number of configurations.

For computational efficiency, the Jastrow factor typically only retains one- and two-body terms,¹¹³

$$J = \sum_i \chi(\mathbf{x}_i) - \frac{1}{2} \sum_{i \neq j} u(\mathbf{x}_i, \mathbf{x}_j), \quad (3.31)$$

where χ describes electron-nuclear correlation and u describes two-electron correlation (including e.g. electron-electron-nuclear correlation). There exist a number of more specific forms for J ,¹³⁸ though one unifying principle is for them to adhere to expected short-range (cusp conditions) and long-range ($1/r$) behaviour.

Φ is typically chosen to be a single Slater determinant.^{113,139–142} In particular, by choosing the HF determinant, typically only a small portion of the energy (that is, the correlation energy) is left for the Jastrow factor to describe, and typically even simple Jastrow factors can do this.¹¹³ However, sometimes this is not enough, for example, when the HF determinant is not sufficient to describe all symmetries of the true wave function.^{143–145}

3.3. Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) starts by performing a Wick rotation (that is, substitute $t \rightarrow i\tau$) on the time-dependent Schrödinger equation, equation 1.1. We then get the imaginary-time Schrödinger equation

$$\frac{\partial}{\partial \tau} \Psi(\mathbf{r}, \tau) = -\hat{H} \Psi(\mathbf{r}, \tau). \quad (3.32)$$

This equation is a diffusion (or heat) equation, and Ψ may be interpreted as the density distribution for a large number of independent particles* (walkers).¹⁴⁶ More specifically, the kinetic energy term, being second order, describes the movement of the walkers whereas the potential energy term describes generation or annihilation of walkers. The solution of this equation is

$$\Psi(\mathbf{r}, \tau) = e^{-(\hat{H}-S)\tau} \Psi(\mathbf{r}, 0), \quad (3.33)$$

where we have introduced the energy shift S which is adapted throughout the simulation every n steps to keep the number of walkers constant, and to estimate the ground state energy.¹⁴⁷ In particular, if we expand $\Psi(\mathbf{r}, 0)$ in terms of the eigenfunctions of \hat{H} , $\{\psi_i\}$, labeled in such a way that $i > j \implies E_i > E_j$, we get

$$\Psi(\mathbf{r}, \tau) = \sum_i a_i e^{-(E_i - S)\tau} \psi_i(\mathbf{r}). \quad (3.34)$$

*Note that the particles, henceforth referred to as walkers, are not the particles being described by the Hamiltonian. One walker represents a configuration of all the particles described by the Hamiltonian.

Here, a_i is the coefficient of the i th eigenfunction, and \hat{E}_i is the corresponding eigenvalue. After the excited-state solutions have decayed, the only possible value of S that stabilises the simulation is $S = E_0$. It is also clear from equation 3.34 that when $S = E_0$, all excited states decay exponentially to zero, and we are left with the ground state for large τ .

We may also express Ψ in terms of its Green's function, in order to get a form that expresses a small time step,

$$\Psi(\mathbf{r}, \tau + \Delta\tau) = \int d^{3N} r' G(\mathbf{r}, \mathbf{r}'; \Delta\tau) \Psi(\mathbf{r}, \tau), \quad (3.35)$$

where the Green's function

$$G(\mathbf{r}, \mathbf{r}'; \Delta\tau) = \langle \mathbf{r} | e^{-\Delta\tau(H-S)} | \mathbf{r}' \rangle \quad (3.36)$$

also obeys the imaginary-time Schrödinger equation, but with the initial condition

$$G(\mathbf{r}, \mathbf{r}'; 0) = \delta(\mathbf{r} - \mathbf{r}'). \quad (3.37)$$

For small time steps, G may be approximated by the Lie-Trotter-Suzuki decomposition,^{148–150} i.e. for operators A, A_k ,

$$A = \sum_k A_k \implies e^{-\alpha A} \approx \left(\prod_k e^{-\alpha A_k / M} \right)^M \quad (3.38)$$

for some value α and for large M . From this, we can get the Green's function,

$$G(\mathbf{r}, \mathbf{r}'; \Delta\tau) = e^{-\Delta\tau(V(\mathbf{r})-S)} \frac{1}{\sqrt{2\pi\Delta\tau}} e^{-(\mathbf{r}-\mathbf{r}')^2/(2\pi\Delta\tau)} + \mathcal{O}(\Delta\tau^2), \quad (3.39)$$

which describes our time evolution.

In DMC, walkers diffuse through configuration space in two stages: diffusion and branching. In the diffusion step, walkers are moved according to the kinetic terms in the Green's function. In the branching step, unfavourable walkers are removed and favourable ones produce new walkers.

For the energy shift estimate, we wish to choose a value such that the overall population does not change too much. If the walkers proliferate too much, then the computational cost becomes infeasible, whereas if too many walkers die, we are left with poor statistics. One way to choose S is by keeping track of the population of walkers N_w for a target population N_w^{target} and adjust,

$$S = S_{\text{old}} + \alpha \ln \left(\frac{N_w^{\text{target}}}{N_w} \right) \quad (3.40)$$

for small α .

More concretely, in pseudocode, the basic DMC algorithm is:

```

Initialise  $N$  walkers in random positions.
until convergence criterion met
  for each walker at position  $\mathbf{r}$ 
    shift position  $\mathbf{r}$  to  $\mathbf{r}'$  according to
       $\hookrightarrow$  transition probability  $G$ 
    evaluate  $q = e^{-\Delta\tau(V(\mathbf{r}')-S)}$ 
    walker survives with probability  $q$ 
    if  $q > 1$ 
      create  $1 - q$  new walkers
       $\hookrightarrow$  (with stochastic rounding)
  update  $S$ .

```

Notice that in contrast to VMC, DMC does not rely ^{on} a trial wave function. There is also a discretisation error incurred by the operator decomposition, though this can be handled via an application of the Metropolis-Hastings algorithm.¹¹⁴ However, in practice a trial wave function is actually necessary for fermions in order to improve the statistics. Since in electronic structure, we focus on fermions (electrons), this is an important feature. The walker distribution is positive, but for fermions this is not the case due to their exchange statistics. In order to still practically use DMC,* an approximation called the fixed node approximation is made, where the nodal structure (i.e. the roots of Ψ) of the distribution is determined by a trial function. Thus, the accuracy of DMC is still limited by the quality of the trial wave function (or more precisely, by its nodal surface). In a Slater-Jastrow-type trial wave function like equation 3.30, only the Slater-component will affect the nodal surface (and hence the accuracy). Thus, the choice of orbitals as well as the number of determinants can be important considerations.

} move to next sentence

3.4. Full Configuration Interaction Quantum Monte Carlo

The main post-Hartree-Fock method used throughout this dissertation is the FCIQMC algorithm.¹⁵¹ For this, there exist a few different implementations,^{152–155} but we focus on NECI.¹⁵² As the name implies, FCIQMC combines concepts from FCI (discussed in section 1.6.1) and QMC concepts already discussed in this chapter. In particular, it implements FCI, retaining many of its desirable properties such as size consistency and accuracy while being able to handle much larger systems thanks to its stochastic framework (however, the scaling remains exponential, like FCI). Like FCI, it is non-perturbative, and so can handle multi-reference problems, and unlike other QMC methods, it does not rely on a trial wave function. It can also be used as a solver in place of FCI, like in a MCSCF calculation.^{156,157}

*Other simple solutions such as assigning a sign to each walker leads to large errors.

3.4.1. Main Concepts

Like DMC, FCIQMC is based on the imaginary-time Schrödinger equation, equation 3.32, repeated here:

$$\frac{\partial}{\partial \tau} \Psi(\mathbf{r}, \tau) = -\hat{H} \Psi(\mathbf{r}, \tau). \quad (3.41)$$

The solution to this equation is

$$\Psi(\mathbf{r}, \tau) = e^{-\tau \hat{H}} \Psi(\mathbf{r}, 0). \quad (3.42)$$

Expanding $e^{-\tau \hat{H}}$ to first order in small imaginary time step $\Delta\tau$ gives us the FCIQMC projector

$$P = 1 - \Delta\tau \hat{H} \rightarrow 1 - \Delta\tau(\hat{H} - S), \quad (3.43)$$

where we have introduced an arbitrary shift S in analogy to DMC. Notice, in particular, that if $S = E_0$ the ground state energy, all other states decay away. Hence, repeated application of P while adjusting S to get as close to E_0 as possible will give us the ground state, as long as our initial state has nonzero overlap with the ground state.

Like FCI, the wave function is expanded in some basis. Throughout this dissertation, SDs are used, though configuration state functions (CSFs) are also possible.¹⁵⁸ The basis will be kept constant, and the coefficients may change with each time step,

$$|\Psi(\tau)\rangle = \sum_i c_i(\tau) |D_i\rangle. \quad (3.44)$$

Like in DMC, we adjust S to approximate the ground state energy E_0 , as $S = E_0$ would project out all states other than the ground state, as long as $\langle \Psi(\tau=0) | \Psi_0 \rangle \neq 0$ where $|\Psi_0\rangle$ is the ground state. Similarly, S affects the walker population. In particular, for walker population N_w , if $S > E_0$ then the population grows, and if $S < E_0$ it decays. If we have a target population N_w^{target} then we keep $S > E_0$ (typically, just by setting $S = 0$) and “turn on” the shift once N_w^{target} is reached. However, unlike DMC, FCIQMC walks ^{CS} can have either negative or positive weight. These weights correspond to the c_i , and the walkers “diffuse” across SD (or CSF) space. The walker population is reported in terms of absolute values, i.e.

$$N_w = \sum_{i \in \{SDs\}} |N_i|, \quad (3.45)$$

where we sum over the walkers across all SDs. Once in variable-shift mode, in practice S is updated every m cycles:

$$S(\tau + m\Delta\tau) = S(\tau) - \frac{\zeta}{m\Delta\tau} \ln \left(\frac{N_w(\tau + m\Delta\tau)}{N_w(\tau)} \right), \quad (3.46)$$

where ζ is a damping parameter.

The key advantage of FCIQMC over ordinary FCI is no longer needing to store the full CI vector, making use of sparsity. Even for relatively large and strongly-correlated systems, only a small subspace of the entire Hilbert space needs to be sampled in FCIQMC, whereas in FCI, the full CI vector must be stored (or if done via dense diagonalisation, the entire H matrix, which is unrealistic for even small systems).

3.4.2. Basic Algorithm

FCIQMC is typically run in three steps: (a) **spawning**, (b) **death/cloning**, and (c) **annihilation**. Each step can be recognised in the application of the projector operator of equation 3.43 onto some state $|\Psi\rangle = \sum_i c_i |D_i\rangle$,

$$P|\Psi\rangle = |\Psi\rangle - \underbrace{\sum_{\substack{i \\ \uparrow (c)}} \sum_{j \neq i} \Delta\tau H_{ji} c_i |D_j\rangle}_{(a)} - \underbrace{\sum_{\substack{i \\ \uparrow (c)}} (\Delta\tau (H_{ii} - S)) c_i |D_i\rangle}_{(b)}, \quad (3.47)$$

where the right-hand side of the equation is the new wave function, $|\Psi(\tau + \Delta\tau)\rangle$. Or, in terms of the coefficients (relabelling $i \leftrightarrow j$ in the (a) terms),

$$c_i(\tau + \Delta\tau) = c_i(\tau) - \Delta\tau (H_{ii} - S) c_i(\tau) - \Delta\tau \sum_{j \neq i} H_{ij} c_j(\tau). \quad (3.48)$$

In practice, we start the algorithm with a population of walkers on a reference determinant (typically, but not always, the Hartree-Fock determinant), and calculate the correlation energy via FCIQMC. Instead of H in the working equations we would use the matrix

$$K_{ij} = H_{ij} - E_{\text{ref}} \delta_{ij} \quad (3.49)$$

where $E_{\text{ref}} := \langle D_{\text{ref}} | H | D_{\text{ref}} \rangle$ is the energy of the reference determinant. However, FCIQMC is capable of adapting to a new reference determinant whenever another determinant exceeds the population of the reference determinant, and is not single-reference like canonical coupled cluster, for example. This shift is primarily a convenience.

The basic FCIQMC algorithm can be summarised by the following pseudocode (with further details below). Note that there are many extensions to the basic algorithm, beyond what is discussed in this dissertation.

```

Initialise  $N_{\text{start}}$  walkers on the reference determinant.
until stop criterion met
  for each occupied determinant  $D_i$ 
    select a random determinant based on  $p_{\text{gen}}$ 
    spawn  $p_s(j|i)$  walkers on  $D_j$ 
  for each occupied determinant  $D_i$ 
    remove  $p_d$  (death/cloning) walkers from  $D_i$ 
  for each occupied determinant  $D_i$ 
    annihilate opposite signed walkers
  calculate the projected energy
  if  $\sum |N_i| \geq N_w$ 
    update  $S$ .

```

Spawning

The first step, spawning, is realised by the application of the off-diagonal matrix elements. In this step, each walker on SD D_i may spawn a new set of walkers to connected* SD D_j . As we are sampling the sum $\Delta\tau \sum_{j \neq i} H_{ij} c_j(\tau)$, as in the Metropolis-Hastings algorithm, it is beneficial (in terms of efficiency) to sample more where H_{ij} is larger. That is, the probability of spawning on D_j from D_i , denoted $p_{\text{gen}}(j|i)$ (the generation probability) should be proportional to H_{ij} . Finally, the spawning probability is

$$p_s(j|i) = \Delta\tau \frac{|H_{ij}|}{p_{\text{gen}}(j|i)}. \quad (3.50)$$

Thus, at each FCIQMC iteration, every determinant D_i spawns $N_i p_s(j|i)$ new walkers on a determinant D_j selected by an excitation generation algorithm.^{159–161} Here, N_i is the (signed) number of walkers on D_i . If $H_{ij} < 0$ then the child walkers are the same sign as the parent; otherwise, they have opposite sign. The original formulation of the algorithm used integer number of walkers, and spawned a walker with probability $p_s(j|i)$ per walker on N_i , but it is more efficient to simply use noninteger values and spawn $p_s(j|i)$ walkers.

Death/Cloning

The death/cloning step is characterised by the diagonal matrix elements. In this step, p_d walkers are removed from the simulation[†] where

$$p_d = \Delta\tau (H_{ii} - S). \quad (3.51)$$

* D_i and D_j are said to be connected if $\langle D_i | H | D_j \rangle \neq 0$.

[†]The original formulation removed (or cloned) walkers with a probability of p_d .

Notice that if p_d is the same sign as the walker, then the number of walkers on D_i increases (hence “cloning”, as opposed to “death”). “Cloning” is especially common in the growth stage of the population, where the population is growing exponentially towards the target N_w .

Annihilation

After spawning and death/cloning, we may have walkers of opposite sign on the same determinant. This is dealt with in the “annihilation” step of the algorithm, realised by combining the terms in equation 3.47. This step, while simple in principle, amounts to cancelling opposite walkers. Determinants without walkers left on them are then removed as well, to keep the storage of $|\Psi\rangle$ sparse. However, so far the steps have all been embarrassingly parallel, but in this step we need to communicate in order to locate determinants on the same determinant, so this step is critical for parallelisation. Nevertheless, efficient implementations such as in NECI is able to scale efficiently to over 40000 cores.¹⁶²

3.4.3. Energy Estimators

The shift S is updated to approximate E_0 , so we already have an estimator for the energy. However, since S is used to also control the population, this estimator can be noisy.¹⁵¹

A more commonly used energy estimator is the projected energy,

$$E_{\text{proj}} = \frac{\langle D_{\text{ref}} | \hat{H} | \Psi \rangle}{\langle D_{\text{ref}} | \Psi \rangle}, \quad (3.52)$$

where we project onto the reference (most occupied) determinant.*

In principle, any wave function could be used in place of the trial determinant in equation 3.52 (so long as it has nonzero overlap with $|\Psi\rangle$). The choice of the reference determinant is for ease of computation, and because the larger the overlap with $|\Psi\rangle$, the smaller the error in the stochastic sampling. Since the reference determinant (by definition) has the largest number of walkers on it, it is the sensible choice for a single-determinant projected energy.

Another energy estimator is the “trial energy”,^{164,165} which is just the projected energy on a multi-determinant wave function instead of $|D_{\text{ref}}\rangle$. In this method, after some time in variable shift mode, a “trial space” is constructed based on the top N_T determinants. The Hamiltonian spanned within the trial space is diagonalised exactly (sparsely or densely), and the eigenvector $|\Psi_{\text{trial}}\rangle$ is stored. Then, the trial projected energy can be calculated as

$$E_{\text{trial}} = \frac{\langle \Psi_{\text{trial}} | \hat{H} | \Psi \rangle}{\langle \Psi_{\text{trial}} | \Psi \rangle}. \quad (3.53)$$

*While it may be tempting to instead use the energy estimator $\langle \Psi | \hat{H} | \Psi \rangle / \langle \Psi | \Psi \rangle$, this comes with a prohibitive cost, and has a bias.^{151,163}

Note that while this is expected to reduce the error in the energy estimator, it also increases the cost by a factor of N_T .

An example of these energy estimators is shown in figure 3.2. As in other QMC methods, these energy estimators have autocorrelation, so in reporting the values, a blocking analysis is required.¹³²

3.4.4. Annihilation Plateaus

FCIQMC has its own kind of sign problem in the form of resolving the correct relative sign structure of the sampled states in the CI vector.¹⁶⁶ This manifests in the form of a so-called “annihilation plateau”, illustrated in figure 3.3.

In this stage of the simulation, walkers are being spawned with incoherent signs, causing a massive amount of walker annihilation. This causes the population to stay relatively constant, until the sign structure is finally resolved and the population continues to grow again. The height of the plateau (i.e. the constant number of walkers where the simulation stagnates) quantifies the difficulty of the sign problem. In particular, for certain large systems, it can be prohibitively expensive to overcome. If we are below the annihilation plateau, then the original formulation of FCIQMC is not reliable, as the correct sign structure is not yet resolved.

3.4.5. The Initiator Approximation

Since incoherent walker proliferation is a major source of the annihilation plateau, a sensible strategy to overcome this is by forcing growth to be coherent. This is achieved by the initiator approximation.¹⁶⁷ An FCIQMC calculation with the initiator approximation is sometimes referred to as i-FCIQMC; however, since the approximation is so useful and ubiquitous, we simply referred to as FCIQMC. Indeed, in this dissertation all FCIQMC calculations used the initiator approximation, unless otherwise specified.

The initiator approximation allows only those determinants (dubbed “initiators”) with at least N_{thresh} walkers to spawn new walkers on unoccupied determinants. If the population is greater than this threshold, the assumption is that the (relative) sign of that state is correct, and hence the spawning event should produce the correct sign.

Since the approximation truncates spawning, the growth phase of the simulation is slower. However, the annihilation plateau disappears. Note that as $N_w \rightarrow \infty$, the initiator approximation becomes exact, as all sampled states are initiators. While there is a small bias introduced by the initiator approximation, the efficiency is greatly improved and the range of potential applications is greatly increased.

3.4.6. Reduced Density Matrix Sampling

As described in section 1.4.2, reduced density matrices are a useful tool for understanding the properties of a system. These can be sampled in FCIQMC by considering the RDMS

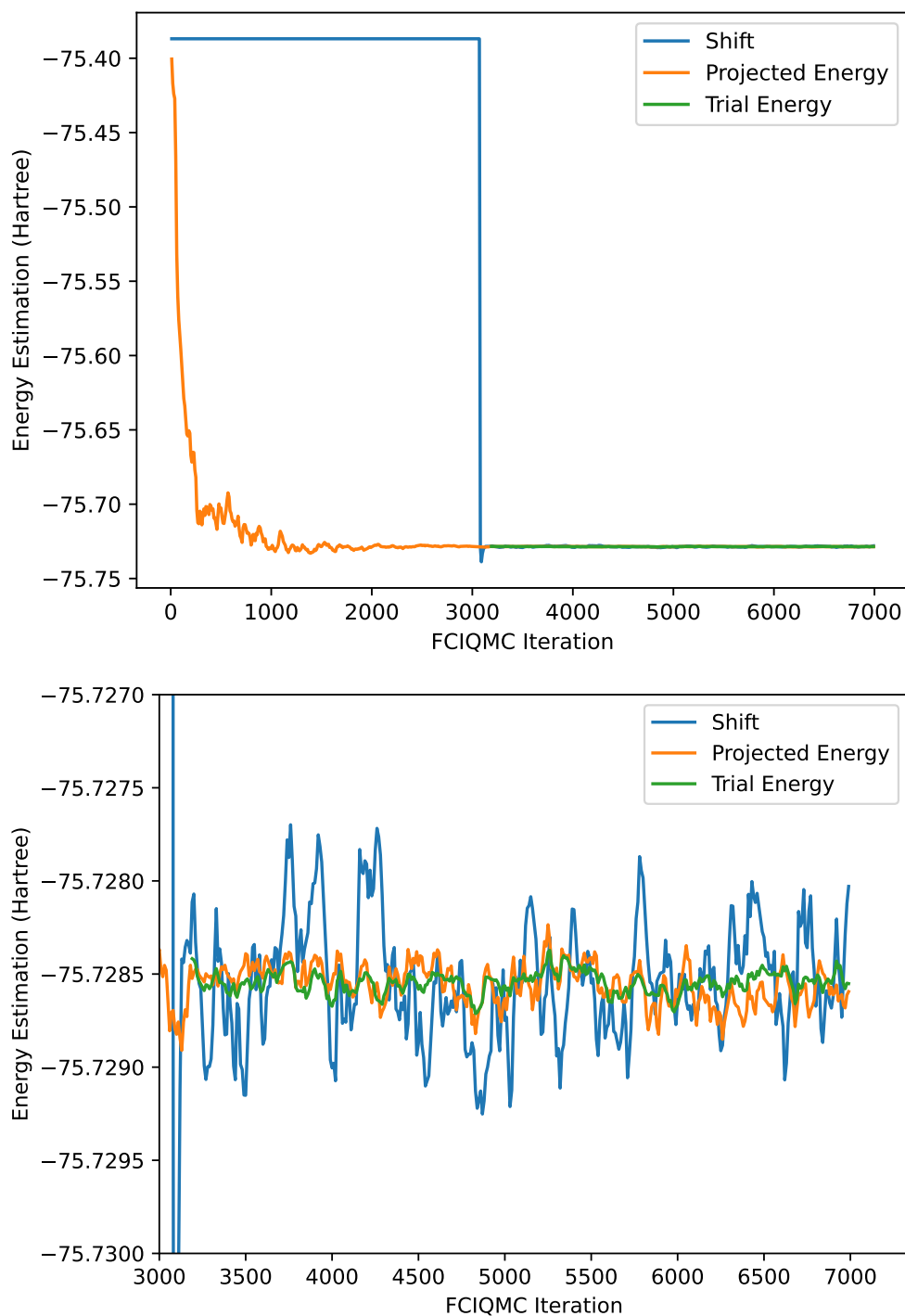


Figure 3.2. An example of the energy estimators used in an FCIQMC simulation. The shift is the most noisy, whereas the trial energy is the least noisy, but only available in variable shift mode and carries the highest cost. In this case, the trial energy is not a substantial improvement from the reference-projected energy. This calculation was done on the C_2 molecule with the cc-pVDZbasis, at equilibrium geometry, 1.2425 Å.

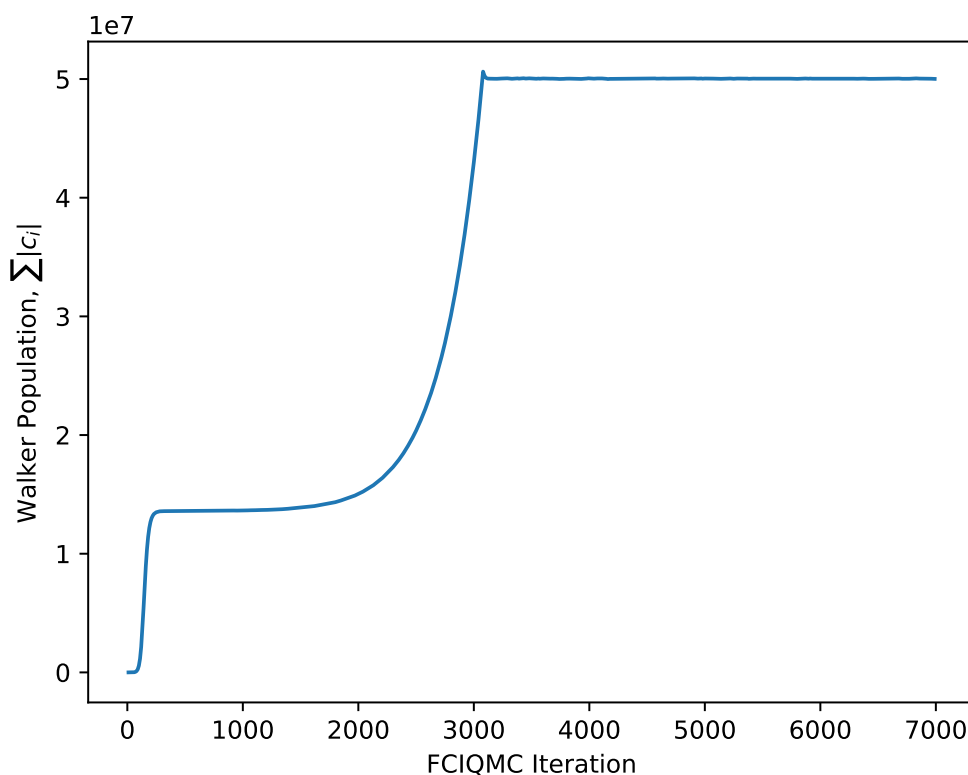


Figure 3.3. An example of an annihilation plateau in the context of a FCIQMC simulation. Despite not yet reaching the target population, the total population is roughly constant before increasing again. This stage is referred to as the annihilation plateau, and the higher the plateau, the more difficult the sign problem is. This calculation was done on the C_2 molecule with the cc-pVDZ basis, at equilibrium geometry, 1.2425 Å. In this particular case, the plateau appears around 1.36×10^7 walkers, and the target number is 5×10^7 .

in the same basis.¹⁶⁸ The 1RDM is given by

$$\gamma_{pq} = \sum_{ij} c_i c_j \langle D_i | a_p^\dagger a_q | D_j \rangle \quad (3.54)$$

and the 2RDM by

$$\Gamma_{pqrs} = \sum_{ijkl} c_i c_j \langle D_i | a_p^\dagger a_q^\dagger a_s a_r | D_j \rangle. \quad (3.55)$$

In FCIQMC, we have the CI coefficients c_i and c_j by the walker distributions. However, using the same simulation for both c_i and c_j introduces a bias, hence to stochastically sample RDMs, we use an independent simultaneous simulation (referred to as a replica). One simulation samples c_i whereas the other samples c_j . Since the statistical errors are independent, the overall error is reduced. Unfortunately this requires twice the computing power, but thankfully since the calculations are independent, they are trivially parallelisable.

Note also that by inserting the Hamiltonian (energy) operator into equation 1.21 and using the replica trick to obtain the RDMs, we can also obtain yet another energy measure in FCIQMC.

3.4.7. Combining TC with Modern Electronic Structure

The FCIQMC algorithm may also be combined with the TC method described in section 2.5.¹⁶⁹ Consider solving for the eigenstates Φ of the TC Hamiltonian \hat{H}_{TC} using the imaginary-time Schrödinger equation,

$$-\frac{\partial}{\partial \tau} \Phi = (\hat{H}_{\text{TC}} - S) \Phi. \quad (3.56)$$

Since \hat{H}_{TC} and \hat{H} are isospectral, we have stationary states for the same values as the non-TC method. That is, we can control the walker population by setting the shift to the ground state energy, $S = E_0$.

The state Φ is described by a linear combination of SDs, as in non-TC-FCIQMC,

$$|\Phi\rangle = \sum_i c_i |D_i\rangle. \quad (3.57)$$

However, since \hat{H}_{TC} is non-Hermitian, the coefficients \tilde{c}_i of the left-eigenvector with the same energy are not necessarily the same,

$$\langle \Phi | = \sum_i \tilde{c}_i \langle D_i |, \quad (3.58)$$

and we must be careful when considering matrix connections $\langle D_i | \hat{H}_{\text{TC}} | D_j \rangle$ and $\langle D_j | \hat{H}_{\text{TC}} | D_i \rangle$. In particular, the probability of spawning a walker from $|D_i\rangle$ to $|D_j\rangle$ may not be the same

as the probability of spawning from $|D_j\rangle$ to $|D_i\rangle$. Furthermore, when considering replicas such as in RDM sampling, the replica needs to be of the adjoint operator, $\hat{H}_{\text{TC}}^\dagger$, in order to target \tilde{c}_i . This is accomplished by a simple transform of the Jastrow factor, $J \rightarrow -J$, though this may cause practical complications. Otherwise, FCIQMC may be extended to TC by simply applying the method directly to \hat{H}_{TC} . An appropriate transformation from \hat{H} is therefore necessary beforehand.

Similar arguments may be made for applying TC to CC methods^{170,171} and density matrix renormalization group (DMRG),¹⁷² and TC variants of these methods have already been developed and successfully applied. In addition to TC-FCIQMC, this also continues to be an active area of research.

Optimizing Jastrow Factors for the Transcorrelated Method

TODO: ...

This chapter is based in large part on the following paper:

Haupt, J. P.; Hosseini, S. M.; López Ríos, P.; Dobrautz, W.; Cohen, A.; Alavi, A. “Optimizing Jastrow Factors for the Transcorrelated Method”. *The Journal of Chemical Physics* **2023**, *158*, 224105

Some images have been reused from this paper (with permission), clearly indicated in their figure labels.

4.1. Introduction

4.2. Jastrow Factor

4.3. Optimisation Strategy

4.3.1. Choosing an Appropriate Sample Size

4.4. Results

TODO: mention details about integration will be in a separate chapter TODO: mention we use a walker number extrapolation already described in another dissertation (cite Muhammedreza)

4.4.1. Neglecting Three-Body Excitations

4.4.2. Basis Set Convergence

4.5. Conclusion and Outlook

TODO: mention this is the way we now optimise Jastrow factors, but there is an important extension to the no-3-body approximation, xTC. Briefly describe.

4.5.1. The xTC Approximation

The Transcorrelated Method for Multi-Reference Problems

TODO: ...

This chapter is based in large part on the following paper:

Haupt, J. P.; Ríos, P. L.; Christlmaier, E. M. C.; Kats, D.; Alavi, A. “Size Consistency and Transcorrelation”. **2024**

Some images have been reused from this paper (with permission), clearly indicated in their figure labels.

5.1. Introduction

TODO: problem description, including the “dip”

5.2. Size Consistency

5.2.1. Jastrow Factor Forms

5.3. CASSCF Ansatz

TODO: present the bad data :(– mention might hint at needing the core (I think Andreea’s paper corroborated this)

5.4. Transcorrelated Trial Wavefunction

5.5. Results

5.5.1. Binding Curves

TODO: mention that we fix the binding curves TODO: also include a table about just “how size-consistent” the results actually are.

5.5.2. Excitation Energies

TODO: mention can target the exact state(s) in question

5.6. Conclusion and Outlook

TODO: make sure to mention the cutoff analysis and constructing Jastrow factors from atomic Jastrow factors. Mention ECPs, since it's already been submitted...

The PyTCHInt Library

TODO: ...

This chapter is based on the following software:

Haupt, J. P.; Ríos, P. L.; Christlmaier, E. M. C.; Filip, M.-A.; Simula, K.; Hauskrecht, J.; Liao, K.; Dobrautz, W.; Guthier, K.; Cohen, A. J.; Alavi, A. *Pythint: Transcorrelated Hamiltonian Integrals Library*, 2024

The basis for how the library works is partially discussed in the following paper and its supplementary material:

TODO: Aron's tchint paper

TODO: May want to move this chapter closer to the start, i.e. just before the two other “original” chapters, to illustrate the tools used for those. Not sure yet.

6.1. Introduction

6.2. Matrix Element Evaluation

TODO: mention the different “intermediate” quantities, RI (no longer really used), xTC

6.3. Deterministic Optimisation

6.4. Performance

TODO: not really sure how I'll write this section, since I don't really have a performance metric to go up against TODO: memory requirements, e.g. full TC, no-3-body, xTC

6.5. Interface

TODO: NECI interface, M7 interface, pythint and interactive HPC

6.6. Conclusion and Outlook

Summary and Outlook

TODO: ... TODO: brief review TODO: outlook: ECPs (if not already published), efficient deterministic optimisation, better understanding of the form of the Jastrow factor and other options, pytorch optimisation, larger molecules, solids, embedding, CUDA

List of Figures

1.1.	Illustration of an <i>s</i> -type orbital and the fitted gaussians (STO- <i>x</i> G means <i>x</i> gaussians are used to approximate the STO). Note that for much of the curve, there is good agreement, but not for the short- and long-range. The short-range behaviour in particular leads to complications, and is the focus of the methods discussed in chapter 2. The long-range behaviour is typically less of a problem, since the integrals we are interested in typically decay with <i>r</i> . Fitting parameters for the gaussians were taken from reference 24.	9
1.2.	A sketch for the hierarchy of theories and basis sets. There are two axes on which to systematically improve until reaching the exact solution to the time-independent Schrödinger equation under the BOA: increasing basis set size (x axis) and increasing the level of theory (y axis). For most of this dissertation, we assume to be at or close to the highest level of theory (the FCI limit), as we mostly employ FCIQMC. We therefore focus on methods of rapidly converging to the CBS limit without the inherently expensive process of adding more basis functions.	14
1.3.	Illustration of a RHF reference and examples of single-, double-, and triple-excitations. The method CISDT, for example, would consider all determinants of the forms illustrated.	15
2.1.	A toy example of the Coulomb cusp. Here, the Fourier expansion $x \approx \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=1}^N \frac{\cos[(2n-1)x]}{(2n-1)^2}$ is plotted for a few values of <i>N</i> , including the exact solution. As can be seen, even for many terms, the Fourier expansion is a poor descriptor in the cusp region. Indeed, the only way to describe it exactly is with an infinite number of terms.	23
3.1.	Our “game of darts”. Points inside the unit circle are coloured blue whereas points outside are orange. Using stochastic sampling, this naive approach uses 5000 randomly generated numbers between -1 and 1 to approximate $\pi \approx 4N_{\text{in}}/N_{\text{out}} \approx 3.1464$	28
3.2.	An example of the energy estimators used in an FCIQMC simulation. The shift is the most noisy, whereas the trial energy is the least noisy, but only available in variable shift mode and carries the highest cost. In this case, the trial energy is not a substantial improvement from the reference-projected energy. This calculation was done on the C ₂ molecule with the cc-pVDZbasis, at equilibrium geometry, 1.2425 Å.	42

- 3.3. An example of an annihilation plateau in the context of a FCIQMC simulation. Despite not yet reaching the target population, the total population is roughly constant before increasing again. This stage is referred to as the annihilation plateau, and the higher the plateau, the more difficult the sign problem is. This calculation was done on the C_2 molecule with the cc-pVDZbasis, at equilibrium geometry, 1.2425 Å. In this particular case, the plateau appears around 1.36×10^7 walkers, and the target number is 5×10^7 43

List of Tables

Bibliography

- (1) Laughlin, R. B.; Pines, D. “The Theory of Everything”. Proceedings of the National Academy of Sciences **2000**, *97*, 28–31.
- (2) Ryden, B., *Introduction to Cosmology*; Cambridge University Press: 2017; 277 pp.
- (3) Jones, L. O.; Mosquera, M. A.; Schatz, G. C.; Ratner, M. A. “Embedding Methods for Quantum Chemistry: Applications from Materials to Life Sciences”. Journal of the American Chemical Society **2020**, *142*, 3281–3295.
- (4) Christlmaier, E. M. C.; Kats, D.; Alavi, A.; Usvyat, D. “Full Configuration Interaction Quantum Monte Carlo Treatment of Fragments Embedded in a Periodic Mean Field”. The Journal of Chemical Physics **2022**, *156*, 154107.
- (5) Helgaker, T.; Jorgensen, P.; Olsen, J., *Molecular Electronic-Structure Theory*; John Wiley & Sons: 2014; 949 pp.
- (6) Born, M.; Oppenheimer, R. “Zur Quantentheorie Der Molekeln”. Annalen der Physik **1927**, *389*, 457–484.
- (7) Hubbard, J. “Electron Correlations in Narrow Energy Bands”. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **1963**, *276*, 238–257.
- (8) Lieb, E. H.; Wu, F. Y. “Absence of Mott Transition in an Exact Solution of the Short-Range, One-Band Model in One Dimension”. Physical Review Letters **1968**, *20*, 1445–1448.
- (9) Liebermann, N. *The FCIQMC Sign Problem in the Real-Space Hubbard Model*, Doctoral Thesis, 2023.
- (10) Motta, M.; Ceperley, D. M.; Chan, G. K. L.; Gomez, J. A.; Gull, E.; Guo, S.; Jiménez-Hoyos, C. A.; Lan, T. N.; Li, J.; Ma, F.; Millis, A. J.; Prokof'ev, N. V.; Ray, U.; Scuseria, G. E.; Sorella, S.; Stoudenmire, E. M.; Sun, Q.; Tupitsyn, I. S.; White, S. R.; Zgid, D.; Zhang, S. “Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods”. Physical Review X **2017**, *7*, 031059.
- (11) Motta, M.; Genovese, C.; Ma, F.; Cui, Z. H.; Sawaya, R.; Chan, G. K. L.; Chepiga, N.; Helms, P.; Jiménez-Hoyos, C.; Millis, A. J.; Ray, U.; Ronca, E.; Shi, H.; Sorella, S.; Stoudenmire, E. M.; White, S. R.; Zhang, S. “Ground-State Properties of the Hydrogen Chain: Dimerization, Insulator-to-Metal Transition, and Magnetic Phases”. Physical Review X **2020**, *10*, 031058.

- (12) Sakurai, J. J.; Napolitano, J., *Modern Quantum Mechanics*; Cambridge University Press: 2017; 569 pp.
- (13) Ascher, U. M.; Greif, C., *A First Course in Numerical Methods*; SIAM: 2011; 574 pp.
- (14) Dunning Jr., T. H. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen". *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- (15) Feller, D.; Peterson, K. A.; Grant Hill, J. "On the Effectiveness of CCSD(T) Complete Basis Set Extrapolations for Atomization Energies". *The Journal of Chemical Physics* **2011**, *135*, 044102.
- (16) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. "Basis-Set Convergence in Correlated Calculations on Ne, N₂, and H₂O". *Chemical Physics Letters* **1998**, *286*, 243–252.
- (17) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. "Basis-Set Convergence of the Energy in Molecular Hartree–Fock Calculations". *Chemical Physics Letters* **1999**, *302*, 437–446.
- (18) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. "Basis-Set Convergence of Correlated Calculations on Water". *The Journal of Chemical Physics* **1997**, *106*, 9639–9646.
- (19) Jensen, F. "The Basis Set Convergence of the Hartree–Fock Energy for H₂". *The Journal of Chemical Physics* **1999**, *110*, 6601–6605.
- (20) Pansini, F. N. N.; Neto, A. C.; Varandas, A. J. C. "Extrapolation of Hartree–Fock and Multiconfiguration Self-Consistent-Field Energies to the Complete Basis Set Limit". *Theoretical Chemistry Accounts* **2016**, *135*, 261.
- (21) Peterson, K. A.; Woon, D. E.; Dunning Jr., T. H. "Benchmark Calculations with Correlated Molecular Wave Functions. IV. The Classical Barrier Height of the H+H₂→H₂+H Reaction". *The Journal of Chemical Physics* **1994**, *100*, 7410–7415.
- (22) Woon, D. E.; Dunning Jr., T. H. "Benchmark Calculations with Correlated Molecular Wave Functions. VI. Second Row A2 and First Row/Second Row AB Diatomic Molecules". *The Journal of Chemical Physics* **1994**, *101*, 8877–8893.
- (23) Boys, S. F. "Electronic Wave Functions - I. A General Method of Calculation for the Stationary States of Any Molecular System". *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1950**, *200*, 542–554.
- (24) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Courier Corporation: 2012; 484 pp.
- (25) Hastie, T.; Tibshirani, R.; Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*; Springer Science & Business Media: 2009; 757 pp.

- (26) Bayes, T.; Price, n. "LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S". Philosophical Transactions of the Royal Society of London **1997**, *53*, 370–418.
- (27) Hartree, D. R. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion". Mathematical Proceedings of the Cambridge Philosophical Society **1928**, *24*, 111–132.
- (28) Fock, V. "Näherungsmethode Zur Lösung Des Quantenmechanischen Mehrkörperproblems". Zeitschrift für Physik **1930**, *61*, 126–148.
- (29) Slater, J. C. "Note on Hartree's Method". Physical Review **1930**, *35*, 210–211.
- (30) Roothaan, C. C. J. "New Developments in Molecular Orbital Theory". Reviews of Modern Physics **1951**, *23*, 69–89.
- (31) Hall, G. G.; Lennard-Jones, J. E. "The Molecular Orbital Theory of Chemical Valency VIII. A Method of Calculating Ionization Potentials". Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **1997**, *205*, 541–552.
- (32) Davidson, E. R. "The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices". Journal of Computational Physics **1975**, *17*, 87–94.
- (33) Lanczos, C. "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators". Journal of Research of the National Bureau of Standards **1950**, *45*, 255.
- (34) Arnoldi, W. E. "The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem". Quarterly of Applied Mathematics **1951**, *9*, 17–29.
- (35) Eade, R. H. A.; Robb, M. A. "Direct Minimization in Mc Scf Theory. the Quasi-Newton Method". Chemical Physics Letters **1981**, *83*, 362–368.
- (36) Roos, B. "A New Method for Large-Scale CI Calculations". Chemical Physics Letters **1972**, *15*, 153–159.
- (37) Olsen, J. "The CASSCF Method: A Perspective and Commentary". International Journal of Quantum Chemistry **2011**, *111*, 3267–3272.
- (38) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. M. "A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach". Chemical Physics **1980**, *48*, 157–173.
- (39) Siegbahn, P.; Heiberg, A.; Roos, B.; Levy, B. "A Comparison of the Super-CI and the Newton-Raphson Scheme in the Complete Active Space SCF Method". Physica Scripta **1980**, *21*, 323.

- (40) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. "The Complete Active Space SCF (CASSCF) Method in a Newton–Raphson Formulation with Application to the HNO Molecule". *The Journal of Chemical Physics* **1981**, *74*, 2384–2396.
- (41) Rayleigh, J. W. S. B.; Lindsay, R. B., *The Theory of Sound*; Courier Corporation: 1945; 548 pp.
- (42) Schrödinger, E. "Quantisierung Als Eigenwertproblem". *Annalen der Physik* **1926**, *385*, 437–490.
- (43) Møller, Chr.; Plesset, M. S. "Note on an Approximation Treatment for Many-Electron Systems". *Physical Review* **1934**, *46*, 618–622.
- (44) Čížek, J. "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods". *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.
- (45) Čížek, J.; Paldus, J. "Correlation Problems in Atomic and Molecular Systems III. Rederivation of the Coupled-Pair Many-Electron Theory Using the Traditional Quantum Chemical Methodst". *International Journal of Quantum Chemistry* **1971**, *5*, 359–379.
- (46) Paldus, J.; čížek, J., *Time-Independent Diagrammatic Approach to Perturbation Theory of Fermion Systems In Advances in Quantum Chemistry*, Löwdin, P.-O., Ed.; Academic Press: 1975; Vol. 9, pp 105–197.
- (47) Shavitt, I.; Bartlett, R. J., *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*; Cambridge Molecular Science; Cambridge University Press: Cambridge, 2009.
- (48) Aoto, Y. A.; Köhn, A. "Internally Contracted Multireference Coupled-Cluster Theory in a Multistate Framework". *The Journal of Chemical Physics* **2016**, *144*, 074103.
- (49) Evangelista, F. A. "Perspective: Multireference Coupled Cluster Theories of Dynamical Electron Correlation". *The Journal of Chemical Physics* **2018**, *149*, 030901.
- (50) Hanauer, M.; Köhn, A. "Pilot Applications of Internally Contracted Multireference Coupled Cluster Theory, and How to Choose the Cluster Operator Properly". *The Journal of Chemical Physics* **2011**, *134*, 204111.
- (51) Jankowski, K.; Paldus, J.; Grabowski, I.; Kowalski, K. "Applicability of Valence-universal Multireference Coupled-cluster Theories to Quasidegenerate Electronic States. I. Models Involving at Most Two-body Amplitudes". *The Journal of Chemical Physics* **1992**, *97*, 7600–7612.

- (52) Jeziorski, B.; Monkhorst, H. J. "Coupled-Cluster Method for Multideterminantal Reference States". *Physical Review A* **1981**, *24*, 1668–1681.
- (53) Köhn, A.; Black, J. A.; Aoto, Y. A.; Hanauer, M. "Improved and Simplified Orthogonalisation Scheme and Connected Triples Correction within the Internally Contracted Multireference Coupled-Cluster Method". *Molecular Physics* **2020**, *118*, e1743889.
- (54) Kats, D.; Manby, F. R. "Communication: The Distinguishable Cluster Approximation". *The Journal of Chemical Physics* **2013**, *139*, 021102.
- (55) Kats, D.; Kreplin, D.; Werner, H.-J.; Manby, F. R. "Accurate Thermochemistry from Explicitly Correlated Distinguishable Cluster Approximation". *The Journal of Chemical Physics* **2015**, *142*, 064111.
- (56) Rishi, V.; Perera, A.; Nooijen, M.; Bartlett, R. J. "Excited States from Modified Coupled Cluster Methods: Are They Any Better than EOM CCSD?" *The Journal of Chemical Physics* **2017**, *146*, 144104.
- (57) Vitale, E.; Alavi, A.; Kats, D. "FCIQMC-Tailored Distinguishable Cluster Approach". *Journal of Chemical Theory and Computation* **2020**, *16*, 5621–5634.
- (58) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. "A Fifth-Order Perturbation Comparison of Electron Correlation Theories". *Chemical Physics Letters* **1989**, *157*, 479–483.
- (59) Woon, D. E.; Dunning Jr., T. H. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon". *The Journal of Chemical Physics* **1993**, *98*, 1358–1371.
- (60) Woon, D. E.; Dunning Jr., T. H. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties". *The Journal of Chemical Physics* **1994**, *100*, 2975–2988.
- (61) Wilson, A. K.; van Mourik, T.; Dunning, T. H. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. VI. Sextuple Zeta Correlation Consistent Basis Sets for Boron through Neon". *Journal of Molecular Structure: THEOCHEM* **1996**, *388*, 339–349.
- (62) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. "R12 Methods in Explicitly Correlated Molecular Electronic Structure Theory". <https://doi.org/10.1080/01442350600799921> **2007**, *25*, 427–468.
- (63) Klopper, W.; Schütz, M.; Lüthi, H. P.; Leutwyler, S. "An Ab Initio Derived Torsional Potential Energy Surface for (H₂O)₃. II. Benchmark Studies and Interaction Energies". *The Journal of Chemical Physics* **1995**, *103*, 1085–1098.

- (64) Grüneis, A.; Hirata, S.; Ohnishi, Y.-y.; Ten-no, S. "Perspective: Explicitly Correlated Electronic Structure Theory for Complex Systems". *The Journal of Chemical Physics* **2017**, *146*, 080901.
- (65) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. "Explicitly Correlated Electrons in Molecules". *Chemical Reviews* **2012**, *112*, 4–74.
- (66) Kato, T. "On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics". *Communications on Pure and Applied Mathematics* **1957**, *10*, 151–177.
- (67) Pack, R. T.; Brown, W. B. "Cusp Conditions for Molecular Wavefunctions". *The Journal of Chemical Physics* **1966**, *45*, 556–559.
- (68) Kurokawa, Y. I.; Nakashima, H.; Nakatsuji, H., *Chapter Two - General Coalescence Conditions for the Exact Wave Functions: Higher-Order Relations for Coulombic and Non-Coulombic Systems* In *Advances in Quantum Chemistry*, Hoggan, P. E., Ozdogan, T., Eds.; Electron Correlation in Molecules – Ab Initio Beyond Gaussian Quantum Chemistry, Vol. 73; Academic Press: 2016, pp 59–79.
- (69) Roothaan, C. C. J.; Sachs, L. M.; Weiss, A. W. "Analytical Self-Consistent Field Functions for the Atomic Configurations $1s^2$, $1s^2 2s$, and $1s^2 2s^2$ ". *Reviews of Modern Physics* **1960**, *32*, 186–194.
- (70) Watson, R. E. "Approximate Wave Functions for Atomic Be". *Physical Review* **1960**, *119*, 170–177.
- (71) Weiss, A. W. "Configuration Interaction in Simple Atomic Systems". *Physical Review* **1961**, *122*, 1826–1836.
- (72) Schwartz, C. "Ground State of the Helium Atom". *Physical Review* **1962**, *128*, 1146–1148.
- (73) Coulson, C. A.; Neilson, A. H. "Electron Correlation in the Ground State of Helium". *Proceedings of the Physical Society* **1961**, *78*, 831.
- (74) Gilbert, T. L. "Interpretation of the Rapid Convergence of Correlated Wave Functions". *Reviews of Modern Physics* **1963**, *35*, 491–494.
- (75) Prendergast, D.; Nolan, M.; Filippi, C.; Fahy, S.; Greer, J. C. "Impact of Electron-Electron Cusp on Configuration Interaction Energies". *Journal of Chemical Physics* **2001**, *115*, 1626–1634.
- (76) Kong, L.; Bischoff, F. A.; Valeev, E. F. "Explicitly Correlated R12/F12 Methods for Electronic Structure". *Chemical Reviews* **2012**, *112*, 75–107.
- (77) Grynberg, G.; Aspect, A.; Fabre, C., *Introduction to Quantum Optics: From the Semi-classical Approach to Quantized Light*; Cambridge University Press: Cambridge, 2010.
- (78) Slater, J. C. "Central Fields and Rydberg Formulas in Wave Mechanics". *Physical Review* **1928**, *31*, 333–343.

- (79) Slater, J. C. "The Normal State of Helium". *Physical Review* **1928**, *32*, 349–360.
- (80) Hylleraas, E. A. "Über den Grundzustand des Heliumatoms". *Zeitschrift für Physik* **1928**, *48*, 469–494.
- (81) Hylleraas, E. A. "Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium". *Zeitschrift für Physik* **1929**, *54*, 347–366.
- (82) Largo-Cabrerizo, A.; Clementi, E. "The Hylleraas-CI Method in Molecular Calculations: Two-electron Integrals". *Journal of Computational Chemistry* **1987**, *8*, 1191–1198.
- (83) James, H. M.; Coolidge, A. S. "The Ground State of the Hydrogen Molecule". *The Journal of Chemical Physics* **1933**, *1*, 825–835.
- (84) Kołos, W.; Wolniewicz, L. "Accurate Adiabatic Treatment of the Ground State of the Hydrogen Molecule". *The Journal of Chemical Physics* **1964**, *41*, 3663–3673.
- (85) Perkins, J. F. "Atomic Integrals Containing R23 λ R31 μ R12 ν ". *The Journal of Chemical Physics* **1968**, *48*, 1985–1988.
- (86) Perkins, J. F. "Atomic Integrals Containing Rijp Correlation Factors with Unlinked Indices". *The Journal of Chemical Physics* **1969**, *50*, 2819–2823.
- (87) Sims, J. S.; Hagstrom, S. "Combined Configuration-Interaction—Hylleraas-Type Wave-Function Study of the Ground State of the Beryllium Atom". *Physical Review A* **1971**, *4*, 908–916.
- (88) Sims, J. S.; Hagstrom, S. A. "One-Center Rij Integrals Over Slater-Type Orbitals". *The Journal of Chemical Physics* **1971**, *55*, 4699–4710.
- (89) Clary, D. C.; Handy, N. C. "Hylleraas-Type Wavefunction for Lithium Hydride". *Chemical Physics Letters* **1977**, *51*, 483–486.
- (90) Clary, D. C.; Handy, N. C. "CI-Hylleraas Variational Calculation on the Ground State of the Neon Atom". *Physical Review A* **1976**, *14*, 1607–1613.
- (91) Boys, S. F. "The Integral Formulae for the Variational Solution of the Molecular Many-Electron Wave Equation in Terms of Gaussian Functions with Direct Electronic Correlation". *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1960**, *258*, 402–411.
- (92) Singer, K. "The Use of Gaussian (Exponential Quadratic) Wave Functions in Molecular Problems - I. General Formulae for the Evaluation of Integrals". *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1960**, *258*, 412–420.
- (93) Mitroy, J.; Bubin, S.; Horiuchi, W.; Suzuki, Y.; Adamowicz, L.; Cencek, W.; Szalewicz, K.; Komasa, J.; Blume, D.; Varga, K. "Theory and Application of Explicitly Correlated Gaussians". *Reviews of Modern Physics* **2013**, *85*, 693–749.

- (94) Lester Jr., W. A.; Krauss, M. "Gaussian Correlation Functions: Two-Electron Systems". *The Journal of Chemical Physics* **1964**, *41*, 1407–1413.
- (95) Bukowski, R.; Jeziorski, B.; Szalewicz, K. "New Effective Strategy of Generating Gaussian-type Geminal Basis Sets for Correlation Energy Calculations". *The Journal of Chemical Physics* **1994**, *100*, 1366–1374.
- (96) Persson, B. J.; Taylor, P. R. "Accurate Quantum-chemical Calculations: The Use of Gaussian-type Geminal Functions in the Treatment of Electron Correlation". *The Journal of Chemical Physics* **1996**, *105*, 5915–5926.
- (97) Pan, K.-C.; King, H. F. "Gaussian Geminals for Electron Pair Correlation". *The Journal of Chemical Physics* **1970**, *53*, 4397–4399.
- (98) Pan, K.-C.; King, H. F. "Electron Correlation in Closed Shell Systems. I. Perturbation Theory Using Gaussian-Type Geminals". *The Journal of Chemical Physics* **1972**, *56*, 4667–4688.
- (99) Szalewicz, K.; Jeziorski, B.; Monkhorst, H. J.; Zabolitzky, J. G. "A New Functional for Variational Calculation of Atomic and Molecular Second-Order Correlation Energies". *Chemical Physics Letters* **1982**, *91*, 169–172.
- (100) Szalewicz, K.; Jeziorski, B.; Monkhorst, H. J.; Zabolitzky, J. G. "Atomic and Molecular Correlation Energies with Explicitly Correlated Gaussian Geminals. I. Second-order Perturbation Treatment for He, Be, H₂, and LiH". *The Journal of Chemical Physics* **1983**, *78*, 1420–1430.
- (101) Wenzel, K. B.; Zabolitzky, J. G.; Szalewicz, K.; Jeziorski, B.; Monkhorst, H. J. "Atomic and Molecular Correlation Energies with Explicitly Correlated Gaussian Geminals. V. Cartesian Gaussian Geminals and the Neon Atom". *The Journal of Chemical Physics* **1986**, *85*, 3964–3974.
- (102) Szalewicz, K.; Zabolitzky, J. G.; Jeziorski, B.; Monkhorst, H. J. "Atomic and Molecular Correlation Energies with Explicitly Correlated Gaussian Geminals. IV. A Simplified Treatment of Strong Orthogonality in MBPT and Coupled Cluster Calculations". *The Journal of Chemical Physics* **1984**, *81*, 2723–2731.
- (103) Tew, D. P.; Klopper, W.; Manby, F. R. "The Weak Orthogonality Functional in Explicitly Correlated Pair Theories". *The Journal of Chemical Physics* **2007**, *127*, 174105.
- (104) Korobov, V. I. "Coulomb Three-Body Bound-State Problem: Variational Calculations of Nonrelativistic Energies". *Physical Review A* **2000**, *61*, 064503.
- (105) Varga, K.; Suzuki, Y. "Precise Solution of Few-Body Problems with the Stochastic Variational Method on a Correlated Gaussian Basis". *Physical Review C* **1995**, *52*, 2885–2905.

-
- (106) Bubin, S.; Varga, K. “Ground-State Energy and Relativistic Corrections for Positronium Hydride”. *Physical Review A* **2011**, *84*, 012509.
- (107) Stanke, M.; Bubin, S.; Molski, M.; Adamowicz, L. “Non-Born-Oppenheimer Calculations of the Lowest Vibrational Energy of HD Including Relativistic Corrections”. *Physical Review A* **2009**, *79*, 032507.
- (108) Hirschfelder, J. O. “Removal of Electron—Electron Poles from Many-Electron Hamiltonians”. *The Journal of Chemical Physics* **1963**, *39*, 3145–3146.
- (109) Boyle, P. P. “Options: A Monte Carlo Approach”. *Journal of Financial Economics* **1977**, *4*, 323–338.
- (110) Johnson, W., *Monte Carlo Methods in Finance: Simulation Techniques for Market Modeling*; HiTeX Press: 2024; 454 pp.
- (111) Kroese, D. P.; Brereton, T.; Taimre, T.; Botev, Z. I. “Why the Monte Carlo Method Is so Important Today”. *WIREs Computational Statistics* **2014**, *6*, 386–392.
- (112) Rosenbluth, M. N.; Rosenbluth, A. W. “Monte Carlo Calculation of the Average Extension of Molecular Chains”. *The Journal of Chemical Physics* **1955**, *23*, 356–359.
- (113) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. “Quantum Monte Carlo Simulations of Solids”.
- (114) Thijssen, J., *Computational Physics*; Cambridge University Press: 2007; 637 pp.
- (115) Metropolis, N.; Ulam, S. “The Monte Carlo Method”. *Journal of the American Statistical Association* **1949**, *44*, 335–341.
- (116) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. “Equation of State Calculations by Fast Computing Machines”. *The Journal of Chemical Physics* **1953**, *21*, 1087–1092.
- (117) Leclerc, G.-L. “Essai D’arithmetique Morale”.
- (118) *Histoire de l’Académie royale des sciences*, 1735; 696 pp.
- (119) Seneta, E.; Parshall, K. H.; Jongmans, F. “Nineteenth-Century Developments in Geometric Probability: J. J. Sylvester, M. W. Crofton, J.-É. Barbier, and J. Bertrand”. *Archive for History of Exact Sciences* **2001**, *55*, 501–524.
- (120) Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*; Clarendon Press: 1989; 412 pp.
- (121) *Monte Carlo Methods in Statistical Physics*; Binder, K., Ed.; Topics in Current Physics, Vol. 7; Springer: Berlin, Heidelberg, 1986.
- (122) Binder, K.; Heermann, D. W., *Monte Carlo Simulation in Statistical Physics*; Cardona, M., Fulde, P., Von Klitzing, K., Queisser, H.-J., Lotsch, H. K. V., red.; Springer Series in Solid-State Sciences, Vol. 80; Springer: Berlin, Heidelberg, 1988.

- (123) Newman, M. E. J.; Barkema, G. T.; Newman, M. E. J.; Barkema, G. T., *Monte Carlo Methods in Statistical Physics*; Oxford University Press: Oxford, New York, 1999; 490 pp.
- (124) Reiher, W. “Hammersley, J. M., D. C. Handscomb: Monte Carlo Methods. Methuen & Co., London, and John Wiley & Sons, New York, 1964. VII + 178 S., Preis: 25 s”. *Biometrische Zeitschrift* **1966**, 8, 209–209.
- (125) Bär, H.-J.; Bopp, P. “M. H. Kalos and P. A. Whitlock: Monte Carlo Methods, Volume I: Basics, John Wiley and Sons, New York, Chichester, Brisbane, Toronto and Singapore 1986, Library of Congress QA298.K35 1986. 186 Seiten Mit Einem Index, Preis: £ 29.00”. *Berichte der Bunsengesellschaft für physikalische Chemie* **1988**, 92, 560–560.
- (126) Bernoulli, J., *Jacobi Bernoulli profess. basil. & utriusque societ. ... Ars conjectandi, opus posthumum : accedit Tractatus de seriebus infinitis, et epistola Gallice scripta de ludo pilæ reticularis*; Basileæ : Impensis Thurnisiorum, Fratrum: 1713; 372 pp.
- (127) James, F. “Monte Carlo Theory and Practice”. *Reports on Progress in Physics* **1980**, 43, 1145.
- (128) Kahn, H. *Modification of the Monte Carlo*. RAND Corporation, 1959.
- (129) Arouna, B. “Adaptative Monte Carlo Method, A Variance Reduction Technique”. **2004**, 10, 1–24.
- (130) Hastings, W. K. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. *Biometrika* **1970**, 57, 97–109.
- (131) Markov, A. A., *Theory of Algorithms*; Academy of Sciences of the USSR: 1954; 468 pp.
- (132) Flyvbjerg, H.; Petersen, H. G. “Error Estimates on Averages of Correlated Data”. *The Journal of Chemical Physics* **1989**, 91, 461–466.
- (133) Cuzzocrea, A.; Scemama, A.; Briels, W. J.; Moroni, S.; Filippi, C. “Variational Principles in Quantum Monte Carlo: The Troubled Story of Variance Minimization”. *Journal of Chemical Theory and Computation* **2020**, 16, 4203–4212.
- (134) Snajdr, M.; Rothstein, S. M. “Are Properties Derived from Variance-Optimized Wave Functions Generally More Accurate? Monte Carlo Study of Non-Energy-Related Properties of H₂, He, and LiH”. *The Journal of Chemical Physics* **2000**, 112, 4935–4941.
- (135) Kent, P. R. C.; Needs, R. J.; Rajagopal, G. “Monte Carlo Energy and Variance-Minimization Techniques for Optimizing Many-Body Wave Functions”. *Physical Review B* **1999**, 59, 12344–12351.

-
- (136) Needs, R. J.; Towler, M. D.; Drummond, N. D.; López Ríos, P.; Trail, J. R. “Variational and Diffusion Quantum Monte Carlo Calculations with the CASINO Code”. *Journal of Chemical Physics* **2020**, *152*, 154106.
- (137) Jastrow, R. “Many-Body Problem with Strong Forces”. *Physical Review* **1955**, *98*, 1479–1484.
- (138) López Ríos, P.; Seth, P.; Drummond, N. D.; Needs, R. J. “Framework for Constructing Generic Jastrow Correlation Factors”. *Physical Review E* **2012**, *86*, 036703.
- (139) Fahy, S.; Wang, X. W.; Louie, S. G. “Variational Quantum Monte Carlo Nonlocal Pseudopotential Approach to Solids: Cohesive and Structural Properties of Diamond”. *Physical Review Letters* **1988**, *61*, 1631–1634.
- (140) Li, X.-P.; Ceperley, D. M.; Martin, R. M. “Cohesive Energy of Silicon by the Green’s-Function Monte Carlo Method”. *Physical Review B* **1991**, *44*, 10929–10932.
- (141) Rajagopal, G.; Needs, R. J.; Kenny, S.; Foulkes, W. M. C.; James, A. “Quantum Monte Carlo Calculations for Solids Using Special k Points Methods”. *Physical Review Letters* **1994**, *73*, 1959–1962.
- (142) Malatesta, A.; Fahy, S.; Bachelet, G. B. “Variational Quantum Monte Carlo Calculation of the Cohesive Properties of Cubic Boron Nitride”. *Physical Review B* **1997**, *56*, 12201–12210.
- (143) Harrison, R. J.; Handy, N. C. “Quantum Monte Carlo Calculations on Be and LiH”. *Chemical Physics Letters* **1985**, *113*, 257–263.
- (144) Lester, W. A., *Recent Advances In Quantum Monte Carlo Methods*; World Scientific: 1997; 244 pp.
- (145) Umrigar, C. J.; Nightingale, M. P.; Runge, K. J. “A Diffusion Monte Carlo Algorithm with Very Small Time-step Errors”. *The Journal of Chemical Physics* **1993**, *99*, 2865–2890.
- (146) Fick, A. “On Liquid Diffusion”. *Journal of Membrane Science* **1995**, *100*, 33–38.
- (147) Kosztin, I.; Faber, B.; Schulten, K. “Introduction to the Diffusion Monte Carlo Method”. *American Journal of Physics* **1996**, *64*, 633–644.
- (148) Lie, S.; Engel, F., *Theorie der Transformationsgruppen*; American Mathematical Soc.: 1970; 672 pp.
- (149) Trotter, H. F. “On the Product of Semi-Groups of Operators”. *Proceedings of the American Mathematical Society* **1959**, *10*, 545–551.
- (150) Suzuki, M. “Generalized Trotter’s Formula and Systematic Approximants of Exponential Operators and Inner Derivations with Applications to Many-Body Problems”. *Communications in Mathematical Physics* **1976**, *51*, 183–190.

- (151) Booth, G. H.; Thom, A. J.; Alavi, A. "Fermion Monte Carlo without Fixed Nodes: A Game of Life, Death, and Annihilation in Slater Determinant Space". *Journal of Chemical Physics* **2009**, *131*, 054106.
- (152) Guthrie, K.; Anderson, R. J.; Blunt, N. S.; Bogdanov, N. A.; Cleland, D.; Dattani, N.; Dobrutz, W.; Ghanem, K.; Jeszenszki, P.; Liebermann, N.; Manni, G. L.; Lozovoi, A. Y.; Luo, H.; Ma, D.; Merz, F.; Overy, C.; Rampp, M.; Samanta, P. K.; Schwarz, L. R.; Shepherd, J. J.; Smart, S. D.; Vitale, E.; Weser, O.; Booth, G. H.; Alavi, A. "NECI: N -Electron Configuration Interaction with an Emphasis on State-of-the-Art Stochastic Methods". *Journal of Chemical Physics* **2020**, *153*, 34107.
- (153) Spencer, J. S.; Blunt, N. S.; Choi, S.; Etrych, J.; Filip, M.-A.; Foulkes, W. M. C.; Franklin, R. S. T.; Handley, W. J.; Malone, F. D.; Neufeld, V. A.; Di Remigio, R.; Rogers, T. W.; Scott, C. J. C.; Shepherd, J. J.; Vigor, W. A.; Weston, J.; Xu, R.; Thom, A. J. W. "The HANDE-QMC Project: Open-Source Stochastic Quantum Chemistry from the Ground State Up". *Journal of Chemical Theory and Computation* **2019**, *15*, 1728–1742.
- (154) Brand, J. *Rimu.Jl*, RimuQMC, 2024.
- (155) Anderson, R. *M7*, 2024.
- (156) Dobrutz, W.; Weser, O.; Bogdanov, N. A.; Alavi, A.; Li Manni, G. "Spin-Pure Stochastic-CASSCF via GUGA-FCIQMC Applied to Iron–Sulfur Clusters". *Journal of Chemical Theory and Computation* **2021**, *17*, 5684–5703.
- (157) Weser, O.; Guthrie, K.; Ghanem, K.; Li Manni, G. "Stochastic Generalized Active Space Self-Consistent Field: Theory and Application". *Journal of Chemical Theory and Computation* **2022**, *18*, 251–272.
- (158) Dobrutz, W.; Smart, S. D.; Alavi, A. "Efficient Formulation of Full Configuration Interaction Quantum Monte Carlo in a Spin Eigenbasis via the Graphical Unitary Group Approach". *The Journal of Chemical Physics* **2019**, *151*, 094104.
- (159) Neufeld, V. A.; Thom, A. J. W. "Exciting Determinants in Quantum Monte Carlo: Loading the Dice with Fast, Low-Memory Weights". *Journal of Chemical Theory and Computation* **2019**, *15*, 127–140.
- (160) Li, J.; Otten, M.; Holmes, A. A.; Sharma, S.; Umrigar, C. J. "Fast Semistochastic Heat-Bath Configuration Interaction". *The Journal of Chemical Physics* **2018**, *149*, 214110.
- (161) Weser, O.; Alavi, A.; Manni, G. L. "Exploiting Locality in Full Configuration Interaction Quantum Monte Carlo for Fast Excitation Generation". *Journal of Chemical Theory and Computation* **2023**, *19*, 9118–9135.

-
- (162) Booth, G. H.; Smart, S. D.; Alavi, A. "Linear-Scaling and Parallelisable Algorithms for Stochastic Quantum Chemistry". *Molecular Physics* **2014**, *112*, 1855–1869.
- (163) Blunt, N. S.; Rogers, T. W.; Spencer, J. S.; Foulkes, W. M. C. "Density-Matrix Quantum Monte Carlo Method". *Physical Review B* **2014**, *89*, 245124.
- (164) Blunt, N. S.; Smart, S. D.; Kersten, J. A. F.; Spencer, J. S.; Booth, G. H.; Alavi, A. "Semi-Stochastic Full Configuration Interaction Quantum Monte Carlo: Developments and Application". *The Journal of Chemical Physics* **2015**, *142*, 184107.
- (165) Petruzielo, F. R.; Holmes, A. A.; Changlani, H. J.; Nightingale, M. P.; Umrigar, C. J. "Semistochastic Projector Monte Carlo Method". *Physical Review Letters* **2012**, *109*, 230201.
- (166) Spencer, J. S.; Blunt, N. S.; Foulkes, W. M. "The Sign Problem and Population Dynamics in the Full Configuration Interaction Quantum Monte Carlo Method". *The Journal of Chemical Physics* **2012**, *136*, 054110.
- (167) Cleland, D.; Booth, G. H.; Alavi, A. "Communications: Survival of the Fittest: Accelerating Convergence in Full Configuration-Interaction Quantum Monte Carlo". *Journal of Chemical Physics* **2010**, *132*, 041103.
- (168) Overy, C.; Booth, G. H.; Blunt, N. S.; Shepherd, J. J.; Cleland, D.; Alavi, A. "Unbiased Reduced Density Matrices and Electronic Properties from Full Configuration Interaction Quantum Monte Carlo". *The Journal of Chemical Physics* **2014**, *141*, 244117.
- (169) Luo, H.; Alavi, A. "Combining the Transcorrelated Method with Full Configuration Interaction Quantum Monte Carlo: Application to the Homogeneous Electron Gas". *Journal of Chemical Theory and Computation* **2018**, *14*, 1403–1411.
- (170) Schraivogel, T.; Cohen, A. J.; Alavi, A.; Kats, D. "Transcorrelated Coupled Cluster Methods". *The Journal of Chemical Physics* **2021**, *155*, 191101.
- (171) Schraivogel, T.; Christmaier, E. M. C.; López Ríos, P.; Alavi, A.; Kats, D. "Transcorrelated Coupled Cluster Methods. II. Molecular Systems". *The Journal of Chemical Physics* **2023**, *158*, 214106.
- (172) Baiardi, A.; Reiher, M. "Transcorrelated Density Matrix Renormalization Group". *The Journal of Chemical Physics* **2020**, *153*, 164115.
- (173) Haupt, J. P.; Hosseini, S. M.; López Ríos, P.; Dobrutz, W.; Cohen, A.; Alavi, A. "Optimizing Jastrow Factors for the Transcorrelated Method". *The Journal of Chemical Physics* **2023**, *158*, 224105.
- (174) Haupt, J. P.; Ríos, P. L.; Christmaier, E. M. C.; Kats, D.; Alavi, A. "Size Consistency and Transcorrelation". **2024**.

-
- (175) Haupt, J. P.; Ríos, P. L.; Christlmaier, E. M. C.; Filip, M.-A.; Simula, K.; Hauskrecht, J.; Liao, K.; Dobrautz, W.; Guthier, K.; Cohen, A. J.; Alavi, A. *Pythint: Transcorrelated Hamiltonian Integrals Library*, 2024.

Appendix (todo)

TODO: Not sure what will be needed in appendices.

Curriculum vitae

Education

- 2020 – 2025** PhD student, Theoretical Chemistry
Max Planck Institute for Solid State Research, Stuttgart, Germany
Thesis: Development of the Transcorrelated Full Configuration
Interaction Quantum Monte Carlo Method
Advisor: Prof Ali Alavi
- 2014 – 2019** Bachelor of Science, Honours Physics and Mathematics, Minor in
Computer Science
University of British Columbia, Vancouver, Canada
Thesis: Effects of Anisotropic Long-Range Hopping on Anderson
Localization
Advisor: Prof Roman Krems

Publications

TODO: Not really sure if there is a point to including a publications list when so few are actually published...

First Author

TODO: publications, I guess: optimization paper, binding curve paper, pythint paper (likely with a "to be submitted")

- (1) Haupt, J. P.; Hosseini, S. M.; López Ríos, P.; Dobrutz, W.; Cohen, A.; Alavi, A. "Optimizing Jastrow Factors for the Transcorrelated Method". *The Journal of Chemical Physics* **2023**, *158*, 224105
- (2) Haupt, J. P.; Ríos, P. L.; Christmaier, E. M. C.; Kats, D.; Alavi, A. "Size Consistency and Transcorrelation". **2024**
- (3) Haupt, J. P. et al. *Pythint: Transcorrelated Hamiltonian Integrals Library*, 2024

Coauthor

TODO: publications, I guess: ECPs, 2nd row atoms, deterministic optimisation, spin-dependent Jastrows
TODO: Pablo's surname doesn't appear correctly in some of these

- (4) Filip, M.-A.; Ríos, P. L.; Haupt, J. P.; Christmaier, E. M. C.; Kats, D.; Alavi, A. "Transcorrelated Methods for Second Row Elements". **2024**
- (5) Simula, K.; Ríos, P. L.; Christmaier, E. M. C.; Haupt, J. P.; Alavi, A. "Effective Core Potentials in the Transcorrelated Method". **2024**