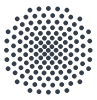


Universität Stuttgart

A real-time algorithm for FCIQMC

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

Vorgelegt von

Kai-Simon Guther

aus Tübingen

Hauptberichter: Prof. Ph.D. A. Alavi

Mitberichter: Prof. Dr. A. Köhn

Prüfungsvorsitzender: Prof. Dr. J. van Slageren

Tag der mündlichen Prüfung: 29. Juli 2020

Max-Planck Institut für Festkörperforschung

Institut für Theoretische Chemie
der Universität Stuttgart

2020

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem
Titel *A real-time algorithm for FCIQMC*
selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Declaration of Authorship

I hereby certify that the dissertation entitled
A real-time algorithm for FCIQMC
is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name/Name: _____

Unterschrift/Signed: _____

Datum/Date: _____

Contents

List of Acronyms

AFQMC Auxiliary Field Quantum Monte Carlo

aug-cc-pVXZ augmented cc-pVXZ

BCH Baker-Campbell-Hausdorff

cc-pVXZ correlation consistent polarized valence X-zeta

cc-pCVXZ correlation consistent polarized core valence X-Zeta

CC Coupled Cluster

CI Configuration Interaction

CS Cauchy-Schwartz

DFT Density Functional Theory

CASSCF Complete Active Space Self Consistent Field Theory

DMC Diffusion Monte Carlo

DMRG Density Matrix Renormalization Group

FCI Full Configuration Interaction

FCIQMC Full Configuration Interaction Quantum Monte Carlo

i-FCIQMC initiator FCIQMC

ST-FCIQMC Similarity-Transformed FCIQMC

GTOs Gaussian-type orbitals

HF Hartree-Fock

MEM Maximum Entropy Method

PCHB pre-computed Heat-Bath

PCPP pre-computed Power-Pitzer

PRNGs pseudo random number generators

RDMs Reduced Density Matrices

STOs Slater-type orbitals

VMC Variational Monte Carlo

Zusammenfassung

Komplexe numerische Methoden sind ein wichtiger Eckpfeiler der theorethischen Chemie, wo sie eingesetzt werden um physikalische und chemische Eigenschaften von Molekülen und Festkörpern vorherzusagen und Einblick in die zugrundeliegenden physikalischen Mechanismen gewähren. Durch die stetige Verbesserung der verfügbaren Hardware und der Entwicklung neuer Methoden werden immer komplexere Systeme zugänglich für eine numerische Betrachtung, wodurch die sich die elektronische Struktur selbst komplexer Moleküle direkt aus den grundlegendsten theoretischen Annahmen ableiten lassen.

In dieser Arbeit wird die zeitabhängige Schrödingergleichung mithilfe des Full Configuration Interaction Quantum Monte Carlo (FCIQMC) Formalismus behandelt. Damit wird ein neuartiger Algorithmus zur Berechnung dynamischer Eigenschaften vorgestellt, der sich sowohl auf Modellsysteme als auch auf ab-initio Systeme anwenden lässt. Die FCIQMC Methode wurde in den letzten Jahren als hochgenaue Methode zur Berechnung von Grundzustandsenergien stark korrelierter Systeme etabliert, wobei ein Formalismus basierend auf stochastischer linearer Algebra verwendet wird um den Speicherbedarf, der auf dieser Ebene der Theorie exponentiell mit der Systemgröße anwächst, zu senken. Dieser Formalismus wird hier an das zeitabhängige Problem angepasst, und eine neue Methode, die Zeitentwicklung eines Quantenzustands zu berechnen wird vorgeschlagen. Dafür wird die Zeitentwicklung entlang einer allgemeinen komplexwertigen Kurve vorgenommen, wodurch die Simulation durch eine simultane Imaginärzeitentwicklung stabilisiert wird. Ausgehend von der zeitabhängigen Lösung der Schröringergleichung wird dann die Greensche Funktion, und damit durch analytische Fortsetzung die Spektralzerlegung und Photoabsorptionsspektren bestimmt.

Der resultierende Algorithmus ist wesentlich speichereffizienter als deterministische Alternativen und kann daher auf größere Systeme angewandt werden, und dabei trotzdem sehr genaue Spektralfunktionen liefern. Um dies zu demonstrieren wird die Methode am Beispiel einer Reihe von ab-initio Systemen sowie dem Hubbard-Modell getestet, einem in der Festkörpertheorie weit verbreiteten Modell. Der Vergleich mit experimentellen

spektroskopsichen Daten sowie numerischen Referenzdaten dient dabei zur Beurteilung der Genauigkeit der berechneten Spektren und Anregungsenergien. Insbesondere für das Hubbard-Modell wird die Rolle von Kontrollparametern zur systematischen Annäherung an die Ergebnisse numerischer exakter Diagonalisierung untersucht, wobei dem neuen Algorithmus zugleich auch Systeme jenseits der Anwendbarkeit deterministischer exakter Diagonalisierung zugänglich sind. Andererseits wird durch die Imaginärzeitentwicklung eine analytische Fortsetzung notwendig, was die Auflösung und Genauigkeit begrenzt.

Im letzten Teil der Arbeit wird dann die Entwicklung und Anwendung der transkorrelierten Methode durch ähnlichkeitstransformierte FCIQMC für ab-initio Systeme diskutiert. Dieser explizit korrelierte Ansatz ist ein vielversprechender Weg, Fehler durch die Unvollständigkeit der Orbitalbasis zu korrigieren und kann direkt in die FCIQMC Methode integriert werden, wobei diese Arbeit sich dann mit der Implementierung der ähnlichkeitstransformierten FCIQMC beschäftigt. Der drastisch reduzierte Fehler durch die Unvollständigkeit der Basis wird unter anderem am Beispiel der Bindungskurve von Be_2 demonstriert.

Abstract

Involved computational methods form a cornerstone of theoretical chemistry in order to make predictions on the physical and chemical properties of molecules and solids of increasing complexity and gain theoretical insight into the underlying physical mechanisms. The availability of increasingly performant hardware and the development of novel methods pushed the limits of accessible systems further and further, allowing to predict the electronic structure of even complex molecules from first principles alone.

Within this thesis, the Full Configuration Interaction Quantum Monte Carlo (FCIQMC) framework is used to solve the time-dependent Schrödinger equation, presenting a novel algorithm for the calculation of dynamic properties both from first principles and for model systems. The FCIQMC method has been established in the last years as a high accuracy method for the calculation of ground state energies of strongly correlated electronic systems, using a stochastic linear algebra framework to push the memory bottleneck deterministic methods at the same level of theory exhibit to higher scales. Here, this formalism is applied to the time-dependent problem, proposing an algorithm to calculate the time-evolution of a quantum state. This is then carried out on a generic complex time contour, stabilizing the procedure by a simultaneous, slow imaginary time evolution. From the time-dependent wave function, Green's functions, spectral decompositions and photo-absorption spectra are then extracted by analytic continuation, granting insight to the excitation spectrum. The resulting algorithm is significantly more memory efficient than deterministic alternatives and thus scalable to larger systems, while still being able to accurately calculate spectral functions at the FCI level of theory. To demonstrate this, the method is benchmarked at the example of a range of ab-initio systems as well as at the Hubbard model, a model used prominently in solid state theory. The comparison to experimental spectroscopic data as well as reference data from Lanczos FCI and excited state FCIQMC is used to gauge the accuracy of the resulting spectra and excitation energies. In particular, for the Hubbard model, systematic improvability towards the FCI result is demonstrated, while also being able

to go beyond the limits of deterministic FCI methods. On the other hand, due to the complex time evolution, analytic continuation is required to regain the spectrum, which limits the spectral resolution and accuracy.

In the last part of the thesis, developments and application of the transcorrelated method by similarity transformed FCIQMC for ab-initio systems are discussed. This explicitly correlated ansatz is a promising approach to treating basis set incompleteness errors, and can be naturally integrated into the FCIQMC method, and this work is addressing the implementation of the similarity transformed FCIQMC. The greatly reduced basis set incompleteness error is demonstrated at the example of the potential energy curve of the Beryllium dimer.

Acknowledgments

My deep gratitude is to Ali Alavi, whose support made the last four years a wonderful time, and I am still amazed by the efforts he made in guiding me, drawing on a seemingly inexhaustible supply of ideas, always full of enthusiasm and always having a minute for discussion, at the same time leaving me space to explore own ideas. I am also very grateful for the numerous opportunities to visit conferences, workshops and summer schools he gave me.

I would also like to gratefully acknowledge the funding by the Max Planck Society and the resources in form of software infrastructure, computing time, journal access and so on that were supplied by the Max Planck Institute for Solid State research (MPI-FKF).

I am grateful to Andreas Köhn and Joris van Slageren for their efforts with being part of my examination committee and to Rotraut Merkle for the external advision within the MPI-FKF PhD committee.

This work would not have been possible without Olle Gunnarsson, being the person who initially perceived the time-dependent FCIQMC and who supplied a great deal of input for me, be it in the form of discussions on method development, reference data, the MEM software or insights into physics in general. I am thankful for learning so much from him, especially his rational and questioning approach left a deep impression on me.

Many thanks goes to my colleagues Alexander 'Sasha' Lozovoi, who enriched so many late afternoons with curious discussions - I really miss the linguistic excursions - and managed to stir some wonder and curiosity about Russian culture and language, and Werner Dobrautz for being an example and sharing his experience, in particularly successfully aiding me finding my way through the `NECI` code, as well as for proofreading parts of my thesis and sharing his useful LaTeXtitlepage package, it has been a great time in our office.

I am very grateful for learning so many things from Ke Liao, about neural networks, Monte Carlo and machine learning, Chinese culture and proper food and his help with

estimating basis set extrapolations, and would also like to thank Nikolay Bogdanov for all the discussions on history, the input on useful improvements of NECI and his essential help for dealing with the Russian railway, and would also like to thank Ke and Nikolay for proofreading. Many thanks also goes to Giovanni Li Manni who put so much effort into teaching me chemistry and whose unique talent in finding symptoms of programming errors has led me to remove more than a few bugs. I am glad we have spent countless hours in discussion over details of methods or results, and his admirable curiosity inspired me to a great deal.

I am indebted to Aron Cohen for the most fruitful debugging session, his great work on the ST-FCIQMC and for inspiring me with his pragmatic approach. My thanks goes also to Hongjun Luo for the fruitful collaboration and sharing his insights on the transcorrelated approach.

Great thanks also to Armin Burkhardt and Armin Schuhmacher for the technical support and keeping the cluster at the FKF running, as well as to Christian Guggenberger for maintaining the Vortex cluster and Markus Rampp for the application support as well as organizing access to SuperMUC-NG. I thankfully acknowledge the Max Planck Computing and Data facility for providing computing time at the supercomputer Cobra as well as the HPC cluster Vortex and the Gauss Center for Supercomputing for providing computing time at SuperMUC-NG under project pr86fe. I would also like to thank Florian Merz for bringing in his expertise to improve NECI .

I would also like to thank Pradipta Samanta for fruitful discussions on the time-dependent algorithm, potential applications and shared-memory hash tables, Oskar Weser for countless discussions on coding practices, motivating me to explore new tools and proofreading, David Tew for insights on explicitly correlated methods, Vamshi Katukuri for proofreading as well as everyone in the Alavi group for the amazing time there.

Of course, I could not imagine this time without my awesome roommates Michael, Ulrike and Matthias with whom I share so many fond memories and on who I could always rely on.

My special gratitude is to my family, who always supported me in pursuing my dreams and would always encourage me to take the next step.

Software authorship declaration

All FCIQMC calculations are carried out using the development version of the `NECI` software [1], where the time-dependent algorithm has been implemented by the author based on a draft by Werner Dobrautz and the support for ab-initio transcorrelated Hamiltonians, the pre-computed heat-bath and pre-computed power-pitzer excitation generation and any other new features and methods presented here have been implemented by the author in the existing `NECI` code base.

The maximum entropy analytic continuation is carried out using a program written by Olle Gunnarsson.

The transcorrelated Hamiltonian is calculated using orbitals generated by PySCF [2] by a program written by Aron Cohen and adapted by the author for distributed memory parallel use, parallel I/O and specific use cases such as counterpoise correction.

All data graphs have been generated using the matplotlib python library [3].

Chapter 1

Introduction

At the beginning of the twentieth century, sparked by new stimuli such as the interpretation of the photoelectric effect [4], the study of black-body radiation [5, 6], the Stern-Gerlach experiments investigating the magnetic moments of silver atoms [7] and the discovery of the Compton effect [8], it became apparent that classical physics was not suited to describe processes at the microscopic scales of atoms and molecules and that other fundamental laws must be behind the perplexing behavior observed in these experiments.

The new physics came in the shape of quantum mechanics, introducing particle-wave dualism [9], where instead of the classical description of a system by a vector $(\mathbf{P}, \mathbf{R})$ in phase space containing the momenta and positions of every single object, a quantum mechanical system would be described by a many-body wave function $|\psi\rangle \in \mathcal{L}_2$, a high-dimensional function ψ of the degrees of freedom of each particle [10, 11]. Observables, which are classically functions of the phase space vector, are now linear, hermitian operators $\hat{A}$ on the Hilbert space $\mathcal{L}_2$. The wave function itself only carries statistical meaning, as the overlap $\langle\phi\,|\,\psi\rangle$ with another state $|\phi\rangle \in \mathcal{L}_2$ is interpreted as the probability of observing the state $|\phi\rangle$ when probing it. Then, measuring an observable $\hat{A}$ is described as probing the eigenstates of $\hat{A}$, each associated with the corresponding eigenvalue as a measurement outcome [10].

Mathematically, a quantum system is governed by the Schrödinger equation for the wave function [12] ,

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\,|\psi(t)\rangle = \hat{H}\,|\psi(t)\rangle\ , \tag{1.1}$$

where $\hat{H}$ is the Hamiltonian operator, which is the hermitian operator corresponding to the energy of the system. Therein lies a conceptual similarity to the classical dynamics,

which, too, is governed by the Hamiltonian function via the Hamiltonian equations of motion, but these by no means imply the Schrödinger equation. For convenience, atomic units are used from now on, in which $\hbar = 1$ [13] unless otherwise noted.

Notable implications of this equation are the norm of the wave function being conserved, this is required for the interpretation of overlaps as probabilities, the conservation of energy and the linearity of the theory [10]. The latter is crucial for the interpretation of ψ as a *wave* function, if $|\psi_1\rangle$ and $|\psi_2\rangle$ are solutions, so is $|\psi_1\rangle + |\psi_2\rangle$, introducing superposition and interference into the theory. In particular, this allows for correlation between different particles.

Initially based on spectroscopic observations, an internal degree of freedom of electrons which couples to the magnetic field had to be introduced [14], the so-called spin. The theoretical foundation thereof came from including special relativity, leading to the Dirac equation [15] that postulates the existence of a new, internal degree of freedom coupling to the magnetic field, explaining the occurrence of the spin. The spin operators obey the algebraic relations of angular momentum operators, and are therefore quantized to half integer values, with the electron having a spin of $\frac{1}{2}$ [10].

An essential difference to the classical description is the indistinguishability of identical particles, which has far-reaching consequences. A central result in quantum theory regarding this is the spin-statistics theorem [16, 17, 18] which states that the wave function is anti-symmetric with respect to exchange of identical particles with half-integer spin and symmetric with respect to exchange of identical integer spin particles. Since electrons fall into the former category, electronic structure theory has to deal with an anti symmetric function, which leads to a manifestation of the infamous sign problem [19] in stochastic numerical treatment. An important corollary of the spin-statistics theorem is the Pauli exclusion principle [20], which states that no two electrons with the same spin can occupy the same state (even though this is an implication of the spin-statistics theorem, historically, it was realized first).

1.1 The many-body problem

The time-dependent Schrödinger equation (1.1) implies that the stationary states are the eigenstates of $\hat{H}$, leading to the time-independent Schrödinger equation

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle\ , \tag{1.2}$$

with an energy eigenvalue E. Since $\mathcal{L}_2$ is of infinite dimension, there is an infinite number of solutions to Eq. (1.2), and solving equation (1.2) for all eigenvalues is equivalent to solving the time-dependent equation (1.1) for all initial states. But for many physical and chemical properties of interest, only the low-energy solutions are relevant, particularly the lowest energy state $|\psi_0\rangle$, which is referred to as ground state. Knowledge of the ground state, and more generally the low energy spectrum of $\hat{H}$ allows to calculate expectation values of observables like dipole moments, bond lengths and geometries, binding and ionization energies, band gaps and absorption spectra, and many more [21, 22, 23].

The operator $\hat{H}$, as introduced above, corresponds to the energy of the system, which is composed of the kinetic terms of all individual particles and their potential energy. The problem considered in quantum chemistry, from a single Hydrogen atom to large proteins, consists of a number of electrons and nuclei. The potential energy is then the Coulomb interaction between the electrons themselves and the electrons and the nuclei, such that the non-relativistic Hamiltonian reads [11]

$$\hat{H} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_I \frac{1}{2m_I}\nabla_I^2 - \sum_{iJ} \frac{Z_J}{|\mathbf{r}_i - \mathbf{R}_J|} - \sum_{IJ} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} + \sum_{ij} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \qquad (1.3)$$

with the electron coordinates $\mathbf{r}_i$ and indices i and the nuclear coordinates $\mathbf{R}_I$, indices I, masses m_I (in atomic units) and charges Z_I.

The analytic solution of the stationary Schrödinger equation for the Hamiltonian (1.3) has only been found for very few special cases, such that in general, one has to rely on numerical solutions, with varying degree of approximations [24].

Still, the determination of even the lowest eigenvalues of $\hat{H}$ would yield deep insight into the nature of a molecule, unraveling the equilibrium electronic structure, and would thus allow to simulate its behavior by purely theoretical means, predicting properties directly from first-principle theory [25]. This also extends to solid-state compounds, which are after all governed by exactly the same Hamiltonian, as they can be viewed as macroscopic molecules with a regular, periodic structure. There, the periodicity allows to consider the whole solid as a, on molecular scales, infinite sequence of a repeated unit cell, from which an effective Hamiltonian for the solid can be constructed [11]. An insightful, qualitative approach to the Hamiltonian (1.3) for solids is the construction of model Hamiltonians that mimic its qualitative properties while simplifying the interaction and reducing the degrees of freedom, prominent representatives being the Hubbard model [26], the Heisenberg model [27] to describe magnetism, the Anderson impurity

model [28] or the BCS model Hamiltonian for superconductivity [29]. Such models can deliver explanations for phenomena such as magnetism or superconductivity, but they always depend on parameters, which in turn either have to be estimated for a specific compound from the solution of the ab-initio Hamiltonian, or determined based on experimental data.

Nevertheless, especially with the tool of second quantization [30], treating model Hamiltonians and the ab-initio Hamiltonian is formally similar. Hence, the same methods can be feasible for both, like for example the Density Matrix Renormalization Group (DMRG) [31] which was originally devised for one-dimensional lattice models but successfully carried over to molecular systems [32, 33],

Seeing from the perspective of the prediction of electronic and chemical properties of compounds that the investment of computational resources to a numerical solution of the stationary Schrödinger equation is worthwhile, the problem requires the deployment of involved numerical algorithms, and a computationally affordable, general black box solution is still not available [34].

The core of the difficulty lies therein that the resulting eigenvalue equation in the functional space $\mathcal{L}_2$ is a high-dimensional non-separable differential equation, that is, the solution can generally not be written as a product of independent single-particle functions, this is tied to the property known as electron correlation. A great variety of different algorithms to approach the problem from different angles exists, these can be sorted into different categories based on the field of application and the underlying formalism [25, 34].

The class of Density Functional Theory (DFT) methods [35, 36] derives from the Hohenberg-Kohn theorem [37] which states that the ground state energy is uniquely determined by the electron density of the ground state, the energy is consequently minimized as a function of the density. While formally an exact formalism is guaranteed to exist by the Hohenberg-Kohn theorem, the exact functional is generally not known and the implementation of DFT hence relies on an effective treatment of electron interactions [38]. The striking advantage of DFT is its high scalability, being applicable to systems with thousands of electrons [39, 40], but it is better suited for weakly correlated systems due to the effective treatment which is difficult to improve systematically [38, 41].

On the other end, there are correlated wave function based methods that explicitly solve for the many-body wave function $|\psi\rangle$. Again, many different flavors of these exist, which differ in the degree and type of approximation made, for example Møller-Plesset Perturbation theory [42, 43, 44], Coupled Cluster (CC) methods [45, 46, 47, 48],

DMRG [49, 50, 51, 31], Complete Active Space Self Consistent Field Theory (CASSCF) [52, 53, 54], Selected Configuration Interaction methods [55, 56, 57, 58], or the Full Configuration Interaction Quantum Monte Carlo (FCIQMC) method [59, 60, 61, 62] which is used and built upon in this work. Many of these can in principle be systematically improved to yield a numerically exact solution, but also here, approximations have to be made to make calculations affordable. The scaling and also the realistic accuracy then depends on the precise nature of these approximations, where the treatment of systems with significant electron correlation requires more involved and thus also more expensive methods [24, 63, 64].

1.2 Context of this work

In this work, a particular approach to the many body problem to aim at the highest levels of accuracy is chosen, the Full Configuration Interaction Quantum Monte Carlo method. As the name suggests, the method at its core is a stochastic version of Full Configuration Interaction (FCI) linear algebra methods, and is therefore designed for applications where an FCI level treatment is required, but the memory cost of deterministic methods is prohibitive. Notably, it can be used to reliably capture the energetics of transition metal complexes in the stochastic CASSCF scheme [65, 66, 67], obtain essentially FCI energies for molecules composed of light atoms [68, 69], evaluate magnetic coupling parameters for solid state models [70] and study lattice models themselves [71, 72], including the bosonic Hubbard model [73]. In this thesis however, with the electronic structure problem in Born Oppenheimer approximation, exclusively fermionic systems are considered.

The solution of the stationary many-body problem for the ground state is already a notoriously difficult problem, and this is true even more so for the original time-dependent problem. As soon as not only ground state properties are sought, but also energy dependent quantities like response functions such as optical conductivities [74], photo-electron spectra [75, 76, 77], in for particular core-electron spectroscopy [78, 79], or the density of states in solids [21], it is inevitable to go beyond the ground state and take many energy eigenstates into account. The time-evolution of a system is also important for the study of reaction dynamics [80, 81], and in the context of non-equilibrium dynamics [82], adding to the class of applications that explicitly require extending the theory beyond the treatment of the equilibrium ground state.

Formally, every function of energy corresponds to a function of time by Fourier trans-

form [83], so two equivalent ways of capturing dynamic properties exist, either in the energy domain or in the time domain. For ab-initio calculations, these correspond to either considering the time-dependent Schrödinger equation or the stationary one for a range of energies. For the latter, the ground state methods can often be extended to treating excited energy eigenstates of $\hat{H}$ [84, 85, 86], but particularly the calculation of spectral functions from this point of view could get affected by the problem of generally not knowing the number of required energy eigenstates and the range of their eigenvalues a priori. Directly approaching the time-dependent problem solves this issue, since the initial state uniquely defines the state at all times and can be tailored to yield the desired property [87, 88]. Furthermore, if strong time-dependent external fields are to be considered, the time-dependent problem has to be dealt with [89, 90].

To this end, powerful tools are available for weakly correlated systems in the form of time-dependent density functional theory [91, 92, 93] and the GW approximation [94, 95, 96], the former being better suited for application to finite systems while the latter finds application in solids [97]. In the case of strong correlation however, the problem becomes more involved and the available methods have severe limitations. Deterministic linear algebra based methods [98, 99] suffer from the infamous memory bottleneck [100] of deterministic Configuration Interaction (CI) methods and are thus strongly limited in their application by the available hardware. For the special case of one-dimensional lattice models, efficient algorithms are available in the form of Time Evolving Block Decimation and related Matrix Product State based methods [101, 102], but only very recently these have found application to ab-initio Hamiltonians [103, 104], and the computational cost of Matrix Product State based methods for quantum chemistry is much higher than for one-dimensional lattice models [105]. For systems with single reference character on the other hand, Equation of Motion Coupled Cluster can give insight into the excitation spectrum [106, 107].

Now, Quantum Monte Carlo methods exhibit a dynamical sign problem in real-time [108], and while the inchworm algorithm resolves this for certain model systems [108], a general solution for ab-initio Hamiltonians is not available. Thus, Quantum Monte Carlo methods commonly operate in imaginary time [109], from which obtaining dynamic properties is an ill-conditioned problem [110, 111].

In this work, a novel approach to the application of Quantum Monte Carlo methods in this context is explored, which is based on the FCIQMC framework of stochastic matrix vector multiplication and sparse wave function representation in a second quantized space. The time-dependent Schrödinger equation is considered on a complex time

contour, capturing dynamic properties via the real-time part of the time evolution, while addressing instabilities of the calculation with the imaginary part. This results in a systematically improvable algorithm for calculating spectral functions via the time-dependent many-body Schrödinger equation that can be applied with high accuracy beyond the limits of deterministic methods.

A ubiquitous tool in quantum chemistry is the use of atomic and molecular orbital basis sets, particularly Gaussian-type orbitals [112] which reduces the electronic structure problem from a high-dimensional coupled partial differential equation to an extremely large matrix eigenvalue problem [100]. For most quantum chemistry methods, including the FCIQMC method, this is a requirement, but a basis set error is introduced on the way [112], owing to the fact that not all possible wave functions are representable in a given basis set.

A major reason behind this basis set error is the Kato cusp condition [113] that demands short-range correlation of electrons and a non-continuity of the derivative of the wave function at coalescence points.

This leads to the notoriously slow convergence of E_0 with respect to the single particle basis [114, 115], curing which requires the explicit introduction of an r_{12} dependency [114, 116]. Several approaches to explicitly include the r_{12} dependence have been devised, such as the Gaussian geminals ansatz [117, 118, 119], where a two particle basis set is used over a single particle basis, inscribing pair correlation into the basis set, or the family of F12 methods [116, 120, 121], which has also been combined with FCIQMC [122].

It is however explicitly incorporating electron correlation via the transcorrelated ansatz of Boys and Handy [123, 124, 125, 126, 127] which recently showed promising prospects in combination with the FCIQMC method [72, 128, 129]. This combination might yet help to overcome the basis set incompleteness error occurring in the method.

The second part of this thesis is therefore dealing with recent developments regarding the combination of FCIQMC with the transcorrelated method addressing ab-initio systems, and will cover technical aspects of the method as well as extensive benchmark calculations.

The structure of this thesis is as follows. First, in chapter 2, an overview of theoretical tools and methods that are relevant to this work is given. Starting with the CI problem and deterministic FCI approaches, basic concepts of Quantum Monte Carlo methods are introduced there, and the FCIQMC algorithm is recapitulated in detail. The chapter closes by discussing the calculation of dynamic properties, specifying how Green's functions are to be calculated from the time-dependent FCIQMC solution.

Chapter 3 will then introduce the time-dependent FCIQMC formalism, first presenting the general idea and then discussing algorithmic details, addressing stability, the issue of correlated data and the generic complex time procedure that is required for application to larger systems. Benchmarks and applications of the method are then presented in chapter 4 at the example of the Hubbard model, both at half filling and doped, the Beryllium and Carbon atoms, the Carbon dimer and a Cu_2O cluster. The chapter closes with considerations on the computational performance of the method.

In chapter 5, the theoretical tools of the transcorrelated ansatz by Boys and Handy are revisited, elaborating on its interface with multi-configurational methods as encountered in the literature. That serves as preparation for chapter 6 which discusses the details of Similarity-Transformed FCIQMC (ST-FCIQMC) and explores its applicability to molecular systems, mainly at the example of the potential energy curve of Be_2.

Lastly, chapter 7 constitutes a summary of the results and suggests possibilities for further development.

Chapter 2

Methods and algorithms

This chapter shall give an overview of the theoretical methods and algorithms used in this work, as well as some related methods to give context. First, the basics of deterministic Full Configuration Interaction methods shall be revisited at the example of the Lanczos method [130] which will serve as a benchmark in Chapter 4, followed by an introduction to stochastic method, starting with a recapitulation of the general concepts and notable Quantum Monte Carlo methods, before addressing some technical details. The FCIQMC method [59, 60, 61, 62] is then discussed in detail to give context to the extensions presented later. Lastly, the dynamic properties of interest for FCIQMC are characterized and the evaluation of spectral functions via real-time evolution is revisited.

2.1 Basic approximations to the quantum chemical problem

As mentioned in section 1.2, the introduction of a basis set of one-electron functions maps the many-body problem to a finite-dimensional eigenvalue problem, together with the Born Oppenheimer approximation [131], this leads to the CI eigenvalue problem [100] which is the starting point for the methods used in this thesis.

2.1.1 Born-Oppenheimer approximation

An important tool and common approximation in quantum chemistry is the Born-Oppenheimer approximation [23, 131], which decouples the nuclear and electronic degrees of freedom. The wave function of the nuclei is then described the effective potential

of the potential energy surface $E_e(\mathbf{R})$. The approximation is justified by the large difference in mass between electrons and nuclei and therefore different energy scale of the electronic and nuclear Hamiltonian. To this end, the wave function ψ, as a function of both the nuclear and electronic degrees of freedom is assumed to factorize into an electronic part ψ_e and a nuclear part ψ_n as

$$\psi(\mathbf{r}, \mathbf{R}) = \psi_n(\mathbf{R})\psi_{e,\mathbf{R}}(\mathbf{r})\,, \tag{2.1}$$

where the electronic wave function ψ_e only depends on the nuclear coordinates $\mathbf{R}$ as parameters, i.e.

$$\nabla_{\mathbf{R}}\psi_e = 0\,. \tag{2.2}$$

The full Hamiltonian is then decomposed into an electronic part $\hat{H}_e$ and a nuclear part $\hat{H}_n$. For the description of the electronic wave function, the nuclear coordinates are assumed to be fixed by the parameters $\mathbf{R}$, and no feedback from the nuclear wave function is assumed. That leads to the purely electronic Schrödinger equation

$$\left[-\sum_i \frac{1}{2m_e}\nabla^2_{\mathbf{r_i}} + \sum_{i,j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}\right]\psi_e(\mathbf{r}) = E_e(\mathbf{R})\psi_e(\mathbf{r})\,, \tag{2.3}$$

with the nuclear charges Z_I and the electronic energy $E_e(\mathbf{R})$. Note that the fermionic statistic requires the wave function $\psi(\mathbf{r})$ to be anti symmetric with respect to exchange of two electrons.

The nuclear wave function is then described by the effective Schrödinger equation

$$\left[-\sum_i \frac{1}{2m_I}\nabla^2_{\mathbf{R_I}} + \sum_{I,J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} + E_e(\mathbf{R})\right]\psi_n(\mathbf{R}) = E_n\psi_n(\mathbf{R})\,, \tag{2.4}$$

with the nuclear masses m_I, and the energy eigenvalue E_n. Included are the nuclear kinetic term, the nuclear-nuclear interaction and the effective electronic energy $E_e(\mathbf{R})$, and the latter is assumed as a static background for the nuclear wave function.

All molecular calculations in this thesis are within Born-Oppenheimer approximation.

2.1.2 Second quantization and orbital bases

Even though the electrons described by equation (2.3) are still correlated, in fact, it can be shown that the solution must contain some correlation [113], the introduction of a single-particle basis carries a number of advantages. One the one hand side, if

the correlation effects are small, an anti-symmetrized product solution can be a valid approximation, this is the notion of Hartree-Fock (HF) theory [132, 133, 134]. One the other side, formulating the problem within a finite-dimensional product space successfully deals with the required anti-symmetry of the electronic wave function [30], and also leads the way to the formulation of efficient multiconfigurational algorithms based on superposition of product states, which include the wave function based methods mentioned in section 1.1.

Given a basis of N single particle wave functions φ_i, called *(spin-)orbitals* [30], an anti-symmetrized n-electron product state ($n < N$) can be expressed as a configuration of n orbitals $\varphi_{i_1}, \ldots, \varphi_{i_n}$, the anti-symmetrization results in the n-electron wave function ψ being written in the form of a Slater determinant [135, 136]

$$\psi(\mathbf{r}_1\sigma_1, \ldots, \mathbf{r}_n\sigma_n) = \frac{1}{\sqrt{n!}}\mathrm{Det}\begin{pmatrix} \varphi_{i_1\sigma_1}(\mathbf{r}_1) & \cdots & \varphi_{i_n\sigma_1}(\mathbf{r}_1) \\ \vdots & \ddots & \vdots \\ \varphi_{i_1\sigma_n}(\mathbf{r}_n) & \cdots & \varphi_{i_n\sigma_n}(\mathbf{r}_n) \end{pmatrix}, \tag{2.5}$$

which is uniquely determined by the combination of orbitals $\varphi_{i_1\sigma_1}, \ldots, \varphi_{i_n\sigma_n}$. The spin-orbitals consist of a *spatial orbital* ϕ and a spin component ξ with fixed spin projection σ

$$\varphi_{i\sigma}(\mathbf{r}) = \phi_i(\mathbf{r})\xi(\sigma), . \tag{2.6}$$

These Slater determinants can be used as a basis for a finite subspace of the full n-particle $\mathcal{L}_2$, within which the solution to the eigenvalue problem is to be sought. A systematic approximation of the full Hilbert space is possible by adding additional orbitals, such that in the limit of $N \to \infty$, the entire Hilbert space is recovered [24]. The development of orbital bases itself is non-trivial and a plethora of potential choices exists [112, 137], notably the families of Slater-type orbitals (STOs) with an exponential radial dependence $\phi \propto \mathrm{e}^{-\alpha r}$ and of Gaussian-type orbitals (GTOs) with a Gaussian radial part $\phi \propto \mathrm{e}^{-\alpha r^2}$ [137]. While the latter converge slower to the complete basis set limit, the calculation of matrix elements between them is more efficient, making them widespread basis sets in quantum chemical application [137, 138]. The selection from the substantial supply of GTOs is again a choice that has to be made based on the particular system of interest [112].

However, independently of the precise choice of the basis set, the resulting Slater determinants can efficiently be described within the formalism of second quantization [30]. The key concept is the introduction of annihilation and creation operators $\hat{a}_{i\sigma}$ and $\hat{a}^\dagger_{i\sigma}$,

which are adjoint to each other and whose action is the removal/addition of an electron from/to the orbital $\varphi_{i\sigma}$. To guarantee fermionic statistics, they have to follow the fermionic anti-commutation relations [11]

$$\{\hat{a}_{i\sigma}, \hat{a}^\dagger_{j\sigma}\} = \delta_{i,j} \tag{2.7}$$
$$\{\hat{a}_{i\sigma}, \hat{a}_{j\sigma}\} = 0 \tag{2.8}$$
$$\{\hat{a}^\dagger_{i\sigma}, \hat{a}^\dagger_{j\sigma}\} = 0\,, \tag{2.9}$$

The Slater determinant (2.5) can then be expressed by subsequently adding electrons to the orbitals $\varphi_{i_1\sigma_1}$ to $\varphi_{i_n\sigma_n}$ as

$$|\psi\rangle = \hat{a}^\dagger_{i_1\sigma_1} \dots \hat{a}^\dagger_{i_n\sigma_1} |0\rangle \ , \tag{2.10}$$

with the 0-electron state $|0\rangle$.

Furthermore, a two-body operator such as the Hamiltonian in equation (2.3) can also be expressed in terms of the annihilation and creation operators as [30]

$$\hat{H} = \sum_{ia\sigma} h^a_i \hat{a}^\dagger_{a\sigma} \hat{a}_{i\sigma} + \frac{1}{2} \sum_{ijab\sigma\sigma'} V^{ab}_{ij} \hat{a}^\dagger_{a\sigma} \hat{a}^\dagger_{b\sigma'} \hat{a}_{j\sigma'} \hat{a}_{i\sigma} + \text{const}\,, \tag{2.11}$$

where, in the case of the electronic structure Hamiltonian, the tensors h and V are given by the integrals [30]

$$h^a_i = \int \mathrm{d}r\, \bar{\phi}_a(\mathbf{r}) \left[-\frac{1}{2m_e} \nabla^2 - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \right] \phi_i(\mathbf{r}) \tag{2.12}$$
$$V^{ab}_{ij} = \int \mathrm{d}r_1 \mathrm{d}r_2\, \bar{\phi}_a(\mathbf{r}_1) \bar{\phi}_b(\mathbf{r}_2) \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\,. \tag{2.13}$$

In the non-relativistic case, $\hat{H}$ conserves the spin projection, such that only the spatial orbitals enter the expression for the integrals

The entire information on the basis, the geometry and the nuclei is now contained in the integrals h and V, and these alone have to be given to specify a system within these approximations.

2.1.3 The CI problem

For a fixed number of electrons, the second quantized Hamiltonian (2.11) can be viewed as a large matrix, and wave functions as finite-dimensional CI vectors [24]

$$|\psi\rangle = \sum_i c_i |D_i\rangle \ , \tag{2.14}$$

summing over all possible Slater determinants $|D_i\rangle$.

Solving the eigenvalue equation for $|\psi\rangle$ posed by the stationary Schrödinger equation in the full expansion (2.14) constitutes the Full Configuration Interaction method [24], which can be approached using techniques for iterative determination of eigenvalues. These methods, including the Davidson-Jacobi [139, 140], Arnoldi [141] and Lanczos [130] algorithms, do not introduce any additional approximations. They are therefore apt to calculate the eigenvalues of a matrix such as the Hamiltonian with the accuracy only limited by the precision of the numerical operations and the iteration count. All these iterative diagonalization techniques scale with the size of the Hamiltonian and thus exponentially in the number of electrons and orbitals, in particular, they require storage of the FCI vectors, or several copies thereof, and thus face a severe memory bottleneck due to the exponential scaling of the Hilbert space with system size [100]. Nevertheless, deterministic exact diagonalization algorithms can be highly valuable due to their reliability [142, 143, 144, 145], and in special cases, the memory requirements can be eased [146].

2.2 Exact diagonalization by Lanczos method

Since the FCIQMC algorithm follows a similar recipe as deterministic FCI methods, understanding the workings of iterative eigensolvers is also beneficial for formulating FCIQMC algorithms.

It is further essential for benchmarking any new algorithm to have access to reliable reference data, and the Lanczos [130] method can supply these and will thus come to use in section 3. Here, an outline of the method and how it can be used to calculate spectral functions shall be given.

The Lanczos method can be envisioned as a Krylov projected [130, 147] version of the power method, so first consider the simple power method [148]. The core idea is to project an initial state onto the eigenstate of an operator $\hat{H}$ with the largest eigenvalue

by repeatedly applying $\hat{H}$. Let

$$|\psi\rangle = \sum_i \langle \psi \,|\, E_i \rangle \, |E_i\rangle \tag{2.15}$$

be an initial state and $|E\rangle_i$ eigenvectors of H with eigenvalues E_i. Applying $\hat{H}$ n-times then yields

$$\hat{H}^n \, |\psi\rangle = \sum_i \langle \psi \,|\, E_i \rangle \, E_i^n \, |E_i\rangle = E_0 \sum_i \langle \psi \,|\, E_i \rangle \left(\frac{E_i}{E_0} \right)^n |E_i\rangle \ . \tag{2.16}$$

Assume the eigenstates to be sorted descendingly by magnitude, i.e. $|E_0| > |E_1| > \ldots > |E_n|$, then

$$\frac{\hat{H}^n \, |\psi\rangle}{|\hat{H}^n \, |\psi\rangle \,|} \to |E_0\rangle \qquad (n \to \infty) \,, \tag{2.17}$$

given that $\langle E_0 \,|\, \psi \rangle \neq 0$. For a finite dimensional Hilbert space, guaranteeing that the ground state has highest absolute value is always possible by shifting the Hamiltonian, so this method is in principle suited to calculate ground state energies. However, the speed of convergence is dictated by the ratio $\frac{E_1}{E_0}$, so for systems with small excitation energies where that ratio is close to 1, convergence will be very slow. Further, the intermediate results $\hat{H}^n \, |\psi\rangle$ are not used in this formulation. By moving towards a Krylov subspace procedure [147], these can be utilized, convergence can be enhanced and multiple states can be targeted in a single calculation, resulting in the well-established Lanczos method for hermitian operators $\hat{H}$ [130] which shall be sketched in the following.

Instead of iteratively applying $\hat{H}$ to a single state, iterative application of $\hat{H}$ is now used to build up a subspace K, referred to as Krylov subspace. Within this subspace, $\hat{H}$ will have a simple form that makes the eigenvalues easily accessible. Start with just $K_0 = \{|\psi_0\rangle\}$, then, in each iteration, the space K is expanded by the next state according to the power method iteration

$$|\phi_{n+1}\rangle = \hat{H} \, |\psi_n\rangle \ . \tag{2.18}$$

Thus, in each iteration, the space K is expanded by one new basis vector, and in the n-th iteration, it is given by

$$K_n = \mathrm{span}\{|\psi_0\rangle \,,\, \hat{H} \, |\psi_0\rangle \,,\, \ldots \,,\, \hat{H}^n \, |\psi_0\rangle\} \,. \tag{2.19}$$

The states $\hat{H}^m \, |\psi_0\rangle$ are not forming an orthonormal basis yet as they are not orthogonal,

thus, instead of the bare vectors $\hat{H}^m |\psi\rangle$, an orthogonal basis $|\psi_m\rangle$ is stored, which is generated by orthogonalizing each new vector $|\phi_{n+1}\rangle$ against the previously obtained basis states $|\psi_m\rangle$ and thus K_n, adding a new state $|\psi_{n+1}\rangle \perp K_n$. In the case $|\phi_{n+1}\rangle \in K_n$, an arbitrary vector orthogonal to K_n is added as a new basis vector.

After n iterations, this results in a n-dimensional subspace K_n in which the operator $\hat{H}$ is tridiagonal, as follows from the orthogonality. For $i < j - 1$, the matrix element of $\hat{H}$ in K_n is

$$H_{ji} = \langle \psi_j \,|\, H \,|\, \psi_i \rangle = \langle \psi_j \,|\, \phi_{i+1} \rangle \ , \tag{2.20}$$

and as $\phi_{i+1} \in K_{i+1}$ and $|\psi_j\rangle \perp K_{i+1}$, it follows

$$H_{ji} = 0 = H_{ij} \, , \tag{2.21}$$

since $\hat{H}$ is hermitian. Thus, only matrix elements H_{ij} with $|i - j| \leq 1$ remain nonzero, making $\hat{H}$ tridiagonal in K_n.

Furthermore, $|\phi_{n+1}\rangle$ is already orthogonal to all vectors $|\psi_i\rangle$, $i < n - 1$, as these overlaps reduce to vanishing matrix elements of $\hat{H}$

$$\langle \psi_i \,|\, \phi_{n+1} \rangle = \langle \psi_i \,|\, H \,|\, \psi_n \rangle = H_{in} = 0 \qquad (i < n - 1) \, . \tag{2.22}$$

Thus, to obtain $|\psi_{n+1}\rangle$, only $|\psi_n\rangle$ and $|\psi_{n-1}\rangle$ are required, the former to obtain $|\phi_{n+1}\rangle$ and the latter for orthogonalization. Constructing $\hat{H}$ in K thus only ever requires storing three vectors, even though information on a range of eigenvalues is obtained. Even though $\hat{H}$ is sparse in quantum chemical applications, it is common to evaluate the matrix elements only on the fly from the coulomb integrals, as storing $\hat{H}$, even in a sparse format, is extremely memory expensive.

The outlined procedure yields a tridiagonal, $n \times n$ matrix $H_{|K}$ after n iterations, while the eigenvalues of $H_{|K}$ are converging to those of the full $\hat{H}$ [149]. Nevertheless, convergence is typically much faster than in a simple power method [150, 151], as the eigenstates can be represented by any polynomial of $\hat{H}$ applied to $|\psi_0\rangle$ now. As $H_{|K}$ is tridiagonal and the size of K is by far smaller than the size of the full Hilbert space even at tight convergence, calculating the eigenvalues of $H_{|K}$ is a routine task [152].

In chapter 3, spectral functions from the Lanczos method supplied by Olle Gunnarsson are used as benchmarks, these are directly obtained by the decomposition of a given state in eigenstates according to equation (2.15). Clearly, obtaining the full decomposition with all eigenstates is impractical, since the number of eigenstates equals the size of

the Hilbert space. Instead, the eigenstates of $H_{|K}$ are to be used as an approximation. The number of theses states is systematically controllable and they converge to the eigenstates on a practicable scale [150]. Knowledge of a few dozen to hundreds of low-energy eigenstates is typically sufficient to evaluate the desired spectral functions, so this approach is suited for calculation of these to arbitrary accuracy [87]. The scheme according to Ref. [87] to extract the spectral function shall be briefly addressed now. The eigenstates of $H_{|K}$ are only available in their representation of the Krylov basis $\{|\psi_i\rangle\}$, but as the first Krylov vector is always the initial state, the overlap of any eigenstate with the initial state is readily available where that this overlap and the corresponding eigenvalues are what is of main interest, such that full storage of the Krylov basis is not required [87].

Generating the initial state for such a calculation however requires nontrivial effort itself, as a specific perturbed state is required for obtaining Green's functions as outlined in section 2.5 which requires knowledge of the ground state. The tridiagonalization scheme and resulting diagonalization of $H_{|K}$ yields the ground state energy E_0 and the ground state $|\Psi\rangle$ in the basis of K

$$|\Psi\rangle = \sum_i c_i \, |\psi_i\rangle \ . \tag{2.23}$$

The basis $\{|\psi_i\rangle\}$ however cannot be stored in a memory efficient way such that knowledge of the coefficients c_i in the Krylov basis cannot be immediately translated to the CI vector in the determinant basis. Hence, to obtain the representation of the ground state in the determinant basis, the coefficients c_i have to be stored and a second, identical Lanczos calculation is run, recalculating the vectors $|\psi_i\rangle$. Then, the ground state $|\Psi\rangle$ can be built on the fly by adding each new Krylov vector with the corresponding coefficient c_i.

Two potential bottlenecks arise, first, the orthogonalization step requires high numerical precision, as floating point errors can lead to a loss of orthogonalization [149]. Second, the memory cost of storing even a single full vector $|\psi\rangle$ quickly becomes prohibitive when considering systems with 20 or more electrons or more extended basis sets including f-orbitals or higher shells. Though massive parallelization can allow storing very large vectors in distributed memory, the steep scaling puts tight bounds on what is computationally feasible [143, 153]. Exploring algorithms that do not exhibit this bottleneck and show improved scaling is thus essential.

One potential approach to this is the selected configuration interaction method [55, 56], which truncates the full Hilbert space to a manageable size. Naturally the question arises

how such a truncation shall be carried out, and a promising recent approach is the Heat-Bath Configuration Interaction [57, 58, 154] that was directly inspired by the FCIQMC Heat-Bath sampling [155].

Another approach to the memory issues is to treat the CI problem on a stochastic basis, which will now be considered more closely. The initiator FCIQMC (i-FCIQMC) method, being an inherently stochastic method, in a sense also implicitly contains aspects of selected CI and can be viewed as combining both approaches.

2.3 Introduction to Quantum Monte Carlo methods

The stochastic solution of the CI problem can be approached in numerous ways, both as variational techniques in the flavor of Variational Monte Carlo (VMC) algorithms [156, 157] as well as projective algorithms like the Diffusion Monte Carlo (DMC) method [158, 159, 160], the Auxiliary Field Quantum Monte Carlo (AFQMC) method [161, 162, 163, 164], or the FCIQMC algorithm [59, 60, 61, 62] or by sampling of the partition function as in continuous-time Quantum Monte Carlo [165]. Conceptually similar to deterministic iterative algorithms, projective Quantum Monte Carlo methods start from some initial state $|\psi(0)\rangle$ and remove all contributions from states other than the ground state $|\psi_0\rangle$ by repeated application of a suited operator $\hat{P}$, until only the ground state remains, that is

$$|\psi_0\rangle = \lim_{n\to\infty} \hat{P}^n |\psi(0)\rangle \ . \tag{2.24}$$

It is the application of this operator which is then performed stochastically, thus overcoming potential bottlenecks of a deterministic scheme such as the drastic memory requirements. While this is the common foundation of the projective Quantum Monte Carlo methods, the choice of the operator employed in the projection scheme and the representation of the wave function distinguish the concrete methods, which vary vastly in their algorithmic details. A common issue of all stochastic methods for Fermionic systems is the manifestation of a sign problem [19], which stems from the necessity to sample a non-positive entity. The subject of this work is the extension of FCIQMC methods, but to give a context and highlight differences and similarities with other flavors of Quantum Monte Carlo, a short recapitulation of DMC and AFQMC methods, the closest relatives of the FCIQMC, shall be given, without claiming to be complete or addressing implementational details of the methods. A general overview of available QMC methods and their particularities can be found in Ref. [109]. A separate introduction to the FCIQMC method is then given in section 2.4.

2.3.1 Monte Carlo Integration

Common to all of the methods addressed here is the evaluation of a relevant integral by stochastic means instead of a deterministic sampling, be it in real space (DMC, real-space VMC), a finite Hilbert space (FCIQMC) or an auxiliary field (AFQMC). Stochastic, or, Monte Carlo integration [166, 167] uses a summation over a random ensemble of points to approximate the value of a given integral

$$I = \int_\Omega \mathrm{d}x\, f(x)p(x) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)\,, \tag{2.25}$$

where $\{x_i\}$ is a set of N random points drawn from the probability distribution p on the integration area Ω. That is, for each point, the probability of drawing the value x as a sample x_i is given by $p(x)$. Note that the integration is not necessarily referring to Lebesgue integration but can refer to any measure. Being a probability distribution, p has to fulfill [168]

- $\int \mathrm{d}x\, p(x) = 1$
- $\forall x:\ 0 \leq p(x)$

The approximate integral can then be read as a mean value of a random variable $y_i = f(x_i)$ which converges to the expectation value

$$\langle y_i \rangle_i \to I \tag{2.26}$$

according to the law of large numbers [169, 170]. Furthermore, the central limit theorem [171, 168] gives an estimate of the error of the sampled integral, which implies that, given the original function f has a finite variance σ_f^2, the average $\langle y_i \rangle_i$ approaches a normal distribution with mean I and variance [167]

$$\sigma_y^2 = \frac{\sigma_f^2}{N}\,. \tag{2.27}$$

In particular, the standard deviation of $\langle y_i \rangle_i$ scales as $N^{-\frac{1}{2}}$, independently of the integration area.

Note that for a general integration

$$\int_\Omega \mathrm{d}x\, f(x) = \int_\Omega \mathrm{d}x\, \frac{f(x)}{p(x)} p(x)\,, \tag{2.28}$$

the choice of p is ambiguous, any probability distribution on the integration area Ω whose support contains the support of f can be chosen. This requirement ensures the fraction $\frac{f(x)}{p(x)}$ is well-defined. The use of a non-uniform p to reduce the variance of y_i is commonly encountered as importance sampling [172, 173], with the ideal choice being $f = p$, as this reduces the variance of y_i to 0 [167].

2.3.2 Diffusion Monte Carlo

A detailed description of the DMC algorithm is not within the scope of this work, which is focused on the related FCIQMC method. Extensive reviews of the DMC method are readily available [109, 174, 175, 176, 177], so only a short sketch of the method shall be given in the following to point out common concepts of the FCIQMC and DMC approaches. To solve the stationary Schrödinger equation and determine the ground state of $\hat{H}$, Diffusion Monte Carlo convergences the imaginary-time evolved wave function

$$|\psi(\tau)\rangle = \mathrm{e}^{-\tau\left(\hat{H}-S\right)} |\psi(0)\rangle \ , \tag{2.29}$$

in the sense of a projection according to (2.24), which connects it to FCIQMC which starts from the same premise. The shift S serves to ensure convergence to the ground state and has to be updated iteratively to approach the ground state energy E_0 [174]. For $S = E_0$, the convergence of $|\psi(\tau)\rangle$ can be readily demonstrated by decomposing $|\psi(0)\rangle$ into eigenstates $|E_i\rangle$ of $\hat{H}$ with eigenvalues $E_i \geq E_0$ as

$$|\psi(0)\rangle = \sum_i \langle E_i \,|\, \psi(0)\rangle \, |E_i\rangle \ , \tag{2.30}$$

then it is

$$|\psi(\tau)\rangle = \sum_i \mathrm{e}^{-\tau(E_i - E_0)} \langle E_i|\psi(0)\rangle \, |E_i\rangle \to \langle E_0 \,|\, \psi(0)\rangle \, |E_0\rangle \qquad (\tau \to \infty) \, . \tag{2.31}$$

That is, while $\langle E_0 \,|\, \psi(0)\rangle \neq 0$, the imaginary time evolution converges to the ground state. As will be seen in section 2.4, on this level, FCIQMC and DMC work in the same way, but the procedure of evaluating the propagation differs greatly.

In DMC, the wave function $|\psi\rangle$ is described in a real-space representation $\psi(\mathbf{r})$ with the n-electron coordinate vector $\mathbf{r}$. Instead of storing the full function, it is sampled by a number N_w of walkers, each of which is associated to a coordinate $\mathbf{r}_i$. The walkers are

distributed according to the probability distribution

$$p(\mathbf{r}, \tau) \propto \psi(\mathbf{r}, \tau)\, \psi(\mathbf{r}, 0)\,, \tag{2.32}$$

commonly referred to as importance sampled wave function [178, 179]. While a direct sampling of the wave function is possible, it is by far less efficient compared to the importance sampling such that the latter form is usually used in practice [178]. The initial wave function can be sampled by the Metropolis-Hastings algorithm [173, 180] for a given initial wave function $|\psi(0)\rangle$ and the subsequent projection is carried out stochastically. With a walker representation, any integral over the importance sampled wave function can be approximated by a sum over the walkers

$$\int \mathrm{d}r\, f(\mathbf{r}) p(\mathbf{r}, \tau) \approx \sum_{i=1}^{N_w} f(\mathbf{r}_i)\,, \tag{2.33}$$

which is nothing but a Monte Carlo integration [166]. When rephrased as an integration, the propagation (2.29) can then be evaluated by summation over the walker population.

While the notion of a walker-representation of the wave function is not essential to the FCIQMC algorithm, it should be noted that it has been influential in the development of the method such that the concept is commonly connected to FCIQMC, too.

The propagation (2.29) is then realized by Trotter decomposition [181] of the exponential

$$\begin{aligned} \mathrm{e}^{-\tau \hat{H}} &\approx \left[\mathrm{e}^{-\frac{\tau}{n}\hat{T}} \mathrm{e}^{-\frac{\tau}{n}\hat{V}}\right]^n \\ &\to \mathrm{e}^{-\tau \hat{H}} \qquad (n \to \infty)\,, \end{aligned} \tag{2.34}$$

with a decomposition of $\hat{H}$ into a kinetic term $\hat{T}$ and a potential term $\hat{V}$ diagonal in $\mathbf{r}$. The action of the full propagator is evaluated in n steps, each time applying this expanded short-time propagator. In the real-space picture, the propagation by one step then reads

$$\psi\left(\mathbf{r}, \frac{\tau}{n}\right) = \int \mathrm{d}r' \underbrace{\left\langle \mathbf{r} \middle| \mathrm{e}^{-\frac{\tau}{n}\hat{T}} \middle| \mathbf{r}' \right\rangle \mathrm{e}^{-\frac{\tau}{n} V(\mathbf{r})}}_{G(\mathbf{r}', \mathbf{r}, \frac{\tau}{n})} \psi(\mathbf{r}', 0)\,, \tag{2.35}$$

the expectation value of the propagator $G(\mathbf{r}', \mathbf{r}, \frac{\tau}{n})$ is nothing but the imaginary time real-space Green's function. For general complex times τ and in a finite basis instead of the real-space representation, this is precisely the object of interest in chapter 3, where FCIQMC will be employed for directly evaluating this object to extract information on

the excited states of $\hat{H}$. For use of the importance sampled wave function, the rescaled Green's function

$$G'(\mathbf{r}', \mathbf{r}, \tau) = \frac{\psi(\mathbf{r}', 0)}{\psi(\mathbf{r}, 0)} G'(\mathbf{r}', \mathbf{r}, \tau)\,, \tag{2.36}$$

has to be used such that the imaginary-time propagation of $\psi(\mathbf{r}, \tau) \cdot \psi(\mathbf{r}, 0)$ can be considered [177]. The rescaled Green's function has the advantage that the diagonal terms do not diverge as they do in the bare Green's function due to the diverging Coulomb potential [178], which is a strong argument in favor of the importance sampling.

As the walkers are distributed according to $p(\mathbf{r}, \tau)$, the integration in equation (2.35) can be carried out as a Monte Carlo integration by summing G' over all walkers, where the action of G' is evaluated stochastically. This is performed in two steps, the so called *diffusion* step where the off-diagonal kinetic part is applied, creating new walkers at different positions, followed by the *branching* step in which the diagonal potential is applied, adjusting the number of walkers at each sampled position [178].

For fermionic systems like those studied in Quantum Chemistry, the wave function has to fulfill the fermionic exchange statistics, i.e. the wave function changes sign when exchanging two fermions. To be able to treat p as a probability distribution, it has to be strictly positive however, requiring $\psi(\mathbf{r}, \tau)$ and $\psi(\mathbf{r}, 0)$ to have the same sign for all $\mathbf{r}$ and τ. This is commonly referred to as the fixed node approximation [159, 182], which constrains the nodal surface, i.e. the subspace for which $\psi(\mathbf{r}, \tau) = 0$ to the initial one, meaning the method can only converge to the correct ground state if the initial nodal surface is correct, generally introducing a fixed-node error associated with this approximation [159].

2.3.3 Auxiliary Field Quantum Monte Carlo

The second Quantum Monte Carlo method that shall be visited here is the AFQMC method from the same family of projective algorithms based on the general projection (2.48) as also DMC and FCIQMC. It is particularly suited for treatment of model systems [161, 183, 184], and is well equipped for evaluation of imaginary-time Green's functions [185, 186]. The latter feature makes it interesting as a reference method for comparison of spectral functions obtained with the time-dependent FCIQMC introduced in chapter 3.

The AFQMC method uses the same projection operator (2.29) as DMC, including a trotter decomposition alike (2.34), but does not operate in real space, but in a discretized space and is formulated in a second quantized form [161]. It is similar in that perspective

to FCIQMC, but the algorithm itself is completely different, in particular approaching the sign problem in a different way and using non-orthogonal determinants [187]. The concept of AFQMC will be briefly sketched based on Ref. [187] to give a view on other QMC methods and highlight what different approaches are taken, for more elaborate reviews, see Refs [164, 187, 188, 189].

Consider the propagator (2.34), but in contrast to the real-space version, the separation

$$\hat{H} = \hat{T} + \hat{V} \tag{2.37}$$

shall be a separation into a one-body operator $\hat{T}$, which contains the kinetic and electron-nuclear Coulomb terms, and a two-body term $\hat{V}$ containing the electron-electron Coulomb interaction.

In a second quantized picture with a finite dimensional Fock space defined by an orbital basis, any operator of the form $\mathrm{e}^{\hat{R}}$ with a one-body operator $\hat{R}$ simply corresponds to an orbital rotation [190]. Therefore, these operators map a Slater determinant to another Slater determinant with different orbitals, allowing for simple evaluation of $\mathrm{e}^{\frac{\tau}{n}\hat{T}}$.

To reduce the action of the potential term to a form that can also be evaluated by orbital rotation, first the operator $\hat{V}$ is rewritten in a quadratic form which can be generally constructed as [187]

$$\hat{V} = \sum_i \lambda_i \hat{x}_i^2 \,, \tag{2.38}$$

with one-body operators $\hat{x}_i$.

Then, a Hubbard-Stratonovich transform [191, 192] is employed to express the operator $\mathrm{e}^{\hat{V}}$ in terms linear in the one-body operators $\hat{x}_i$ [162]

$$\mathrm{e}^{-\frac{\lambda_i}{2}\hat{x}_i^2} = \int \frac{\mathrm{d}\sigma}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\sigma^2}{2}} \mathrm{e}^{\sqrt{-\lambda_i}\sigma\hat{x}_i} \,, \tag{2.39}$$

which can again be expressed as orbital rotations. This transforms the interacting problem to a Gaussian integral over non-interacting ones, weighted with a newly introduced variable σ. The integral over σ is then evaluated stochastically [161, 162], in the case of $\lambda_i > 0$, the exponential operator $\mathrm{e}^{\sqrt{-\lambda_i}\sigma\hat{x}_i}$ is complex however, leading to a phase problem [164] when sampling the integral. This can be addressed in a fashion similar to the fixed node approximation in DMC by a phaseless approximation which uses a trial wave function to fix the phase in the stochastic integration [164]. Since the phase problem is ubiquitous in AFQMC for ab-initio systems [164, 193], the phaseless approximation is usually employed [188, 194, 195].

Notably, the phase problem does not occur in its full severity in many model systems [188], which instead only exhibit a sign problem that can be addressed by the constrained path approach [187] which employs a trial wave function to guide the sign of the propagated wave function.

The FCIQMC approach is thus very different in spirit from the both AFQMC and DMC, as it does not require an external input in the form of a trial wave function, but is self-contained, at the cost of a scaling that is in principle exponential [196].

2.3.4 Pseudo random number generation

As trivial as it may sound, a fundamental requirement of any stochastic method is the availability of random numbers, which itself is a very non-trivial issue. Two possible approaches to random number generation were already discussed by von Neumann [197], namely firstly a generation by physical processes which are random according to the current state of theory, or at least chaotic, and secondly a generation by iteration of a mathematical function exhibiting ergodic behavior. He identified the latter as not actual random numbers, but deterministic sequences of numbers that could be used in the place of true random numbers. The functions generating these sequences are commonly referred to as pseudo random number generators (PRNGs) [198].

Modern hardware usually supports physical random number generation based on thermal noise [199, 200], which is essential for security relevant applications which rely on the true randomness of the supplied numbers [201]. However, their performance is inferior to that of PRNGs [199], and other than cryptographic applications, numerical Monte Carlo calculations do not rely on the series of random numbers being not reproducible [198]. On the contrary, for the sake of testing a method and its implementation, being able to reproduce the exact same sequence of random numbers is even desirable, as it allows to precisely reproduce a specific sampling and thus potential errors that would otherwise only occur at random.

For the sake of performance, Quantum Monte Carlo methods thus rely on PRNGs [198, 202, 203], and the PRNG chosen in the `NECI` FCIQMC implementation used in this work is the Mersenne Twister [204], specifically the double precision SIMD[1] oriented fast Mersenne Twister (dSFMT) implementation by Saito and Matsumoto [208, 209]. For further reading on random and pseudo random number generators and their comparison, the reader shall be referred to the literature [204, 210, 211, 212, 213, 214, 215].

[1] Single Instruction, Multiple Data, a hardware-level parallelization technique [205, 206, 207]

2.3.5 Alias sampling

In practice, all Monte Carlo methods require a scheme to draw random values from a given probability distribution. For a uniform distribution, this boils down to the aforementioned problem of generating (pseudo) random numbers, for which a host of methods exists [204, 210, 215]. In concrete application however, usually non-uniform distributions are sampled for reduced variance by means of importance sampling, and techniques of drawing from these are required. In the general case, the Metropolis-Hastings scheme [173, 180] can be employed even if the normalization of the distribution is unknown. If, however, the distribution is known in advance and a discrete variable is to be sampled, very efficient schemes like the table method [216, 217] or the alias method [218, 219] are available. By no means can or should an exhaustive discussion of this field be the scope of this work, for a general review, the reader shall be referred to the literature [220]. Since the `NECI` FCIQMC program uses mainly the alias method for sampling, besides the inverse transform method which maps uniform to non-uniform random numbers directly using the cumulative distribution function [220], the former shall be briefly recapitulated here.

Given a discrete probability distribution

$$p: \{1, \dots, N\} \to [0, 1], \tag{2.40}$$

which assigns each number i from 1 to N a probability $p(i)$. As p is a probability distribution, the average value of p_i is $\frac{1}{N}$. For non-uniform p, there will be some values of i for which $Np(i) > 1$, i.e. which are more probable than the average, and some which for which $Np(i) < 1$, i.e. which are less probable than the average.

The basic idea behind the alias sampling method is to draw a value i uniformly and then randomly select between i with a bias $b(i)$ and another value $a(i)$, the alias value. This effectively increases the probability of sampling $a(i)$ and reduces the probability of sampling i and thus allows to shift weight from those values which are less probable than the average to those which are more probable. Of course, in practice, the biases b to chose the alias value and the alias values a themselves have to be carefully adjusted to yield the desired distribution p.

Therefore, the tables a and b have to be precomputed, this process only has to be performed once for each distribution p, but also means the method is only really efficient if the distribution is reused often. The initialization scheme implemented in `NECI` by the author is the one given by Vose [219], so that setup of a and b shall be described here.

Initially, set

$$a(i) = i \tag{2.41}$$

$$b(i) = p(i)\,, \tag{2.42}$$

corresponding to a uniform distribution. By subsequent refinement of a and b, the probability distribution p is then achieved. Therefore, separate the values $\{1, \ldots, N\}$ into two subsets,

$$L = \{i | Nb(i) > 1\} \tag{2.43}$$

$$S = \{i | Nb(i) \leq 1\}\,, \tag{2.44}$$

with L representing those values which need to be sampled more often and S those values that need to be sampled less often. Now, start taking pairs of values $i \in L$ and $j \in S$ and set

$$a(j) = i\,, \tag{2.45}$$

increasing the probability of sampling i by $1 - p(j)$ and reducing the probability of sampling j to $p(j)$, which is the desired value. Hence, j does not need further adjustments and is removed from S. The increase in the probability of sampling i can either result in a probability larger than $p(i)$ or not, which is reflected in an update of $b(i)$ as

$$b(i) = b(i) + b(j) - 1\,, \tag{2.46}$$

reducing $b(i)$ by the newly acquired weight $1 - p(j)$. This might, if the new probability of sampling i is too high now, result in i being moved from L to S. This procedure is repeated until no values in S are left.

Initially, $\sum_i b(i) = N$ and each iteration reduces this sum by 1 via the removal of one value from S and the update (2.46). Since all values $i \in L$ fulfill $b(i) > 1$, after each number of steps n, at least one value has to be contained in S since else

$$N - n = \sum_i Nb(i) = \sum_{i \in L} Nb(i) > \sum_{i \in L} 1 = N - n \tag{2.47}$$

which is a contradiction [219]. Therefore, the whole procedure for initialization requires $\mathcal{O}(N)$ steps. In particular, each value i except for the last one visited in this scheme will be removed from S at some point and then be sampled with the desired probability

$p(i)$. The last value also has to have the correct probability then, since the distribution p is normalized [219].

2.4 The basic FCIQMC algorithm

After addressing the general notion of Quantum Monte Carlo methods, the focus shall now be on the FCIQMC method [59, 60, 61, 62], which will then be extended to treat the time-dependent Schrödinger equation.

The FCIQMC algorithm as designed in Ref. [59] and updated with ideas from Refs. [60, 61] constitutes the basic imaginary-time FCIQMC algorithm and its core concepts which shall be outlined in this section, albeit using a walker-free formulation as encountered in Ref. [221]. A useful recent perspective on the FCIQMC method is given in Ref. [222] in the context of numerical linear algebra from the point of view of the Fast Randomized Iteration method [223]. A useful summary in the context of stochastic Multiconfigurational Self-Consistent Field theory can be found in Ref. [224]. The original formulation of FCIQMC uses a walker representation of the wave function similar to DMC methods [59], but the walker-free numerical linear algebra perspective is more useful from an implementational point of view, and, but that might only be an opinion, easier to understand. Both of the two major FCIQMC implementations available, `NECI` [196, 1], which is used in this work, and `HANDE` [225], use an implementation close to the framework summarized here.

To formulate the algorithm, three components are required:

1. A projection operator $\hat{P}$ that can be employed in the iterative scheme
2. A memory-efficient representation of the wave function
3. A scheme to stochastically apply $\hat{P}$ to the wave function

For the FCIQMC method, the choice of $\hat{P}$ is

$$\hat{P} = \left(1 - \Delta\tau(\hat{H} - S)\right) , \tag{2.48}$$

where $\Delta\tau$ is a step-size parameter commonly referred to as time-step, and S is a constant shift. For now, assume that S is the ground state energy. The desired property of projecting onto the ground state can be easily verified, consider an initial state

$$|\psi(0)\rangle = \sum_i \langle E_i \,|\, \psi_0\rangle \, |E_i\rangle \ , \tag{2.49}$$

where $|E_i\rangle$ are the eigenstates of $\hat{H}$ with eigenvalue E_i, in an ascending order, such that $S = E_0$. Then it is

$$\begin{aligned}\hat{P}^n |\psi(0)\rangle &= \left(1 - \Delta\tau(\hat{H} - S)\right)^n |\psi(0)\rangle \\ &= \sum_i \langle E_0 \,|\, \psi(0)\rangle \left(1 - \Delta\tau(E_i - S)\right)^n |E_i\rangle \\ &\to \langle E_0 \,|\, \psi(0)\rangle \, |E_0\rangle \qquad \Delta\tau < \frac{2}{E_{\max} - E_0}\,,\end{aligned} \tag{2.50}$$

since for sufficiently small time-step $\Delta\tau < \frac{2}{E_{\max} - E_0}$, it holds

$$|1 - \Delta\tau(E_i - S)| < 1 \qquad \text{for } i > 0\,. \tag{2.51}$$

At this point, the existence of a largest eigenvalue $E_{\max}$ is assumed, and, in fact, this algorithm requires a bound spectrum to work. While the electronic Hamiltonian is not bound in general, employing a finite basis set does immediately put a bound on the energy since the resulting Hamiltonian is finite-dimensional, thus making this assumption valid in any finite basis set. Of course, in practice E_0 is not known in advance and hence S cannot be set to the required value, but for convergence, it is sufficient that S converges to E_0, making iterative updating of S to approach the correct value viable, as is discussed below. The role of the shift in FCIQMC is largely similar to the analogous energy offset in DMC shortly mentioned in section 2.3.2. A striking advantage of this iterator is that it is first order in $\hat{H}$, making the iterations relatively easy. This advantage will be lost when moving to a real-time evolution, as it becomes necessary to retreat to higher order $\hat{P}$.

To put the choice of $\hat{P}$ into scope, it is useful to consider the time-dependent Schrödinger equation

$$\mathrm{i}\frac{\partial}{\partial t} |\psi\rangle = \hat{H} |\psi\rangle \,, \tag{2.52}$$

with the formal solution

$$|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}t} |\psi(0)\rangle =: \hat{U}(t) |\psi(0)\rangle \,, \tag{2.53}$$

for an arbitrary initial state $|\psi(0)\rangle$. Now, consider an imaginary time evolution with an imaginary time $\tau = \mathrm{i}t$, meaning that

$$|\psi(\tau)\rangle = \mathrm{e}^{-\tau\hat{H}} |\psi(0)\rangle = \hat{U}(\tau) |\psi(0)\rangle \,, \tag{2.54}$$

which is also how the FCIQMC propagator was initially motivated [59]. Separating the full evolution from 0 to τ in n steps of size $\Delta\tau$ involves no approximation, such that one can also write

$$|\psi(\tau)\rangle = \hat{U}^n(\Delta\tau)\,|\psi(0)\rangle\ . \tag{2.55}$$

Expanding $\hat{U}$ to first order in $\Delta\tau$ now yields the FCIQMC projector $\hat{P}$ as

$$\hat{U}_1 = (1 - \Delta\tau\hat{H})\,, \tag{2.56}$$

up to the shift S. Since shifting the Hamiltonian by a constant does not change the eigenstates and only offsets all eigenvalues, the physics of the shifted Hamiltonian are the same and the FCIQMC projector can be identified with the first-order imaginary time propagator $\hat{U}_1$, with the initial step-size parameter $\Delta\tau$ now having the literal interpretation as a time-step. This connection shall be revisited when the real-time algorithm is formulated, as it creates an intuitive connection between the iterative FCIQMC ground state search and the time-dependent Schrödinger equation.

2.4.1 Wave function representation

While the open questions of how S is chosen and how $\hat{P}\,|\psi\rangle$ is evaluated will be revisited, for now, consider the representation of $|\psi\rangle$ itself. The FCIQMC is formulated in second quantization, implicitly capturing the exchange statistics of the wave function, but requiring a finite basis set, as already required when picking the projection operator. Even though the CI vector defining the wave function is now finite, the scaling problem prohibits storage of the entire CI vector in all but a few cases. To overcome the memory bottleneck, FCIQMC makes use of the sparsity of the CI vector, as even for highly entangled systems with strong correlation, only a small part of the full Hilbert space is important for the low-energy behavior [57, 59, 86]. Here, a stochastic rounding of the CI coefficients is used to significantly reduce the required memory, making memory consumption a non-issue for FCIQMC. Define the stochastic round function as

$$R(x) = \begin{cases} x & \text{if } |x| \geq c_{\min}, \\ \operatorname{sgn}(x)\, c_{\min} & \text{if } |x| < c_{\min} \text{ with prob. } \frac{|x|}{c_{\min}} \\ 0 & \text{else,} \end{cases}\ , \tag{2.57}$$

for some minimal coefficient value $c_{\min}$. This function rounds a coefficient x with absolute value smaller than $c_{\min}$ stochastically either to $c_{\min}$ or to 0, weighted with the absolute value of x, such that $R(x)$ averages to x. The full wave function at any given imaginary time τ is now compressed using R, yielding the approximate FCIQMC wave function

$$|\psi(\tau)\rangle = \sum_i R(c_i(\tau))\,|D_i\rangle\ . \tag{2.58}$$

Since for reasonable normalizations and/or choices of $c_{\min}$, most of the coefficients are below the threshold, and only a small selection of the coefficients has to be stored at any given point, essentially removing the memory bottleneck from the algorithm. This mechanism of compression automatically represents the more important parts of the wave function exactly, while still retaining access to the whole Hilbert space.

As changing $c_{\min}$ and adjusting the norm of $|\psi\rangle$ leads to the same effect of increasing the accuracy of this compression, $c_{\min}$ is typically set to 1 and the L_1-norm

$$N_w = \|\psi(\tau)\|_1 = \sum_i |R(c_i(\tau))|\,, \tag{2.59}$$

is used as a control parameter to tune the accuracy of the compression. For historical reasons, it is typically referred to as the walker number N_w [59], and this term will also be used here. With increasing N_w, more and more coefficients exceed $c_{\min}$, until eventually the rounding becomes exact when no coefficients below the threshold are left, which justifies the role as a control parameter.

2.4.2 Evaluating the FCIQMC projector

With a compressed, memory efficient wave function available, the next step is to evaluate $\hat{P}\,|\psi\rangle$ both computationally cheap and under conservation of the sparsity of the wave function. This is achieved by separating the action of $\hat{P}$ into three steps labeled *spawning*, *death* and *annihilation*

$$\begin{aligned}\hat{P}\,|\psi(\tau)\rangle = &\underbrace{\sum_i \left(1 - \Delta\tau(H_{ii} - S(\tau))\right) C_i(\tau)\,|D_i\rangle}_{\text{(a) Death step}} - \\ &\underbrace{\sum_i \sum_{j\neq i} \Delta\tau H_{ji} C_i(\tau)\,|D_j\rangle}_{\text{(b) Spawn step}}\ .\end{aligned} \tag{2.60}$$

The separation of the application of $\hat{H}$ into a diagonal and an off-diagonal part is similar to the effect of the Trotter decomposition in DMC, but no approximation is connected to it here.

The *spawning* step consists of applying the off-diagonal part of $\hat{H}$ and carries the main computational cost. The issue here is the evaluation of the sum over determinants connected to a given $|D_i\rangle$

$$\Delta \left|\psi\right\rangle_i := \sum_{j \neq i} \Delta\tau H_{ji} C_i(\tau) \left|D_j\right\rangle \,, \tag{2.61}$$

since all connected determinants have to be considered here, not only those present in the compressed wave function. A full evaluation is prohibitively expensive as that would constitute a deterministic matrix-vector multiplication with the full Hamiltonian, which can technically only be done for limited Hilbert space sizes. Instead, $\Delta \left|\psi\right\rangle_i$ is evaluated stochastically for each i by sampling the sum over j with a number M_i of random samples $j(k)$, $k = 1, \ldots, M_i$. The full sum is then approximated as

$$\Delta \left|\psi\right\rangle_i \approx \tilde{\Delta} \left|\psi\right\rangle_i = \Delta\tau C_i(\tau) \sum_{k=1}^{M_i} \frac{H_{j(k)i}}{p_g\left(j(k)|i\right)\, M_i} \left|D_{j(k)}\right\rangle \,, \tag{2.62}$$

where $p_g\left(j(k)|i\right)$ is the probability to pick $j(k)$ when choosing a single sample for given i. This sampling requires generating M_i random determinants $|D_j(k)\rangle$ connected to $|D_i\rangle$, that is with $H_{j(k)i} \neq 0$, which in turn is what constitutes the computationally most expensive part of the FCIQMC scheme. The sampling probabilities p_g are in principle arbitrary, but to prevent a bias, the relation

$$H_{ji} \neq 0 \Rightarrow p_g(j|i) \neq 0 \,, \tag{2.63}$$

must hold, or else relevant terms of the Hamiltonian will be neglected. In section 2.4.5, choices of p_g and how to realize them will be discussed.

The normalization $\frac{1}{p_g(j(k)|i)\, M_i}$ guarantees the correct limit

$$\tilde{\Delta} \left|\psi\right\rangle_i \to \Delta \left|\psi\right\rangle_i \qquad (M_i \to \infty) \,, \tag{2.64}$$

as in the limit of large M_i, each possible $|D_j\rangle$ will be picked $p_g(j|i)\, M_i$ times. In practice, a convenient choice for M_i is $|C_i(\tau)|$, stochastically rounded to the next integer. This choice ensures that the contributions from more important determinants are automatically taken into account with greater accuracy, while also guaranteeing exactness in the

Algorithm 1 Spawning step of the FCIQMC algorithm

loop
  for all occupied determinants $|D_i\rangle$ **do**
    select number of samples M_i
    for all $k = 1, \ldots, M_i$ **do**
      Choose a random determinant $|D_{j(k)}\rangle$ with $H_{i(k)} \neq 0$ with probability $p_g\left(j(k)|j\right)$
      Calculate $p_s\left(j(k)|i\right) = \Delta\tau \frac{|H_{j(k)i} C_i(\tau)|}{M_i p_g(j(k)|i)}$
      (Optionally) Stochastically round $p_s\left(j(k)|i\right)$
      Update the spawn vector $\tilde{\Delta}\,|\psi\rangle \to \tilde{\Delta}\,|\psi\rangle + p_s\left(j(k)|i\right)|j(k)\rangle$
    end for
  end for
end loop

limit of $N_w \to \infty$ which then implies $M_i \to \infty$. Further, the evaluation of $\tilde{\Delta}\,|\psi\rangle_i$ also becomes simpler since $|C_i(\tau)|$ does not enter explicitly anymore, albeit its sign does.

To ensure the sparsity of the wave function, it is in principle sufficient to perform the stochastic rounding of the coefficients only once at the end of each iteration, but for the parallelizability of concrete implementations, it is desirable to also keep

$$\tilde{\Delta}\,|\psi\rangle = \sum_i \tilde{\Delta}\,|\psi\rangle_i \tag{2.65}$$

as compact as possible, and thus, a stochastic round is typically performed for each sample. That can be viewed as accepting each move with a spawning probability

$$p_s(i \to j) = \Delta\tau \frac{|C_i(\tau) H_{ji}|}{M\, p_g(j|j)}\,. \tag{2.66}$$

The new contributions from the single samples are also referred to as *spawns*, and the full off-diagonal contribution $\tilde{\Delta}\,|\psi\rangle$ as *spawn vector*. The spawning step is summarized in algorithm 1. The spawn vector is kept separate for now and then added back to the wave function later in the *annihilation* step.

The following *death* step constitutes the application of the diagonal part of $\hat{H}$ and can be performed in a straightforward way. Each coefficient is now updated according to the diagonal matrix element and the shift S, with a subsequent stochastic round to eliminate negligible coefficients. That is, for each occupied determinant i, the update

$$C_i(\tau) \to S\left(1 - \Delta\tau(H_{ii} - S) C_i(\tau)\right)\,, \tag{2.67}$$

is performed, after the spawning step is completed. This operation has a low cost compared to the spawning step and is embarrassingly parallel, and thus does not significantly affect the scaling or performance of the algorithm.

After the spawn vector has been generated and the death step has been performed for all determinants, in the *annihilation step*, the spawn vector is added back to the current wave function to create the wave function for the next iteration

$$|\psi(\tau + \Delta\tau)\rangle = \sum_i C_i(\tau) |D_i\rangle - \tilde{\Delta} |\psi\rangle \ , \tag{2.68}$$

using the updated $C_i(\tau)$ from after the death step. When parallelizing the algorithm with a distributed memory model, the wave function is stored distributed across different processors, requiring a prior communication of $\tilde{\Delta} |\psi\rangle$ here. The annihilation step then serves as a synchronization between different processors. There is no inherent approximation to this step, but coefficients of opposing sign in the wave function and the spawn vector might cancel, thus removing determinants from the list of occupied determinants. To simplify this operation, the wave function is typically stored using a hash table, which allows for quick lookup of occupied determinants and check for occupation, then, the annihilation step scales linearly in the number of walkers.

So far, it was assumed S to be the ground state energy, which obviously can not be guaranteed from the beginning. However, S fulfills another purpose: It controls the walker number N_w. Recall equation (2.48), if now $E_0 - S < 0$, at least one of the terms in equation (2.48) diverges, then

$$\|\hat{P}^n |\psi(0)\rangle\| = \sum_i |d_i| |(1 - \Delta\tau(E_i - S)|^n \to \infty \Rightarrow N_w \to \infty \, . \tag{2.69}$$

In other words,

$$S > E_0 \Rightarrow N_w \to \infty \, . \tag{2.70}$$

On the other hand, if $E_0 - S > 0$, all terms in equation (2.48) converge to 0, so analogously

$$S < E_0 \Rightarrow N_w \to 0 \, . \tag{2.71}$$

That means the right value of S can be obtained by maintaining a constant N_w, as this implies $E_0 = S$, and vice versa. In projective FCIQMC, the shift is thus updated iteratively to maintain constant N_w. In practice, this means adjusting S every m iterations

as [59]

$$S \to S - \frac{\xi}{m\Delta\tau} \log\left(\frac{N_w(\tau + m\Delta\tau)}{N_w(\tau)}\right), \tag{2.72}$$

where ξ is a control parameter to adjust the speed of adjustment of S.

The notion of converging to $S = E_0$ will be lost in the real-time algorithm, and while the mechanisms making S control N_w are still in place, the role of S will have to be re-evaluated.

2.4.3 Energy measures

The shift S was already introduced as an energy measure, but it usually carries a lot of noise since it is used in the population control [59].

The more commonly used measure is therefore the projected energy [59]

$$E_P(\tau) = \frac{\left\langle D_R \middle| \hat{H} \middle| \psi(\tau) \right\rangle}{\langle D_R \,|\, \psi(\tau) \rangle}, \tag{2.73}$$

with a reference determinant $|D_R\rangle$, which is taken to be the most populated determinant, usually the Hartree-Fock determinant. It corresponds to the energy of $|\psi(\tau)\rangle$ if it was an eigenstate of $\hat{H}$, but only depends on the population of the reference and the single and double excitations from the reference, so strictly speaking, only the exact coefficients of these are required. All FCIQMC energies reported in this work are projected energies.

A more general version of this measure can be used by using a multi-reference trial wave function $|\psi_T\rangle$ instead of $|D_R\rangle$, to reduce the stochastic noise as more coefficients are used in the energy expression [196]. The trial wave function $|\psi_T\rangle$ is generated on the fly by taking a snapshot of the leading determinants of the FCIQMC wave function $|\psi(\tau_T)\rangle$ for a fixed τ_T after equilibration.

The resulting energy measure is a correlated quantity, the averaging is therefore performed using the renormalization group method of blocking analysis [226] to obtain the averaged energy and the statistical uncertainty.

Neither the shift, nor the projected or trial energy are variational measures, their advantage lying in being cheaply computable. At additional cost, the variational energy can also be obtained if two independent populations $|\psi_a\rangle$ and $|\psi_b\rangle$ are available [227, 228].

2.4.4 Initiator adaptation

A major milestone for the FCIQMC method was the introduction of the i-FCIQMC [60] which helps establishing the sign structure of the wave function by restricting the possible contributions to $\Delta|\psi\rangle$. A key issue with the FCIQMC algorithm as presented originally [59] is that it relies on the correctness of sign(c_i) when generating $\Delta|\psi\rangle$. For coefficients well above $c_{\min}$, this is a non-issue, but if a determinants obtains weight by stochastic error and ends up with a coefficient of $c_i = \pm c_{\min}$ of incorrect sign which is then propagated [60, 229]. It turns out for these stochastic errors to sufficiently average out to establish the correct sign structure of the wave function, a system specific minimal walker number is required, which often is inaccessible for larger systems [59, 60]. In the dynamics, this manifests in the form of a so called annihilation plateau, as the walkers number increase slows down significantly once the required N_w is reached before accelerating again [59]. This slowdown is due to an increase in cancellation of weights in the annihilation step once sufficiently many determinants carry nonzero weight. As the wave function and thus the resulting energy is incorrect below these walker numbers, this puts tight limits on the systems accessible to the base method [59, 60]

The initiator adaptation [60] is an efficient approximation tackling the problem of propagating sign errors while introducing only minimal error. To prevent erroneous spawns, the spawning step is modified by dismissing all samples of determinants $|D_{j(k)}\rangle$ selected in when evaluating the sum over k in equation (2.62) that fulfill the following criteria

- They are generated from a determinant $|D_i\rangle$ for which $|c_i(\tau)| \leq c_{\text{init}}$ and $|D_{j(k)}\rangle$ carries no weight in $|\psi(t)\rangle$
- The determinant $|D_{j(k)}\rangle$ is only selected as a sample once in this iteration in total

where c_{init} is the so-called initiator threshold which is introduced as a second control parameter next to N_w. Still, in the limit of $N_w \to \infty$, the method becomes exact for any $c_{\text{init}} > 0$ as then, all coefficients fulfill $|c_i(\tau)| > c_{\text{init}}$, while for $c_{\text{init}} < c_{\min}$, the initiator method reduces to the original FCIQMC method.

Employing the initiator adaptation greatly increases the range of systems that can be reliably treated by FCIQMC at the cost of introducing a minor systematic error from the neglect of certain matrix elements of $\hat{H}$. The i-FCIQMC method is well established as a state of the art version of FCIQMC in successful application [65, 66, 67, 68, 69, 70, 230].

2.4.5 Excitation generation

At the heart of the FCIQMC method is the stochastic evaluation of the sum (2.62), to which an essential component is the selection process of the coupled determinants $|D_j\rangle$. From an algorithmic point of view, the details of the selection process itself are irrelevant as long as it does faithfully reproduce the sum, that is, each determinant $|D_j\rangle$ with $H_{ji} \neq 0$ has to be choosable with a computable probability $p_g(j|i)$.

From an implementational point of view however, this process mainly determines the performance of an implementation, cf. also Fig. 6.4 and Appendix A, an efficient procedure is therefore highly desirable.

In the original formulation, the sampling was performed uniformly under consideration of symmetry constraints [59], but since then, a number of non-uniform schemes have been suggested [155, 196, 231] which use the matrix elements of the Hamiltonian, or a derived quantity, as weights in the selection process. For a useful overview of different schemes of excitation generation, see also Ref. [231]. For the calculations presented later in this thesis, different methods were employed based on availability in the `NECI` program.

While there is a formulation of Heat-Bath Sampling [155] that treats single and double excitations on the same ground, most methods, and all of those that are implemented in `NECI` , differentiate between single and double excitations. To this end, whenever a determinant $|D_j\rangle$ is to be selected, a random choice between single and double excitation is made, with a bias p_{doub} towards double excitations, which then has to be accounted for in the generation probability.

As neither the contribution from single nor double excitations should be preferenced in the evaluation of the sum, the bias towards single excitation is set such that [59]

$$\max_{\text{singles}} \left(\frac{H_{ij}}{p_g(j|i)} \right) = \max_{\text{doubles}} \left(\frac{H_{ij}}{p_g(j|i)} \right) , \tag{2.74}$$

to put equal weight on singles and doubles. In practice [196], that works by logging the value of

$$\gamma_{\text{sing/doub}} = \max_{\text{sing/doub}} \left(\frac{H_{ij}}{p_g(j|i)} \right) p_{\text{sing/doub}} , \tag{2.75}$$

with $p_{\text{sing}} = 1 - p_{\text{doub}}$. Setting

$$p_{\text{doub}} = \frac{\gamma_{\text{sing}}}{\gamma_{\text{sing}} + \gamma_{\text{doub}}} , \tag{2.76}$$

then fulfills equation (2.74).

This scheme typically leads to a high bias towards double excitations $p_{\text{doub}} > 0.9$, as there are much more double excitations than single excitations, which reduces $p_g(j|i)$ for double excitations. To leverage that, $p_g(j|i)$ is then increased for these by picking a high double excitation bias, while reducing $p_g(j|i)$ for single excitations. While the matrix elements of the latter are typically larger than for double excitations, that is not sufficient to make up for the greatly reduced $p_g(j|i)$.

Based on the choice between double and single excitation, the determinant $|D_j\rangle$ is then randomly selected from the pool of corresponding excitations, that is, for a single (double) excitation one (two) occupied and one (two) empty orbitals are chosen.

Since p_{doub} is usually quite large and double excitations are then more of a concern, consider the case of a double excitation. The most efficient way of selecting the orbitals is by (exact) Heat-Bath sampling [155], which weights each excitation $(i,j) \to (a,b)$ with the matrix element $H_{ij}^{ab} := \left\langle a_a^\dagger a_b^\dagger a_j a_i D_i \,\middle|\, H \,\middle|\, D_i \right\rangle$. Then, the acceptance probability is constant $p_s(i \to j) \propto \Delta\tau \frac{|C_i(\tau)|}{M}$, which is independent of the selection j. This choice for $p_g(j|i)$ corresponds to the optimal choice in the sense of the importance sampling encountered in section 2.3.1, as a non-uniform $p_g(j|i)$ is nothing but using importance sampling for the stochastic evaluation of the sum in (2.62).

The original Heat-Bath Sampling requires an $\mathcal{O}(n_{\text{el}})$ time [155], where n_{el} is the number of electrons. With the perspective of scalable rapid excitation generation, a variant in the form of a pre-computed Heat-Bath (PCHB) sampling is suggested to reduce the cost to $\mathcal{O}(1)$, at the cost of further approximating the Heat-Bath probability distribution. This variant of the method originally proposed by Holmes *et al* [155] was previously published in Ref. [196]. Instead of splitting the selection process into four steps where subsequently the spin-orbitals i, j, a, and b are selected, only two steps are used where first a pair of occupied spin-orbitals (i,j) is selected followed by a pair of spin-orbitals (a,b).

The selection of occupied orbitals, which is what brings the $\mathcal{O}(n_{\text{el}})$ cost in the Heat-Bath sampling is then done uniformly at constant cost [2]. This selections works exactly the same as in the Cauchy-Schwartz excitation generation [196], including a bias toward parallel-spin spin-orbitals p_p that is used to account for the difference in matrix elements between parallel-spin and opposite-spin excitations.

For each possible ordered pair (i,j), a pre-computed alias sampler [218, 219] as addressed in section 2.3.5 is prepared at the beginning of the calculation, which is seeded

[2] In the `NECI` implementation, this formally scales as $\mathcal{O}(n_{\text{el}})$ for technical reasons, but does not pose a bottleneck.

with the probability distribution

$$p(ab|ij) = \frac{|H_{ij}^{ab}|}{\sum_{a'b'} |H_{ij}^{a'b'}|}\,, \tag{2.77}$$

for ordered pairs (a, b). Since each of these tables requires $\mathcal{O}(N^2)$ memory, N being the number of orbitals, which amounts to total cost of $\mathcal{O}(N^4)$, it is more efficient to use spatial orbitals here to cut the memory cost. The matrix element H_{ij}^{ab} does depend on the spin of the orbitals however, as it determines if exchange integrals appear in the matrix element.

There are three cases to be considered. Either all orbitals have the same spin, then the matrix element contains both direct and exchange terms. Or i and a can have the same spin, then only the direct term is used, or i and b have the same spin, in that case, only the exchange term is used. That means, three alias samplers for spatial orbital pairs (a_s, b_s) are prepared for each ordered pair of spatial orbitals (i_s, j_s), one for each of the three options. The relative spin of the occupied orbitals is already fixed in their selection, that is the role of the bias p_p. Should they have the same spin, the first sampler is to be used that corresponds to parallel-spin excitations. In the case of different spins, the desired probability distribution is split up in a way analogous to the original scheme in [155]. Consider orbitals $a = a_s\alpha$, $b = b_s\alpha$, $i = i_s\alpha$, and $j = j_s\beta$, then rewrite

$$p(ab|ij) = \underbrace{\frac{\sum_{a_s,b_s} |H_{i_s\alpha j_s\beta}^{a_s\alpha b_s\beta}|}{\sum_{a'b'} |H_{i_s\alpha j_s\beta}^{a'b'}|}}_{b_d(i_s j_s)} \underbrace{\frac{|H_{i_s\alpha j_s\beta}^{a_s\alpha b_s\beta}|}{\sum_{a'_s b'_s} |H_{i_s\alpha j_s\beta}^{a'_s\alpha b'_s\beta}|}}_{p_d(a_s b_s|i_s j_s)}\,. \tag{2.78}$$

For fixed (i, j), the term $b_d(i_s j_s)$ is a constant that does not depend on a and b, but only on their relative spin, while the second term $p_d(a_s b_s|i_s j_s)$ only depends on the spatial orbitals for fixed relative spin. For the exchange case, an analogous distribution $p_e(a_s b_s|i_s j_s)$ is defined, the prefactor there corresponds to $1 - b_d(i_s j_s)$. The spin-orbital distribution $p(ab|ij)$ can then be sampled by first deciding between direct and exchange case with a bias $b_d(i_s j_s)$ towards the direct case, then spatial orbitals are sampling from the corresponding distribution p_d/p_e and the spins attached according to the earlier decision. The distributions p_d/p_e are pre-computed and can therefore have no knowledge of which orbitals are occupied in a given determinant. If therefore one of the selected orbitals (a, b) is occupied, the excitation has to be rejected at this point. This could potentially negatively impact performance, but even in worst case scenarios where $N = n_{\mathrm{el}}$, the method still performs efficiently, cf. Appendix A.

Since no probability distribution has to be constructed on the fly and alias sampling itself is performed in constant time [218], the excitation generation can then be performed in constant time, while still achieving efficient weighting with the matrix elements, making this scheme very competitive in terms of performance.

2.4.6 Time-step determination

While in theory, the time-step is limited by the range of the eigenvalues of $\hat{H}$, in practice, a stable calculation often requires smaller time-steps to prevent large fluctuations from very large spawning events. Therefore, the time-step in the `NECI` implementation is interlinked with the excitation generation scheme to ensure the dynamics does not get out of hand [232, 233].

The conventional method first described by Smart [232] sets the time-step $\Delta\tau$ to enforce the two conditions

$$\max_{ij} p_s(i \to j) < b \tag{2.79}$$

$$\max_{i} \Delta\tau(H_{ii} - S) < 1\,. \tag{2.80}$$

The first inequality, with a given parameter b, ensures that no massive spawning events occur that would greatly increase the variance of the stochastic summation (2.62) and circumvent the initiator criterion by instantly creating initiators. The second inequality ensures the death step cannot lead to instantaneous sign flips, which are also a possible source of instability [232].

Clearly, not all possible values entering the maxima on the left hand sides of the criteria (2.79) and (2.80) are known at the beginning of a calculation, instead, the maxima are taken over all values encountered so far in a calculation and updated dynamically, by which means the time step can then be updated during run-time [232].

The recently proposed histogram search by Dobrautz *et al* [233] is a generalization of this scheme and uses a modification of the criterion (2.79), arguing that a few outliers will not endanger stability, but might require significantly smaller time-steps. The maximum is then replaced by a high percentile, and a histogram of the occurring values $\frac{p_s(i\to j)}{\Delta\tau}$ is created at run-time to evaluate that percentile.

2.4.7 Semi-stochastic FCIQMC

A notable extension to the FCIQMC method was the development of semi-stochastic FCIQMC [61, 234] which identifies a subspace $D \subset \mathcal{H}$ of the full Hilbert space $\mathcal{H}$ spanned by essential determinants and treats the evaluation of $\hat{H}\,|\psi\rangle$ in D deterministically. These determinants are identified as the most populated determinants in the FCIQMC wave function after equilibration. The semi-stochastic extension includes two alterations to the standard FCIQMC scheme.

First, no stochastic rounding is ever performed for determinants in D, and their coefficients are stored even when zero.

Second, the propagation (2.60) is split in two parts

$$\hat{P}\,|\psi(\tau)\rangle = \Big(1 - \Delta\tau(\hat{H}_{|D} - S)\Big)\,|\psi_D(\tau)\rangle + \Big(1 - \Delta\tau(\hat{H}_R - S)\Big)\,|\psi(\tau)\rangle\ , \tag{2.81}$$

where $\hat{H}_{|D}$ is the restriction of $\hat{H}$ on D, $\hat{H}_R = \hat{H} - \hat{H}_{|D}$, and $|\psi_D(\tau)\rangle$ is the FCIQMC wave function in the subspace D. The first term which involves the evaluation of $\hat{H}_{|D}\,|\psi_D(\tau)\rangle$ is then carried out by deterministic matrix-vector multiply, while the second term is treated with the FCIQMC scheme outlined above. Note that the latter also involves excitations between determinants in D and in the stochastic space. As D is small compared to the full Hilbert space and explicitly sampling only excitations to the stochastic space would be computationally extremely involved, all types of excitations are considered in the spawning step and those between determinants in D are then discarded to prevent double counting.

While typical dimensions for D are in the range of 10^4 to 10^5 [196], by efficient parallelization of the constriction of D, deterministic subspaces of up to at least $3 \cdot 10^7$ are feasible.

The semi-stochastic extension can greatly reduce the noise of the energy measures with little computational overhead and is therefore routinely used [69, 235, 236]. It is especially useful when targeting high-accuracy such as in ST-FCIQMC discussed in Chapters 5 and 6, in particular, the results there make use of the semi-stochastic extension.

2.4.8 Excited-state FCIQMC

The projective ground state search of FCIQMC can be extended to excited states using an orthogonal projector technique [85, 237] which allows for parallely calculating the lowest n eigenstates of $\hat{H}$ from n parallel FCIQMC calculations which are orthogonalized

against each other. Therefore, the basic algorithm with n parallelly evolved states is modified by adding an extra step that applies an instantaneous projector $\hat{O}_m$ to the m-th state $|\psi_m\rangle$, orthogonalizing it against all previous states $|\psi_{m'}\rangle$, $m' < m$. It is defined as [85]

$$\hat{O}_m = 1 - \sum_{m'<m} \frac{|\psi_{m'}\rangle \langle\psi_{m'}|}{\langle\psi_{m'} \,|\, \psi_{m'}\rangle}, \tag{2.82}$$

to remove all contribution parallel to any lower state from $|\psi_m\rangle$. The first state will be unmodified and thus converge to the ground state, while the second state will be orthogonalized to $|\psi_1\rangle$ at every iteration, thus converging to the lowest energy state orthogonal to $|\psi_1\rangle$, which is the first excited state, and higher states converging to higher excited states. Therefore, this approach is suited for calculating a fixed number of adjacent states while the high-energy part of the spectrum is difficult to reach.

Compared to the real-time approach to dynamic properties, directly calculating excited states with FCIQMC is less general but can yield greater accuracy in specific cases, particularly when the energy differences are to small to be resolved by the spectral decomposition. On the other hand, it is not feasible if a high number of states or higher excited states are contributing.

2.5 Green's functions and spectral functions

In a non-interacting picture, a system is entirely described by the single particle spectrum consisting of the eigenvalues ϵ_i of the single particle Hamiltonian $\hat{H}_1$. Be it in the form of a band structure or orbital energies, the single particle energies provide a complete insight into the physics of a non-interacting system [22], as any many body eigenstate can be readily constructed as product state, and properties such as the density matrix and partition function can be analytically evaluated [83]. The nature of interaction however is to render the Hamiltonian non-linear in the number of electrons, thus necessarily correlating electrons. This might be taken into account in the wave function ansatz, which will be the topic of Chapters 5 and 6, but first consider the notion of single particle states themselves.

Consider now an interacting system, that is, the eigenstates are not product states anymore and while formally a single particle basis can be used to express the problem, there exists no such basis to decouple the electrons [22], so a single particle spectrum $\{\epsilon_i\}$ can only carry meaning as an approximation to the many-body spectrum, potentially as a starting point for perturbative treatment [238].

A common concept to attribute meaning to single particle quantities, especially in the context of solid state physics, is the quasi-particle spectrum [22, 239, 240], which describes many-body excitations with removal and addition of electrons to the system, and can be interpreted as available many-body energy levels, for example as a band structure of a solid [21]. Closely connected to the quasi-particle spectrum is the density of states, which determines optical, electrical and thermal properties of a compound [83].

As the quasi-particle spectrum is given by the energies of excitations in the presence of a background of electrons, it is defined by the energy levels associated with adding or removing a particle to the many-body ground state [240], which can be interpreted as the energies of a single quasi-particle or hole in the presence of the other particles. In principle, these energies could be obtained by simply calculating a sufficiently large set of eigenvalues and eigenstates of the Hamiltonian, but given the difficulty one has with even calculating only the ground state, that path is reserved for a limited system size [88, 241]. The more direct approach however is via the Green's function formalism [242, 243, 244], which itself can be used to entirely formulate quantum mechanics equivalently without the concept of wave functions [242].

Here however, Green's functions shall be evaluated using a wave function based technique, consider therefore the causal Green's function, which is defined as [22, 245]

$$G_{ij}(t) = -\mathrm{i}\left[\Theta(t)\left\langle \hat{a}_i(t)\hat{a}_j^\dagger(0)\right\rangle + \Theta(-t)\left\langle \hat{a}_j^\dagger(0)\hat{a}_i(t)\right\rangle\right], \tag{2.83}$$

where $\hat{a}_{i,j}$ are the single-particle basis annihilation operators, $\hat{a}_{i,j}^\dagger$ the corresponding creation operators, Θ the Heaviside step function and the expectation value is with respect to the ground state $|\psi_0\rangle$. This definition is to be read in the Heisenberg picture [246, 247] where the time-dependence of an operator $\hat{O}$ is given as

$$\hat{O}(t) = \mathrm{e}^{-\mathrm{i}\hat{H}t}\hat{O}\hat{e}^{\mathrm{i}\hat{H}t}. \tag{2.84}$$

Being a time-dependent quantity, the Green's function can readily be obtained by solution of the time-dependent Schrödinger equation, the established procedure [248] as used in Ref. [249] is given in the following. Physically, the Green's function describes the response of the system to removing/adding a particle from a single-particle state j and putting it back/removing it after a time t to a state i. As the time-evolution of the system is governed by the Hamiltonian, this response carries information on all involved energy levels, allowing it to describe the (inverse) photoemission spectra for $t > 0$ ($t < 0$). For $t > 0$ ($t < 0$), the Green's function is here labeled G_{ij}^+ (G_{ij}^-), and the

time-evolution is carried out with one electron more (less) than present in the ground state. In the following, the case $t > 0$ is considered, the procedure for $t < 0$ is analogous. Inserting the time-evolution of $\hat{a}_i$, the Green's function reads

$$\begin{aligned} G_{ij}^{+}(t) &= -\mathrm{i}\left\langle \psi_0 \middle| \mathrm{e}^{\mathrm{i}\hat{H}t}\hat{a}_i\mathrm{e}^{-\mathrm{i}\hat{H}t}\hat{a}_j^{\dagger}\psi_0 \right\rangle \\ &= -\mathrm{i}\mathrm{e}^{\mathrm{i}E_0 t}\left\langle \hat{a}_i\psi_0 \middle| \mathrm{e}^{-\mathrm{i}\hat{H}t}\hat{a}_j\psi_0 \right\rangle \\ &= -\mathrm{i}\mathrm{e}^{\mathrm{i}E_0 t}\left\langle \left[\hat{a}_i^{\dagger}\psi_0\right](0) \middle| \left[\hat{a}_j^{\dagger}\psi_0\right](t) \right\rangle , \end{aligned} \tag{2.85}$$

with the ground state energy E_0. So evaluating the Green's function boils down to calculating the time-evolution of a perturbed state $|\hat{a}_j^{\dagger}\psi_0\rangle$. Furthermore, this time-evolution can be reduced to a spectral decomposition of the perturbed state. Formally, one can express $|\hat{a}_j^{\dagger}\psi_0\rangle$ in an eigenbasis $|E_n^{+}\rangle$ with eigenvalues E_n^{+}

$$\hat{a}_j^{\dagger}\left|\psi_0\right\rangle = \sum_n \left\langle E_n^{+} \middle| \hat{a}_j^{\dagger} \middle| \psi_0 \right\rangle \left|E_n^{+}\right\rangle . \tag{2.86}$$

Note that the eigenvalues E_n^{+} are the eigenvalues of $\hat{H}$ with one additional electron compared to $|\psi_0\rangle$. The time-evolution is then given by

$$\left[\hat{a}_j^{\dagger}\left|\psi_0\right\rangle\right](t) = \sum_n \left\langle E_n^{+} \middle| \hat{a}_j^{\dagger} \middle| \psi_0 \right\rangle \mathrm{e}^{-\mathrm{i}E_n^{+}t}\left|E_n^{+}\right\rangle . \tag{2.87}$$

Inserting into equation (2.85) yields

$$G_{ij}^{+}(t) = -\mathrm{i}\sum_n \overline{d}_{i,n}^{+}\left\langle E_n^{+} \middle| \hat{a}_j^{\dagger} \middle| \psi_0 \right\rangle \mathrm{e}^{-\mathrm{i}\left(E_n^{+}-E_0\right)t} , \tag{2.88}$$

and thus the Green's function contains both information on the spectrum of $\hat{H}$ as well on the spectral decomposition of the states $|\hat{a}_{i,j}^{\dagger}\psi_0\rangle$. The Green's function can also be expressed as a Fourier transform of the spectral function A_{ij} as [245]

$$G_{ij}^{+}(t) = -\mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}\omega A_{ij}^{+}(\omega)\mathrm{e}^{-\mathrm{i}\omega t} . \tag{2.89}$$

The spectral function A^{+} is nothing but the spectral decomposition of the initial state

$$A_{ij}^{+}(\omega) = \sum_n \overline{d}_{i,n}^{+}\left\langle E_n^{+} \middle| \hat{a}_j^{\dagger} \middle| \psi_0 \right\rangle \delta\left(\omega - \left(E_n^{+} - E_0\right)\right) . \tag{2.90}$$

The case $t < 0$ works analogously and gives rise to a spectral function A^-_{ij}, it has the form

$$A^-_{ij}(\omega) = \sum_n \overline{d}^-_{i,n} d^-_{j,n} \delta\left(\omega + (E^-_n - E_0)\right) , \tag{2.91}$$

thus if $|\psi_0\rangle$ is the ground state of $\hat{H}$ with respect to the number of electrons, too, the support is limited to $\omega < 0$, while A^+ will have a support on $\omega > 0$. The full spectral function is then [83]

$$A_{ij} = A^+_{ij} + A^-_{ij} , \tag{2.92}$$

and describes the quasi-particle spectrum of the system, that is, the peaks of A_{ij} correspond to quasi-particle excitations levels involving a transfer from single-particle state j to i. A common choice of the single-particle basis in solid state models are momentum eigenstates, in that case the Green's and spectral functions are diagonal and only A_{ii} is nonzero, due to momentum conservation [22]. Ultimately, the goal is to obtain the spectral function as the information on the present energy levels can be readily obtained from it. To do so from the solution of the time-dependent Schrödinger equation, first the Green's function has to be calculated on the fly by evaluating the overlap of the time-evolved state with a reference state, yielding $G^{\pm}_{ij}(t)$, and then an inverse transformation of $G^{\pm}_{ij}$ is required, a task that can either be performed by Fourier transform or, more generally, by analytic continuation [110, 245]. Since the Green's function carries the same information as a decomposition of a given state into the eigenbasis, calculating all possible Green's functions, including higher order versions, carries the same amount of information as a diagonalization of the Hamiltonian [250]. A real time-evolution offers an elegant route here to extract targeted information from the Hamiltonian, without the need for a full diagonalization.

For a molecular system, photo-absorption processes are more often of interest [77], but they can be analyzed in an analogous procedure, using the expectation value

$$\left\langle \hat{a}^\dagger_j(t)\hat{a}_i(t)\hat{a}^\dagger_i(0)\hat{a}_j(0) \right\rangle = \left\langle \hat{a}^\dagger_i\hat{a}_j\psi_0 \,\middle|\, \mathrm{e}^{-\mathrm{i}(\hat{H}-E_0)t} \,\middle|\, \hat{a}^\dagger_i\hat{a}_j\psi_0 \right\rangle , \tag{2.93}$$

which describes the process of performing a single excitation from an orbital j to an orbital i, evolving for some time t and performing the reverse single excitation. This models the absorption of a photon at time 0 and re-emission at time t, while all excited states with nonzero overlap with this singly excited state are captured in the spectrum.

A procedure analogous to the treatment of Green's functions outlined above then leads

to the spectral decomposition of the singly excited state $\hat{a}_i^\dagger \hat{a}_j |\psi_0\rangle$,

$$A(\omega) = \sum_n |d_n|^2 \delta(\omega - (E_n - E_0)) \ , \tag{2.94}$$

where

$$d_n = \left\langle E_n \middle| \hat{a}_i^\dagger \hat{a}_j \psi_0 \right\rangle \ , \tag{2.95}$$

with the eigenstates $|E_n\rangle$ with eigenvalues E_n of $\hat{H}$. Since the single-excitation operator is particle number conserving, these are now the eigenstates at the same number of electrons as $|\psi_0\rangle$. The function A now constitutes a spectral decomposition of the initial state, as it contains the information of which eigenstates $|E_n\rangle$ contribute with which weights d_n to the initial state, and further contains information on the corresponding energies E_n. Using a singly excited state $\hat{a}_i^\dagger \hat{a}_j |\psi_0\rangle$ as an initial state is simply an important example, but the spectral decomposition can be calculated for any initial state with this technique.

To summarize, this procedure can be performed for any initial states, yielding the spectral decomposition thereof. The cases of particular interest are outlined here, being the calculation of Green's functions and spectral decomposition of excitations.

2.5.1 Analytic continuation

The transformation (2.89) from the real-frequency function A to the time domain function G is dictated by the Kernel

$$K(\omega, t) = \mathrm{e}^{-\mathrm{i}\omega t} \, , \tag{2.96}$$

so solving Eq. (2.89) for A when G is known requires inversion of the Kernel. When discretizing the integral in Eq. (2.89), this amounts to inversion of a corresponding matrix K, and for real t, the inverse problem is well-defined and treatable by means of inverse Fourier Transform [251].

On the other hand, the case of an imaginary t has been of considerable interest in the context of Quantum Monte Carlo methods [252, 253, 254] which are typically evaluating imaginary-time properties. There, the Kernel K is singular as it decays exponentially for large ω, meaning the time-signal G is agnostic to high-energy features of A, and A can thus not be uniquely determined from a numerically obtained G [110, 245].

The approaches to this problem are known as analytic continuation methods, with noteworthy representatives being the average spectrum method [255, 256, 257], analytic

continuation by Padé approximation [258, 259, 260], and the Maximum Entropy Method (MEM) [261, 262, 263, 110, 111, 245, 264].

In this work, the maximum entropy method is used, which has its origins in statistical mechanics [261, 262] and has also found notable application in image reconstruction [265] and machine learning [266, 267]. It is based on Bayesian Inference to determine the spectral function A that is most probable to yield a given Green's function G [110]. Here, spectral decompositions and diagonal Green's functions, for which A is positive are considered, and the idea of the MEM following Ref. [245] is summarized.

Consider the probability $P(A|G)$ to obtain some spectral function A from a given Green's function G and the probability $P(G|A)$ to obtain a certain G with a given A. Bayes theorem then states that [168]

$$P(A|G)P(G) = P(G|A)P(A)\,. \tag{2.97}$$

The goal of the MEM analytic continuation is now to find the function A that maximizes

$$P(A|G) = \frac{1}{P(G)} P(G|A)P(A) \tag{2.98}$$

for a given G. The prefactor $\frac{1}{P(G)}$ does not depend on A and therefore does not enter the optimization. As this is a statistical method, a set of samples G_i of G is required, these are assumed to be randomly drawn according to the probability function $P(G|A)$ [245]. The central limit theorem [171] states that the distribution of G_i is normally distributed in the limit of many samples, justifying the ansatz for $P(G|A)$ [245, 268]

$$P(G|A) = \mathrm{e}^{-\frac{1}{2}(\langle G_i\rangle - G(A))C^{-1}(\langle G_i\rangle - G(A))}\,, \tag{2.99}$$

with the covariance matrix C of the samples G_i and $G(A)$ given by the transformation (2.89).

For the ansatz for $P(A)$, the normalization and positivity of A are employed to interpret it as a probability distribution. Then, the probability $P(A)$ has the form

$$P(A) \propto \mathrm{e}^{\alpha \int \mathrm{d}\omega \left[A(\omega)\left(1-\ln\left(\frac{A(\omega)}{M(\omega)}\right)\right) - M(\omega)\right]}\,, \tag{2.100}$$

with a default model M, which would be the spectrum in the absence of data [245]. This functional form follows alone from statistical arguments [263, 269], the exponent can be viewed as entropy relative to the default model [245].

The spectral function A is then chosen to maximize $P(A|G)$, where multiple schemes for determining the parameter α are available, notably the historic [270], classic [268] or Bryan's [271] method. The data presented in Chapter 4 is obtained using the classic method, but those results do not strongly depend on this choice.

A non-singular covariance matrix C is required to evaluate $P(G|A)$ in equation (2.99), as the inverse C^{-1} enters the expression. If only few samples are available, this might become problematic as the inversion becomes ill-defined. One potential approach which is common in MEM analytic continuation is ignoring the off-diagonal elements of C [245], which might not correctly account for noise in the analytic continuation. This can then be approached by separating the data into batches which are treated independently, and subsequently averaging over these batches [245].

Another potential approach that is suggested here and has also been used in Ref. [249] is to impose a minimum value $\sigma_{\min}$ on the diagonal elements of C to regularize the covariance matrix by overestimating the noise in the data. The resulting spectra can be affected by this approximation in their details, mainly manifest as blurring of high-energy features. This can then be systematically evaluated for several possible values of $\sigma_{\min}$ that achieve an invertible C, to analyze the dependency of the resulting spectra. This regularization scheme is used when obtaining the MEM spectra in chapter 4, using values $\sigma_{\min} = 10^{-4} \sim 10^{-6}$ which are preferably low but still yield a non-singular C, see also Fig. 4.5.

While the MEM analytic continuation traditionally have been designed for use with the imaginary-time kernel, they can be generalized straightforwardly for general complex time contours $t = t(\tau)$ [249] with a real parameter τ.

Chapter 3

Stochastic time-evolution algorithm

A broad range of applications for electronic ground state calculations, including the evaluation of potential energy surfaces and geometry optimization, the study of ionization potentials, spin properties and many more. There are however notable problems that require knowledge of dynamic properties of a given system, encompassing a range of spectroscopic applications, including among others the determination of band structures, optical conductivities and photo-absorption spectra.

The established algorithm for excited state calculation with FCIQMC [85, 237], cf. Section 2.4.8, relies on simultaneous convergence of a set of orthogonal states, which will converge to a set of lowest energy states. This makes obtaining dynamic properties, that is, properties that depend on energy, difficult, because a high number of states is required unless only the lowest energy states are contributing, i.e. a purely low-energy property is considered. Furthermore, when calculating spectral functions, the number and distribution of states appearing in a spectrum is usually not known *a priori*, which makes obtaining a spectral function by subsequent calculation of single states inefficient and costly. A more direct approach to calculating dynamic properties is by considering time-dependent quantities, whose time-dependence is dictated by the time-dependent Schrödinger equation, which are linked to the corresponding energy-dependent property by Fourier transformation. This carries the advantage that the states which are relevant for a given problem, and only these, are taken into account implicitly in the calculation. Furthermore, a specific dynamic property can be targeted directly by precisely defining the time-dependent quantity to be calculated. On the flip side, solving the time-dependent Schrödinger equation is more involved than the stationary variant.

The algorithm and its details have been developed in collaboration with Olle Gun-

narsson and Ali Alavi, with the original design being conceived by Olle Gunnarsson, and previously published in K. Guther, W. Dobrautz, O. Gunnarsson, A. Alavi, Phys. Rev. Lett. 121, 056401 (2018) [249], figures reproduced from this publication are marked as such.

In this chapter, the real-time evolution via FCIQMC calculation is described, followed by details of the algorithm and the generalization to arbitrary complex-time contours using small exemplary calculations.

3.1 A real-time quantum Monte Carlo algorithm

In section 2.4 the basic FCIQMC algorithm was outlined, making the connection to the imaginary time Schrödinger equation

$$-\frac{\partial}{\partial\tau}\left|\psi\right\rangle = \hat{H}\left|\psi\right\rangle \ . \tag{3.1}$$

As discussed in section 2.4, the basic FCIQMC method can be formulated as an iterative solution to the imaginary time Schrödinger equation using a first order expanded propagator and applying it stochastically to a sparsely sampled wave function. To do so, the FCIQMC method uses a recipe to efficiently evaluate $\hat{H}\left|\psi\right\rangle$. Starting here, the formulation of a real time algorithm to treat the full time-dependent Schrödinger equation seems straightforward. In a naive formulation, starting from the FCIQMC propagator (2.56), one simply has to move back to real times $t = -\mathrm{i}\tau$ and employ the propagator

$$\hat{U}_1 = (1 - \mathrm{i}\Delta t\hat{H})\,, \tag{3.2}$$

within the FCIQMC framework described in section 2.4. In principle, this already captures the notion of the real time FCIQMC method, as repeatedly applying $\hat{U}_1$ yields an approximate time-evolution of an initial wave function $\left|\psi(0)\right\rangle$. However, there are a number of caveats here that make this simple algorithm practically unusable, and a more involved approach is required, which shall be discussed in this section, leading to a stable time-dependent algorithm.

As adding a constant shift to the Hamiltonian does not change the eigenvectors and only shifts the spectrum by a constant, therefore, arbitrary shifts μ may be applied to $\hat{H}$. When referring to $\hat{H}$ in this chapter, this always refers to the shifted Hamiltonian $\hat{H} - \mu$. In the context of Green's functions, a natural choice is $\mu = E_0$, directly yielding the spectral function without subsequent need of re-aligning the frequency axis, but

different choices might be advantageous for technical reasons, which is addressed in section 3.3. These deviant choices of μ are then corrected for by subsequently offsetting the spectrum by the difference, yielding the spectral function as defined in section 2.5.

Also note that standard imaginary time FCIQMC works with real Coulomb integrals, and since the propagator is then entirely real, the wave function can also be represented using real coefficients. Since the real time propagation inadvertently introduces a complex phase to the CI coefficients, they have to be represented as complex numbers. In practice, that requires storing both a real and an imaginary part of the CI coefficient per determinant, essentially doubling both the memory and the computational cost of any operation.

3.1.1 Integrator selection

When targeting the ground state with imaginary time evolution, using the first order expanded propagator $\hat{U}_1$ does not change the convergence behavior, and still leads to the same state as the exponential propagator. Even though the trajectory changes, this is of no concern for a ground state calculation. For a real time calculation, the wave function along the entire trajectory is of relevance, and the deviation from the exponential propagator is of concern.

The error connected to the expansion of the propagator can be estimated by considering an energy eigenstate, for which the exact propagation is known. Let $|\psi\rangle$ be an eigenstate of $\hat{H}$ with energy E, then the first order propagator with time $t = n\,\Delta t$ yields

$$|\psi(t)\rangle = (1 - \mathrm{i}\Delta t E)^n\,|\psi(0)\rangle = \mathrm{e}^{n\ln(1-\mathrm{i}\Delta t E)}\,|\psi(0)\rangle\ , \tag{3.3}$$

while the exact time-evolution would be

$$|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}n\Delta t E}\,|\psi(0)\rangle\ . \tag{3.4}$$

An estimate for the error can be given by expanding the logarithm [249] to second order in $\mathrm{i}\Delta t E$

$$n\ln(1 - \mathrm{i}\Delta t E) \approx -\mathrm{i}Et + E^2 t\frac{\Delta t}{2}\,, \tag{3.5}$$

meaning the first order propagation leads to approximately

$$|\psi(t)\rangle \approx \mathrm{e}^{-\mathrm{i}tE}\mathrm{e}^{E^2 t\frac{\Delta t}{2}}\,|\psi(0)\rangle\ . \tag{3.6}$$

The phase error is hence of at least second order in Δt and thus plays a negligible role on reasonable integration times t, but unitarity of the propagation, which manifests in norm conservation, is violated exponentially with increasing t with an exponent linear in Δt. This induces a quick exponential growth of the norm, which directly translates to an increase in memory cost over time that becomes unmanageable.

While this argument applies to energy eigenstates, it can be qualitatively extended to linear combinations of these and thus arbitrary wave functions, as each eigenstate will suffer from an exponential norm error. In the general case of a linear combination, the relative weight of the states will change however, since the exponent is quadratic in E, and the original norm error cannot be cured by simple normalization, it now also affects the structure of the wave function. The impact this has in practice is illustrated in Fig. 3.1, where real time FCIQMC calculations with a first order and a second order propagator are compared, highlighting the difference in behavior in both the norm and the Green's function $\left\langle \hat{a}_k^\dagger(t)\hat{a}_k \right\rangle$ for $k = (0,0)$, which contains information on the full wave function.

At first, the appearance of an exponential error in the norm as demonstrated here leaves little hope to cure the error, since any finite order propagator will suffer from this problem. However, the exponent scales with Δt, such that for a higher order expansion, the exponent of the norm violating term will scale with higher orders of Δt, leading to negligible error on the scale of relevant values of t.

Consider a second order expansion of the propagator

$$\hat{U}_2 = 1 - \mathrm{i}\Delta t\hat{H} - \frac{(\Delta t)^2}{2}\hat{H}^2 \,. \tag{3.7}$$

Applying it n times to the energy eigenstate $|\psi\rangle$ yields

$$\begin{aligned} |\psi(t)\rangle &= (1 - \mathrm{i}\Delta t E - \frac{(\Delta t)^2}{2}E^2)^n \,|\psi\rangle \\ &= \mathrm{e}^{n\ln\left(1-\mathrm{i}\Delta t E - \frac{(\Delta t)^2}{2}E^2\right)} \,|\psi\rangle \,. \end{aligned} \tag{3.8}$$

Again, by expanding the logarithm in Δt, an estimate for the phase and normalization

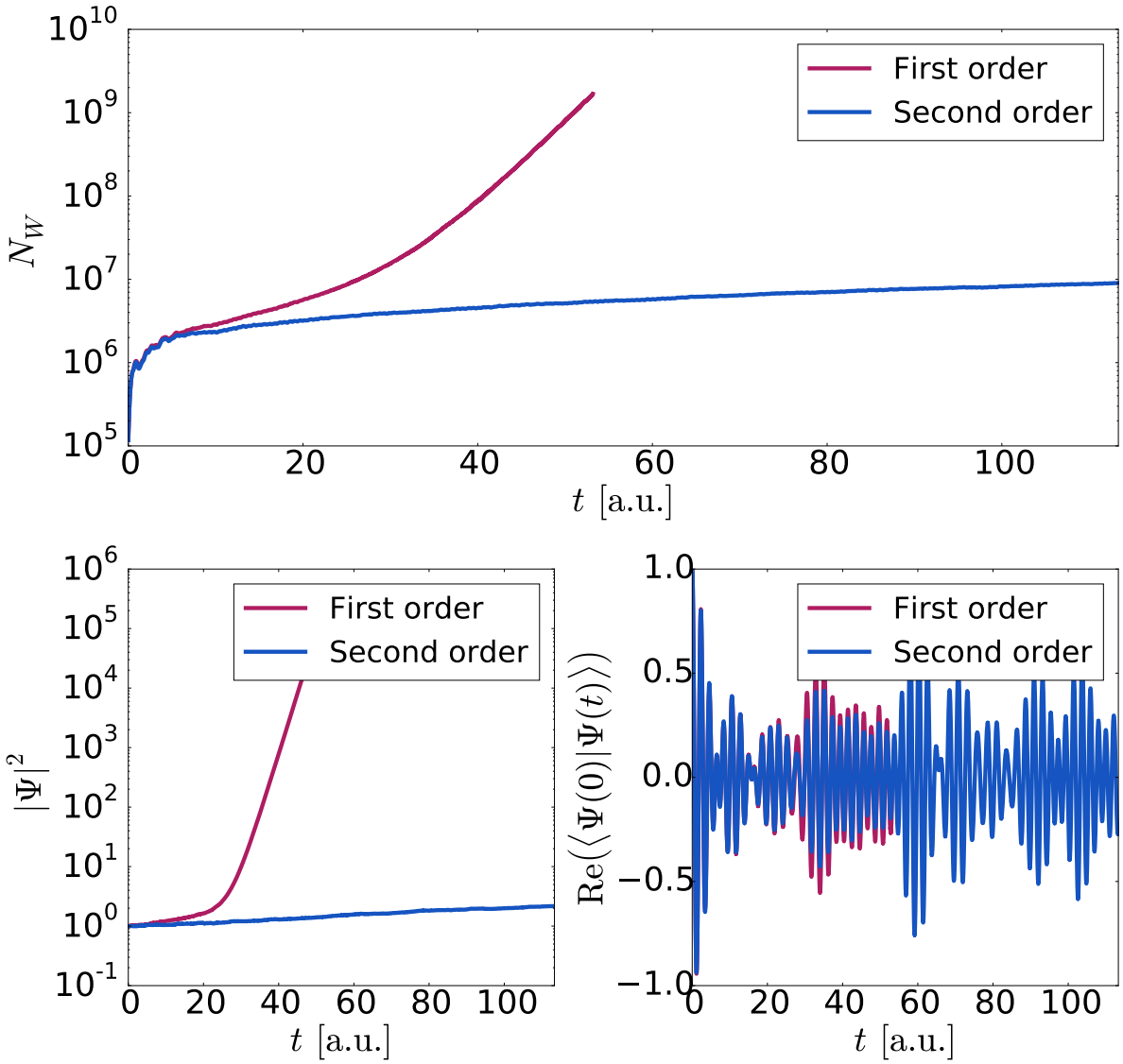


Figure 3.1: L_1 (N_w) and L_2 ($|\psi|^2$) norms of the wave function and Green's function over time for a first order and second order propagator in the real time FCIQMC scheme. The considered system is a 10-site Hubbard model at $k = (0, 0)$ and $U/t = 1$. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), supplementary material, Copyright 2018 American Physical Society.

error can be made, the expansion to second order in $\mathrm{i}\Delta t E + \frac{(\Delta t)^2}{2}E^2$ yields

$$n\ln\left(1 - \mathrm{i}\Delta t E - \frac{(\Delta t)^2}{2}E^2\right) \approx -\mathrm{i}tE - \frac{t\Delta t}{2}E^2 - \frac{n}{2}\left(\mathrm{i}\Delta t E + \frac{(\Delta t)^2}{2}E^2\right)^2$$
$$= -\mathrm{i}tE - \mathrm{i}t(\Delta t)^2E^3 - \frac{t}{2}(\Delta t)^3E^4\,. \tag{3.9}$$

While the phase error is still second order in Δt, the exponent of the norm violation

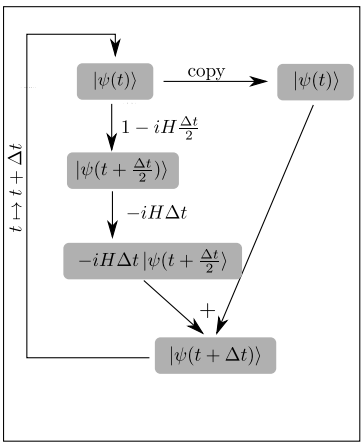


Figure 3.2: Illustration of the second order Runge-Kutta implementation using two half-steps. The annihilation step in the second half-step is performed using a backup of the initial wave function. In particular, the death step of the second half step is not carried out in place, but treated like a second spawn step. The wave function after one half-step is labeled $|\psi(t + \frac{\Delta t}{2})\rangle$, as this is the wave function that is created by propagating by $\frac{\Delta t}{2}$ with the first order propagator $\hat{U}_1$.

is cubic in Δt, leading to a much slower violation of unitarity. Not only can this effect be neglected on the scales of reasonable calculations, but even turns out to be small compared to the effect of stochastic errors, which will be discussed in section 3.1.2.

Second order Runge-Kutta Integrator

Before turning to the impact of the stochastic nature of the evaluation of $\hat{H}\,|\psi\rangle$ in FCIQMC, a few considerations regarding the implementation of the integrator shall be made. Foremost, it is desirable to keep the repeated applications of $\hat{H}$ per iteration low, as each induces both computational and memory overhead and thus significantly increases the cost per iteration. So while it is possible to use a higher order integrator, the simplest integrator of sufficient accuracy should be used, and second order integration has proven to be sufficiently accurate on reasonable time scales. The method employed here shall be a second order Runge-Kutta [272, 273] method , the explicit midpoint method [274], in the following simply referred to as second order Runge-Kutta. Since it is desirable not to explicitly evaluate $\hat{H}^2$, the operator $\hat{U}_2$ is decomposed into

$$\hat{U}_2 = \underbrace{1 - (\mathrm{i}\hat{H}\Delta t)}_{\text{second half-step}} \underbrace{(1 - \frac{\mathrm{i}}{2}\hat{H}\Delta t)}_{\text{first half-step}} . \tag{3.10}$$

This corresponds to two first order half steps, first with half the time-step, following by a half-step by the full time step. In the second half-step, the annihilation step is performed with the original wave function from before the first half-step, such that a reset is required after the second death step, this is illustrated in Fig. 3.2. That is, additional memory to store a second copy of the wave function is required. Since FCIQMC is generally not memory-bound, that is not an issue however.

Third order Verlet Integrator

Now, a third order symplectic Verlet integrator [275] could be implemented with the same computational cost by re-using $|\psi(t-\Delta t)\rangle$ to obtain $|\psi(t+\Delta t)\rangle$. For the formulation according to [275], consider the third order integrator

$$\hat{U}_3 = 1 - \mathrm{i}\Delta t\hat{H} - \frac{(\Delta t)^2}{2}\hat{H}^2 + \mathrm{i}\frac{(\Delta t)^3}{6}\hat{H}^3\,, \tag{3.11}$$

which translates to the time-propagation

$$|\psi(t+\Delta t)\rangle = \left(1 - \mathrm{i}\Delta t\hat{H} - \frac{(\Delta t)^2}{2}\hat{H}^2 + \mathrm{i}\frac{(\Delta t)^3}{6}\hat{H}^3\right)|\psi(t)\rangle \tag{3.12}$$

$$|\psi(t-\Delta t)\rangle = \left(1 + \mathrm{i}\Delta t\hat{H} - \frac{(\Delta t)^2}{2}\hat{H}^2 - \mathrm{i}\frac{(\Delta t)^3}{6}\hat{H}^3\right)|\psi(t)\rangle\ , \tag{3.13}$$

as the original propagation is unitary, it is valid to assume the evolution by $-\Delta t$ is governed by the same equation as the evolution by Δt. This argument is the core idea of the symplectic formulation of the integrator and allows to evaluate the third order integrator at the same computational cost as a direct second order integrator like the Runge-Kutta integrator discussed in section 3.1.1 [275].

Adding the two equations (3.12) and (3.13) eliminates the cubic term, yielding

$$|\psi(t+\Delta t)\rangle + |\psi(t-\Delta t)\rangle = \left(2 - (\Delta t)^2\hat{H}^2\right)|\psi(t)\rangle\ , \tag{3.14}$$

or, solving for $|\psi(t+\Delta t)\rangle$,

$$|\psi(t+\Delta t)\rangle = \left(2 - (\Delta t)^2\hat{H}^2\right)|\psi(t)\rangle - |\psi(t-\Delta t)\rangle\ . \tag{3.15}$$

Which formulates the third order symplectic Verlet integrator. For a deterministic algorithm, it is more accurate at the same cost, and thus should be preferred over the aforementioned Runge-Kutta integrator. The implementation is done by directly evaluating $\hat{H}^2|\psi\rangle$ by repeatedly applying $\hat{H}$. Since the wave function of the last iteration is required, too, the memory and computational cost per iteration is then the same as in the second-order Runge Kutta implementation.

However, as an exemplary calculation displayed in Fig. 3.3 shows, stochastic errors accumulate when evolving with a verlet integrator, rendering the time-evolved wave function useless after a short propagation even in simple cases.

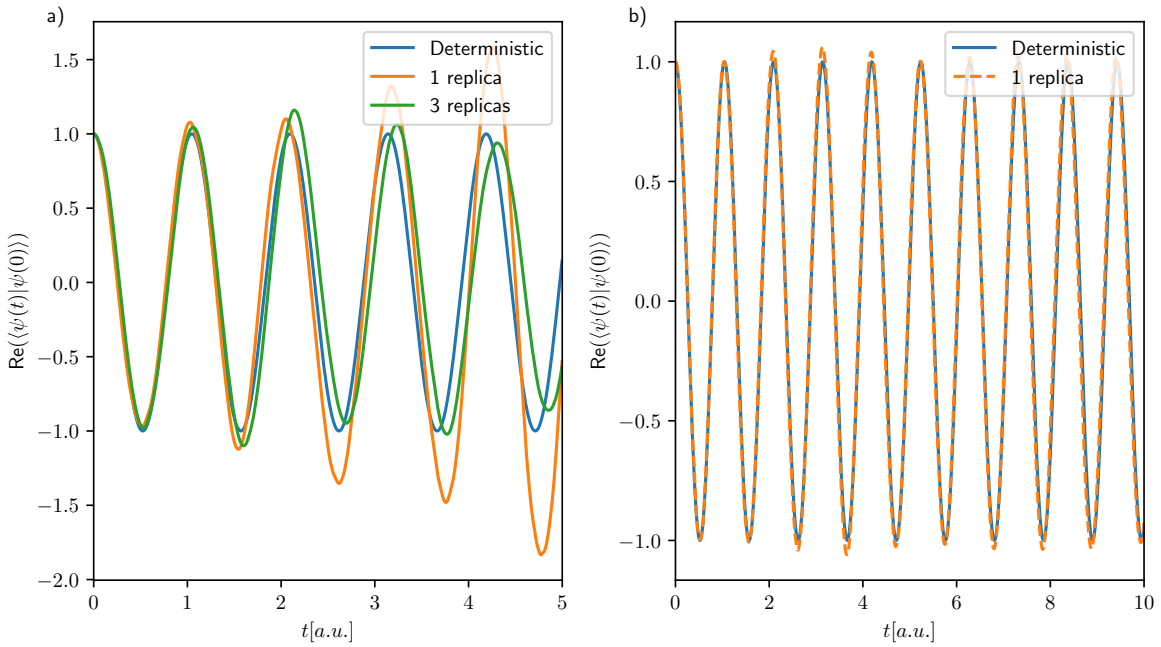


Figure 3.3: Green's function of the two-site Hubbard model, calculated with the stochastic method using a a) Verlet and b) Runge-Kutta propagator, compared to the corresponding deterministic integration. Averaging over several parallel calculations remediates the errors to some extent, but the Verlet integration nevertheless leads to an accumulation of stochastic errors.

3.1.2 Impact of Stochastic Propagation

The algorithm presented so far could as well describe a deterministic procedure, performing a numerical integration of the time-dependent Schrödinger equation, were it not for the severe memory bottleneck deterministic algorithms face. This can be approached in various ways, by approximating the many-body Hilbert space like in CASSCF or multi-reference CI approaches, by specifying the form of the wave function as for example in Coupled Cluster or DMRG methods, or by sampling the wave function in a stochastic manner, leading to the family of QMC methods as described in section 2.3. Using the stochastic FCIQMC approach means two things here: first, the wave function shall be sampled and represented in a sparse format instead of represented as a dense CI vector and second, the evaluation of the action of the Hamiltonian on the instantaneous wave function is sampled stochastically.

In particular, having a sampled wave function has serious implications for a real time algorithm. First, in contrast to imaginary time calculations, any results obtained from a real time evolution crucially depend on the initial state $|\psi(0)\rangle$, as was addressed in detail in section 2.5. As the initial state has to be given in the form of an FCIQMC

wave function, that is, sparsely sampled, with coefficients being rounded stochastically, it is inevitably an approximate description to the intended state in all but trivial cases. If, say, the task of calculating the time evolution of an eigenstate is to be performed, which is trivial from a theoretical point of view, this eigenstate has to be supplied in the FCIQMC format using the non-normalized approximate finite-norm representation described in 2.4.1, meaning the initial CI coefficients will not exactly match the CI coefficients of the eigenstate, even if they were known. In practice, this is even more problematic, since the exact form of the desired initial state is not known in general. Then, the initial state itself has to be calculated using the imaginary time FCIQMC algorithm, which leads to a stochastic error associated with all CI coefficients, not just those affected by the stochastic round, since the initial FCIQMC calculation uses a stochastic sampling of the Hamiltonian. Thus, one can assume any non-trivial initial state to carry some stochastic error $|\delta\psi(0)\rangle$.

In addition to that, the time-evolution itself is also stochastic in nature, thus introducing further stochastic errors alongside the time-evolution. This is visible very prominently when employing the third-order symplectic Verlet integrator discussed in section 3.1.1, and is the reason behind the resulting failure. In order to keep stochastic errors manageable, it is essential for the errors of subsequent iterations to be statistically independent, which is not the case for the Verlet integrator. There, the first order term is estimated from the difference in the wave function between the last two time steps, and not calculated independently, meaning that the change in $|\psi(t)\rangle$ between time steps is correlated such that stochastic errors in $|\psi(t)\rangle$ accumulate quickly.

In the ground state calculations performed with imaginary time FCIQMC, any properties are averaged over imaginary time to eliminate noise, a technique that can even be extended to sampling the expectation value of arbitrary two-body operators [276] and reduces the errors arising from the approximate evaluation of the Hamiltonian to the point of negligibility. Obviously, it is not possible to average over subsequent iterations when performing time evolution, as different iterations correspond to different times and thus physically different wave functions. Note that this restriction does not apply to a converged imaginary time evolution, since the state is stationary then and subsequent iterations actually correspond to the same physical state. The inability to average over iterations means that only the instantaneous wave function can be used to extract any properties. The noise in the resulting properties can then be addressed by averaging over several populations, commonly referred to as replicas, which are evolved in parallel. Using several parallel populations is an established concept in other appli-

cations of the FCIQMC algorithm [276, 277] and will turn out to be indispensable for more involved real time calculations. This conceptually simple approach starts from a number of independent FCIQMC wave functions $|\psi_a(0)\rangle$ and evolves each of them into a time-dependent wave function $|\psi_a(t)\rangle$, such that properties can now be calculated by averaging over the different time-evolved states. Depending on the application, a large number of states might be required, while for some as few as two are sufficient. The computational cost increases linearly with the number of replicas, which adds to the overall increased cost of the time-dependent FCIQMC compared to the conventional, imaginary time algorithm.

There is a caveat connected to the evaluation of properties which is less straightforward, but can be the root of subtle errors. Calculating properties such as Green's functions is essentially evaluating a function f of the instantaneous and the initial CI coefficients, yielding an instantaneous value of the property. Concretely, for a Green's function, this boils down to evaluating

$$f\left(\mathbf{c}_a(t), \mathbf{c}_a(0)\right) = \mathbf{c}_a(0) M \mathbf{c}_a(t)\,, \tag{3.16}$$

where M is a matrix defining the precise Green's function and $\mathbf{c}_a$ is the normalized CI vector of $|\psi_a\rangle$. In the special, but significant, case of the diagonal elements of a Green's function, or the calculation of a spectral decomposition of an excitation, it is $M = \mathbb{I}$ and thus

$$f\left(\mathbf{c}_a(t), \mathbf{c}_a(0)\right) = \sum_i c_{a,i}(0)\, c_{a,i}(t)\,. \tag{3.17}$$

To illustrate the caveat, the example of a Green's function is used, but the described issue arises for any property which depends quadratically, or to a higher even order, on the CI coefficients. Each instantaneous CI coefficient can be assumed to carry some stochastic errors

$$c_{a,i}(t) = c_i^{\mathrm{determ}}(t) + \delta c_{a,i}(t) \tag{3.18}$$

where $c_i^{\mathrm{determ}}(t)$ is the coefficient a deterministic integration would yield and does not depend on a. This also holds for $t = 0$ except for trivial cases, as the initial wave function is typically a result of an FCIQMC calculation itself, as discussed above. Inserting into equation 3.17 yields

$$\begin{aligned} f\left(\mathbf{c}_a(t), \mathbf{c}_a(0)\right) = \sum_i & c_i^{\mathrm{determ}}(0)\, c_i^{\mathrm{determ}}(t) + c_i^{\mathrm{determ}}(0)\delta c_{a,i}(t) \\ & + c_i^{\mathrm{determ}}(t)\delta c_{a,i}(0) + \delta c_{a,i}(0)\delta c_{a,i}(t)\,. \end{aligned} \tag{3.19}$$

For uncorrelated $|\psi_a\rangle$, averaging over a will reduce the terms linear in $\delta c_{a,i}$ to 0 in the limit of infinitely many replicas. This is not true, however, for the quadratic term, as $\delta c_{a,i}(0)$ and $\delta c_{a,i}(t)$ are correlated, especially for small t. Then, the quadratic terms accumulate when averaging over a, might lead to subtle errors in the short-time Green's function. In most cases, these are negligible, but there are examples in which these auto-correlations introduce spurious high-energy features into the spectra, so care has to be taken. A similar problem is faced when evaluating properties like the reduced density matrices in the imaginary time FCIQMC [276] or in the Krylov projected FCIQMC [277], where it is solved by evaluating these expectation values between two different replicas. The same solution can naturally be used here, and since multiple replicas are required anyway, there is no additional cost associated. In the end, all combinations of different replicas can be used, such that for example a Green's function is evaluated as

$$G(t) = \sum_{a \neq a'} \sum_{i} c_{a,i}(0) c_{a',i}(t) \,. \tag{3.20}$$

Noteworthy in this context is the requirement for the initial states to be sign-coherent, that is, each of the independent replicas has to have the same overall sign of the wave function and the leading determinants have to have the same sign, else, the states $|\psi_a(0)\rangle$ and $|\psi_b(0)\rangle$ for $i \neq j$ cannot be interpreted as different snapshots of the same physical state and the overlaps $\langle \psi_a(t) \,|\, \psi_b(0) \rangle$ which are used in the end are meaningless. Severe sign-incoherence can be detected by considering the averaged overlap of the initial replicas. Consider a set of replicas $|\psi_a\rangle$, then the averaged overlap is

$$G(t) := \frac{\sum_{a \neq b} \langle \psi_a(t) \,|\, \psi_b(0) \rangle}{\sum_{a \neq b} |\, \langle \psi_a(0) \,|\, \psi_b(0) \rangle \,|} \,. \tag{3.21}$$

In the absence of stochastic fluctuations between the replicas, i.e. if all replicas are identical, it is

$$G(0) = \frac{\langle \psi(0) \,|\, \psi(0) \rangle}{|\psi(0)|^2} = 1 \,, \tag{3.22}$$

and this holds for as long as all overlaps $\langle \psi_a(0) \,|\, \psi_b(0) \rangle > 0$. If ever any of these overlaps becomes negative, the normalization of G is lost, which serves as an indicator of severe sign-incoherence between the replicas. When referring to the averaged Green's function in the following, the entity G is meant. It is used as an input for Fourier transform, which does not rely on a detailed statistical analysis. Maximum Entropy analytic continuation on the other hand requires more detailed statistical measures and is thus based on the

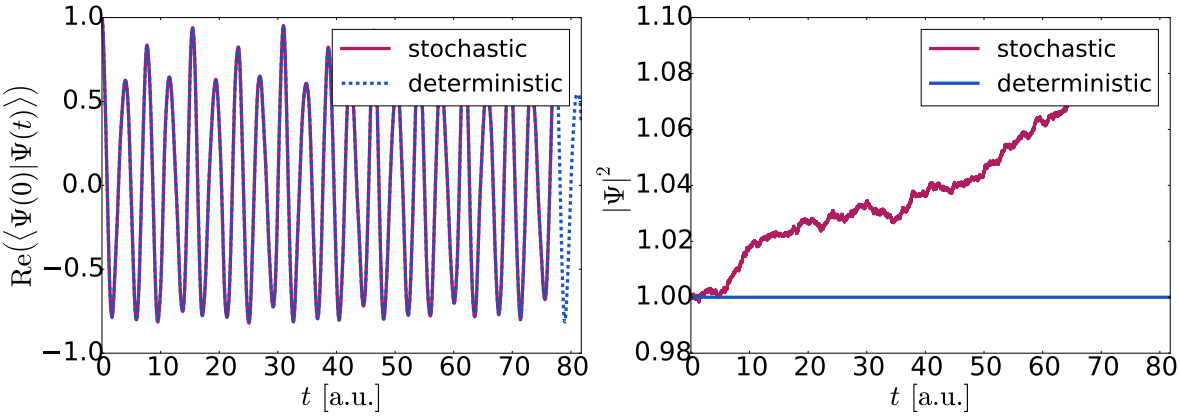


Figure 3.4: Green's function and norm of the wave function as a function of time for an exemplary calculation using the 10-site two-dimensional Hubbard model at $U/t = 1$ with $k = (0, 0)$. While the Green's function is resistant to stochastic errors, the norm grows over time well beyond the integration error. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), supplementary material, Copyright 2018 American Physical Society.

individual samples $\frac{\langle\psi_a(t)\,|\,\psi_b(0)\rangle}{\langle\psi_a(0)\,|\,\psi_b(0)\rangle}$.

It can happen that in the preparing imaginary time FCIQMC calculation, one state converges to $|\psi_0\rangle$ and another one to $-|\psi_0\rangle$,both sampling the sign structure of the state faithfully. In that case, the latter is simply multiplied by -1 to equalize the sign of the states. The implementation uses the sign of the leading determinant to gauge if two replicas have different overall sign.

To summarize, the stochasticity of the time-evolution is addressed by running multiple replicas in parallel and obtaining any property by averaging over these replicas.

3.2 Norm control

In section 3.1, the growth of the L_1-norm N_w, in the following just referred to as norm, due to the integrator and stochastic errors was discussed. It turns out that for more non-trivial systems, the issue of non-unitarity is extremely severe at the accessible N_w. In a sense, that can be interpreted as a manifestation of a dynamical sign problem. In the real-time dynamics, what is the FCIQMC death step in the imaginary time algorithm effectively manifests as a second spawn step, such that the only channel by which the population can decrease is the annihilation step. That on the other hand relies on a sufficiently dense sampling to perform actual annihilation, that is, elimination of occupied determinants, a circumstance which manifests in the plateau phase in the original, non-initiator method [59, 229].

This is highly problematic since the computational cost per iteration scales linearly with the norm, as does the memory cost. An exemplary calculation displayed in Fig. 3.5 shows the rapid increase of the norm with time, up to the point where the continuation of the calculation is unfeasible. While for very small systems, this growth is still manageable and extended calculations are possible, this is in general not the case and it is essential to limit the norm of the wave function. This can in essence only be done by introducing a damping that suppresses the growth of the wave function, but various ways of introducing such a damping can be thought of. The most promising approaches shall be discussed here

3.2.1 Static damping

One of the simplest form of damping that can be envisioned is a static, exponential damping of the form

$$\hat{U}(t) \mapsto \mathrm{e}^{-St}\hat{U}(t) = \mathrm{e}^{-\mathrm{i}(\hat{H}-\mathrm{i}S)t}\,, \tag{3.23}$$

with a damping parameter S. Clearly, this mapping exponentially suppresses the wave function and thus can be used to counter the norm growth, but at no point is unitarity recovered. In fact, recovering unitarity cannot be achieved by any damping scheme, since the damping has to be unbound, meaning it cannot conserve the L_2-norm and therefore has to violate unitarity. Another specific issue of a static damping is that there is no recipe to estimate the required value of S, making it inefficient and impractical to use. While it might though still serve the purpose of slowing down the norm growth in calculations in combination with other schemes, it has the advantage of its implications being

straightforward. The mapping in equation (3.23) can be understood as an imaginary shift to the Hamiltonian, thus resulting in the spectrum being shifted into the complex plane. For the dynamics, this directly implies an exponential decay of the wave function. For the spectral function which is composed out of delta functions at the eigenenergies, that implies a broadening [278]. The delta functions can be expressed as [83]

$$\delta(x-\omega) = -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\mathrm{Im}\left(\frac{1}{x-\omega+\mathrm{i}\epsilon}\right). \tag{3.24}$$

Adding an imaginary part S to the eigenenergies ω thus results in Lorentzian peaks instead of the sharp delta peaks, the spectral functions then takes the functional form

$$A(\omega) = \sum_n a_n \frac{\pi S}{S^2+(\omega-E_n)^2}, \tag{3.25}$$

with some weights a_n. In particular, a small damping S can be applied without algorithmic changes required and will only result in a slight broadening of the spectrum.

3.2.2 Dynamic damping

In imaginary time FCIQMC, the issue of a growing norm also arises, although for different reason. There, it is a property of the propagator to exponentially increase the norm unless $E_0 - S \geq 0$, while in the real time algorithm, the exact propagator is unitary and as seen in section 3.1 violation thereof comes to a small part from the expansion of the propagator and mainly from stochastic error.

That is good news in the sense that in the limit of infinite norm, where the stochastic evaluation of the action of the Hamiltonian is exact, the non-unitarity will solely be due to the expansion of the propagator and thus negligible on relevant time scales. This is also the reason why, for small systems, extremely long-time calculations are possible, as the norm required to reduce the stochastic error in the application of $\hat{H}$, to the point where unitarity violation is not significant anymore, is well within reach. For systems that go beyond only a few electrons however, the required wave function norms for that are completely out of reach, such that any attempt such as the one in Fig. 3.5 simply leads to rapidly growing norm and eventually failure.

Now, even though the root of the problem is different, the solution used in imaginary time FCIQMC, namely applying a dynamic shift to the Hamiltonian, can be ported to the real time algorithm. In practice, that means applying an exponential damping to

the wave function, in the form

$$|\psi(t)\rangle \mapsto \mathrm{e}^{-St}\,|\psi(t)\rangle\ , \tag{3.26}$$

with a shift S. That reduces the norm by an exponential factor and thus completely eliminates the associated growing computational cost. Since the damping factor is diagonal, it is not applied stochastically. Any observables will be affected by the same damping, and an advantage of this form is that observables can be readily corrected by applying the inverse mapping, e.g.

$$\left\langle \hat{a}_i^\dagger(t)\hat{a}_i \right\rangle \mapsto \mathrm{e}^{St} \left\langle \hat{a}_i^\dagger(t)\hat{a}_i \right\rangle\ . \tag{3.27}$$

As the required value of S is not known a-priori, it can be chosen to be time-dependent and updated iteratively to keep a constant norm, a technique that is successfully applied in the imaginary time FCIQMC. For a time-dependent t, the form of the inverse mapping (3.27) is then modified to

$$\left\langle \hat{a}_i^\dagger(t)\hat{a}_i \right\rangle \mapsto \mathrm{e}^{\int_0^t \mathrm{d}t' S(t')} \left\langle \hat{a}_i^\dagger(t)\hat{a}_i \right\rangle\ . \tag{3.28}$$

Even though the shift technique can be applied successfully in some cases, a fundamental problem becomes apparent here. The expectation value $\left\langle \hat{a}_i^\dagger(t)\hat{a}_i \right\rangle$ of the exact wave function is a quantity oscillating in the unit circle S^1, so it has an absolute value of an order of magnitude of 1, as can be seen from equation (2.88). The prefactor $\mathrm{e}^{\int_0^t \mathrm{d}t\, S(t')}$ is exponential in time, so the calculated overlap must be exponentially decaying, or at least, that would be the case for the exact wave function. The attempt to stochastically evaluate an exponentially decaying property, however, ultimately leads to vanishing signal to noise ratio, as stochastic noise is not exponentially decaying, eventually leaving only random noise to be rescaled, yielding an exponentially growing random signal instead of the desired Green's function. In essence, this is the same problem as the ill-definition of the inverse transformation in the analytic continuation, a connection that becomes more clear in section 3.3.

But before moving on to energy-dependent damping, a quick note on the inverse application shall be made. If, for some reason, values $S < 0$ can be chosen, the circumstances would invert, and instead of scaling up the noise in the overlap $\langle\psi(0)\,|\,\psi(t)\rangle$, it would be scaled down exponentially. Then, the Green's function could be resolved with astonishing resolution, essentially down to machine precision. In combination with an

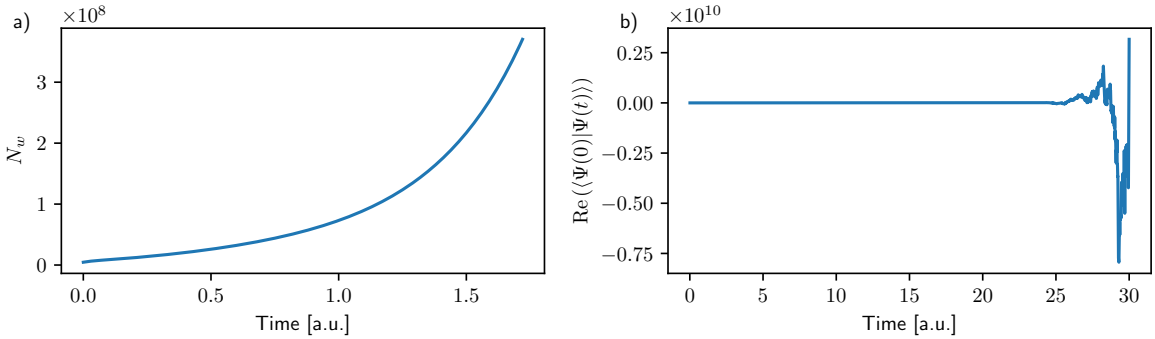


Figure 3.5: a) Growth of the L_1-norm of the wave function (N_w) over time in an uncontrolled calculation of a singly excited state of C_2 in a cc-pVTZ basis set. After only a few atomic units of time, the calculation has to be stopped due to excessive growth of the norm. b) Overlap $\mathrm{Re}\left(\langle\psi(t)\,|\,\psi(0)\rangle\right)$ over time for the same system, but with a dynamic damping $S(t)$ applied to keep the norm constant. While the calculation can now be carried out for an arbitrary time, the vanishing signal to noise ratio leads to unusable results.

energy-dependent damping, this technique can then be used to extend the time window for which a well resolved Green's function is available, see also Fig. 3.8.

3.3 Complex time contours

It is typically the low-energy properties of a system that are of special interest, as these govern the chemical and physical properties under non-extreme conditions. It might thus seem natural to try an approach which is more suited to describe the low-energy part of the spectrum at the cost of reduced accuracy in the description of the high-energy part. In section 3.2.2, the application of a damping was discussed, which is suited to suppress the rapid increase in the wave function norm over time. Introducing an energy dependence to the damping allows to increase the accuracy of the low-energy description, while also giving access to tools specifically designed for this purpose, the family of analytic continuation methods, which then replace the simple rescaling suggested in section 3.2.2.

An elegant way to achieve energy-dependent damping is to introduce a complex time rotation

$$t = \mathrm{e}^{-\mathrm{i}\alpha}\tau = \cos(\alpha)\tau - \mathrm{i}\sin(\alpha)\tau\,, \tag{3.29}$$

where α is a control parameter determining the time trajectory. The time evolution is then carried out alongside the trajectory $t(\tau)$, with τ parametrizing the curve. Internally, the algorithm thus uses $\Delta\tau$ and α to determine the next complex time point

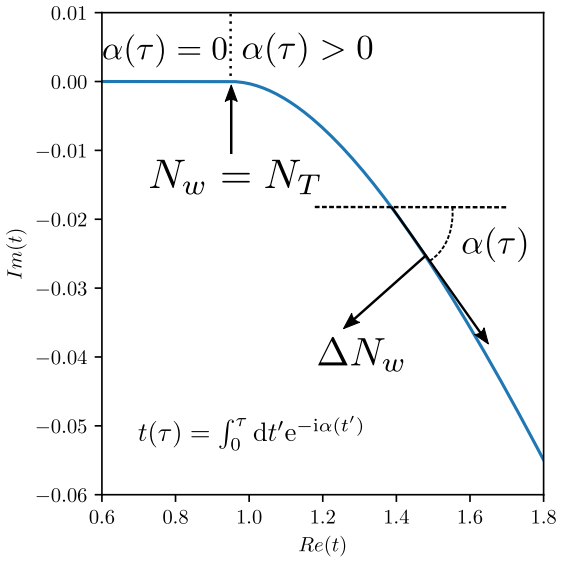


Figure 3.6: Illustration of the complex time contour evaluation, starting from $\alpha = 0$ until the norm exceeds a threshold value N_T, from when α is increased according to the change of the norm ΔN_w to maintain constant N_w.

t. Calculating the wave function as a function of τ is to switch to the complex-time propagator

$$\hat{U}(\tau) = \mathrm{e}^{-\mathrm{i}\hat{H}\cos(\alpha)\tau}\mathrm{e}^{-\hat{H}\sin(\alpha)\tau}\,, \tag{3.30}$$

which naturally contains the exponentially damping, imaginary time propagator. This comes at the cost that the resulting Green's functions also decay exponentially, while still carrying the information contained in the unitary propagator.

The goal of introducing the complex rotation by α is ultimately to maintain a constant norm, meaning a value of α has to be chosen that is neither too small to successfully suppress the norm growth, nor a value that is too big, in which case the norm vanishes. There is no prediction for the correct value in a specific case, such that the value has to be determined dynamically in practice. This is done by introducing a time-dependence

$$\alpha = \alpha(\tau)\,, \tag{3.31}$$

meaning that t now has the form

$$t(\tau) = \int_0^\tau \mathrm{d}t' \mathrm{e}^{-\mathrm{i}\alpha(t')}\,, \tag{3.32}$$

which is discretized with a time step $\Delta\tau$ to

$$t(\tau = n\Delta\tau) = \sum_i^n \Delta\tau \mathrm{e}^{-\mathrm{i}\alpha(i\Delta\tau)}\,. \tag{3.33}$$

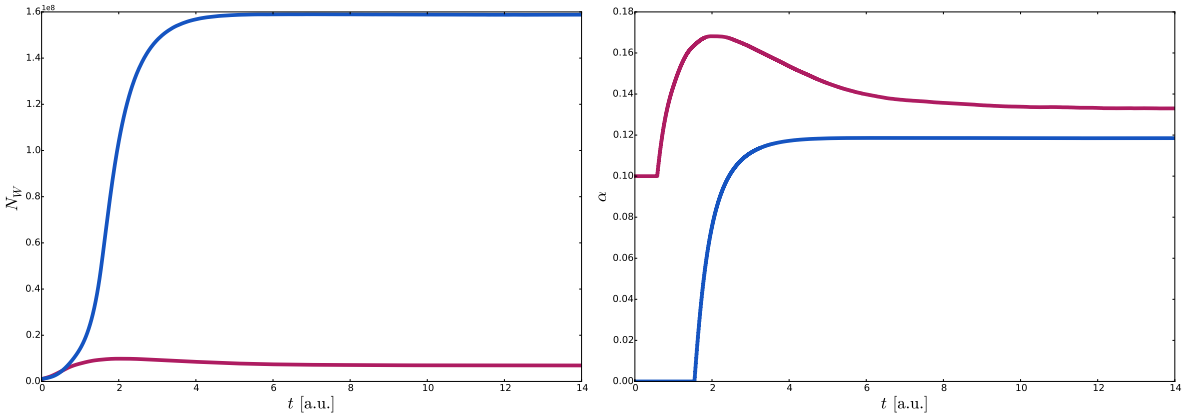


Figure 3.7: L_1-norm and time rotation parameter α over time for an exemplary calculation of the $k = (0, 0)$ Green's function of the 18-site Hubbard model with $U/t = 2$ for two calculations only differing in the choice of $\alpha(\tau)$. When starting at $\alpha = 0$ (blue), the norm initially grows rapidly, which is suppressed once α is allowed to vary, while maintaining a higher value of α from the beginning keeps the norm low overall (red). In particular, by increasing the norm, the necessary values of α decrease, and the time contour moves closer towards the real axis.

The parameter τ still carries the interpretation of the total elapsed time in the sense of the curve length of the time trajectory, while also serving to parametrize it, thus being the parameter used in the transformation employed for the analytic continuation. To guide α to control the norm, $\alpha(\tau)$ is determined iteratively, using an update analogous to the shift in the imaginary time FCIQMC. Every n iterations, α is updated according to

$$\alpha(\tau + n\Delta\tau) = \alpha(\tau) + D \operatorname{atan}\left(\frac{N_w(\tau + n\Delta\tau)}{N_w(t)} - 1\right), \tag{3.34}$$

with a control parameter D that can be chosen to adjust the reaction speed of α to changes in the L_1-norm N_w. Reasonable values for D are in the range $D \sim 0.1 \ldots 1$, the larger the chosen value, the more curvature in the time trajectory is allowed to follow fluctuations in N_w.

The resulting Green's functions are now functions of a complex time τ, but as the analytic continuation only requires for evaluation of the transformation from A to G according to equation (2.89), no further changes are required here, the complex time parametrized with equation (3.32) can be readily inserted into equation (2.89). Note that the exponential decay of the kernel

$$K(\omega, t) = \mathrm{e}^{-\mathrm{i}\omega \int_0^t \mathrm{d}t' \mathrm{e}^{-\mathrm{i}\alpha(t')}} \tag{3.35}$$

is given by α, the smaller the values of $\alpha(t')$, the slower the exponential decay. It has

been demonstrated by Olle Gunnarsson that the information in the spectral function that can be reconstructed from the complex time Green's function away from $\omega = 0$ increases with decreasing α [249]. At $\omega = 0$, there is no decay and thus the inverse transformation is feasible at all values of α.

The evolution alongside a complex time contour successfully cures the problem of the growing norm and allows to use well established analytic continuation techniques for reverting back to real frequencies. In contrast to imaginary time techniques, this approach can also accurately capture the spectrum away from $\omega = 0$, is systematically improvable and allows to recover the full spectrum in the limit $\alpha \to 0$.

Often, a high number of parallel FCIQMC calculations has to be carried out, and while the `NECI` implementation supports running with up to 10 replicas, that might not be sufficient and the memory overhead associated with a high replica count is undesired. To gather statistics it is thus usually most efficient to run a number of parallel FCIQMC calculations with a number of replicas each. Since the time trajectory is now determined at run-time, to ensure all replicas are using the same curve $\tau(t)$, `NECI` supports reading and writing the values $\alpha(\tau)$ at run-time during a calculation. One FCIQMC calculation can then be used as reference for $\tau(t)$, outputting the value of α each iteration, while the other instances are continuously reading this output. While the exact values of α that are determined at run-time depend on the precise trajectory and thus the seed of the random number generator, the required range of α to maintain a constant norm is a general property of a system and thus the same for all instances. This approach is not fail safe as the update scheme does not guarantee $\alpha < \frac{\pi}{2}$, being designed for handling small values of α. Nevertheless, as long as $\mu < E_0$, a suitable value of $0 \leq \alpha \leq \frac{\pi}{2}$ exists, since $\alpha = \frac{\pi}{2}$ corresponds to imaginary time evolution and in that case, a value $\mu < E_0$ is equivalent to a shift $S < E_0$ and thereby an exponential decay of the wave function.

The imaginary part of the time evolution naturally imposes a time scale on which a calculation can be carried out, as the Green's function now decays exponentially. To increase the total time of a calculation with a fixed initial N_w, the maximum N_w that is maintained has to be increased, and the imaginary part of time only slowly converges to zero, meaning the method is still relatively expensive compared to ground state FCIQMC calculation.

In the subsequent sections, for all time evolutions carried out in pure real time, the inverse transformation is carried out by Fourier transform and for complex time contours by maximum entropy analytic continuation. The Fourier transform is carried out using the Matlab Fourier transform functionality [279]. For the analytic continuation, it is

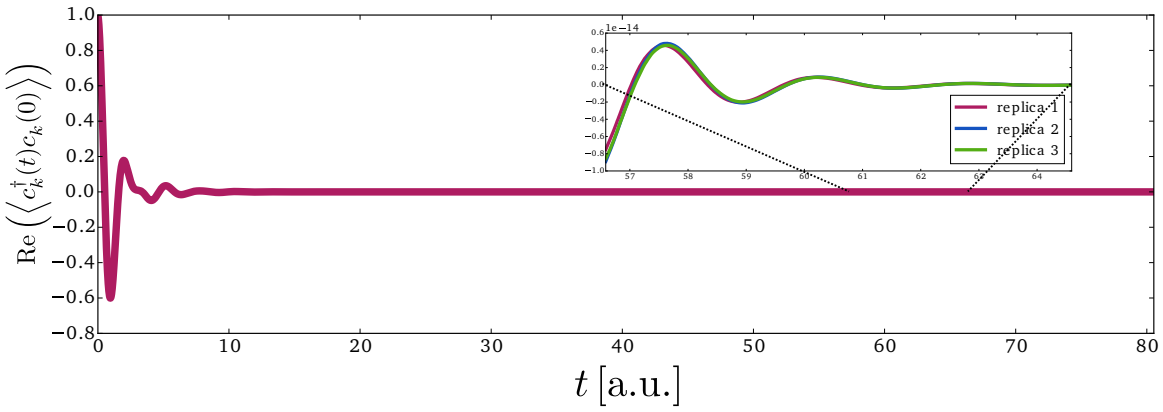


Figure 3.8: Green's function of the 18-site Hubbard model with $U/t = 2$ at $k = (0, 0)$ on an automatically determined time trajectory, using a positive shift of $S = 2.5$. Having such a high value of S allows for very high resolution of the Green's function, at the cost of increased α.

advantageous to have the relevant part of the spectrum residing around $\omega = 0$, as it is most accurate there. The arbitrariness of the energy offset μ in the real time algorithm is useful now, as μ can be set to a suitable value to shift the part of the spectrum which is of interest towards $\omega = 0$. Achieving this is possible by first performing a real time calculation with an initial guess for μ, like the ground state energy, to obtain a first spectrum, which can then be used to estimate an optimized value of μ. Some caution has to be made not to shift the spectrum such that the shifted Hamiltonian $\hat{H}$ obtains negative eigenvalues, as that would revert the effect of the energy dependent damping into an amplification. While shifting the spectrum towards 0 also potentially increases the required values of α, doing so when viable still allows to put focus on specific features of the spectrum, compared to simply increasing the norm of the wave function. Nevertheless, requiring analytic continuation means that details of the spectrum are not necessarily reliably recovered, which is particularly a concern in the examples in section 4.1.

3.3.1 Higher order damping

The complex time rotation is a special form of an energy-dependent damping which can be expressed in an easily accessible form, but other approaches could also prove useful. In cases where a small damping is sufficient, a viable ansatz is a minor damping quadratic in energy, which can efficiently reduce the growth of N_W while introducing

only a very slow decay. Particularly, the special case

$$\mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \mapsto \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t}\mathrm{e}^{-\frac{(\Delta t)^2}{2}\hat{H}^2}\,, \tag{3.36}$$

can be very useful, since it seamlessly integrates into the real-time evolution by modification of the integrator

$$1 - \mathrm{i}\Delta t\hat{H} - \frac{(\Delta t)^2}{2}\hat{H}^2 \mapsto 1 - \mathrm{i}\Delta t\hat{H} - (\Delta t)^2\hat{H}^2\,, \tag{3.37}$$

but a generic prefactor could also be realized as easily, so the approximation can be made systematic. As this efficiently damps higher energy states, long-time stability can be reached as seen in Fig. 3.9, potentially much more efficiently than with a static damping, while achieving a similar effect. As long as the prefactor of the damping exponent is small, the occurrence of the quadratic time step ensures the approximation to the correct evolution is small, such that the damping can be neglected in the transformation to the frequency domain, only resulting in a broadening of the spectra, similar to the static damping discussed in section 3.2.1, see Fig. 4.2. The time-evolution (2.87) is modified in this aspect to

$$|\psi(t)\rangle = \sum_n \langle E_n \,|\, \psi(0)\rangle \, \mathrm{e}^{-\mathrm{i}(E_n - \mathrm{i}\frac{E_n^2\Delta t}{2})t} \, |E_n\rangle \;, \tag{3.38}$$

such that the Fourier transform now reads

$$A(\omega) = \sum_n \langle E_n \,|\, \psi(0)\rangle \, \frac{\pi E_n^2 \Delta t}{\frac{E_n^4(\Delta t)^2}{2} + 2(\omega - E_n)^2}\,, \tag{3.39}$$

translating to an energy-dependent broadening. So, this ansatz ultimately serves the same purpose as the complex time contour, and can in fact be combined, using the complex Δt to reduce the required imaginary time. As it relies on the broadening being small, it is viewed as an auxiliary approach here to improve the stability, though it could potentially be developed into a standalone scheme.

3.4 Initiator adaptation

The initiator method [60] addressed in section 3.4 can be of great use for the real time algorithm, serving the purpose of preventing proliferation of sign errors while at the same time restricting the norm growth by neglecting unimportant spawns. The real time algorithm comes with excellent prerequisites for using the initiator adaptation, as

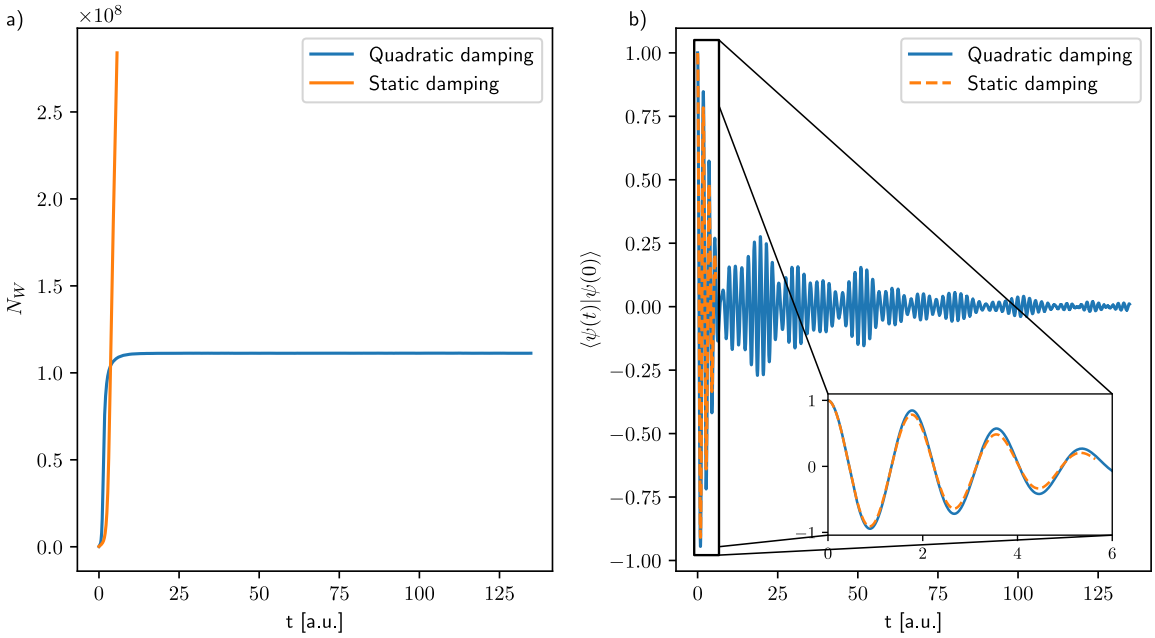


Figure 3.9: a) Norm N_W over time for the damping quadratic in energy and a static damping with $S = 0.08125$, which corresponds to a decay on the same time scales, at the example of a 16-site square lattice Hubbard model with 15 electrons at $k = (0, 0)$. This corresponds to the spectra shown in Fig. 4.2. b) Corresponding Green's function for the two calculations, demonstrating that the effect of the two schemes is comparable.

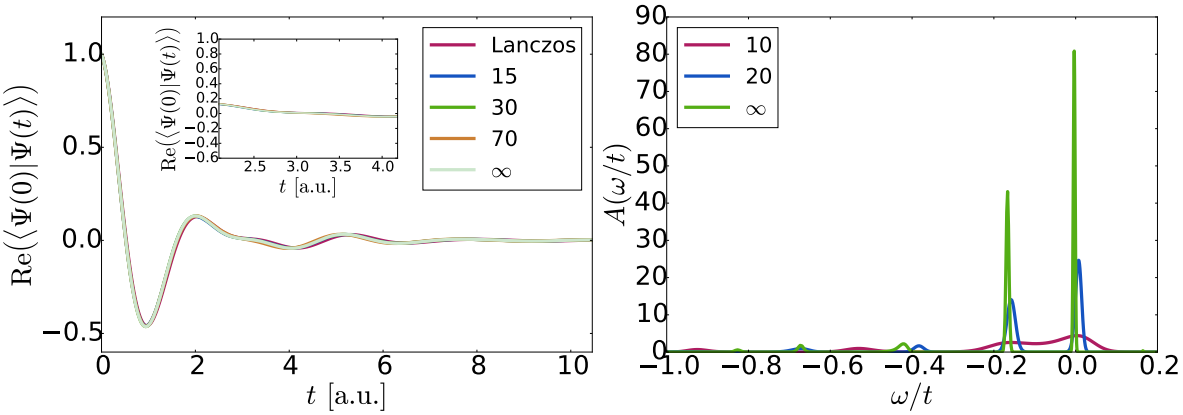


Figure 3.10: a) Green's function of the 18-site tilted Hubbard model at $k = (0, 0)$ and $U/t = 2$ for different initiator thresholds and $\alpha = 0.2$ in comparison to the same function obtained using the Lanczos method. b) Photo-absorption spectra of the Carbon dimer for different initiator thresholds in a cc-pVTZ basis set. Due to different time contours required for lower thresholds, the resulting spectrum is then broader. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), supplementary material, Copyright 2018 American Physical Society.

the initial wave function already contains a large number of occupied determinants and the quantity of interest is the overlap with an already occupied space, which is unaffected

by the approximations made when introducing initiators in first order. That means very high thresholds up to infinity can be used without introducing significant error, as is shown in Fig. 3.10. In the case of the photo-absorption spectrum of C_2, a slight shift in the spectrum can be observed when increasing the initiator threshold, but shifting the entire spectrum has no effect on energy differences and thus does not negatively impact the results. In the following examples, initiator thresholds of 20 up to infinity are thus regularly used. Note however, that an infinite initiator threshold by no means implies a restriction to the initial space, that, in fact, would be insufficient to generate accurate spectra. Instead, new determinants can now only be populated by simultaneous spawn by at least two parent determinants, which is a frequent event.

Chapter 4

Benchmarks of the stochastic time-evolution

After introducing the algorithm itself, the next step is to verify its applicability by studying benchmark systems, where the comparison to available results is possible. Of great interest as an application are model systems on the one hand side, where the calculation of band structures leads the way to characterizing thermal and optical properties, and atomic and small molecular ab-initio systems on the other hand, for which the characterization of excitations can give insight into the nature of high energy states.

First, a prototypical model system, the two-dimensional Hubbard model shall be considered as a benchmark system, using spectra obtained with the Lanczos method and the AFQMC method as benchmarks for the accuracy of the method.

Then, a set of ab-initio systems is considered, where the various techniques for real and complex time evolution are tested and methods for characterizing excitations by means of spectral decomposition are discussed.

4.1 Application to the Hubbard model

A prominent model system in solid state physics describing a whole range of phenomena is the Hubbard model[26], which serves as a simple, but general model for interacting electrons in a crystal and is commonly used to study the interplay of kinetic energy and Coulomb repulsion in the many-body problem [280]. It has been used to describe ferro- and antiferromagnetism, Mott Insulators as well as superconductivity [281, 282, 283], and has been used to study the mechanisms behind high-temperature superconductivity

in Cuprates [284]. The Hamiltonian in real space is given by

$$\hat{H} = -t \sum_{<i,j>,\sigma} \hat{a}^\dagger_{j,\sigma} \hat{a}_{i,\sigma} + \hat{a}^\dagger_{i,\sigma} \hat{a}_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} \,, \tag{4.1}$$

where $\sum_{<i,j>}$ goes over nearest neighbors, using periodic boundary conditions, t is a parameter for the kinetic energy of the electrons and U describes the Coulomb repulsion for electrons at the same location. The only parameter with physical relevance is the ratio U/t, with small values of U/t corresponding to a weakly interacting metal with a band structure close to a tight binding model, while large values of U/t describe a Mott insulator, that is an interaction-driven insulator, in contrast to a band insulator, whose low-energy physics is governed by magnetic excitations [283]. In the limit of large U/t, the Hubbard model can then effectively be described by the Heisenberg model [27, 283, 285], as electrons are pinned to the physical sites, with the kinetic terms being captured perturbatively. The regime of intermediate U/t however, where neither of these approximations allow a perturbative treatment, is particularly challenging, but also insightful for understanding the competition between kinetic energy and Coulomb repulsion [286].

Being translationally invariant, the Hamiltonian is momentum conserving, which is why the momentum space formulation will be used here. In momentum space, the Hamiltonian reads

$$\hat{H} = -2t \sum_{\mathbf{k},\sigma} \epsilon_{\mathbf{k}} n_{\mathbf{k},\sigma} + U \sum_{\mathbf{k},\mathbf{k}',\mathbf{q},\sigma,\sigma'} \hat{a}^\dagger_{\mathbf{k}+\mathbf{q},\sigma} \hat{a}^\dagger_{\mathbf{k}'-\mathbf{q},\sigma'} \hat{a}_{\mathbf{k}',\sigma'} \hat{a}_{\mathbf{k},\sigma} \,, \tag{4.2}$$

with the dispersion relation $\epsilon_{\mathbf{k}}$ which is given by

$$\epsilon_{\mathbf{k}} = \sum_i \cos(\mathbf{k} \cdot \mathbf{a}_i) \,, \tag{4.3}$$

with the lattice vectors $\mathbf{a}_i$ [11]. The lattices used here are square lattices with lattice vectors $\{\mathbf{a}_i\} = \{(4,0),(0,4)\}$ (16-site), $\{\mathbf{a}_i\} = \{(3,3),(3,-3)\}$ (18-site) and $\{\mathbf{a}_i\} = \{(3,3),(-5,3)\}$ (24-site). The specifications of the lattices are given in Appendix B

Despite its simplicity, an analytic solution is only known in special cases [287], and the numerical treatment is generally challenging since strong interactions make a mean-field description difficult and require methods suited for strongly correlated systems while the Fermionic nature makes it challenging for Quantum Monte Carlo methods [286]. The two-dimensional case is of special interest, as it is commonly used to describe the

low energy physics of the Copper-Oxygen planes in Cuprates [284], and is especially challenging to treat numerically, as mean-field descriptions such as dynamical mean field theory become more accurate only in higher dimensions [280], while the tools available in one dimension where the powerful field theoretical method of Bosonization [288] and the density matrix renormalization group algorithm [31] are available. Hence, the two-dimensional Hubbard model shall serve as a first benchmark system to apply the real time FCIQMC.

4.1.1 Case study: The two-dimensional Hubbard model as a benchmark system

For initial benchmarking beyond the simple tests performed for the algorithm design in section 3.1, it is essential to have reference values available. Therefore, as a proof of concept, a tilted half-filled 18-site square lattice momentum space Hubbard model is considered amongst others, where reference values for both the Green's function and the spectral function are available from exact diagonalization using the Lanczos method. The reference Lanczos and AFQMC results were supplied by Olle Gunnarsson, these results have been partly published in Ref. [249], the corresponding figures are marked as such. Note that the particle numbers in this section always refer to those of the unperturbed initial state, the time evolution to obtain the photo-emission Green's function is always carried out with one electron less, or one electron more in the case of inverse photo-emission.

To gauge the accuracy of the time-evolution, the Green's function and the frequency domain analogon, the spectral function, as introduced in section 2.5 can be used. While a direct comparison of the time-evolved wave function is technically not possible in all but the most trivial cases and certainly not for an 18-site Hubbard model, comparing the resulting properties is straightforward. Since these correspond to directly observable properties in the form of (inverse) photo emission spectra, they are also the objects the theory aims to predict, and thus the suited quantities for benchmarking.

Due to momentum conservation, the Green's function $G_{\mathbf{k},\mathbf{k}'}$ is diagonal in $\mathbf{k}$ [22], so the second momentum can be dropped. As a non-trivial example, the Green's function

$$G^{-}_{(0,0)}(\tau) = \left\langle \hat{a}^{\dagger}_{(0,0)}(\tau)\hat{a}_{(0,0)} \right\rangle , \tag{4.4}$$

is considered. Time is denoted τ to distinguish from the Hubbard parameter t. Note that this is not a low-energy property as $k = (0, 0)$ is far below the Fermi surface, thus it

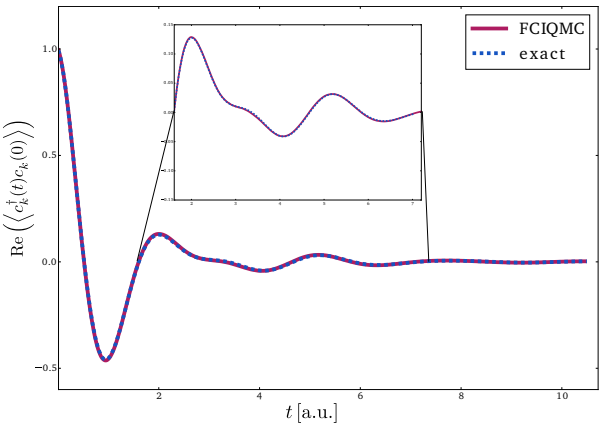


Figure 4.1: Green's function $\left\langle \hat{a}_{\mathbf{k}}^\dagger(t)\hat{a}_{\mathbf{k}} \right\rangle$ of the 18-site Hubbard model for $k = (0,0)$ along the complex time trajectory given by constant $\alpha = 0.2$, calculated using the time-dependent FCIQMC method and exact diagonalization with the Lanczos algorithm. The FCIQMC results are averaged over 4 replicas and basically exactly reproduce the deterministic result.

is inaccessible by perturbative means. In the non-interacting case, the time signal would be

$$G^{-,0}_{(0,0)}(\tau) = \mathrm{e}^{-4ti\tau}\,, \tag{4.5}$$

and the resulting spectrum would thus consist of a single delta function residing at $4t$, which is precisely the single-particle energy of the $k = (0,0)$ state, in the non-interacting case, the interpretation of the spectral function as quasi-particle excitation levels can be transferred to the actual single-particle levels, as quasi-particles and actual electrons coincide in that case. With the introduction of interactions, a splitting of the single delta peak into several quasi-particle levels can be expected as excitonic states emerge.

The interacting Green's function is depicted alongside a complex time contour in Fig. 4.1. The comparison with results obtained using the Lanczos method shows that the real time FCIQMC method can reproduce the Lanczos results, which are highly accurate, essentially exactly. This exact matching of the complex-time Green's function is regularly achieved for the two-dimensional Hubbard model.

Here, the Lanczos Green's functions was calculated by Olle Gunnarsson using the procedure from Ref. [87]. The spectral function is obtained using the Krylov subspace procedure recapitulated in section 2.2, which is first used to calculate the eigenvectors and eigenenergies of $\hat{H}$. Then, the spectral function is calculated using equation (2.90), from which the Green's function is subsequently obtained via equation (2.89), inserting the complex time $\tau = \mathrm{e}^{0.2\mathrm{i}}\, t$. This procedure is equivalent to performing the time-evolution with knowledge of the decomposition of the initial state into eigenstates and employing equation (2.87). That justifies the interpretation as performing the exact time evolution, up to the accuracy of the Lanczos eigenvectors and eigenvalues.

However, as was already discussed in section 2.5.1, the numerically obtained Green's function on a complex contour does not carry the full information of the real frequency

spectral function. In that sense, the last step in the procedure to obtain the Lanczos Green's function, namely evaluating exact time-evolution, removes information, and while the comparison demonstrates the accuracy of the time evolution, the actual information on the spectrum has yet to be extracted. When considering the FCIQMC Green's function, that means that the additional step of analytic continuation is required to regain that information, where a full recovery of the spectral function cannot be expected and the details of the reconstructed spectrum are not necessarily reliable.

Therefore, maximum entropy analytic continuation is used to extract the spectral function from the Green's function time series data. The Green's functions of the individual replicas are used as input for the analytic continuation, with the overlap being calculated with the initial states of any other replica. Due to the finite time range of the calculated Green's function and the exponential decay, the continued spectrum is broadened in comparison to the delta functions of the exact spectrum. For comparison of the weights, a Lorentzian broadening is hence applied to both the continued spectrum and the spectrum calculated via the Lanczos method in order to obtain comparable spectra. This comparison is displayed in Fig. 4.2 for different lattice sizes. Excellent agreement between the Lanczos and FCIQMC determined spectra can be observed for the 18-site case on this scale, while the the finite run-time and thus limited spectral resolution is not sufficient to fully resolve the intricate details of the 16-site spectrum. In contrast to the 18-site example, for the 16-site square lattice, the Hubbard model is an open-shell system, the spectrum is thus more detailed, requiring higher resolution. Here, a damped real-time evolution is feasible, so the spectrum can be obtained both by Fourier transform and MEM analytic continuation. The comparison emphasizes that while the MEM spectrum correctly reconstructs the low-energy part and qualitative aspects of the spectrum, information on the details is lost in the process. Increasing the broadening in Fig. 4.2b) verifies that the spectrum is in qualitative agreement with the Lanczos results with a visible broadening and it can be expected that with longer run-times, the spectrum could be reproduced more faithfully. It is technically not possible for the FCIQMC method to obtain the exact form of the spectrum composed of delta functions, due to the approach via explicit time evolution, but that obstacle is of minor relevance in practice as any observed spectrum always comes with a broadening, caused by a multitude of effects such as Doppler broadening, inhomogenous broadening or simply the natural line width [289]. Note that for the 16-site and 18-site Hubbard models, the photo-emission spectra are displayed with a mirrored frequency axis, effectively showing $A^{-}(-\omega)$ from equation 2.91. The displayed function can then be interpreted as giving the energy that

is required to create a specific quasi-particle excitation.

The method is not limited to a momentum space treatment, and could also be applied to the real-space version of the Hubbard model, too, which is more useful at higher values of U/t, where the inter-site hopping can be treated more and more perturbatively. Then, translational symmetry of the model translates to a dependence of the Green's function G_{ij} only on the distance Δr between the two sites i and j, which is equivalent to the Green's function being diagonal in momentum space since the real space and momentum space Green's functions are related by Fourier transform. In conclusion, every real space Green's function contains contributions from all momenta.

Analogously, the spectra for all momenta can be obtained to yield the full dispersion relation $A(\omega, \mathbf{k})$, which is equivalent to knowledge of the real space Green's function for all distances and encodes the entire quasi-particle excitation spectrum, thus, the band structure. The momentum-resolved spectra for the 18-site cluster with $U/t = 1$, where a band structure is still meaningful as the interactions are not dominant yet, is given in Fig. 4.3. The spectral function displays the expected time reversal symmetry, that is $A(\omega, (k_x, k_y)) = A(\omega, (\pm k_x, \pm k_y))$. Notable are especially the spectra of $k = (\pm 1, \pm 1)$ and $k = (\pm 1, 0)/k = (0, \pm 1)$, which correspond to the lowest energy quasi-particles and show very high agreement with the quasi-particle levels predicted by mean-field theory at $\omega = 0.5$ and $\omega = 1.5$, while the high energy part of the spectrum clearly goes beyond mean-field theory, which predicts one distinguished quasi-particle level per momentum [290]. Also note that the entire part of the spectrum above the Fermi surface, i.e. with momenta above $k = (\pm 1, \pm 1)$, is a purely multiconfigurational effect, in a single configurational picture, A^- would have no weight at these momenta as they are unoccupied in the mean-field determinant [11]. The ability to treat each momentum on a separate basis is a great advantage of the real time FCIQMC approach, as this allows to capture these high energy parts of the spectrum on the same level of accuracy, even though they carry less weight than the mean-field part.

The systematic study of the same cluster at $U/t = 2$ for different walker numbers displayed in Fig. 4.6 confirms that the trajectory automatically chosen according to equation (3.34) moves closer to the real axis the higher the walker number N_w is, leading to more detailed time-signals $\langle \psi(0) \,|\, \psi(t) \rangle$ that decay slower and carry more information. As established by Gunnarsson [249], the maximum entropy analytic continuation resolves more details the closer to the real axis the time trajectory is. Note that details of the spectrum can depend on regularization schemes, see Fig. 4.5, meaning there is an additional uncertainty in the spectra as indicated in Fig. 4.5. From the resulting

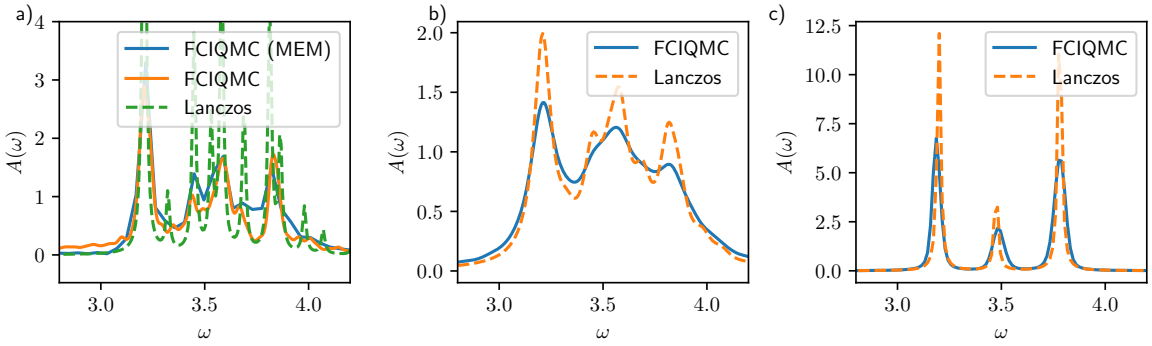


Figure 4.2: Photo-emission spectral function A^- of a half-filled square lattice Hubbard model at $U/t = 1$, with a,b) 16-sites and c) 18-sites, comparing the $\mathbf{k} = (0, 0)$ photo-emission spectral function obtained using FCIQMC and the Lanczos method. A Lorentzian broadening with FWHM of 0.02 is applied to the MEM spectra in a) and c) in order to create comparable spectra while b) shows the same MEM spectrum as a), but with a broadening of 0.1. The spectrum in c) corresponds to the $k = (0, 0)$ case in Fig 4.3. While the 18-side mode uses a complex time contour, the 16-site spectrum is calculated entirely for real times, so both the Fourier transformed spectrum and the MEM spectrum are given. Both additionally make use of the quadratic damping discussed in section 3.3.1.

real-frequency spectra, two things can be observed. First, as already mentioned, the spectra become sharper the higher the walker number, which can be attributed to the reduced values of α. Second, there is a small systematic error in the position of the peaks in the spectrum associated with smaller walker numbers, which is unsurprising as imaginary-time FCIQMC too has to converge energies with respect to walker number, so a dependency on the walker number that gradually vanishes at high N_w can be expected, which then manifests as an error in the Green's function. All in all, the Lanczos spectrum can be reproduced accurately with comparably small effort, particularly in terms of memory, while also drawing on the high parallel efficiency of FCIQMC which enables larger calculations such as for the 24-site model discussed below.

The Hubbard model at intermediate interactions U/t is particularly challenging, so it is also unsurprising that the same system with $U/t = 4$ becomes more challenging and requires more walkers to qualitatively recover the spectrum, see therefore Fig. 4.4 which shows the $k = (0, 0)$ inverse photo-emission spectrum of that system obtained with $N_w = 1.4 \cdot 10^8$. While the integrated weight agrees reasonably well and the energy distribution of the spectrum is recovered, the details cannot be resolved as they would require a much finer frequency resolution, which in turn requires smaller values of α and therefore larger N_w.

On the other hand, FCIQMC does not suffer from the severe memory bottleneck affecting deterministic exact diagonalization techniques, such that accurate calculation

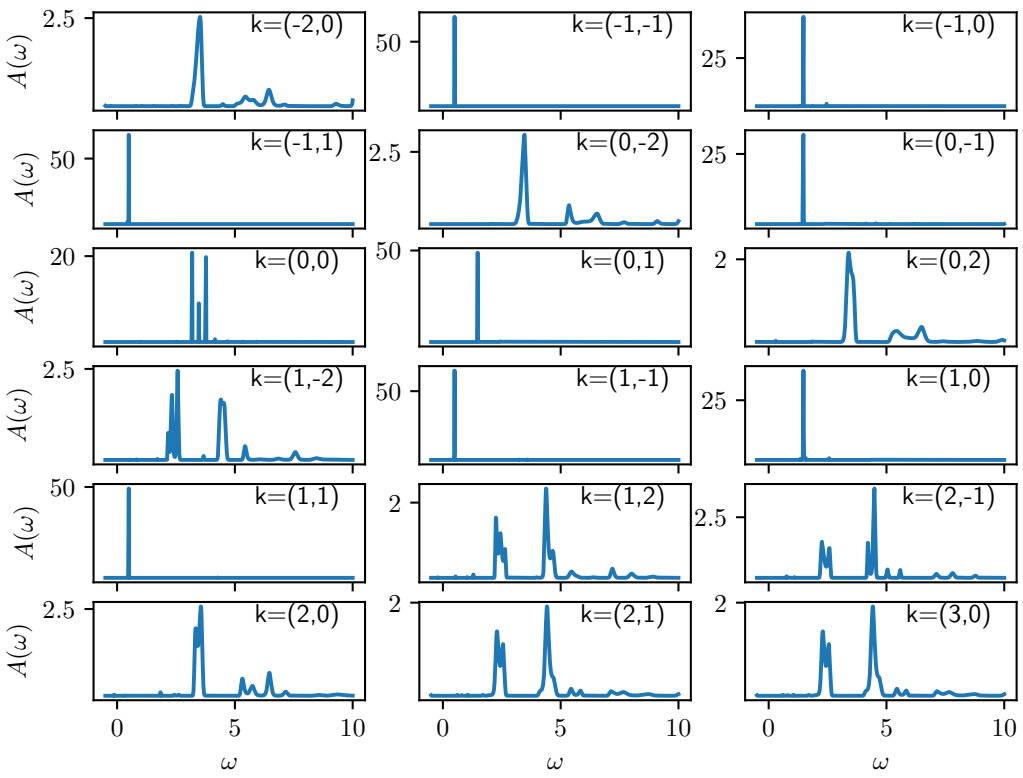


Figure 4.3: Full spectral function $A^-(\omega, \mathbf{k})$ of the 18-site Hubbard model with $U/t = 1$ and spin $-1/2$ calculated by real time FCIQMC. Momenta are given in units of $2\frac{\pi}{L_{x,y}}$ with $L_{x,y} = 5, 3$, respectively. The Fermi surface is between $k = (\pm 1, \pm 1)$ and $k = (\pm 1, \pm 2)$, and being photo-emission spectra, the spectra with momenta below the Fermi surface carry the main weight. These calculations use complex-time evolution in combination with quadratic damping as described in section 3.3.1.

of these spectra is possible with much less than 20 GB of distributed (!) memory, and single replicas can be carried out with as little as 800 MB, more than 70 times less than the requirements of the Lanczos algorithm. In addition, the high parallelizability of the FCIQCM algorithm can shine here, allowing to scale up to ten thousands of processors in parallel. Since the real time FCIQMC has only negligible requirements for shared memory, and scaling to that many processors also comes with a proportional scaling of the accessible RAM, memory is a non-issue here. The shared memory requirements of other FCIQMC extensions will become a topic in Chapter 6, as these can make the FCIQMC method memory-bound.

Thus, Hubbard clusters beyond the reach of deterministic exact diagonalization methods are in reach for this algorithm, as is shown by the example of a doped 24-site square

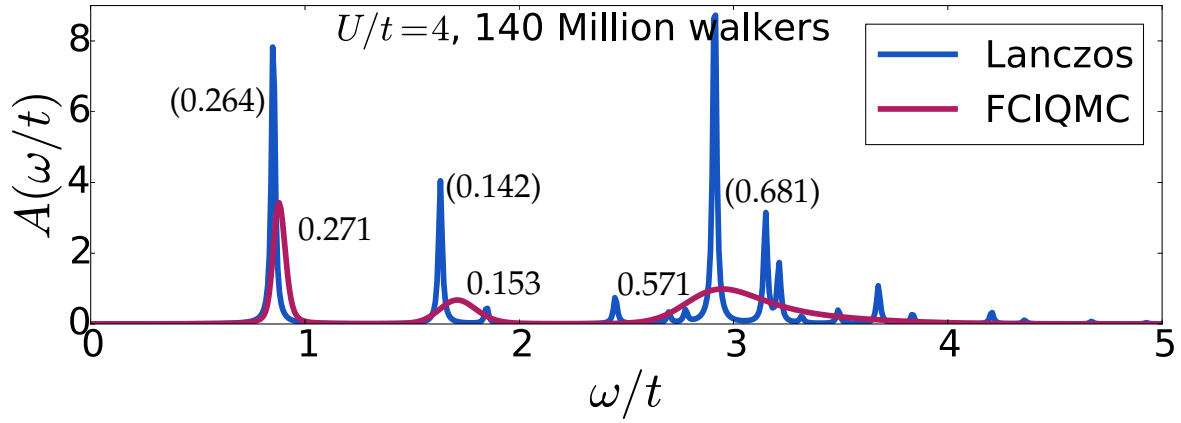


Figure 4.4: Photo-emission spectrum of a half filled 18-site Hubbard cluster at $U/t = 4$ for $k = (0, 0)$. Both spectra have been broadened by convolution with a Lorentzian of FWHM 0.02. The integrated weights of the FCIQMC spectrum are given, alongside the weights of the peaks of the Lanczos spectrum.

lattice cluster with lattice vectors (3,3) and (-5,3), a closed-shell system with a Hilbert space of substantial size. For this lattice, the photo-emission and inverse photo-emission spectra at $U/t = 4$ are displayed in Fig. 4.6, compared with a spectrum of the same system obtained by Olle Gunnarsson using the Hirsch-Fye AFQMC method [183]. While the spectra are consistent, the AFQMC method cannot resolve peaks which are accessible for the new FCIQMC method. The non-constrained AFQMC was used,which is exact but suffers from a sign problem [19, 291, 187], hence the calculation was carried out at finite temperature $T = 0.2t$ to make it affordable.

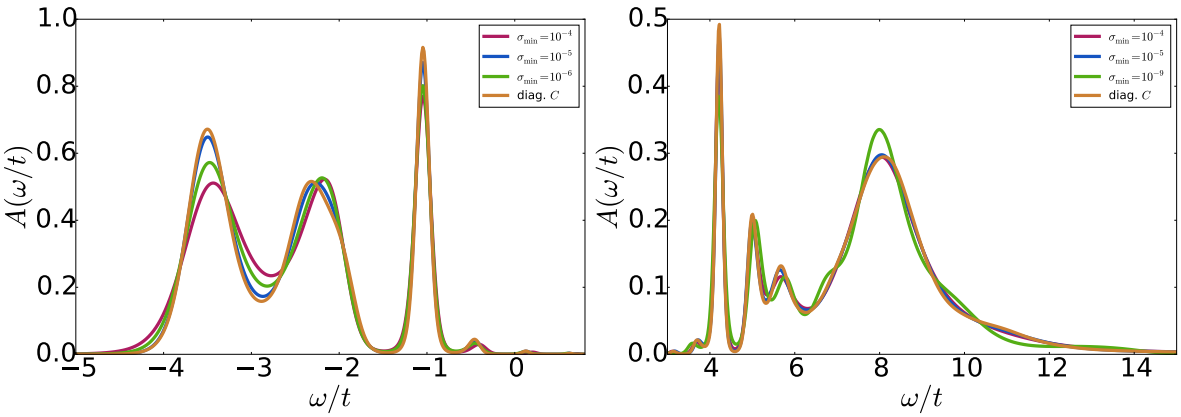


Figure 4.5: MEM analytic continuation of FCIQMC photo-emission (left) and inverse photo-emission (right) spectra of a 24-site Hubbard model with 22 electrons for different values of $\sigma_{\min}$ and neglecting off-diagonal C. A cutoff of $\sigma_{\min} = 10^{-9}$ corresponds to no approximation but yields a singular C, a continued photoemission spectrum could not be obtained in this case. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), Copyright 2018 American Physical Society.

Three reasons for the higher resolution of the FCIQMC method have been identified by Olle Gunnarsson in Ref.[249]. First, the FCIQMC method works on a complex time trajectory close to the real axis, while AFQMC works in a pure imaginary time framework, such that the analytic continuation can resolve less details of the spectrum, especially at higher energies. Second, in contrast to AFQMC, the FCIQMC method allows for an arbitrary choice of the zero of energy. This is crucial here, as the transformation between time-dependent Green's function and frequency dependent spectral function is invertible at $\omega = 0$ and no loss of information is caused by the analytic continuation, allowing for a more accurate resolution of the spectrum simply by shifting it towards 0. Third, the new method allows for separate calculation of inverse photo-emission and photo emission spectrum, which enables it to efficiently treat cases like the spectrum shown in Fig. 4.6, where the two carry vastly different weight. The corresponding calculations are then carried out using independent states at the same level of accuracy, and there is no difficulty arising in resolving a small weight on the inverse photo-emission spectrum, while, if both are addressed in the same calculation as in AFQMC, it becomes difficult to extract the small weight fraction of the inverse photo-emission spectrum.

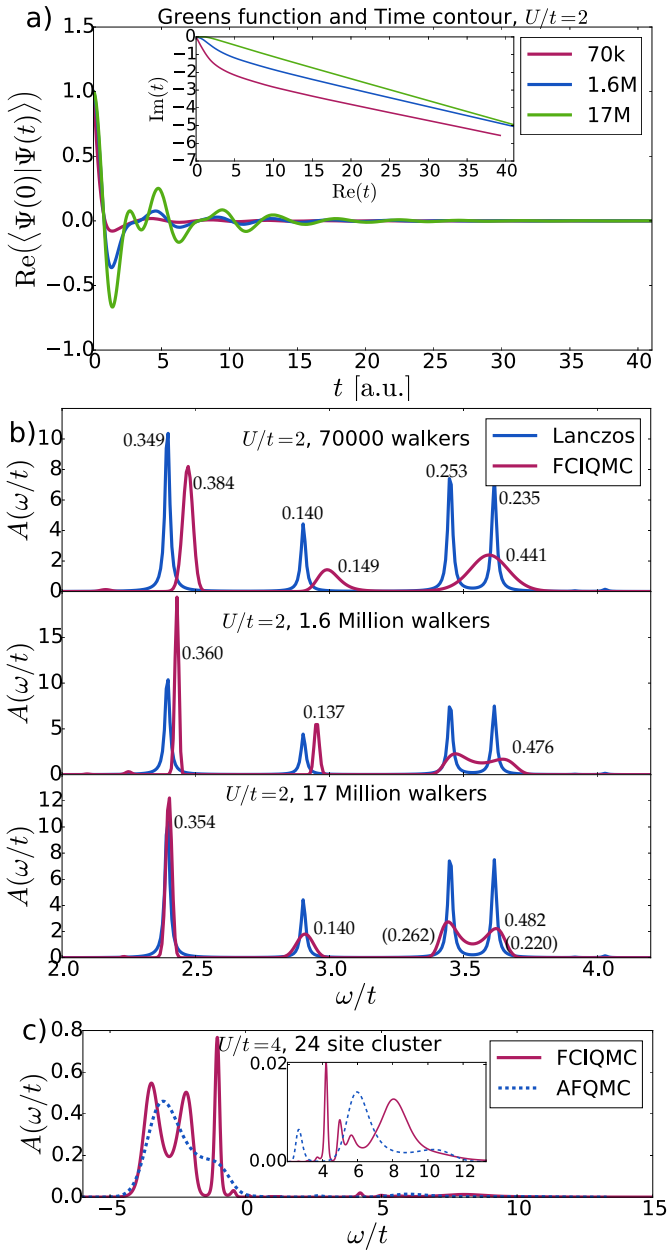


Figure 4.6: Green's and spectral functions of 18-site and 24-site square lattice Hubbard model at $U/t = 2$ and $U/t = 4$ respectively. a) Real part of the Green's function $\langle \hat{a}^{\dagger}_{(0,0)}(t)\hat{a}_{(0,0)}\rangle$ of a half-filled 18-site Hubbard model at $U/t = 2$ for different FCIQMC populations with corresponding complex-time trajectories. The Green's function is averaged over 5 replicas. Higher populations regularly allow for smaller values of α in the time rotation. The initial state is the same in the three cases. b) Inverse photo-emission spectra corresponding to the Green's functions in a) in comparison with the reference spectra obtained using the Lanczos method. A Lorentzian broadening with FWHM 0.02 is applied to all spectra for comparability. The integrated weights of the resonances are given with each peak, the weights of the Lanczos spectrum are indicated in the first figure in brackets, showing good agreement. c) Photo-emission and inverse photo-emission spectra for a 24-site tilted square lattice Hubbard model with 22 electrons[a] in comparison to the same property calculated with AFQMC with $U/t = 4$ at $k = (0, 0)$. The inverse photo-emission spectrum is given in the inlay, as it only carries little weight. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), Copyright 2018 American Physical Society.

[a] I.e. the time-evolution is carried out with 21 and 23 electrons, respectively

4.2 Application to ab-initio systems

After seeing the time-dependent FCIQMC method applied to a model system, the next step will be to take it to ab-initio systems, where slightly different requirements are in place. In addition of the spectral functions A that represent quasi-particle excitation spectra and are mainly of interest in the solid state, the spectral decomposition of excitations is of prime interest, reproducing photo-absorption spectra. On a number of systems, the applicability of the method to ab-initio Hamiltonians shall be demonstrated here. The examples of the Beryllium and Carbon atoms and the Carbon dimer all use Hartree-Fock orbitals calculated with `Molpro` [292] using correlation consistent Gaussian basis sets [293, 294], while the active spaces considered for the Cu_2O cluster in section 4.2.4 were provided by Nikolay Bogdanov using PySCF [2].

4.2.1 Case study: The Beryllium atom

The Beryllium atom shall serve as a first example to study the applicability of the method to ab initio systems. The basis set incompleteness error is less severe for an atom of this size, the comparison with experimental data can be readily made and the system is small enough to be treated by the pure real time algorithm without the need for the introduction of a complex time rotation, in these examples, the quadratic damping discussed in section 3.3.1 is used. Studying such a system is essential as a proof of concept for the algorithm, since the sparse representation of the wave function raises the question if accurate calculations are possible at all with the available computational resources.

When running several replicas in parallel and averaging the resulting signal over them, one of the best measures of the accuracy of the evolution is the coherence of the replicas. The validity of the arguments for the reliability of the instantaneous overlap brought forward in section 3.1 at a given scale can be directly checked from the coherence of the replicas. If independent calculations yield the same overlaps over extended time, the description of a single population is sufficiently accurate to warrant a stable time evolution with no serious stochastic errors introduced into the overlap by the stochastic sampling, as the result does not depend on the concrete sampling. If however the overlap starts to differ between replicas at some point, it is a sign that stochastic errors are taking over and the results can no longer be trusted. The coherence of the replicas, in fact, is an important estimate for the stochastic error introduced along the time evolution. When determining the time up to which the overlap can be used to extract a spectral function,

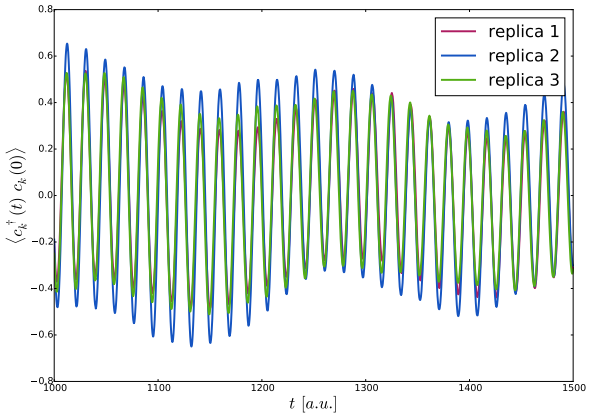


Figure 4.7: Overlap with the initial state for three independent, parallel time evolutions of an excited state of the Beryllium atom in a cc-pVDZ basis set. Even after extremely long times, the replicas still show coherence, demonstrating the stability of the algorithm.

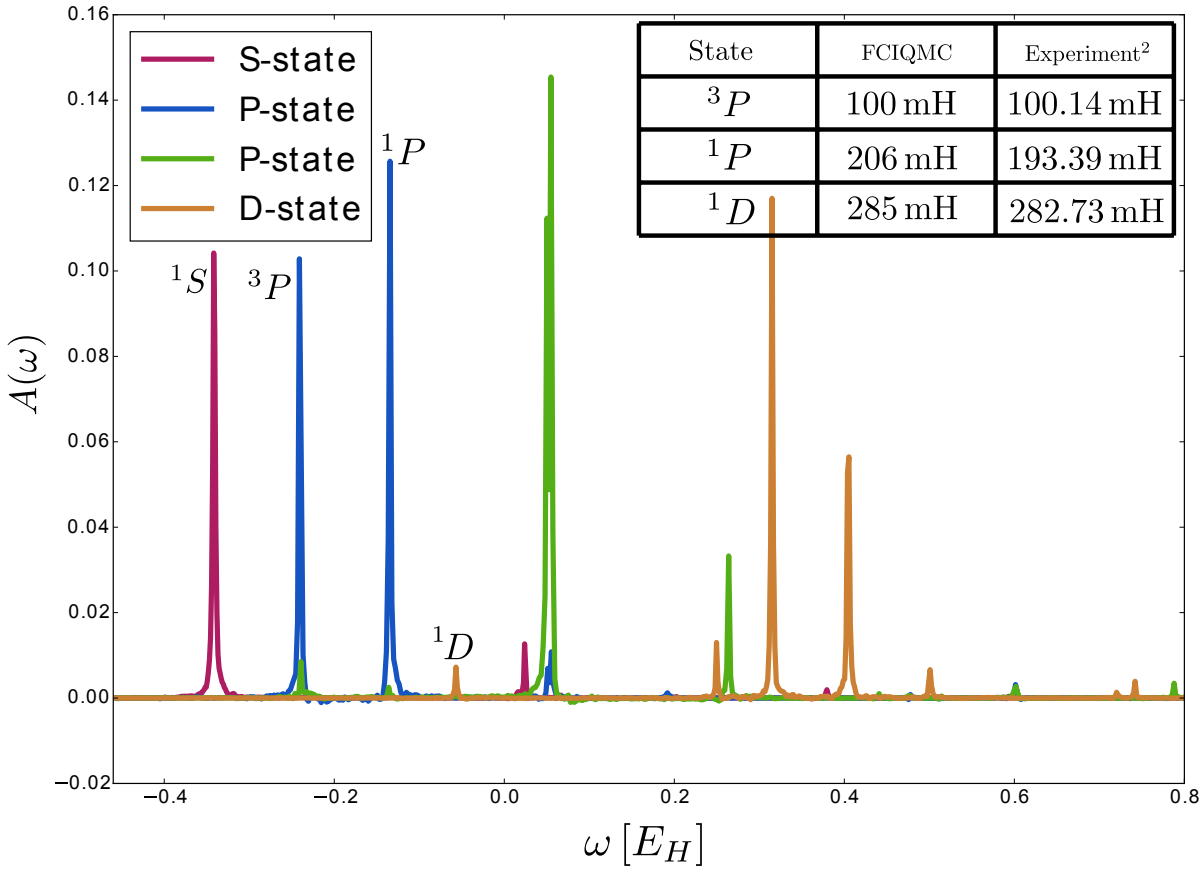
the loss of coherence of replicas marks the point when the overlap loses its meaning, as it is then depending on the path taken by the stochastic processes instead of the actual propagator. Remarkably, independent replicas can stay coherent for surprisingly long times, as is displayed in Fig. 4.7, which impressively shows the validity of the stochastic evaluation of the propagator in principle.

One possible way of preparing the initial state is the basic technique used in the Green's function calculations of the Hubbard model. The initial state is then created to probe the inverse photo-emission of the cation

$$|\psi(0)\rangle = \hat{a}_i^\dagger \, |\psi_0^{N-1}\rangle \ , \tag{4.6}$$

with a suited creation operator $\hat{a}_i^\dagger$ to populate an orbital of the desired type. It is well known for the Beryllium atom to have low-energy excited states of P and D characteristics [295], so adding the electrons to the $2p$ orbitals can be expected to result in states with significant contributions from these low-energy P eigenstates of spectroscopic relevance, the D states however are problematic for this approach, since the D lowest state has $1s^2 2p^2$ character [295]. While, thanks to the multiconfigurational character of the created initial state, it is still possible as seen in Fig. 4.8, more flexible methods should be pursued. Besides the option to start from a single determinant, applying excitation operators on a previously calculated FCIQMC wave function, of which Green's functions are a special case, can be generalized to a flexible tool, and gives a clear intuition of what to expect from the spectral decomposition of the initial state. These different approaches are compared in section 4.2.3

Since applying excitation operators even to spin-pure states creates states of mixed spin, the resulting spectral decompositions can be used to calculate spin gaps, as the



State	FCIQMC	Experiment[2]
^3P	100 mH	100.14 mH
^1P	206 mH	193.39 mH
^1D	285 mH	282.73 mH

Figure 4.8: Spectral decomposition of excited states of the Beryllium atom, displaying low-energy P and D states, calculated with pure real time FCIQMC in a cc-pVDZ basis set. The time evolutions started from the cation ground state with an electron added to a p or d orbital, corresponding to inverse photo-emission. Time evolution has been carried out for 3000 atomic units of time. The experimental data for comparison is taken from [295].

initial state will then consist of the different spin states of a given excitation, each of which appears as a different peak in the resulting spectrum. This can be readily seen for the P states of Beryllium in Fig. 4.8, the P states of the Carbon atom in Fig. 4.10 or the single-excitation of the Carbon dimer in Fig. 4.11.

To identify the individual spin states, they can be subsequently eliminated from the spectrum by preparing initial states with increased M_s quantum number by introducing spin flips in the initial state preparation. Doing so eliminates all contributions from lower spin eigenstates, only leaving the higher spin states in the spectrum. For example, preparing the P state of Beryllium considered above with $M_s = 1$ removes one of the peaks, which can therefore be identified as the $S = 0$ state, this is displayed in Fig. 4.9.

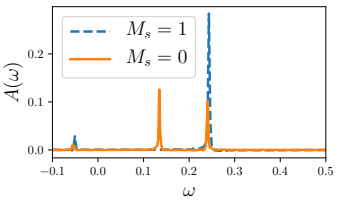


Figure 4.9: Spectral decomposition of a spin mixed P state and a ${}^3P$ state of the Beryllium atom, the latter prepared with $M_s = 1$. The spin mixed state is the same as used in Fig. 4.8, both are calculated using the pure real time FCIQMC method in a cc-pVDZ basis set. From the comparison of these spectra, the higher state can be identified as the triplet state, the lower one as the singlet state.

4.2.2 Case study: The Carbon atom

When increasing the system size, a pure real time evolution becomes less and less feasible. However, to some extent, it can still be made possible by applying a static damping that is sufficiently small not to adversely affect the resolution of the spectrum as introduced in section 3.2.1. As an example to consider damped real time evolution over extended times, the carbon atom is chosen, being a well studied system with readily available spectroscopic reference data from experimental studies. To observe the atomic multiplet, it is sufficient to perform two real time calculations, whose results are presented in Fig. 4.10. Both use an initial state representing an inverse photo-emission process, starting from the cation ground state calculated with imaginary time FCIQMC, then adding an electron to a p orbital. Adding the electron to the occupied p orbital inadvertently leads to a singlet state, which can be expected to mix an S and D state in a ratio of 1:2 from angular momentum addition [10, 296]. In fact, the resulting spectral decomposition in Fig. 4.10 reproduces this very accurately, with some deviation due to the multiconfigurational treatment. If the electron is added to an unoccupied p orbital, the resulting state has $L_z \neq 0$ and thus no contribution of S nature, but will both have triplet and singlet admixture. Assuming the cation state to have $m_s = \frac{1}{2}$, in a single configurational picture, the p electrons will be described e.g. as

$$|p_z(\alpha)p_x(\beta)\rangle = \underbrace{\frac{1}{2}\left(|p_z(\alpha)p_x(\beta)\rangle + |p_z(\beta)p_x(\alpha)\rangle\right)}_{{}^3P} + \underbrace{\frac{1}{2}\left(|p_z(\alpha)p_x(\beta)\rangle - |p_z(\beta)p_x(\alpha)\rangle\right)}_{{}^1D}, \tag{4.7}$$

assuming the electron is added in the p_x orbital. Analogous states would result had the new electron be put into the p_y orbital. The first state can be identified as a triplet P state, while the second can be identified as a singlet D state, and should have the same energy as the singlet D state observed with the other initial state. This is indeed observed in Fig. 4.10, while the two states do notably carry different weight, implying the actual states have a more multiconfigurational character. The excitation energies themselves are then determined by a local Lorentzian fit to the spectrum, carried out

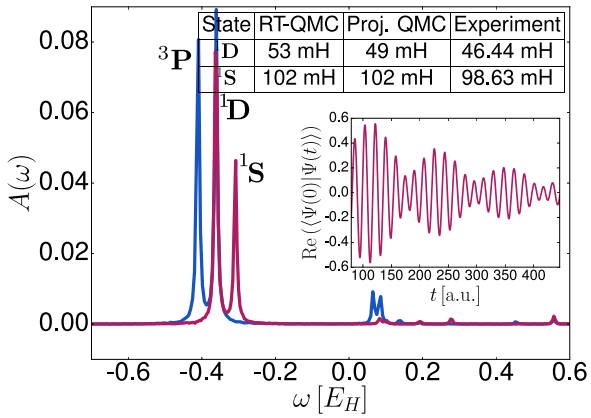


Figure 4.10: Spectral decomposition of two different initial states both created by addition of a p electron to the cation ground state obtained by real time FCIQMC using a cc-pVTZ basis set.The $1s$ core electrons have been frozen. For the first initial state, the electron has been added to an unoccupied p orbital (blue), for the second to the occupied (red). The time evolution has been carried out for 1600 atomic units of time, yielding a frequency resolution of 3.9 mH, and the axis is zeroed at -37.3706 H, the cation ground state energy calculated using imaginary time FCIQMC. The experimental values are taken from [297]. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), Copyright 2018 American Physical Society.

using the Matlab software [279].

4.2.3 Case study: The Carbon dimer

The two following examples serve to show the applicability to more involved systems which are well beyond the reach of deterministic methods or a pure real time evolution, requiring running FCIQMC calculations along a complex contour. First, consider the Carbon dimer at equilibrium distance, a typical strongly correlated molecule [298], where different methods of preparing the initial state and different basis sets are compared. The equilibrium distance is taken from Ref [299]. Spectroscopic data is readily available [300], which can be used to characterize the observed excited states and gauge the accuracy of the method. Consider the excited states $^3\Sigma_u^+$, $^1\Sigma_u^+$, $^1\Pi_u$ and $^3\Pi_u$, whose energies are to be calculated and the corresponding excitations characterized using complex-time FCIQMC. Here, the 1s core electrons are treated as frozen as no 1s core excitations are considered. These results have been previously published in Ref. [249]. As all of these states can be targeted by imaginary time FCIQMC excited state calculations, these are additionally performed for comparison.

To obtain the spectra describing the Σ_u and Π_u excited states, first, the $^1\Sigma_g^+$ ground state is calculated using imaginary time FCIQMC, and the initial state for the time evolution is prepared by applying a single excitation operator $\hat{a}_i^\dagger \hat{a}_j$ corresponding to a $2\sigma_u \rightarrow 3\sigma_g$ or $1\pi_u \rightarrow 3\sigma_g$ excitation, respectively, simulating photo-absorption. As single excitation operators are non-spin conserving, the resulting states are a mixture

of both singlet and triplet, even though the original ground state is a singlet. In the context of real time FCIQMC, this is an advantage, as a single calculation can now capture both the singlet and the triplet state. The calculation is carried out along a complex time trajectory with values of $\alpha \in [0.1, 0.2]$, which allows to recover the real frequency spectra with good resolution. The excitation energies are again obtained by locally fitting a Lorentzian to the spectrum.

In addition, (inverse) photo-emission spectra are obtained for a cc-pVTZ basis set by calculating the ground state of the anion (cation), then removing (adding) an electron from the $1\pi_u$ (to the $3\sigma_g$) orbital. The resulting spectra have weight at the same excited states as the photo-absorption spectra, but the stochastic error is significantly lower for the inverse photo-emission spectra compared to the photo-emission. A careful choice of the initial state can then help to increase the accuracy of the results, so it is useful to consider various techniques to prepare initial states targeting a given set of excited states.

The resulting energies spectra and energies in Fig. 4.11 show reasonable agreement, the basis set dependence is in the range expected from previous FCIQMC studies of relative energies [301, 230, 298], while an overestimation of the excitation energies compared to experiment can be explained from the neglect of changes in geometry between ground state and excited states, which is known to be significant [86]. Note also that the pVQZ basis set is required to correctly capture the order of the states with this method. The energies are obtained from the spectra using a local Lorentzian fit performed with the Matlab software [279].

4.2.4 Case study: Cu_2O cluster

Last, consider the Cu_2O cluster of Sr_2CuO_3, a representative of the Cuprates and antiferromagnetic quasi low-dimensional materials [302]. The copper-oxygen planes are believed to be responsible for the antiferromagnetism and are commonly studied using the Hubbard model, being a good example of a strongly correlated system [303, 304, 305]. The strong electron-electron correlation induces an antiferromagnetic spin-spin coupling that can be described by the Heisenberg model [70, 306, 307]. This antiferromagnetic coupling is known to involve virtual electron transfer between the copper 3d and the oxygen 2p orbitals [70, 308]. The role of charge transfer excitations depending on the used active space is considered here, for an active space containing only the $2p_y$, $3p_y$ orbitals of oxygen and the $3d_{z^2}$, $3d_{x^2-y^2}$, $4d_{z^2}$ and $4d_{x^2-y^2}$ orbitals of copper, as well as a larger active space containing all 2p and 3p orbitals of the oxygen and all 3d and

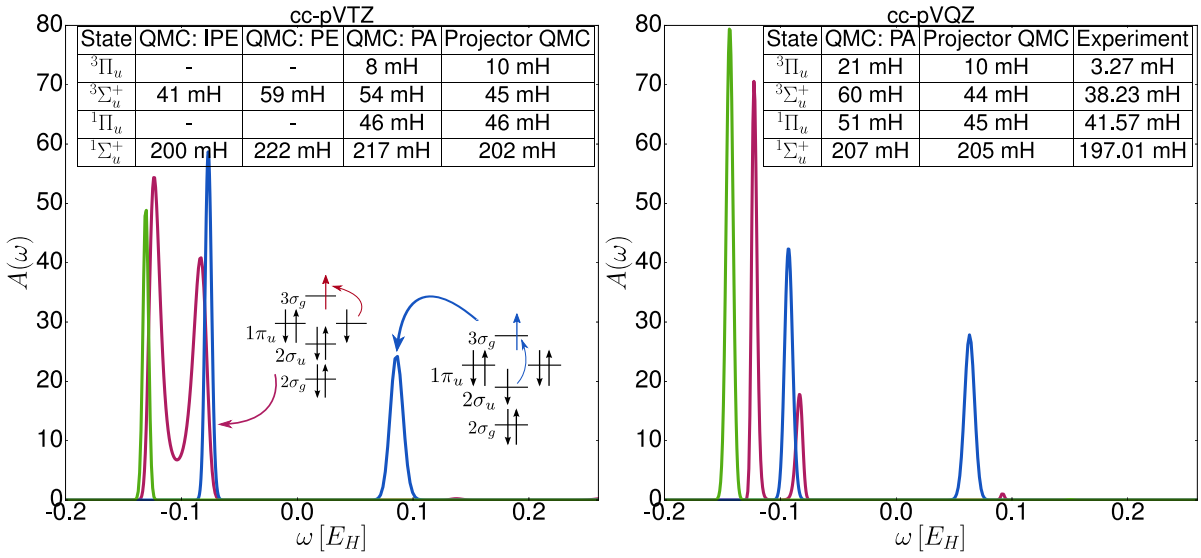


Figure 4.11: Photo-absorption spectra of C_2 in a cc-pVTZ (left) and cc-pVQZ (right) basis set corresponding to a single excitation from $2\sigma_u$ to $3\sigma_g$ (blue) and from $1\pi_u$ to $3\sigma_g$ (red), probing selected excited states. To obtain excitation energies, the FCIQMC ground state's spectral decomposition is also shown in green, with the peak corresponding to the ground state energy. The energy axis is zeroed to $-75.649\,\mathrm{H}$. The same excited states have also been probed for the pVTZ basis set using initial states corresponding to (inverse) photo-emission from the anion (cation) for comparison, and imaginary time FCIQMC excited state energies are given in addition. States are characterized according to and energies compared to the experimental data from Ref. [300]. Reproduced from K. Guther *et al.*, Phys. Rev. Lett. 121, 056401 (2018), Copyright 2018 American Physical Society.

4d Orbitals of the coppers. The active spaces and basis functions used here were supplied by Nikolay Bogdanov. The dominant peaks in the photo-absorption spectra can be interpreted as the excitation energies required to transfer electrons between orbitals, thus giving a notion of relative orbital energies in a truly multiconfigurational picture. These excitation spectra are given in Fig. 4.12 for the active space of 24 electrons in 26 orbitals.

Next, compare the excitations between the two $3d_{x^2-y^2}$ (which are the HOMO/LUMO) of the coppers and the $2p_y$ of the oxygen between a small active space of 8 electrons in 10 orbitals and the larger active space of 24 electrons in 26 orbitals. While the $2p \rightarrow 3d$ excitations are unaffected by the inclusion of additional basis functions, the $3d \rightarrow 3d$ excitation feels a significant impact from the additional basis functions. The states considered are all of the form $\hat{a}_i^\dagger \hat{a}_j \left|\psi_0\right\rangle$, with $\left|\psi_0\right\rangle$ being the ground state as obtained by imaginary time FCIQMC calculations. By applying the real time FCIQMC method on a complex time contour to calculate spectral decompositions of these states, the charge transfer process between oxygen and copper can be quantified, the resulting spectra are shown in Fig. 4.13.

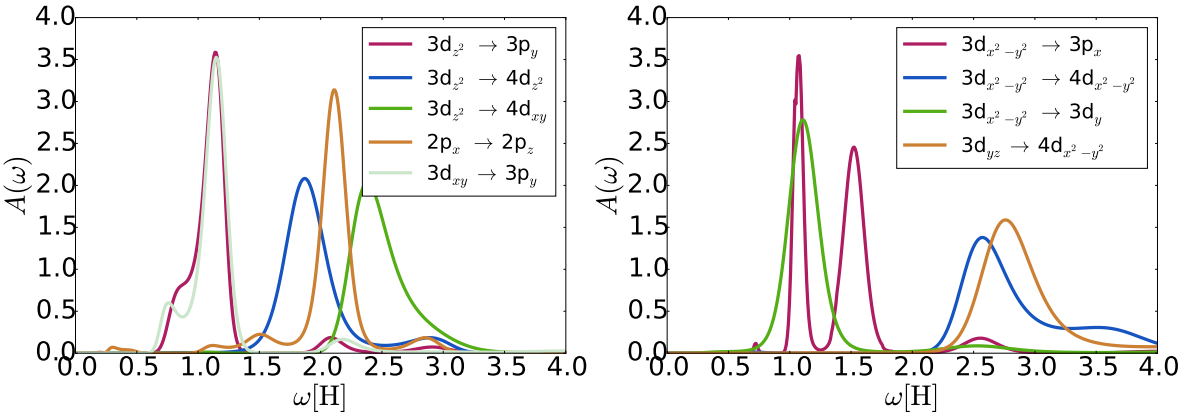


Figure 4.12: Photo-absorption spectra for the Cu_2O cluster of Sr_2CuO_3 in a (24,26) active space for various excitations between copper d and oxygen p orbitals, revealing that most of the considered excitations are high-energy excitations.

4.2.5 A note on performance

From a computational point of view, the core components of the complex time evolution with FCIQMC are essentially the same as for imaginary time FCIQMC, as most of the extra steps are extremely cheap. Profiling of a minimal example of the Beryllium atom in a cc-pVDZ basis set with $N_w = 1.5 \cdot 10^4$ and $n_{\text{rep}} = 5$, where n_{rep} is the number of replicas on four CPU cores of an Intel i7-4790 CPU@3.6 GHz, for the original imaginary time FCIQMC and real-time FCIQMC shows that in both cases, the spawn step require the most resources, accounting for 48% and 57% of the total iteration time excluding initialization, with the rest mainly accounting for communication and annihilation, only 8% of the iteration time is spent on time-dependent specific operations. This is surely only a minimal example and the precise numbers will not generally hold, but it gives an impression of how the cost of real- and complex-time iterations constitute.

When disregarding the specific operations, the overhead of a naive implementation versus single-replica i-FCIQMC is therefore a factor of $4 \cdot n_{\text{rep}}$, the factor of 4 stemming from the second order integrator and the complex coefficients, each contributing a factor of 2. In practice, the `NECI` implementation can optimize the replica and complex coefficient overhead to some degree by performing certain operations collectively for all replicas[1], at the cost of nonlinear increase of the required memory. In the previously visited example of C_2 with $N_w = 1.8 \cdot 10^7$ and $\Delta\tau = 0.001$, the time per iteration thus only increases by a factor of ~ 6.5 from single-replica i-FCIQMC to $n_{\text{rep}} = 4$ time evolution.

[1]This is a general, previously existing optimization in `NECI` and not newly introduced in the context of complex-time evolutions

The overall cost in terms of CPU hours is difficult to compare, since the resulting properties are not equivalent, so a concrete direct comparison thereof under controlled conditions would be extremely involved. It can however be estimated that, depending on the desired accuracy, the time-dependent approach can be much more expensive to converge. To get a rough estimate, consider the real-time calculations used for Carbon in Fig. 4.10 which used almost 60 times more CPU time than the corresponding excited state calculations. This already includes the previously discussed overhead between $n_{\mathrm{rep}} = 5$ of the time-dependent and $n_{\mathrm{rep}} = 3$ of the excited states, and much shorter calculations might be feasible with larger time steps or less replicas, so it is not a definite overhead. Furthermore, acquiring additional samples of the Green's function is embarrassingly parallel, such that even large numbers of samples can be calculated efficiently.

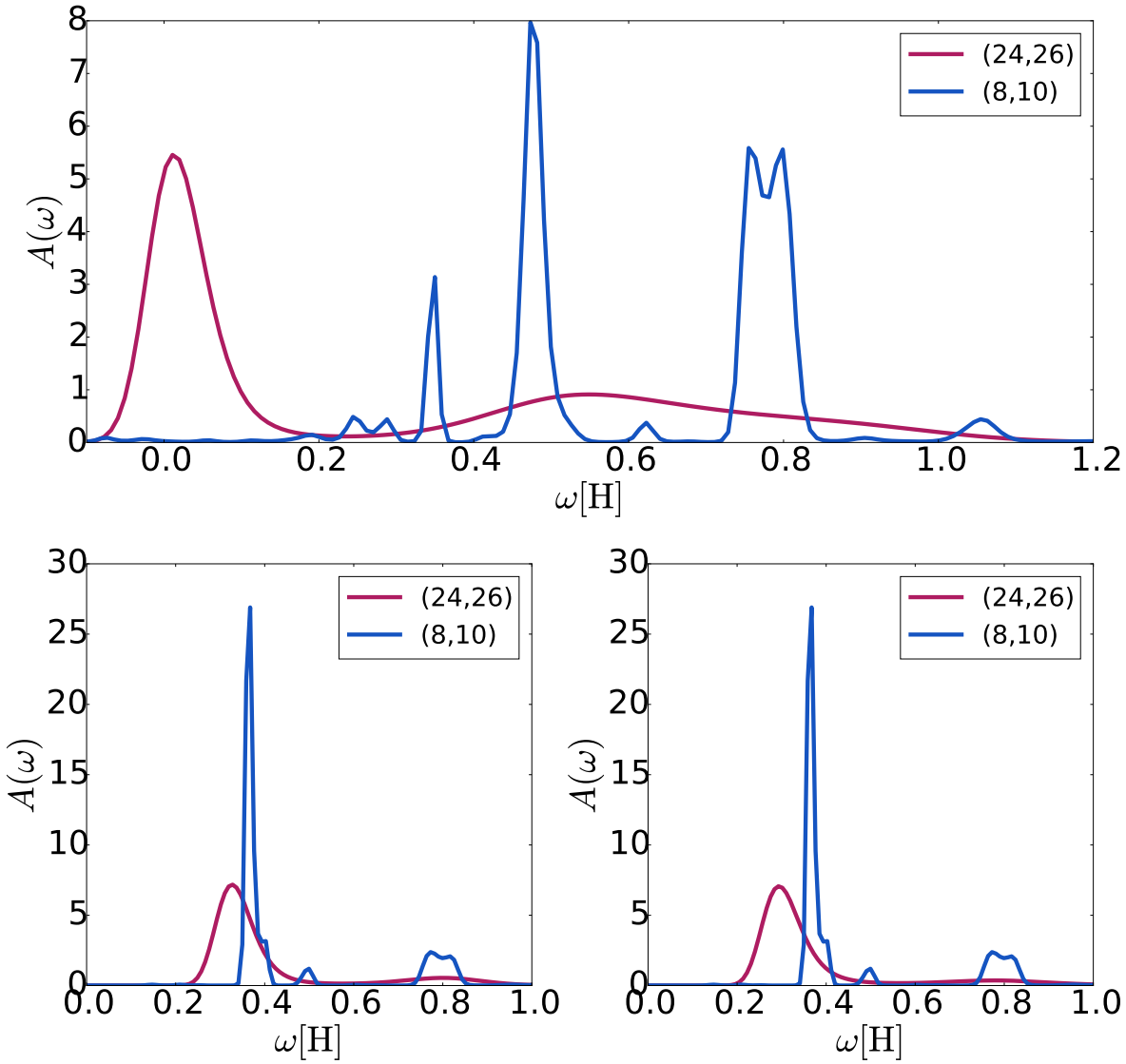


Figure 4.13: Spectral decomposition of a singly excited states compared between a small active space of 8 electrons in 10 orbitals and a larger active space of 24 electrons in 26 orbitals calculated using a complex time contour. a) Single excitation between the $3d_{x^2-y^2}$ orbitals of copper, showing significant impact of the additional p and d orbitals. b), c) Single excitations between oxygen $2p_y$ and copper $3d_{x^2-y^2}$ orbitals, showing little impact of increasing the active space size. The difference in peak width between the space sizes is due to the increased computational cost of preforming calculations in a larger active space, a higher imaginary part of time is required then, leading to broader spectra.

Chapter 5

A recapitulation of the transcorrelated method

The remainder of this work will be directed towards the application of ST-FCIQMC[1] [72, 128, 129] on the base of the transcorrelated ansatz by Boys and Handy [123, 124, 125, 126, 127, 309]. First, the transcorrelated method in general and ST-FCIQMC specifically will be put into context of explicitly correlated methods, followed by a recapitulation of the similarity transformed Hamiltonian and its numerical treatment. Correlation consistent Gaussian basis sets [293, 294, 310] are used as single particle basis in the examples in this chapter and for the results in the next chapter.

5.1 Explicit correlation by similarity transformation

Formulating the stationary Schrödinger equation as the CI problem with a finite basis is most commonly done by introduction of single-particle basis functions, representing the full wave function in a product basis made of Slater Determinants (SD) or Configuration State Functions (CSF) [100]. When solving the CI problem for the ground state $|\psi\rangle$ of the Hamiltonian, unless the true solution to the Schrödinger equation happens to be representable in the chosen basis by chance, an error will be introduced as the Hamiltonian is only considered on a subspace D of the full $\mathcal{L}_2$ space of square-integrable functions [112]. Since any solution of the Schrödinger equation in a finite basis set $|\psi\rangle \in D$ is also an element of $\mathcal{L}_2$, the corresponding eigenvalue, the CI ground state

[1] ST-FCIQMC always refers to similarity transformed FCIQMC including the initiator approximation and the semi-stochastic extension

energy E_0, is an upper limit to the true ground state energy [24]. In this sense, the basis set discretization is variational and systematically improvable by enlarging D, albeit at slow convergence rate when systematically adding higher and higher angular momentum states [114, 311].

However, the exact ground state solution Ψ is known to fulfill the cusp condition [113, 312], which generally reads [113]

$$\left(\frac{\partial\bar{\Psi}}{\partial r}\right)_{r=0} \propto \Psi(r=0)\,, \tag{5.1}$$

with $\bar{\Psi}$ being the spherical average of Ψ and r an electron-electron or electron-nuclear distance.

This was later shown to imply the specific form of the electron-electron cusp [312]

$$\Psi = \left(1+\frac{1}{2}r_{12}\right)\Psi(r_{12}=0) + O(r_{12}^2) \tag{5.2}$$

$$\Psi = \left(1+\frac{1}{4}r_{12}\right)\mathbf{r}_{12}\cdot\mathbf{w}[\Psi] + O(r_{12}^3)\,, \tag{5.3}$$

with the electron-electron distance $\mathbf{r}_{12}$ and a system-dependent vector $\mathbf{w}$. The first equation holds for same-spin electrons while the latter for opposite-spin electrons, analogous conditions also apply for the electron-nuclear distances [312]. Two important conclusions follow from the cusp condition. Firstly, the first derivative of the wave function has a discontinuity for all $r = 0$ [113], which is a property entirely absent in the commonly used Gaussian basis sets [112]. Secondly, the wave function around the electron-electron coalescence points $r_{12} = 0$ inherently depends on the inter-electron distance r_{12}, such that pairs of electrons are necessarily correlated at close distances.

At the core of the transcorrelated ansatz to capture the necessary short-range correlation lies the nonlinear Jastrow-type wave function ansatz [313]

$$|\psi\rangle = \mathrm{e}^{-\hat{\tau}}\,|\phi\rangle\ , \tag{5.4}$$

where $\hat{\tau}$ is a symmetric pair correlator

$$\hat{\tau} = \sum_{i<j}\hat{u}(\mathrm{r}_i,\mathrm{j})\,, \tag{5.5}$$

with a symmetric function $\hat{u}$. The Jastrow ansatz is commonly used in VMC [314] and DMC [158, 315] and has been studied in the context of the random phase approximation

[316]. It is known that choosing a $\hat{u}$ with the short-range behavior

$$\left(\frac{\partial \hat{u}}{\partial r_{12}}\right)_{r_{12}=0} = c\,, \tag{5.6}$$

with $c = \frac{1}{2}$ for opposite-spin electrons and $c = \frac{1}{4}$ for same-spin electrons fulfills the cusp condition [316]. In the transcorrelated ansatz, the Jastrow correlation factor is absorbed into the Hamiltonian by the similarity transformation [123]

$$\hat{H}_{st} = \mathrm{e}^{-\hat{\tau}} \hat{H} \mathrm{e}^{\hat{\tau}}\,, \tag{5.7}$$

resulting in an effective Hamiltonian $\hat{H}_{st}$ which is not hermitian anymore as e^{τ} is generally not unitary and the method is inherently non-variational, but conserves the spectrum of $\hat{H}$ [123].

Historically, the transcorrelated method was developed for single Slater determinant ansatzes for $|\psi\rangle$, for which it arguably cannot outperform DMC, but Tenno suggested a perturbation theory based on $\hat{H}_{st}$ [317], leading the way to a multi-reference treatment that might end in an FCI accuracy treatment at the complete basis set limit. Luo later suggested a variational technique [318] dealing with the non-hermiticity, but even though the transcorrelated method is promising in terms of basis set convergence, treating the effective Hamiltonian with FCI accuracy only became possible in combination with FCIQMC [129], with the ST-FCIQMC putting this goal into reach.

5.2 The transcorrelated approach for ab-initio systems

While similarity transformations has also been successfully applied to model systems [72], the transcorrelated ansatz was originally designed for the electronic ab-initio Hamiltonian [123], and that application shall be the focus of this work. In the following, a recapitulation of the formulation of the similarity transformed electronic Hamiltonian and the treatment of the arising integrals as encountered in the literature [123, 128, 129, 319] shall be given. The resulting Hamiltonian can be treated within a CI framework using a finite single particle basis set.

5.2.1 The similarity transformed electronic Hamiltonian

The similarity transformation (5.7) results in a dressed Hamiltonian $\hat{H}_{st}$ consisting of the bare Hamiltonian $\hat{H}$ plus a non-perturbative correction, a sketch of whose formulation according to Refs. [123, 129, 317] is presented here. Employing the Baker-Campbell-Hausdorff (BCH) formula [23] on the similarity transformed Hamiltonian (5.7) yields

$$\begin{aligned}\hat{H}_{st} &= \mathrm{e}^{-\hat{\tau}}\hat{H}\mathrm{e}^{\hat{\tau}} \\ &= \hat{H} + \sum_{n>0}\frac{1}{n!}[\hat{H},\hat{\tau}]_n\,. \end{aligned} \tag{5.8}$$

As $\hat{\tau}$ is a function of the $\hat{\mathbf{r}}$-operators, it commutes with the Coulomb term of the Hamiltonian and the commutator reduces to the commutator of $\hat{\tau}$ with the kinetic term. Now, the commutator of ∇ with any function f of $\hat{\mathbf{r}}$ is

$$[\nabla, f] = (\nabla f) + f\nabla - f\nabla = (\nabla f)\ , \tag{5.9}$$

and thus does not contain any derivative operator anymore. In particular, $[\nabla^2,\tau]_2$ is a function of $\hat{\mathbf{r}}$ and thus commutes with $\hat{\tau}$, meaning the BCH expansion in equation (5.8) terminates at the second order. Note that this is a particular property of the ab-initio electronic Hamiltonian, the BCH expansion of other similarity transformed Hamiltonians such as for the Hubbard model using a Gutzwiller transformation does not necessarily terminate [72]. Inserting the kinetic term of the Hamiltonian, the BCH expansion reads

$$\begin{aligned}\hat{H}_{st} =& \hat{H} - \sum_i\left(\frac{1}{2}[\nabla_i^2,\tau] + \frac{1}{2}[(\nabla_n\tau)\nabla_n + \nabla_n(\nabla_n\tau),\tau]\right) \\ =& \hat{H} - \sum_i\left(\frac{1}{2}(\nabla_n\tau)\nabla_n + \frac{1}{2}\nabla_n(\nabla_n\tau)\right. \\ &\left. + \frac{1}{2}\left(\nabla_n\left((\nabla_n\tau)\tau\right) - \tau\nabla_n(\nabla_n\tau) + (\nabla_n\tau)^2\right)\right) \\ =& \hat{H} = \sum_i\left(\frac{1}{2}\nabla_n^2\tau + \tau(\nabla_n\tau) + \frac{1}{2}(\nabla_n\tau)^2\right)\ . \end{aligned} \tag{5.10}$$

Now using the functional form (5.5) with the expansion (5.10) leads to the explicit

expression for the transcorrelated Hamiltonian [123]

$$\begin{aligned}\hat{H}_{st} =& \hat{H} - \frac{1}{2}\sum_{nm} \nabla_n^2 \hat{u}_{nm} + 2\left(\nabla_n \hat{u}_{nm}\right) + \left(\nabla_n \hat{u}_{nm}\right) \\ & - \frac{1}{2}\sum_{nml} \nabla_n \hat{u}_{nm} \cdot \nabla_n \hat{u}_{nk}\,, \end{aligned} \tag{5.11}$$

where the first sum corresponds to a dressing of the coulomb interaction with the two-body operator

$$\hat{K}_{nm} := \nabla_n^2 \hat{u}_{nm} + 2\left(\nabla_n \hat{u}_{nm}\right) \cdot \nabla_n + \left(\nabla_n \hat{u}_{nm}\right)^2\,, \tag{5.12}$$

and it is this dressing that can cancel the $\frac{1}{r}$-singularity of the Coulomb operator [317, 319]. The second sum is a newly arising three-body interaction

$$\hat{L}_{nml} := \nabla_n \hat{u}_{nm} \cdot \nabla_n \hat{u}_{nl}\,, \tag{5.13}$$

which arises from the second order commutator

$$(\nabla_n \tau)^2 = \frac{1}{4}\left(\sum_{ml} \nabla_n u_{ml}\right) \cdot \left(\sum_{pq} \nabla_n u_{pq}\right) = \sum_{nlq} (\nabla_n u_{nl}) \cdot (\nabla_n u_{nq})\,, \tag{5.14}$$

since both terms u_{qn} and u_{nq} appear in the full sum, the factor $\frac{1}{4}$ disappears.

For use in the FCIQMC formalism, which works in second quantization, the operators $\hat{K}$ and $\hat{L}$ are represented by four-index, respectively six-index, tensors K and L in a second quantized form based on an orbital basis $\{\phi_i\}$ as [317]

$$\hat{K} = \frac{1}{2}\sum_{nm} \hat{K}_{nm} = \sum_{ijab\,\sigma\sigma'} K_{ij}^{ab} \hat{a}_{a\sigma}^\dagger \hat{a}_{b\sigma'}^\dagger \hat{a}_{j\sigma'} \hat{a}_{i\sigma} \tag{5.15}$$

$$\hat{L} = \frac{1}{2}\sum_{nml} \hat{L}_{nml} = \sum_{ijkabc\,\sigma\sigma'\sigma''} L_{ijk}^{abc} \hat{a}_{a\sigma}^\dagger \hat{a}_{b\sigma'}^\dagger \hat{a}_{c\sigma''}^\dagger \hat{a}_{k\sigma''} \hat{a}_{j\sigma'} \hat{a}_{i\sigma}\,, \tag{5.16}$$

introducing the integrals

$$K_{ij}^{ab} = S\left(\left\langle \phi_a \phi_b \,\middle|\, \hat{K}_{12} \,\middle|\, \phi_i \phi_j \right\rangle\right) \tag{5.17}$$

$$L_{ijk}^{abc} = S\left(\left\langle \phi_a \phi_b \phi_c \,\middle|\, \hat{L}_{123} \,\middle|\, \phi_i \phi_j \phi_k \right\rangle\right)\,, \tag{5.18}$$

the symmetrization operator S performs symmetrization with respect to permutation of orbital pairs (a,i), (j,b) and (k,c) to respect the fermionic anti-commutation relation.

In practice, the integrals of K_{ij}^{ab} can be added to the Coulomb integrals, resulting in a

Table 5.1: FCIQMC reference and total energies in H for the ground states of different molecules in a cc-pVTZ basis set with and without the inclusion of the three-body operator $\hat{L}$ using the a 35-term correlator $u(\mathbf{r}_1, \mathbf{r}_2)$ supplied by Hongjun Luo for the dimers and the correlator by Schmidt and Moscowitz [320] for the Neon atom, respectively. For Be_2, an exponential $\bar{r}$ was used, the other cases use a polynomial $\bar{r}$, cf. section 6.2.1. For comparison, estimates for the experimental values for the energy are given, from Ref. [321] for Ne, while the values given for the dimers were obtained from the atomic energies in Ref. [321] and the experimentally determined well depths from Refs. [322] (Li_2),[299] (C_2) and [323, 324] (Be_2). Equilibrium geometries were used for the dimers, using the empirical values from Refs [322] (Li_2), [299] (C_2) and [323] (Be_2). The total energy for Ne reproduces the result given in Ref. [129].

System	Ne	Li_2	C_2	Be_2
ref. Energy (excl. L)	-128.9062	-15.0466	-75.7010	-29.2863
corr. Energy (excl. L)	-0.1418	-0.0325	-0.319	-0.1038
ref. Energy	-128.7971	-14.9636	-75.5820	-29.2343
corr. Energy	-0.1121	-0.0316	-0.3249	-0.1031
tot. Energy	-128.9092	-14.9952	-75.9069	-29.3374
Expt. value	128.9376	-14.994926	-75.9226	-29.3389

dressed interaction, and breaking hermiticity with the first order derivative $(\nabla_n \tau) \cdot \nabla_n$. Still, the FCIQMC algorithm with the projector (2.48) does not require hermiticity of $\hat{H}$ and thus nevertheless converges to the lowest energy right eigenvector as it does not rely on $\hat{H}$ being hermitian [128, 129]. The dressing $\hat{K}$ therefore introduces only minor issues on the technical side which are exclusively related to FCIQMC excitation generation.

However, as an analysis of FCIQMC ground state energies with and without the three-body interaction matrix elements, which is given in table 5.1, shows, the inclusion of the three-body terms can have crucial effect and they cannot be neglected when high energy accuracy is required. As the method is targeted at sub-millihartree accuracy, simply neglecting L would introduce uncontrollable and potentially serious errors. This in turn requires the FCIQMC implementation to support triple excitations, store the integrals corresponding to $\hat{L}$ and use extended Slater-Condon rules for evaluation of matrix elements. Furthermore, the number of integrals now scales as $\mathcal{O}(N^6)$, and substantial effort is required to obtain the integrals in the first place. For the uniform electron gas, an effective two-body interaction has been proposed [128], but for general ab-initio systems, no such form is available.

The parametrizations of the function u used here have been determined by Hongjun Luo using Variational Monte Carlo using functional forms suggested by Boys and Handy [125] and extensions thereof, which have then been implemented for integral calculation by Aron Cohen and the author.

5.2.2 Numerical evaluation of the integrals

The coulomb integrals for GTOs can be obtained routinely with high efficiency [138], the program PySCF[2] is used here therefor, but the matrix elements of the transcorrelated Hamiltonian, that is, the two-body dressing integrals W and particularly the three-body interaction terms L require additional effort and have to be evaluated numerically, done here by approximating the integrals with a discrete sum over a finite grid, which is implemented in a program written by Aron Cohen and extended for specific applications by the author, such as an exponential form of $\bar{r}$, separate grids for terms linear and quadratic in u, distributed memory parallelization, or the addition of zero-charged 'ghost' atoms for counterpoise correction. The used recipe for integral calculation designed by Aron Cohen is described in the following according to Ref. [129].

Two entities are to be evaluated in the numerical integration. First, consider the two-body interaction terms K from equation (5.17). In order to circumvent the numerically expensive inclusion of second derivatives of τ, partial integration is carried out to shift the derivative to the wave function

$$\begin{aligned} &\int \mathrm{d}r_1\, \mathrm{d}r_2\, \nabla_1^2 u(\mathbf{r}_1, \mathbf{r}_2) \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2) \bar{\phi}_a(\mathbf{r}_1) \bar{\phi}_b(\mathbf{r}_2) \\ &= -\int \mathrm{d}r_1\, \mathrm{d}r_2\, \nabla_1 u(\mathbf{r}_1, \mathbf{r}_2) \cdot \nabla_1 \left(\phi_i(\mathbf{r}_1) \bar{\phi}_a(\mathbf{r}_1) \right) \phi_j(\mathbf{r}_2) \bar{\phi}_b(\mathbf{r}_2)\,, \end{aligned} \tag{5.19}$$

yielding an expression for K containing only u and its first derivatives, which are evaluated numerically. The integration is then carried out explicitly on a grid.

The three-body interaction integrals (5.18) are explicitly given by [129]

$$L_{ijk}^{abc} = S\left[\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \mathrm{d}\mathbf{r}_3\, \bar{\phi}_a(\mathbf{r}_1) \bar{\phi}_b(\mathbf{r}_2) \bar{\phi}_c(\mathbf{r}_3) \left(\nabla_1 u(\mathbf{r}_1, \mathbf{r}_2)\right) \cdot \left(\nabla_1 u(\mathbf{r}_1, \mathbf{r}_3)\right) \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_1)) \phi_k(\mathbf{r}_1) \right], \tag{5.20}$$

where S is the symmetrizer with respect to permutation of the pairs (a, i), (b, j) and (c, k) already encountered in section 5.2.1, which is equivalent to permuting the coordinates $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$. Directly evaluating these six-index integrals would therefore require computing $\mathcal{O}(N^6)$ nine-dimensional integrals. Already Boys and Handy noticed that the integrations over $\mathbf{r}_2$ and $\mathbf{r}_3$ are independent and can be carried out separately [123], yielding intermediate tensors

$$\mathbf{y}_j^b(\mathbf{r}_1) = \int \mathrm{d}\mathbf{r}_2\, \bar{\phi}_b(\mathbf{r}_2) \nabla_1 u(\mathbf{r}_1, \mathbf{r}_2) \phi_j(\mathbf{r}_2)\,, \tag{5.21}$$

of which only $\mathcal{O}(N^2)$ many have to be evaluated, each at the cost of one three-dimensional integral. The calculation of the six-index integrals can then be broken down to evaluating an $\mathcal{O}(N^6)$ number of three-dimensional integrals

$$L_{ijk}^{abc} = \int \mathrm{d}\mathbf{r}_1 \, \bar{\phi}_a(\mathbf{r}_1)\phi_i(\mathbf{r}_1)\mathbf{y}_j^b(\mathbf{r}_1) \cdot \mathbf{y}_k^c(\mathbf{r}_1) \, . \tag{5.22}$$

Additionally, the terms L_{ijk}^{abc} have a high symmetry, reducing the effective cost of calculation and storage as, for real integrals, it is [129]

$$\begin{aligned} L_{ijk}^{abc} &= L_{\epsilon(ijk)}^{\epsilon(abc)} \\ &= L_{ajk}^{ibc} \, , \end{aligned} \tag{5.23}$$

where $\epsilon(ijk)$ is an arbitrary permutation of ijk. In total, that makes for 48-fold symmetry of the L-tensor, greatly easing the cost associated with handling it [129].

A possible concern is a grid dependence of the resulting integrals, but convergence with respect to the number of grid points proves to be rapid [129]. It can be quantified by determining the reference and FCIQMC correlation energy for different grid sizes. The grid dependence of these energies for the Neon atom in a cc-pVQZ basis set is shown in figure 5.1, revealing a weak grid dependence of the total energy. The reference energy displays a monotonous decrease with increasing grid size, but the same is not true for the correlation energy, such that the total energy features a minor non-monotonous grid dependence. As this dependence is of similar scale for similar systems, the remaining grid dependence of energy differences can be expected to be negligibly small.

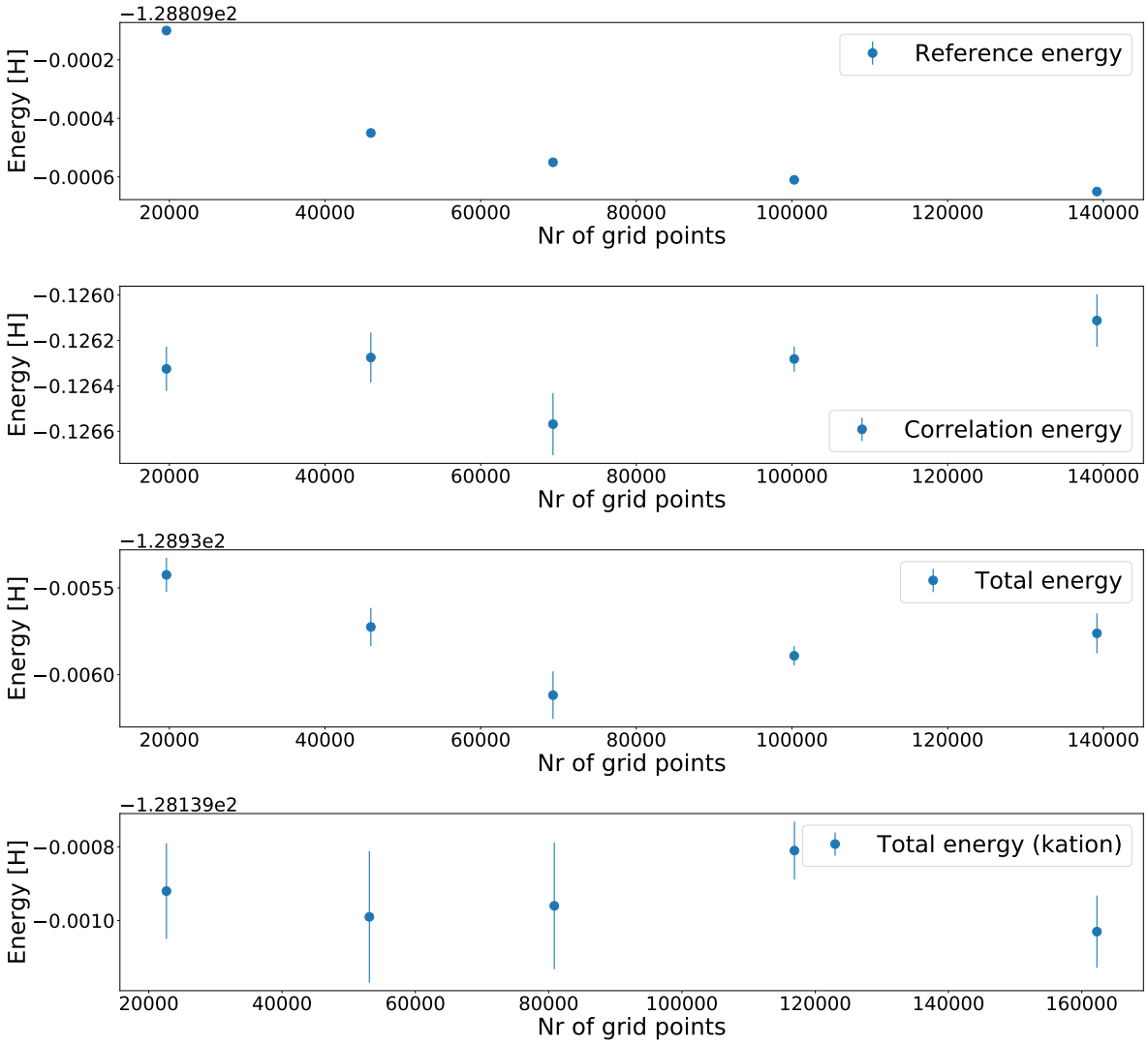


Figure 5.1: Dependence of reference and correlation energy of the Neon atom for different grid sizes used for evaluating the integrals of the transcorrelated Hamiltonian. The difference in total energy between a grid of 20000 points and a grid of 160000 points is 0.33 mH for the neutral atom and 0.10 mH for the cation.

Chapter 6

ST-FCIQMC for molecular systems

With the general framework of the similarity transformed Hamiltonian in the transcorrelated ansatz revisited, this chapter shall address the development of the ST-FCIQMC and cover the authors contribution to it, before demonstrating the high yield of the method at the example of the potential energy curve of Be_2. The ab-initio ST-FCIQMC was developed in collaboration with Aron Cohen, Hongjun Luo and Ali Alavi, and it has been previously published in Ref. [129]. Here, new results from the same collaboration and implementational details are presented. All correlation factors used were supplied by Hongjun Luo resulting from VMC optimization unless noted otherwise.

6.1 Solving the similarity transformed Hamiltonian using FCIQMC

The basic i-FCIQMC algorithm [59] revisited in section 2.4 does not rely on the hermiticity and can thus be applied to the similarity transformed Hamiltonian [72, 128, 129], although the procedure in practice becomes more involved compared to the basic algorithm due to the addition of three-body interactions. This affects the excitation generation, the evaluation of matrix elements and naturally the storage of the Hamiltonian. These points shall be addressed here, explaining how the implementation in `NECI` deals with them. The implementation of the support for the similarity transformed Hamiltonian by the author in `NECI` has been used previously in Ref. [129].

On a side note, the treatment of the Hubbard model with ST-FCIQMC also requires the inclusion of three-body interaction, but does not give rise to the issues addressed in the following and uses a separate, independent implementation specific to the Hubbard

model [72].

6.1.1 Excitation generation and time-step control

The change in excitation generation has serious impact on the computational performance, but does not introduce new bottlenecks on its own, excitation generation in the presence of three-body interactions is simply less efficient than the optimized schemes for two-body interactions. First, three-body excitations are required, which is realized by adding a new bias towards triple excitations to the overall excitation generation scheme as visited in section 2.4.5.

From an implementational point of view, triple excitations are best implemented on top of the existing excitation generators, such that the natural approach is the introduction of an analogous bias p_{trip} towards triple excitations. The triple excitation bias is determined in a fashion analogous to the single/double excitation bias p_{doub} such that

$$\max_{\text{triples}} \left(\frac{H_{ij}}{p_g(j|i)} \right) = \max_{\text{singles,doubles}} \left(\frac{H_{ij}}{p_g(j|i)} \right) . \tag{6.1}$$

Now, the process of sampling excitations in the evaluation of $\hat{H}\,|\psi\rangle$ can be performed by first deciding if a triple excitation shall be attempted, which happens with probability p_{trip}. If no triple excitation is attempted, an existing excitation generator with the double excitation bias p_{doub} in its original meaning is called, and subsequently the probability is adjusted as

$$p_g(j|i) \to (1 - p_{\text{trip}})\, p_g(j|i) \,. \tag{6.2}$$

The triple excitation bias itself when determined according to equation (6.1) is of the same order as the singles bias $(1 - p_{\text{doub}})$, in particular, $p_{\text{trip}} \ll p_{\text{doub}}$, exemplary values are given in table 6.1. That means even though there are many more triple excitations than double excitations, the matrix elements of $\hat{L}$ corresponding to triple excitations are much smaller than the double excitation matrix elements, which results in the double excitations still being dominant even in the presence of three-body interactions. Note at this point that the contribution of $\hat{L}$ to the two body matrix elements is essential, which contributes to the difficulty of efficiently sampling the similarity transformed Hamiltonian.

The triple excitations are sampled uniformly in this work, it shall be noted that an almost identical scheme for randomly choosing triple excitations has been developed independently by Anderson *et al.* in the context of sampling higher-body Reduced

Density Matrices (RDMs) [325]. To pick a random triple excitation $(i, j, k) \rightarrow (a, b, c)$, first uniformly select a pair of occupied orbitals (i, j) with spins σ_i, σ_j and a given bias towards parallel-spin selection. By convention, if $\sigma_i = \sigma_j$, pick k with the same spin, else, with a random spin with a potential bias towards α/β spin. While a 3-tuple of electrons could be chosen directly, this scheme allows for optimal integration in the existing framework in `NECI` without negatively impacting performance. Then, two unoccupied orbitals are chosen uniformly subsequently, and a third unoccupied orbital is chosen uniformly such that the triple excitation respects potential symmetries of the Hamiltonian.

Some care has to be taken when evaluating the generation probability p_{gen} which is required for the FCIQMC spawning step, since a given triple excitation $(i, j, k) \rightarrow (a, b, c)$ can be generated in multiple ways by choosing the occupied and/or unoccupied orbitals in a different order. Each path can have a different probability due to different symmetry requested in the last step, so all have to be considered separately, but this is encountered in any uniform scheme, not just for triple excitations, [59] and is more of a technicality.

Next, consider the sampling of the single/double excitations, for which established techniques can be used. Regarding the correctness of the algorithm, this choice does not matter, any unbiased scheme can be used and will converge to the correct result. The efficiency of the implementation however largely depends on this choice, so an analysis of the options here shall come next.

The single/double excitation generation schemes (in the following simply referred to as excitation generation) considered here are the PCHB [155, 196], pre-computed Power-Pitzer (PCPP)[231], Cauchy-Schwartz (CS) [196] and uniform [59] excitation generation implementations of `NECI` , of which only PCHB and PCPP have been implemented by the author. These methods themselves all have been successfully used in i-FCIQMC calculations before, e.g. in Refs. [66, 68, 69, 196, 231, 298].

Exemplary acceptance rates for these existing excitation generators are compared in Fig.6.1 (see also Fig.6.4), indicating that existing non-uniform generation schemes may become extremely inefficient for the similarity transformed Hamiltonian. A gain can still be obtained in principle from non-uniform selection, but great care has to be taken when choosing the excitation generation method, and to some extent also the time-step. While neglecting the triple excitations in the dynamics could be considered, there is no point in doing so as they are computationally cheap and do not limit the available time-steps or the acceptance rate, since else, p_{trip} would dominate. It will be demonstrated however, that the low acceptance rates stem from the time-step to a large degree and

non-uniform excitation generation is not per se inefficient for ST-FCIQMC.

Problematic for the optimized non-uniform schemes is the non-negligible presence of the three-body operator $\hat{L}$. The commonly used PCHB method from the family of heat-bath methods [155] is assuming the matrix element $H_{ab}^{rs} = H_{ij}$ of a double excitation

$$|D_j\rangle = a_{r\sigma}^\dagger a_{s\sigma'}^\dagger a_{j\sigma'} a_{i\sigma} |D_i\rangle \tag{6.3}$$

does not depend on the determinant $|D_i\rangle$ but only on the involved orbital indices a, b, r and s to use precomputed lookup tables for efficient sampling. This assumption however is broken by the inclusion of $\hat{L}$, as the matrix element H_{ij} now also contains contractions of the L-tensor over occupied orbitals. Therefore, the PCHB flavor used here slightly varies from the approximate heat-bath version used in conventional i-FCIQMC by using the double excitation matrix elements with respect to a reference determinant. This is a generalization that does not introduce any alteration to the algorithm for a two-body Hamiltonian, the `NECI` implementation in fact does not differentiate here and always uses the matrix element of a double excitation from a reference, which just happens to be independent of the reference in the two-body case.

The CS scheme on the other hand deeply relies on the validity of the Cauchy-Schwartz inequality

$$|H_{ab}^{rs}|^2 \leq |H_{aa}^{rr}||H_{bb}^{ss}|\,, \tag{6.4}$$

which is broken by both the three-body terms of $\hat{L}$ and the dressing of the Coulomb terms by $\hat{K}$.

On the other hand, generating a weighted probability distribution $p_g(j|i) \propto |H_{ji}|$ on the fly is very inefficient for the predominant double excitations since evaluating H_{ij} is an $\mathcal{O}(n)$ operation, making the generation of the distribution an $\mathcal{O}(n^3N^2)$ operation, compared to the $\mathcal{O}(1)$ uniform and precomputed generation schemes, such that precomputed schemes should always be preferred. Even though the CS scheme can generally achieve high acceptance rates in i-FCIQMC, it is not an $\mathcal{O}(1)$ operation as it requires creating the probability distribution for sampling on the fly [196].

While approximate methods such as PCHB, CS or PCPP may fail to capture some part of the matrix elements, if the problematic matrix elements are few, a technique that can be employed to increase the time-step and thus efficiency per iteration is the histogram tau-search[233], using a high percentile instead of the maximum value of the ratios $\frac{H_{ij}}{p_g(j|i)}$. As introduced by Dobrautz [233], this neglects outliers that could otherwise negatively impact the time-step, leading to much faster convergence with respect to iteration at the

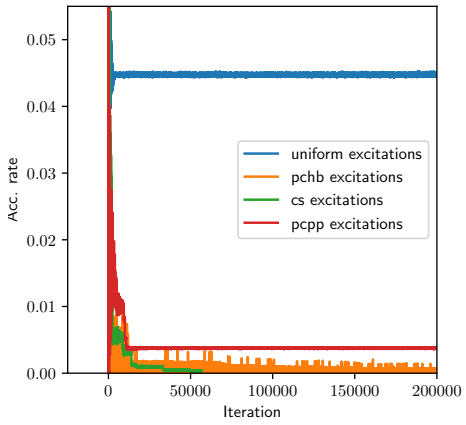


Figure 6.1: Acceptance rates over iteration for different excitation generation methods for the similarity transformed Hamiltonian of Neon in a cc-pVDZ basis set. The time-step is automatically determined based on the maximum scheme outlined in section 2.4.6. The walker number is set to $5\,10^6$, but only uniform and PCPP excitation generation reach that number in the given time. Note that these are calculations for profiling purpose only. The labeling scheme refers to the single/double excitation generation scheme, triple excitations are sampled uniformly in all cases. The uniform method is by far best suited of the available ones, but still only achieves very low acceptance rates.

Table 6.1: Triple excitation bias p_{trip} and time-steps $\Delta\tau_{st}$ for different excitation generators for ST-FCIQMC calculations of Be_2 in an aug-cc-pVTZ basis set with $N_w = 3 \cdot 10^6$ compared to the time-steps $\Delta\tau$ for the analogous i-FCIQMC calculations. Time-steps are determined with 99.99 percentile histogram tau-search, except for the CS (max τ), which is determined using maximum-based time-step determination. Notably, for i-FCIQMC, the time-step is consequently larger than for ST-FCIQMC.

ex. gen.	p_{trip}	$\Delta\tau_{st}$ [10^{-3} au]	$\Delta\tau$ [10^{-3} au]
uniform	$8.6 \cdot 10^{-3}$	0.21	0.36
CS	0.10	2.82	5.14
CS (max τ)	0.263	0.02	3.31
PCHB	0.015	0.30	6.45
PCPP	0.152	2.83	3.02

same computational cost per iteration, as shown in Fig 6.2 and table 6.1. This can cure the problems of the non-uniform excitation generators, greatly increasing the efficiency of the sampling, but usually not back to the level of the original i-FCIQMC, see therefore section 6.1.3.

Since the time-step τ is determined in an automated fashion based on the ratios $\frac{H_{ij}}{p_g(j|i)}$, the choice of the excitation generator directly impacts the time-step, particularly in ST-FCIQMC as the data in table 6.1 indicates, giving it twofold impact on the performance. First, it dictates the computational effort per iteration, and second, the number of iterations required to reach a given imaginary time and therefore convergence of the wave function. As maintaining a constant ratio $\frac{H_{ij}}{p_g(j|i)}$ is more difficult for the similarity transformed Hamiltonian for the reasons outlined above, the utilized time-steps are smaller than for equivalent i-FCIQMC calculations, see also table 6.1, thus requiring

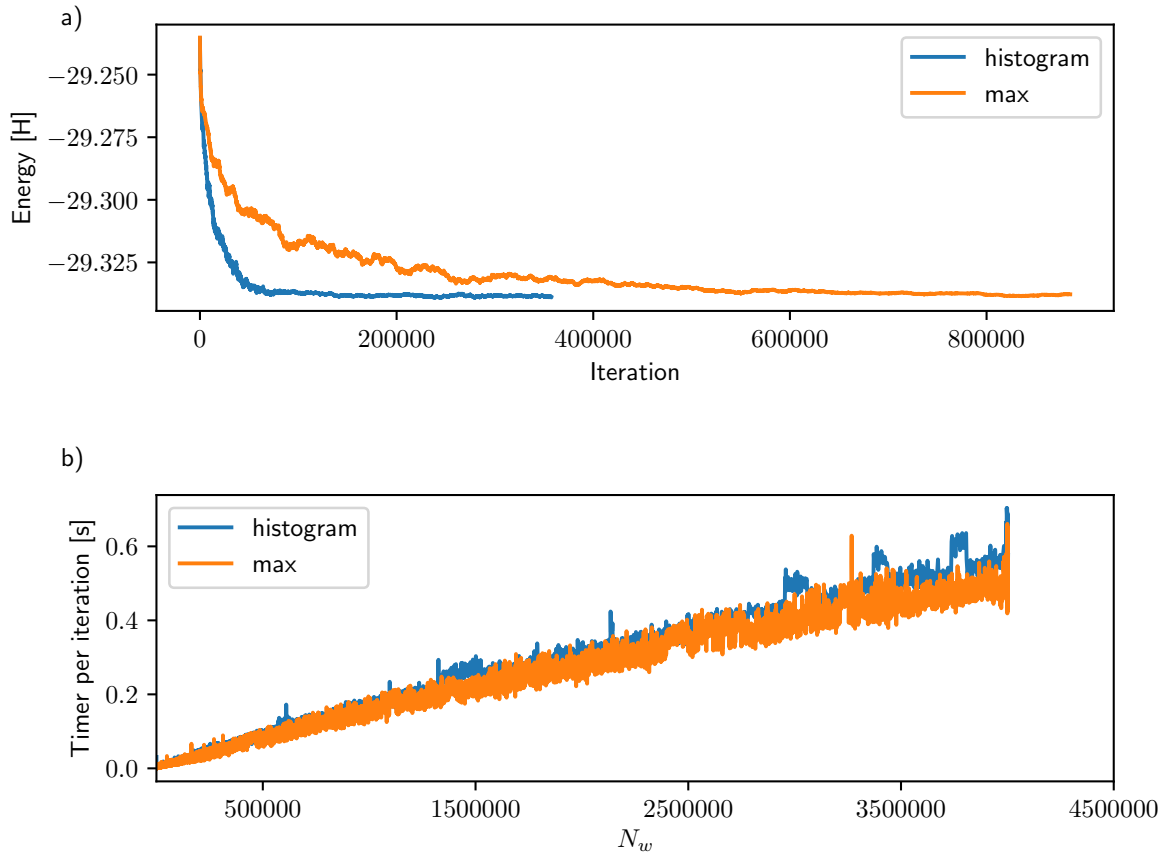


Figure 6.2: a) Energy over iteration with maximum based and 99.99-percentile histogram based time-step search for the similarity transformed Be_2 in an aug-cc-pVTZ basis set. b) Time per iteration over walker number for the same system, the overhead due to increased spawns per iteration from the larger time-step is minimal, the computational effort comparable. The calculation was carried out on a single node of 2 Intel Xeon E5-2698 v4@2.20 GHz processors with 20 CPU cores each.

a large number of iterations for convergence, often in the range $10^5 - 10^6$ iterations such as in the exemplary calculation in Fig. 6.2. This comes with the benefit of much smaller basis sets required for accurate description of a given system and therefore only comparably few walkers required for convergence, as shown in table 6.2, demonstrating the rapid convergence of the transcorrelated FCIQMC with respect to the basis set. For a comprehensive study of the convergence of atomic energies and ionization potentials see Ref. [129], the results here shall only emphasize the increased computational efficiency by requiring reduced N_w. In particular, the similarity transformation itself does not adversely affect FCIQMC convergence.

Table 6.2: Ground state energies and walker numbers required for convergence for the Neon atom in different basis sets with (st) and without similarity transformation. The similarity transformed calculations use the SM17[320] correlation factors, all calculations use an initiator threshold of 2, the given walker numbers are estimates from the dynamics where the population is grown up to $N_w = 5 \cdot 10^6$ and then verified in separate calculations. For an extensive analysis of atomic energies using ST-FCIQMC see Ref. [129], the given numbers are from independent calculations but reproduce results given there. The complete basis set limit corresponds to the value given in Ref. [321] as extrapolation from experimental findings.

Ne	pVDZ (st)	pVTZ (st)	pVQZ (st)	pVQZ	est. CBS
E_0	-128.8366	-128.9093	-128.9352	-128.8769(1)	-128.9376
N_w	10^5	$5 \cdot 10^5$	10^6	10^6	

6.1.2 Integral storage and access

The additional $\mathcal{O}(N^6)$ memory cost for storing the six-index tensor L is clearly the restricting factor for the applicable systems, as it imposes tight limits on the number of orbitals that can be treated. For example, storing L for a first row dimer in a cc-pVTZ basis set with 60 spatial orbitals requires $\sim$ 7.6 GB of shared memory, while for an aug-cc-pVTZ basis set with 92 spatial orbitals, this already grows to $\sim$ 97 GB of shared memory. Problematic in this context is that this cost cannot be distributed across compute nodes, each node requires access to the full tensor L, since the stochastic process allows for no prediction which elements will be required and communication across nodes for matrix elements would cripple the parallel performance. It is true that those elements of L which enter the matrix elements of double excitations are required more often, but these still constitute a large part of L such that a separate treatment would not be efficient. On a 256 GB RAM machine, this imposes a hard limit of 108 spatial orbitals on ST-FCIQMC for a dense storage of L, without taking into account any other memory requirements of the method such as for the two-body integrals.

What can be used as an advantage is the sparsity of L by storing only the nonzero entries of the six-index tensor and a lookup table for accessing entries. Two memory layouts for L were implemented in `NECI` , a dense format where all values of L_{ijk}^{abc} are stored in a single contiguous array, including entries that are 0. Given indices i, j, k, a, b, c, the corresponding element L_{ijk}^{abc} can then be obtained by computing a combined index

$$I(i,j,k,a,b,c) = A + \frac{B(B-1)}{2} + \frac{C(C-1)(C-2)}{6}\,, \tag{6.5}$$

where the three values $A < B < C$ are the sorted combined indices $C(a,i)$, $C(b,j)$ and

Table 6.3: Sparsity of L and computational overhead of the sparse scheme for different molecules in cc-pVTZ basis sets. The sparsity given is the ratio of zero elements to total elements, while the slowdown given is measured as the ratio of the average time per iteration using the sparse scheme and the dense scheme with uniform excitation generation. For Ne, the Schmidt and Moscowitz factor [320] was used, the factors for Li_2 and Be_2 were supplied by Hongjun Luo, using a polynomial $\bar{r}$ for Li_2 and exponential $\bar{r}$ for Be_2. For Li_2 and C_2, only single core calculations for profiling purposes were performed using $N_w = 10^3$ and $N_w = 10^4$, respectively. The calculations for Ne and Be_2 were run on 40 cores in parallel with $N_w = 4 \cdot 10^6$ and $N_w = 3 \cdot 10^6$, respectively. All calculations were performed on Intel Xeon Gold 6138 processors @2.00 GHz.

System	Sparsity	Slowdown
Ne	87.5 %	1.8
Li_2	79.3 %	1.6
C_2	88.5 %	1.8
Be_2	79.7%	2.3

$C(k, c)$ according to

$$C(n, m) = \min(n, m) + \frac{\max(n, m)\,(\max(n, m) - 1)}{2}\,, \tag{6.6}$$

such that I is symmetric with respect to all symmetries of L. In the dense layout, I is directly used for addressing the matrix elements, which induces no overhead for lookup and is thus faster than the sparse alternative. Therefore, when the memory is available, the dense layout should be used for faster matrix element access and thus notably faster iterations.

The matrix elements themselves can then be constructed from the stored tensor L by summing over all permutations of orbitals with their corresponding sign, analogously to the exchange term in the two-body matrix elements. The matrix element of an excitation $\left\langle \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_c^\dagger \hat{a}_k \hat{a}_j \hat{a}_i D \,\middle|\, \hat{L} \,\middle|\, D \right\rangle$ with spin-orbitals a, b, c, i, j, k is then given by

$$L(a, b, c, i, j, k) = \sum_{\epsilon} (-1)^{\pi(\epsilon)} L_{ijk}^{\epsilon(abc)} \delta_{\sigma(i)\sigma(\epsilon(a)} \delta_{\sigma(j)\sigma(\epsilon(a))} \delta_{\sigma(k)\sigma(\epsilon(c))}\,. \tag{6.7}$$

Here, $\sigma(x)$ is the spin of a spin-orbital x and $\epsilon(abc)$ is a permutation of abc with parity $\pi(\epsilon)$. The sum contains six terms, but only for the same-spin case all six contribute, if one of the orbitals is of a different spin than the others, only two terms are nonzero.

As opposed to the computational cost, the memory cost can be significantly reduced by only storing the nonzero elements of L. As seen in table 6.3, the sparsity can be significant and hence the memory cost notably reduced, for the aforementioned Be_2 integrals in the aug-cc-pVTZ basis set, the sparse format only requires $\sim$ 24 GB of

memory at a sparsity of $\sim 92\%$. The reason the memory saving is much less than the sparsity of the matrix elements would imply is the requirement for a lookup table in order to be able to locate a given element L_{ijk}^{abc} in the now non-contiguous storage. The lookup table is realized as an open addressing hash table with linear probing [326], storing the index $I(i,j,k,a,b,c)$ of any nonzero element together with L_{ijk}^{abc}. To read an element, first the the index I is computed and then used to calculate a hash value. For each hash value, the position of the first corresponding element is stored, and the range between this position and the position of the first element of the next hash value is searched for the matching index. If it is found, the corresponding matrix element is returned, else it is 0. The required memory for the sparse layout grows as $2n_{\mathrm{L}} + n_{\mathrm{H}}$, where n_{L} is the number of nonzero entries of L and n_{H} is the hash table size, i.e. the range of the hash function, which should be of the same order as n_{L} for good performance. A larger value of n_{H} reduces the number of conflicts in the hash table and thus speeds up the lookup, but also increases the memory cost. In contrast to e.g. the list of occupied determinants, this table only needs to be set up at initialization and requires no future read aces, which is, in addition to its better suitability for MPI-3 shared memory[1] and faster access times, the reason this form of a hash table is used here over the linked list implementation already present in `NECI` . Potentially, more involved hashing algorithms for faster lookup could be explored, but the current implementation is already sufficiently efficient not to drastically increase the computational cost of an iteration, see also Table 6.3. Note that the time per iteration includes any operation during an iteration, not only the lookup of three-body matrix elements. The computational overhead of the lookup itself is thus higher, direct profiling with the Intel VTune Amplifier [327] for Li_2 shows a computational overhead of 2.7 for the lookup itself. The sparsity may be artificially increased by introducing tight cutoffs on the elements of L, which can greatly benefit the memory requirements at the cost of a mild loss of accuracy, as shown in table 6.4. Notable is also the reduction in sparsity of the untruncated L compared to the cc-pVTZ basis set in Table 6.3, meaning in addition to the increased memory cost, the sparsity can also reduce for larger basis sets.

The integrals of L are supplied in a separate file to the one- and two-body terms, the I/O of these integrals in both `NECI` and the integral calculation supports ASCII and HDF5 [328] file formats, where HDF5 should be strongly preferred for its increased performance on parallel systems. The coulomb integrals only support ASCII format

[1]The `NECI` implementation uses MPI-3 shared memory windows for handling integrals, originally implemented by Florian Merz and further expanded by the author

Table 6.4: ST-FCIQMC Ground state energy and sparsity of the L tensor for the Neon atom in a cc-pVQZ basis set for different cutoffs imposed on the matrix elements of L. The SM17 correlation factor [320] is used here, reproducing the result from Ref. [129] in the absence of a cutoff. Large parts of L can in fact be neglected without significantly affecting the energy, but care as to be taken as L cannot be generally neglected. In this example, both reference and correlation energy strongly depend on L in principle.

Cutoff	GS Energy [H]	Sparsity of L
None	-128.9355	51.9%
$1e-6$	-128.9354	71.5%
$1e-5$	-128.9352	93.6%
$1e-4$	-128.9363	99.6%
$1e-3$	-128.9639	99.9%
$1e-2$	-129.0759	99.9%

in the form of the FCIDUMP format [329], but play only a minor role for the I/O performance as most integrals belong to L. The HDF5 support was newly added based on general HDF5 I/O routines of `NECI` .

6.1.3 ST-FCIQMC Performance

Naturally, the question of computational cost of using the similarity transformed Hamiltonian arises, and in section 6.1.1 a potential impact on the sampling and the time-step was addressed. As the sparse format induces a significant total overhead, the handling of the three-body terms obviously also carries non-negligible cost itself.

To get an overview of the overhead connected to handling the three-body interaction, consider the time per iteration at fixed walker number on fixed hardware, which is compared for several systems between a simple ST-FCIQMC and conventional i-FCIQMC in Fig. 6.3 for uniform excitation generation. Clearly, ST-FCIQMC is not only more expensive in terms of cost per iteration, this cost also increases with the number of electrons as can also be seen in Fig.6.3, since the evaluation of a double excitation matrix element scales as $\mathcal{O}(n)$ in the presence of three-body interactions and double excitations dominate the overall cost.

For the full picture, not only the time per iteration has to be considered, which only accounts for the direct overhead associated with the three-body integrals, but also the impact on the time step, since what ultimately defines the performance of an FCIQMC scheme is how much imaginary time can be propagated with a given walker number in a given time as that determined how fast an implementation converges in terms of CPU time. The time per iteration gives the wall clock time on a given hardware

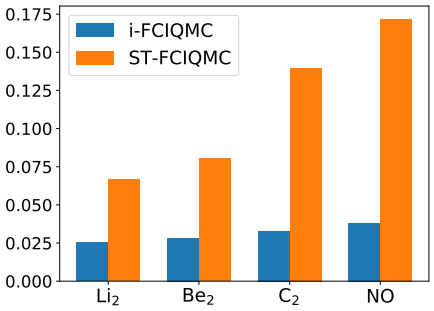


Figure 6.3: Time per iteration for different systems with uniform excitation generation compared between ST-FCIQMC and i-FCIQMC, obtained on four CPU cores of an Intel Xeon E5-2680 v2@2.80 GHz, each with $N_w = 10^5$ and fixed $\Delta\tau = 10^{-4}$. Since uniform generation is used, the only impact from the ST-FCIQMC in timing is in the matrix element evaluation, i.e. accounting for the three-body interaction. These timings can therefore be interpreted as measuring the computational cost of treating L.

required for performing a single iteration, so by dividing by the time-step, the time required for propagating a single atomic unit of time with this iteration speed is given. Surely the time per iteration is not constant, so this is not the actual time required for a single atomic unit of time in a concrete calculation, but nevertheless estimates the effort required for moving forward a set imaginary time. Since the time per iteration scales roughly linearly with the walker number [62], cf also Fig. 6.2, the same can be expected of the time per imaginary time, so consider this timing over the walker number as a measure of performance of an implementation. This is compared between ST-FCIQMC and conventional i-FCIQMC in Fig.6.4 b) and d), demonstrating that ST-FCIQMC is more expensive than conventional FCIQM in terms of required CPU time, but at manageable level. Of particular interest is here the comparison of different excitation generation schemes, which is what ultimately determines the performance of an FCIQMC implementation. As illustrated in section 6.1.1, excitation generation faces different issues in ST-FCIQMC than in conventional i-FCIQMC, and it is thus unsurprising that different schemes are favorable in ST-FCIQMC. The most performant one in the considered example is the PCPP method for ST-FCIQMC, in contrast to conventional i-FCIQMC, where PCHB performs best. As shown in Fig.6.4 a) and c), the PCPP excitation generator achieves similar acceptance rates in the ST-FCIQMC compared to i-FCIQMC, which is remarkable given its single-reference nature.

To give more concrete timings and put these into context, consider for the Be_2 example in Fig. 6.4 with the most performant excitation generator (PCPP for ST-FCIQMC and PCHB for i-FCIQMC). To reduce the statistical error of the ST-FCIQMC to $\sigma = 4\,\mu\mathrm{H}$ a total run-time of $T \sim 339$ CPU hours was required. The corresponding efficiency in the definition of Holmes *et al.* [155] as $\eta = \frac{1}{\sigma^2 T}$ is $\eta = 7626$ (T in seconds), which not drastically lower than $\eta = 59337$ of the corresponding i-FCIQMC calculation, considering

that the impact of the excitation generator is larger than the difference, see Appendix A. Even though ST-FCIQMC is more expensive in terms of cost per calculation, it can compensate this by compensating substantial basis set errors. ST-FCIQMC can therefore be seen as an incredibly efficient method for getting energies close to the complete basis set limit. This however does not take into account the time required for optimizing the correlation factor and calculating the integrals, and especially the latter, with the current implementation, can require CPU times similar to the ST-FCIQMC itself for larger basis sets. Further optimizations to the implementation like the inclusion of symmetry in the integral calculation might significantly reduce this cost, however.

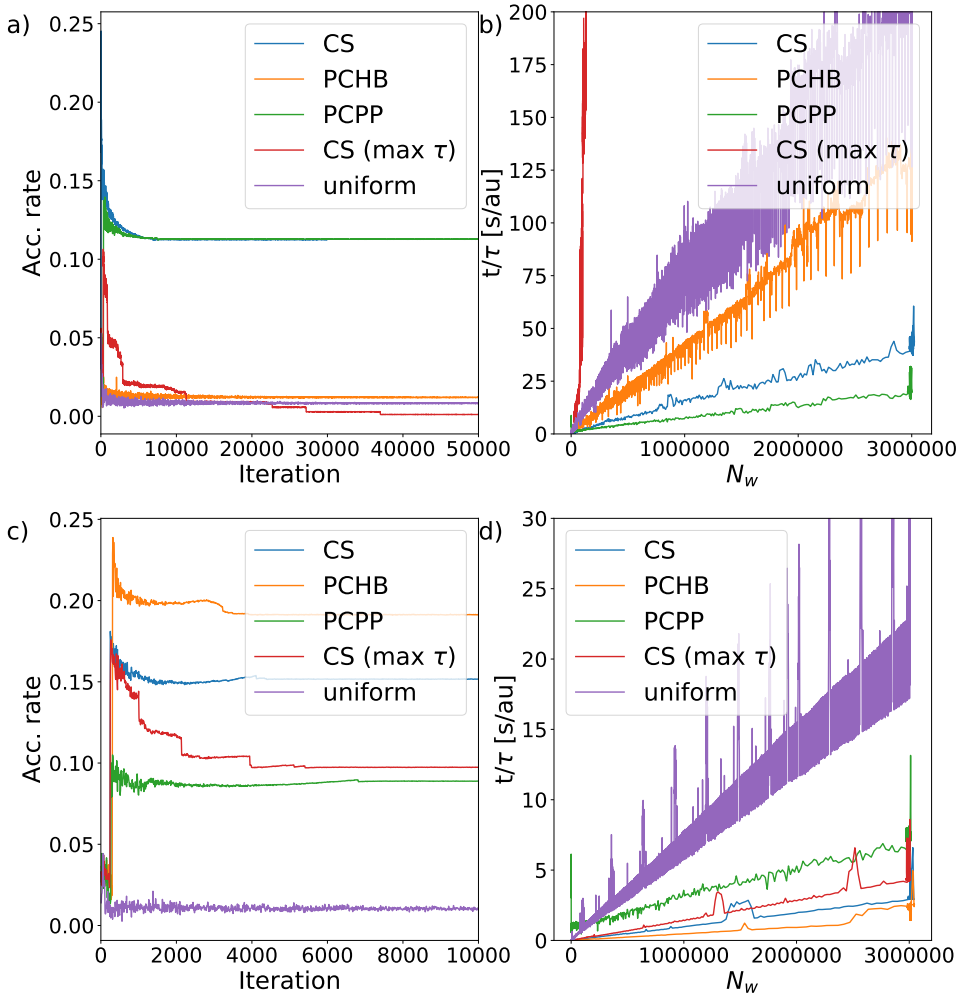


Figure 6.4: a) Acceptance rate over iteration of several excitation generators for ST-FCIQMC of Be_2 in an aug-cc-pVTZ basis set with a target population $N_w = 3 \cdot 10^6$ and the time-step determined using 99.99-percentile based histogram tau-search with the exception of the CS (max τ), which uses maximum-based tau-search. b) Wall clock time per iteration per imaginary time over the instantaneous walker number for the same calculations, indicating the total computational cost per imaginary time propagated with each walker number. The peak at target walker number can be attributed to the overhead occurring when switching to variable shift mode such as core-space initialization. Note that this measure is not averaged and thus the values at the individual walker numbers are not necessarily precise and can carry significant noise, they serve to give an overview of the performance of different schemes only. c) Acceptance rate over iteration for the same excitation generation schemes for the analogous conventional i-FCIQMC calculations. d) Wall clock time per iteration per imaginary time over instantaneous walker number for the analogous i-FCIQMC calculations. The peaks in the uniform curve can be attributed to load-balancing steps that take extra time, which is significant in this measure only for small time-steps and fast iterations, only matching the uniform i-FCIQMC case. All calculations were performed on a single node of 2 Intel Xeon E5-2698 v4@2.20 GHz processors with 20 CPU cores each.

6.2 Case study: Binding curve of Be_2

The Beryllium dimer was already briefly visited in this chapter and is in fact ideally suited to explore the potential of the method. Displaying significant electron correlation and basis set dependence, an extremely accurate method is required to correctly capture the weak bond [330]. Even though the bond is only weak, it is much stronger than for other closed shell dimers due to the splitting of the s and p orbitals being extremely small[330, 331], which, together with the large required basis sets [332, 333] makes calculation of the $^1\Sigma_g$ ground state potential energy curve is a challenging problem requiring substantial computational effort [333, 334, 335]. In the following, the ST-FCIQMC is used to the obtain the potential energy curve and vibrational states of Be_2, gathering useful insight for the application of transcorrelated FCIQMC to molecular systems.

Early spectroscopic data predicted a well depth of 800 cm^{-1}[336], which was later refined to estimates of 929(2) cm^{-1}[323] and 934.9(0.4) cm^{-1}[324] based on the experimental data by Merritt [323], whose fit and the corresponding spectroscopic values will be used as benchmark data in this section to compare the ST-FCIQMC results to.

6.2.1 Correlation factors

The purpose of the correlation factor in the presented method is to remove the singularity from the Coulomb interaction and thereby the cusp condition for the wave function. That alone does not uniquely define the correlation factor however, and a great deal of freedom remains. To some extent, this is used in the VMC optimization of the factors determining the function u within a fixed functional form, but ambiguity remains. It is well known from VMC and DMC studies that the form of u is essential [314, 337, 338, 339]. To correctly describe correlation in a molecule, the electron coordinates with respect to each nucleus have to enter u [338]. For a dimer consisting of two atoms A and B, u is then a function

$$u = u(\mathbf{r}_{1A}, \mathbf{r}_{1B}, \mathbf{r}_{2A}, \mathbf{r}_{2B})\,. \tag{6.8}$$

The polynomial parametrization of Boys and Handy [124, 320] is therefore extended to molecular systems as

$$u(\mathbf{r}_{1A}, \mathbf{r}_{1B}, \mathbf{r}_{2A}, \mathbf{r}_{2B}) = \sum_{nmlXY} c^{XY}_{nml} \bar{r}^l_{12} \left(\bar{r}^n_{1X}\bar{r}^m_{2Y} + \bar{r}^m_{1Y}\bar{r}^n_{2X}\right)\,, \tag{6.9}$$

where X, Y are iterating over the atoms A and B and $\bar{r}_{iX}$ is the transformed distance $r_{iX} = |\mathbf{r}_{iX}|$ between electron i and nucleus X and $r_{12} = |\mathbf{r}_{1X} - \mathbf{r}_{2X}|$ the inter-electronic distance. The concrete form is then given by the values the exponents n, m and l assume. This functional form of the correlator including $\bar{r}$ and the optimized parameters c_{nml}^{XY} used in the integral generation for the results presented here have been suggested and supplied by Hongjun Luo and implemented for integral calculation by Aron Cohen and the author. Individually optimized factors were used for each geometry.

Essential for removing the cusp condition is the transformation of the electron-electron and electron-nuclei distances $r \mapsto \bar{r}$ for the arguments of τ in order to fulfill eq. (5.6). A sufficient condition is

$$\partial_r \bar{r} \to 1 \qquad (r \to 0) \tag{6.10}$$

$$c_{001} = \frac{1}{2} \tag{6.11}$$

to capture the opposite-spin cusp condition. For the same-spin cusp condition, $c_{001} = \frac{1}{4}$ would be required, thus, to fulfill both, a spin dependency had to be introduced into u. This possibility is shortly explored in section 6.2.1, but as significant overhead is attached to a spin-dependence and spin-independent factors already yield high-accuracy results, it was not considered further beyond the analysis in section 6.2.1.

In previous works [125, 129, 320], the function $\bar{r}$ was commonly chosen as

$$\bar{r} = \frac{r}{1+r}\,, \tag{6.12}$$

which is well-suited for atomic systems [129], but as displayed in Fig. 6.5, can be problematic for application to molecules. The long-range behavior can introduce correlation effects between electrons at different molecules into the integrals, violating size-consistency. This hinders correct description of the wave function particularly at stretched geometry where the long-range behavior of the correlator becomes increasingly inept to capture the now increasingly local correlation. Instead, correlation between electrons at different nuclei is included via the polynomial form (6.9) of $u(\mathbf{r}_1, \mathbf{r}_2)$.

A local form of $\bar{r}$ suggested by Hongjun Luo as

$$\bar{r} = \frac{1}{\alpha}\left(1 - \mathrm{e}^{-\alpha r}\right)\,, \tag{6.13}$$

with additional parameters α can be employed. In fact, two additional parameters are introduced this way, one value α_{ee} for electron-electron and one value α_{en} for electron-

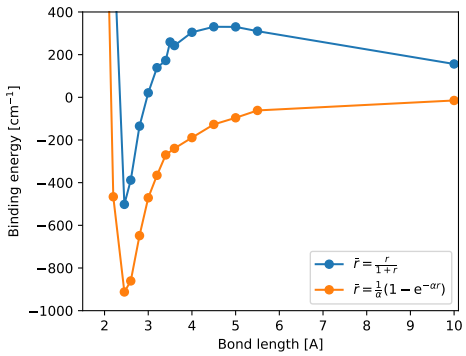


Figure 6.5: Binding curve of Be_2 in a cc-pVTZ basis set for 35-term correlation factors using different $\bar{r}$. All energies are with respect to the dissociated system, the atomic energy for $\bar{r} = \frac{r}{1+r}$ is taken from [129]. Note that both curves do not incorporate counterpoise correction such that a basis set superposition error remains, which shifts the curve for $\bar{r} = \frac{1}{\alpha}(1 - e^{-\alpha r})$ towards the experimental results.

$\bar{r}$	cc-pVDZ	cc-pVTZ
local	-14.6645	-14.6666
long-range	-14.66793	-14.66749

Table 6.5: FCIQMC ground state energies of the Beryllium atom using the local form (6.13) of $\bar{r}$ compared with literature values for the long-range form (6.12) from [129].

nucleus distances.

As the results in Fig. 6.5 indicate the local form is much better suited than the long-range form for capturing the correlation at stretched geometries, which exhibits systematic size inconsistency and fails to reproduce the correct dissociation at large distances. As a side note, also the ground state energy calculation for C_2 at equilibrium distance profits from the exponential $\bar{r}$, yielding an energy of $-75.9156(2)\,\mathrm{H}$ compared to the $-75.9069(3)\,\mathrm{H}$ of the polynomial $\bar{r}$, cf. also Table 5.1 for the literature value of the CBS estimate. The corresponding atomic energies of Be on the other side agree well between the local $\bar{r}$ used here and the long-range version used in [129] once a cc-pVTZ basis set is used, as listed in table 6.5. While the molecular system is highly sensitive to the long-range behavior of $\bar{r}$, this does not hold for the atomic one, strengthening the interpretation of the difficulty in capturing the energy at stretched geometries as spurious long-range correlation induced by the correlator. In this study, the same parameters α are used at all bond distances, an attempt at point-wise optimization only made a very minor influence, but future investigations might also use point-wise optimize these.

The correlation factors were obtained from single reference VMC, one might also try to bootstrap the correlation factors in a self-consistent formalism where the FCIQMC ground state wave function is used as a trial wave function in VMC to obtain re-optimized factors. For Be_2 in a cc-pVTZ basis set with polynomial $\bar{r}$, using re-optimized factors supplied by Hongjun Luo without self-consistency only changes the ground state energy from $-29.3363(3)\,\mathrm{H}$ to $-29.3359(5)\,\mathrm{H}$, so only a minor effect can be expected if any at all. This might warrant future research to increase accuracy even further, but has not

been considered further so far.

Using a more refined polynomial expansion of the correlation factor suggested by Hongjun Luo including 120 terms for the molecule and 44 terms for the atom can lead to a greatly improved binding energy with a cc-pVTZ basis set at equilibrium distance of 891 cm^{-1}compared to the 786 cm^{-1}obtained from the 35 term (17 term for the atom) correlation factor used here, with a manageable cost increase of the integral calculation which can be offset by parallelization. The best empirical estimates for the binding energy at equilibrium distance are 929(2) cm^{-1}[323] and 934.9(0.4) [324], while the most accurate ab-initio estimates to date are 935(10) cm^{-1}[332] coming from CCSD(T) in an aug-cc-pCV7Z basis set with FCI correction from an aug-cc-pVQZ basis set and relativistic corrections and 944(10) cm^{-1}[333] from DMRG calculations extrapolated and including corrections for basis set incompleteness, basis set superposition error and relativistic effects. The high quality of the ST-FCIQMC results given the small basis set used is very promising for future applications, especially given that accurate calculations are much cheaper in smaller basis sets.

Spin dependent correlators

As mentioned above, already Pack and Brown proved that the cusp condition depends on the spin of the involved electrons [312], which would raise the need for a spin-dependent u [316]. While conceptually not different, this induces a significant resource overhead as the operators $\hat{K}$ and $\hat{L}$ come in two flavors each then, and the opposite-spin version of the tensor L has reduced symmetry, since there now is an electron distinguishable by spin.

The spin-dependent cusp condition could be imposed by a posterior multiplication of u with a factor $\frac{1}{2}$ for same spin electrons (an approach suggested by David Tew), but even though this is conceptually very simple, it induces significant overhead in terms of computation, memory and implementation. The permutational symmetry of L is reduced from 6-fold to 2-fold, tripling the memory requirement, additionally, K has to be separated into terms linear in u and those quadratic in u, both of which have to be available at run-time, tripling the cost of the two-electron integrals, since for spin-independent u, the standalone matrix K is not required at run-time. The evaluation of three-body matrix elements is also slowed down by a factor of three, as they now have

to be computed as

$$\left\langle \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_c^\dagger \hat{a}_k \hat{a}_j \hat{a}_i D \,\middle|\, \hat{L} \,\middle|\, D \right\rangle = \\ s(a,b,c)L(a,b,c,i,j,k) + s(b,a,c)L(b,a,c,j,i,k) + s(c,a,b)L(c,a,b,k,i,j) \tag{6.14}$$

with the function L from Eq. (6.7) and the spin-dependent factor

$$s(x,y,z) = \begin{cases} \frac{1}{4} & \text{if } \sigma(x) = \sigma(y) = \sigma(z) \\ 1 & \text{if } \sigma(z) \neq \sigma(x) \neq \sigma(y) \\ \frac{1}{2} & \text{else} \end{cases} . \tag{6.15}$$

Problematic with the a posteriori adaptation is however that the correlation factor is not optimized in the presence of the prefactor. Separate optimization of the factors for same- and opposite spin would further increase the overhead, as separate functions L for same- and opposite spin case would be required in equation (6.14).

Furthermore, as Filippi and Umrigar noted, the correlator will then introduce spin-contamination [340], to avoid which, they suggest discarding the same-spin cusp condition, which is justified by the electron density vanishing at the coalescence of two same-spin electrons due to the Pauli exclusion principle. Hence, spin-independent u targeting the opposite-spin cusp condition are used here.

6.2.2 Basis set dependence

The similarity transformation is apt to remove the cusp of the wave function, but the right eigenvector $|\phi\rangle$ of $\hat{H}_{st}$ for which the FCIQMC now solves has to be represented in some finite basis to be treatable within FCIQMC, and the arguments in section 5.1 for a finite basis set error still hold. The dependency on the selection of $\bar{r}$ was seen in Fig. 6.5, the convergence with respect to the basis set used in the FCIQMC shall be considered next.

Since the potential energy curve is obtained with respect to the atomic energy, the comparability of molecular and atomic basis set is of a concern here, where a potential basis set superposition error has to be taken into account. The counterpoise correction [341, 342] is used to consider these, which uses a molecular basis set in the presence of an uncharged ghost atom with the same geometry as the molecule to describe the atom. This requires a dedicated evaluation of the corrected atomic energy at each point in the potential energy curve. While converging the FCIQMC for these is extremely cheap,

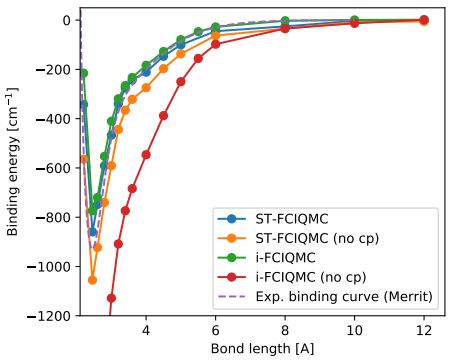


Figure 6.6: Potential energy curve of Be_2 in an aug-cc-pVTZ basis set calculated with (ST-FCIQMC) and without (i-FCIQMC) similarity transform, both with and without counterpoise correction. A 35-term correlation factor was used. The corresponding total energies are given in Table 6.6.

the basis set is nevertheless of the same size as the molecular basis set, thus requiring the same effort for calculating and storing the three-body interaction terms as for the molecular calculations.

In comparison to standard i-FCIQMC results, the basis set superposition error appearing in ST-FCIQMC is extremely small, as can be seen in Fig. 6.6, where the corrected and uncorrected potential energy curves for i-FCIQMC and ST-FCIQMC are compared. While the standard i-FCIQMC suffers from massive basis set superposition error which is of the order of magnitude of the binding energy itself, the ST-FCIQMC yields already reasonable results even without counterpoise correction. After accounting for the basis set superposition error, the i-FCIQMC curve also reasonably agrees to the experimental results, though it does still underestimate the binding energy more than ST-FCIQMC. The total ground state energies, which are given in Table 6.6 (without counterpoise correction), reveal that i-FCIQMC still massively relies on error cancellation between atom and dimer, which could become problematic as the convergence with respect to basis set is typically faster for atoms than for molecules. The ST-FCIQMC energy at equilibrium distance of $-29.33835(2)$ H on the other hand is remarkably close to the high-accuracy result of -29.338657 H obtained by Sharma *et al.* by means of extrapolating DMRG in a cc-pCVQZ-F12 basis set [333], emphasizing that accurate total energies can be obtained by means of similarity transform even in moderate basis sets. Note that in the DMRG energy reported in Ref. [333] is at a bond distance of 2.45 A, while here, the equilibrium distance given by Merritt [323] of 2.4536 A was used, so a slight discrepancy can be expected. These values agree excellently with the estimated CBS of -29.3389 H obtained from the empirically determined well depth [323, 324] and the extrapolation of the atomic energy from experimental data [321].

An evaluation of the counterpoise correction and comparison between the corrected and uncorrected ST-FCIQMC energies displayed in Fig. 6.7 reveals a significant effect at

Table 6.6: Ground state energies of Be_2 in an aug-cc-pVTZ basis set in Hartree calculated with standard i-FCIQMC and ST-FCIQMC, using $4{\cdot}10^6$ walkers and a semi-stochastic core space of 5000 determinants. Given in brackets is the statistical error, which tends to be larger for ST-FCIQMC due to the increased time-step, but to a manageable extent. The distance of 2.4536 A corresponds to the equilibrium distance given in Ref. [323].

Bond distance [A]	i-FCIQMC	ST-FCIQMC
2.2	-29.260330(2)	-29.33612(4)
2.4536	-29.2599633(9)	-29.33835(2)
2.6	-29.258088(9)	-29.33775(3)
2.8	-29.255681(5)	-29.336922(4)
3.0	-29.254033(1)	-29.336241(6)
3.2	-29.253030(4)	-29.33557(7)
3.4	-29.252416(2)	-29.33521(1)
3.6	-29.252007(2)	-29.33501(2)
4.0	-29.251382(3)	-29.33479(3)
4.5	-29.250659(2)	-29.33444(3)
5.0	-29.2500313(4)	-29.3341742(8)
6.0	-29.2493377(4)	-29.33383(1)
8.0	-29.2490481(1)	-29.333707(5)
10.0	-29.248951(2)	-29.33358(1)
12.0	-29.248884(1)	-29.333574(8)

short bond-lengths, considering that the entire potential energy curve resides in a range of less than 1000 $cm^{-1}$and thus a very high accuracy is required to correctly resolve its features. In particular, adding core functions to the basis set essentially removes the basis set superposition error, indicating that the problem is entirely related to the core electrons.

Comparing the corrected and uncorrected potential energy curves using a cc-pVTZ basis set for correlated and frozen core electrons confirms this finding in Fig. 6.7b). Freezing the core electrons in the atomic calculation removes the basis set superposition error, and the resulting potential energy curve using frozen core molecular calculations is close to the counterpoise corrected fully correlated curve. Note that in conventional i-FCIQMC, where, unless freezing the core electrons, the minimum of the potential energy curve is $r_e \leq 2.2\,\text{A}$ instead of the correct $r_e \sim 2.45$ which is fixed when adding the counterpoise correction.

This opens three potential approaches to treat the basis set superposition error, first, by freezing the core electrons, removing the sensitivity to additional degrees of freedom for these, but introducing additional approximation by neglecting core-correlation. Second, by expanding the basis set with core functions, which inherently offer the degrees of

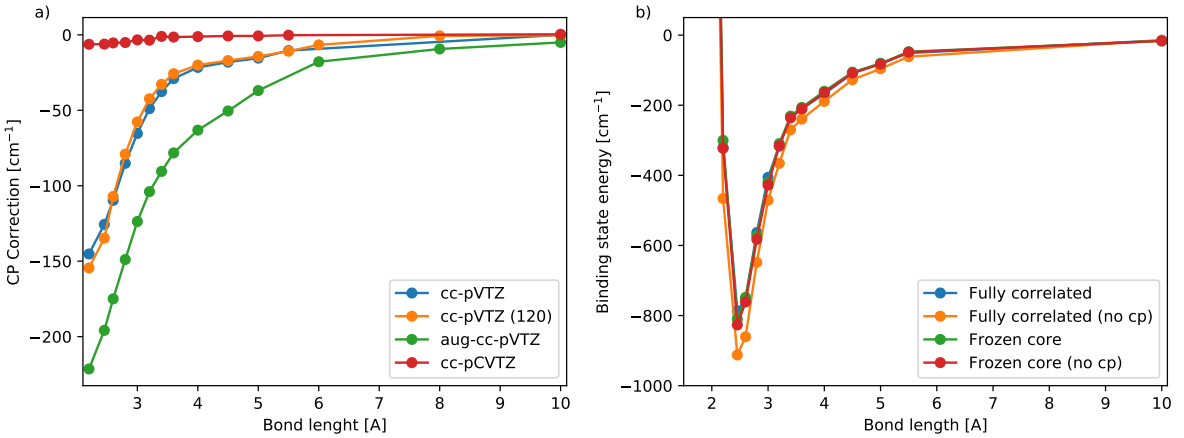


Figure 6.7: a) Counterpoise correction for different basis sets for the ground state potential energy curve of the Beryllium dimer treated with ST-FCIQMC, using the exponential form of $\bar{r}$. The given energies are the correction to the potential energy curve which is twice the difference in the atomic energies. Atomic 17 term correlation factors were used. b) ST-FCIQMC potential energy curves for the Beryllium dimer in a cc-pVTZ basis set with correlated and frozen core electrons in the presence (cp) and absence (no cp) of counterpoise correction. The corresponding atomic energies are also calculated with correlated/frozen core electrons, respectively. 35-term correlation factors were used.

freedom differing between the molecular and atomic basis set for the atom. This however increases the size of the basis set, which is associated with steep memory costs in the presented formulation of ST-FCIQMC and should thus be avoided. Lastly, the atomic energies can be counterpoise-corrected, suppressing, but potentially slightly overshooting, the basis set superposition error [342]. The latter essentially doubles the number of required ST-FCIQMC calculations, but generally at lower cost since atomic calculations are easier to converge. In the following, the counterpoise correction is chosen to address the basis set superposition error for the given reasons.

A direct comparison of the potential energy curves for cc-pVDZ and cc-pVTZ basis sets, with and without core functions, reveals that the cc-pVDZ basis sets are insufficient to converge the curve at all distances short of dissociation, even on a qualitative level. A comprehensive comparison of the potential energy curves for different DZ and TZ basis sets in Fig. 6.8 to the empirical curve fitted by Merritt *et al.* to their experimental findings for the vibrational energy levels [323] clearly demonstrates that a triple zeta basis set is required to correctly describe the bonding. Based on this finding, the counterpoise-corrected aug-cc-pVTZ potential energy curve will be used for comparison of vibrational levels in the next section. An important take here is that even when using the similarity transformed Hamiltonian, the basis set still needs to offer a minimal number of orbital degrees of freedom to correctly describe bonding.

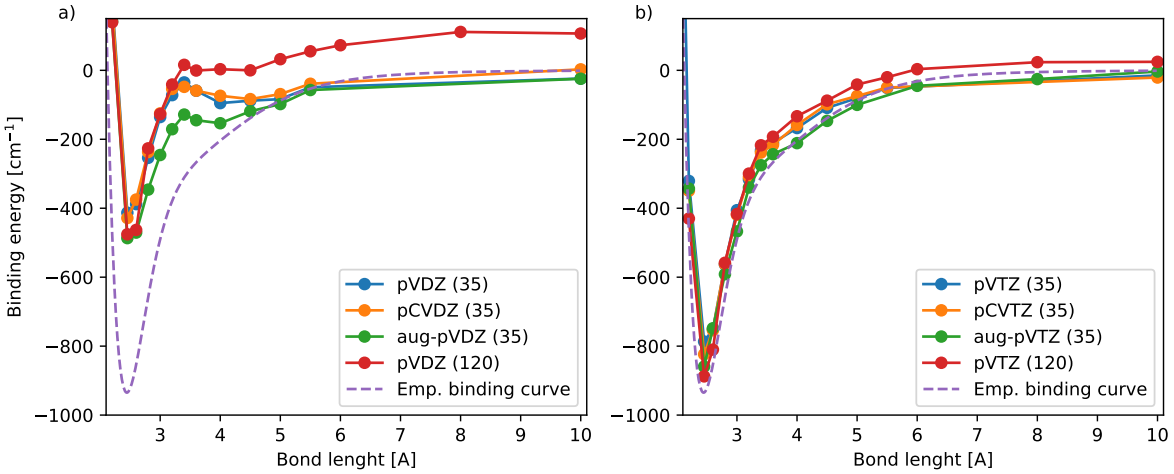


Figure 6.8: a) Ground state potential energy curve of $Be_2$for cc-pVXZ, cc-pCVXZ, and aug-cc-pVXZ (a) X=D, b) X=T) basis sets obtained using ST-FCIQMC. All given with respect to the counterpoise-corrected atomic energy at the same geometry. The number of terms of the correlation factor is given in brackets, the atomic calculations used corresponding 17 (44) term factors. Due to the inclusion of counterpoise correction, the addition of core function in the cc-pCVXZ basis sets has little impact. Additionally, the empirical curve according to Ref. [323] is given for comparison.

6.2.3 Vibrational energy levels

The potential energy curve can now be used to extract vibrational energy levels, which can be compared to experimental results to gauge the basis set convergence. The Beryllium dimer has $D_{\infty h}$ point group symmetry which reduces the number of degrees of freedom for the geometry to one, the inter-nuclear distance R [289]. Working in Born-Oppenheimer approximation, the vibrational states are obtained by solving the effective nuclear Schrödinger equation mentioned in section 2.1.1 using the electronic energies $E_e(R)$ of the potential energy curve. With only a single degree of freedom, the nuclear equation can be readily solved for all eigenvalues using the Numerov method [343] with an interpolated $E_e(R)$.

For the Beryllium dimer, eleven vibrational states are documented [323] with the highest state being extremely weakly bound and thus challenging to capture in ab-initio calculations [324]. The vibrational energies are computed from the ST-FCIQMC ground state energies for the aug-cc-pVTZ basis set with the 35 term Jastrow factor given in Fig. 6.8 using the `VIBROT` module of `OpenMolcas` which fits $E_e(R)$ to the supplied electronic ground state energies using the Numerov method [344]. The resulting vibrational energies show good agreement with the experimental observation, with deviation below 30 cm^{-1}for the first four states and up to 100 cm^{-1}for the higher levels, which can be attributed to the underestimation of the well depth in ST-FCIQMC by ~ 70 cm^{-1}, which

Table 6.7: Energy levels of vibrational states obtained from the ST-FCIQMC ground state potential energy curve, equilibrium distance r_e, and well depth D_e in comparison to the experimentally observed vibrational levels and fitted parameters r_e and D_e by Merritt *et al.* [323].

	$\nu = 0$	1	2	3	4	5	6	7	8	9	10
FCIQMC											
$G(\nu)$	149	363	512	619	671	718	759	789	809	821	826
$\Delta G_{\nu+\frac{1}{2}}$	213	149	107	51	47	41	30	20	11	6	
r_e [A]	2.459										
D_e	859.7										
Expt.											
$G(\nu)$	122.3	345.0	519.4	640.4	717.1	773.8	821.1	860.0	890.5	912.2	924.9
$\Delta G_{\nu+\frac{1}{2}}$	226.6	174.5	121.0	76.7	56.7	47.3	38.9	30.5	21.7	12.7	
r_e [A]	2.453										
D_e	929.7										

indicates that the complete basis set limit is not fully reached to spectroscopic accuracy yet. In particular, the existence of a weakly bound $\nu = 10$ state is correctly predicted, indicating the potential energy curve qualitatively correctly captures the spectrum of the molecule. This demonstrates that the basis set error can be overcome to a high degree by similarity transformation at a low computational overhead, albeit with additional memory costs.

6.3 Outlook: Dynamic properties from ST-FCIQMC

The consideration of the evaluation of dynamic properties by FCIQMC in Chapter 4 made clear that basis set convergence is an issue there, likely even more so than for ground state properties. A potential solution to this could be the combination of complex-time FCIQMC with ST-FCIQMC to address basis set issues. The non-hermiticity of $\hat{H}$ has important implications here, however, since the real-time evolution itself is not unitary anymore for the similarity-transformed Hamiltonian $\hat{H}_{st}$, which manifests itself in the property that the backwards evolution $t \mapsto -t$ additionally requires the non-trivial inversion $\hat{\tau} \to -\hat{\tau}$.

Given the ground state $|\psi_0\rangle$ a of $\hat{H}$, the evaluation of Green's functions thereby also requires the dual vector

$$\langle \phi_0 | = \langle \psi_0 | \, \mathrm{e}^{\hat{\tau}} \,, \tag{6.16}$$

adding to the complexity of the method. This could in practice be realized by a separate

FCIQMC calculation solving for the left eigenvector $\langle\phi_0|$ of $\hat{H}_{st}$, the solution of which could then be used in the evaluation of the overlaps yielding the Green's function. As the simple example in Fig. 6.9 shows, this can in principle yield reasonable results. For the FCI eigenstate, the expected overlap would be constant $\langle\psi_0(0)\,|\,\psi_0(t)\rangle = 1$, to which the deviation is small and can be well described by a minor offset in energy, as the fit in Fig. 6.9 demonstrates, attributing it to the energy zero not being exactly at the ground state energy. The stochastic noise in the real part appears more severe, which might be due to the real part being quadratic in t to lowest order, such that only the higher orders of the propagator which are more difficult to sample contribute. This is surely only a minimalistic example which does not strictly require knowledge of the right eigenvector, but it is indicating that the ST-FCIQMC could be combined with the time-dependent FCIQMC formalism elaborated in Chapters 3 and 4.

The practicability of that approach could be the subject of future research, potentially allowing to treat the dynamic properties with greatly reduced basis set dependency.

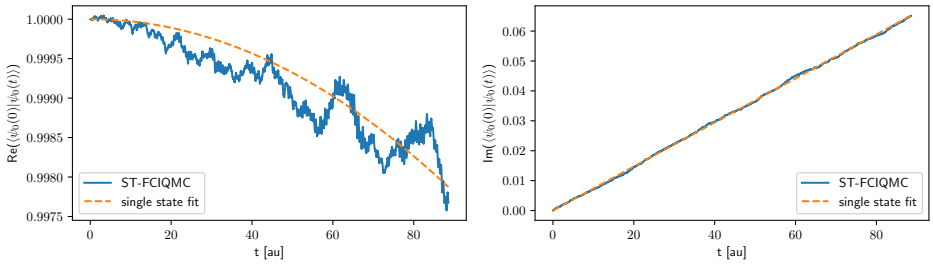


Figure 6.9: Real-time evolution of the overlap $\langle\psi_0(0)\,|\,\psi_0(t)\rangle$ with ST-FCIQMC of the ground state of Be in a cc-pVTZ basis set. The displayed overlap is averaged over 3 replicas, and still exhibits stochastic noise on the scale of the plot. The walker number scales up from 10^6 to $1.5\cdot10^7$ over the full calculation. The energy zero is set to $E_R = -14.6672\,\mathrm{H}$, which is the average projected energy of $|\psi_0\rangle$. The dotted line is a fit with the theoretical overlap of an eigenstate $\langle\psi_0(0)\,|\,\psi_0(t)\rangle = \mathrm{e}^{-iEt}$ yielding $E = 0.74\,\mathrm{mH}$. In particular, $E > 0$ indicates that the state $|\psi_0\rangle$ has an energy slightly above the energy zero of E_R

Chapter 7

Summary and Conclusion

In this thesis, new developments of the FCIQMC algorithm were explored, opening new fields of application to this stochastic FCI method.

A novel approach to the time-dependent Schrödinger equation for strongly correlated electron systems was introduced, giving FCIQMC a way of calculating spectral functions. The FCIQMC framework of a sparsely sampled wave function in Slater determinant space and Carlo evaluation of the action of the Hamiltonian was used to numerically calculate the time-dependent wave function. A complex-time formalism was employed to deal with the steep increase in computational cost over time. By switching from a unitary real-time description to the propagation alongside a complex contour, stability of the numerical calculation is maintained at the cost of an exponentially decaying time-signal due to the energy-dependent exponential damping imposed by the imaginary time part of the propagation. Coupling the complex time contour to the sparse representation of the wave function, results in a systematically improvable scheme where the imaginary part of time vanishes in the FCI limit of FCIQMC. From the complex-time wave function, corresponding Green's functions can then be obtained and by means of analytic continuation, the real-frequency spectral function is recovered.

This method has been applied to model and ab-initio systems to demonstrate its applicability and has been benchmarked against established algorithms and experimental results.

At the example of a half filled 18-site Hubbard model, the method was directly comparison to the Lanczos exact diagonalization algorithm, highlighting its memory efficiency and demonstrating the ability to accurately reproduce spectral functions at the FCI level. The scalability beyond the application range of deterministic FCI methods was demonstrated at the example of a doped 24-site Hubbard model, where the method

has the advantage of systematic improvability towards real-time evolution compared to other, purely imaginary-time, stochastic methods.

At the examples of the Beryllium and Carbon atoms as well as the Carbon dimer and a CuO_2 cluster, the new method has been tested on ab-initio systems to calculate photoelectron spectra and characterize excitations. For the first three, the resulting excitation energies were compared to both experimental data and analogous results obtained using excited-state FCIQMC, while the last one served to illustrate how the method can be used to characterize excitations. Here, the time-dependent formalism has the advantage over direct excited state calculations of being able to target specific excitations, even if the contributing states are higher excited states.

Future applications could lie in core-electron spectroscopy, where that advantage is particularly useful, as well as the treatment of time-dependent Hamiltonians such as strong dynamic fields. For the latter, further investigations might be required regarding the impact of a complex time evolution, also exploring potential modifications to the method. A challenge of the presented method is its high computational cost in conjunction with high walker numbers required for long-time calculations.

A persistent issue in many ab-initio methods, including the new complex-time FCIQMC, is the basis set error. An explicitly correlated ansatz in the form of ST-FCIQMC to accelerate basis set convergence of the imaginary-time FCIQMC method was therefore addressed, outlining the algorithmic details on the FCIQMC side.

The application of the ST-FCIQMC for molecular systems was demonstrated mainly at the example of the Beryllium dimer, for which the binding curve was constructed for different basis sets, obtaining highly accurate results with small basis sets. Compared to experiment and highly involved calculations with very large basis sets, the resulting dissociation energy is slightly underestimated, but the entire binding curve can be reproduced at all bond distances efficiently even with small basis sets, making these calculations computationally cheap. The vibrational spectrum extracted from the calculated curve is in reasonable agreement with the experimental data, capturing all vibrational states within the very small energy range of the curve.

This demonstrates that the ST-FCIQMC is apt to produce very accurate energies approaching the complete basis set limit at comparably low cost. The major bottleneck of the current scheme is the treatment of the integrals of the three-body interaction. These are expensive to calculate and store, imposing tight limits on the application to larger orbital spaces with the available hardware.

Here, further developments regarding the storage of the three-body interaction can be

expected, allowing to access larger molecules and basis sets. Furthermore, the interface with the time dependent method can be further enhanced, and possibilities of using the transcorrelated Hamiltonian for the calculation of Green's function could be subject of future research.

Appendix A

Notes on the performance of the NECI FCIQMC implementation

Any algorithm is just as performant as its implementation, and analyzing the performance of the NECI implementation by profiling and scaling analysis can give insight in the computational cost of the single steps of FCIQMC in practice and provide measures for its parallelizability. Here, in a first step, the role of the excitation generation is considered, which turns out to be decisive for efficiency, cf. also Sections 2.4.5 and 6.1.1 , before turning towards the parallel efficiency which is essential for large scale application.

The MPCDF HPC Performance Monitoring System [345] reports an arithmetic intensity [346, 347] of NECI for the cases considered in Fig. A.1 of ~ 0.05, and while the precise value is system dependent, the order of magnitude is reproducible across different applications. This means, NECI is strongly memory bandwidth bound. The monitoring system also reports $\sim 60\%$ cache misses, which adds up to the required memory bandwidth. In combination, this strongly indicates that NECI would profit from more efficient memory access and data reuse, which is of course challenging given the stochastic nature of the FCIQMC method.

Of the steps of the FCIQMC scheme, it is the spawning step which is most effortive, in all examples in this section except for the highest scaling in Fig. A.2 it amounts for more than 50% of the iteration time, in many ST-FCIQMC calculations of Section 6 even for more than 75% up to over 90%. The precise timings strongly depend on the system and choice of parameters, so no general assessment can be made, but the prevalence of the spawning step seems common to most calculations. Furthermore, the death step

is extremely cheap and is of little concern from the performance point of view. The annihilation step can contribute significantly such as in the example of Fig. A.1, but often amounts for less than 20% of the total iteration time, such as in the example of Fig. A.2. A greatly variable cost is associated to the communication of the spawns, this is studied in Section A.2.

As a side note, the overhead of the semi-stochastic feature is negligible unless extremely large semi-stochastic spaces are considered, for benchmark calculations of the Chromium dimer in a cc-pVDZ basis set on 480 cores of Intel Xeon E5-269 v4@2.20 GHz processors, with $N_w = 10^8$, even for semi-stochastic spaces of dimension 10^6, only $\sim 2.3\%$ of the iteration time is spent on the deterministic part of the propagation, including communication. Only for larger semi-stochastic spaces, the communication becomes problematic, increasing to $\sim 31.1\%$ of the iteration time for a space of dimension $3 \cdot 10^7$ on 1280 cores. That means, semi-stochastic spaces of dimension of order of magnitude 10^6 can and should be routinely used if the FCIQMC wave function supports it. The developments that enable this scaling are the implementation of the fast generation method [154] by Nick Blunt and the optimized parallelization by the author.

A.1 Excitation generation

Since excitation generation is the computationally most demanding part, it should be the focus of optimization.

Consider the minimal example of Dodekacene in a (50,50) active space (the same system as in Table A.2) with $N_w = 2 \cdot 10^5$ on 8 cores of an Intel i7-4790 CPU@3.6 GHz using the Intel Vtune Amplifier [327]. The profiling reveals that the selection of the double excitations in the Cauchy-Schwartz excitation generator constitutes the bottleneck, amounting for 63% of the total time, with the matrix element access making up for 24% and the rest mainly coming from looping over electrons and orbitals. Since only 0.9% of CPU time is spent on single excitations, from a performance point of view, considering double excitations is sufficient. So if an excitation generator ran in constant time without evaluating matrix elements on the fly, and still use an efficient weighting, it could easily double the speed of a calculation.

A promising candidate to achieve this is the Heat-Bath sampling by Holmes *et al.* [155], but in its original form, it is still computationally as expensive as the Cauchy-Schwartz scheme. Removing the expensive on-the-fly evaluation of a matrix element based probability distribution leads to the PCHB sampling addressed in Section 2.4.5 and Ref. [196].

Table A.1: Efficiency for the Be_2 i-FCIQMC calculations considered in Fig 6.4 for different excitation generators.

Excit. gen	Efficiency
CS	27979.12
PCHB	59337.58
PCPP	1100.78
Uniform	1215.86

Table A.2: Efficiency, average time per iteration (TPI) and acceptance rate for the Cauchy-Schwartz and PCHB excitation generator for an exemplary calculation of Dodekacene in a (50,50) active space with $N_w = 2 \cdot 10^7$ on 64 cores of two nodes using Intel(R) Xeon(R) Gold 6138 processor @2.00 GHz with 20 CPU cores each. The active space and integrals have been supplied by Giovanni Li Manni.

Excit. gen	Efficiency	TPI [s]	Acc. Rat
CS	0.01731	1.572	0.06037
PCHB	0.09592	0.4014	0.06037

As the efficiency comparisons in Tables A.1 and A.2 show, it is substantially faster while achieving acceptance rates comparable to the Cauchy-Schwartz scheme, cf also Fig. 6.4.

For the Dodekacene example, the PCHB scheme reduces the cost of excitation generation by a factor of ~ 6.5, more than doubling the overall performance in the minimal example. At medium scales such as in Table A.2, the speedup is even larger, as the time spent in excitation generation is larger, while achieving comparable acceptance rates.

A.2 Parallel efficiency

To measure parallel efficiency, several approaches exist. Here, first consider the weak scaling [348, 349], that is, the parallel performance with constant problem size per CPU core. Since FCIQMC scales linearly [62], the required time per iteration is ideally constant when increasing the number of cores N_p and the walker number N_w at the same rate. In Fig. A.1, the average time per iteration for constant N_w/N_p at different scales is given, showing a steep early rise in the time per iteration between a single node of 40 cores and 16 of these, equaling 640 cores. This has two reasons, one is the steep increase in communication cost, since `NECI` requires a global communication step to distribute the spawn vector across cores, which does not scale up ideally, and becomes a bottleneck when many cores are used. The second reason is a steep increase in the cost of the annihilation step, which however gradually reduces again at larger scales and could not be

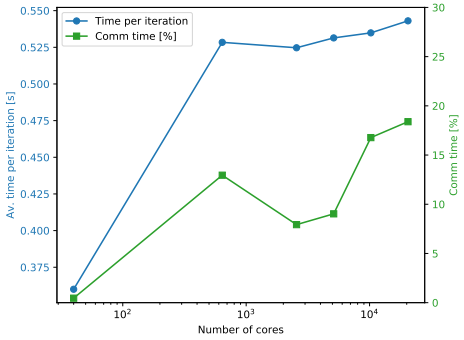


Figure A.1: Average time per iteration and percentage of time spent in communication for NECI FCIQMC calculations of Butadiene in an ANO-L-pVDZ basis set with constant $N_w/N_p = 1.5 \cdot 10^5$ and PCHB excitation generation. The number of cores ranges from 40 to 20480. These calculations were performed on the HPC system Cobra, using Intel Xeon Gold 6148@2.4 GHz processors with 40 CPU cores per node and 100 Gb/s Intel OmniPath node interconnect, the integral file has been supplied by Ali Alavi.

reproduced with other systems, so this might not be a principal bottleneck. So while the total efficiency here is only $\sim 66\%$, this is almost entirely due to the scaling from 1 to 16 nodes, adding further computational resources then scales extremely well, the efficiency from 16 to 512 nodes is then $\sim 97\%$. In that sense, NECI greatly profits from additional computational resources at large scale, as these allow increasing N_w basically linearly with the number of CPU cores. Nevertheless, reducing the substantial communication overhead between the nodes would further increase the performance, and enable the code to run on even larger machines.

The other parallel efficiency measure considered here is the strong scaling [349, 350], that is, the speedup achieved by adding additional CPU cores for a fixed problem size. This study has been previously partly published in Ref. [196], where the parallel efficiency of NECI has also been considered. Such a scaling for NECI is given in Fig A.2, which shows a reasonable speedup of 20.8 from 640 to 20480 cores, which is $\sim 65\%$ of the ideal speedup of 32. The limiting factor here is clearly the global spawn communication, which becomes increasingly important at large scales simply because it does not speed up at the same rate as the other components, up to the point that the communication with 20480 cores takes more absolute time than with 256 cores, which becomes a limiting factor for the strong scaling. A few technical notes on this profiling: Only the time per iteration was accounted for, not the initialization, which itself scales less efficiently. The given times were directly obtained from NECI internal timers, in particular, the communication time accounts only for the communication of the stochastic spawn vector excluding synchronization. Load imbalance only plays a minor role at all considered scales, see Ref. [196].

At larger scales, the cost of communication becomes more problematic and eventually negates the parallel speedup as seen in Fig. A.2b), which indicates efficient scaling up

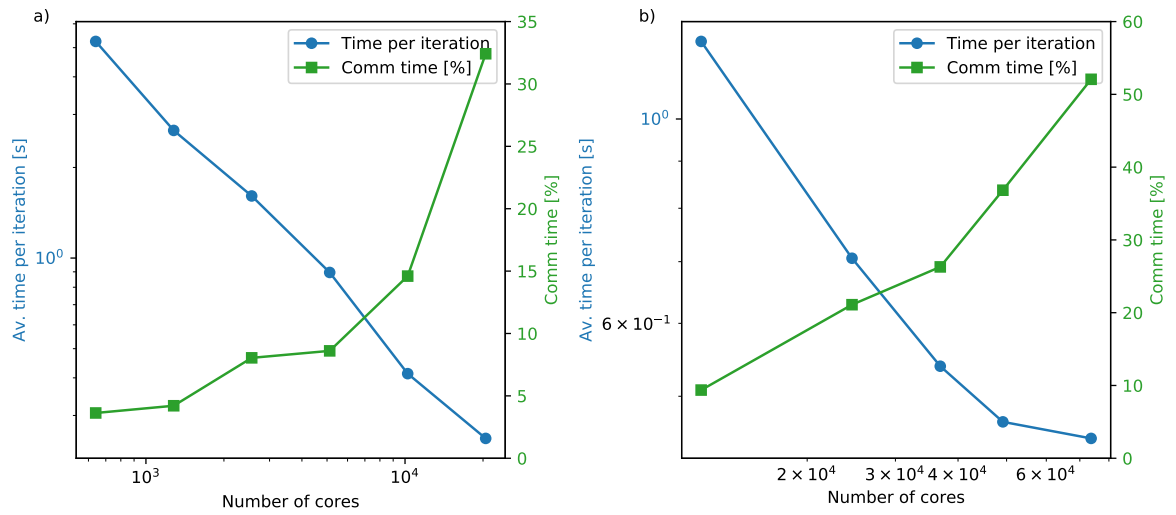


Figure A.2: Average time per iteration and percentage of time spent in communication `NECI` FCIQMC calculations of Cr_2 in a cc-pVDZ basis set. a) With constant $N_w = 1.6 \cdot 10^9$, PHCB excitation generation and a semi-stochastic space of dimension $N_c = 5 \cdot 10^4$. The number of cores ranges from 640 to 20480. These calculations were performed on the HPC system Cobra, using Intel Xeon Gold 6148@2.4 GHz processors with 40 CPU cores per node and 100 Gb/s Intel OmniPath node interconnect. b) With constant $N_w = 6.4 \cdot 10^9$, PCHB excitation generation, no semi-stochastic space and disabled load balancing, the number of cores ranges from 12288 to 73728, the calculations were performed on the HPC system SuperMUC-NG, using Intel Xeon Platinum 8174 processors with 48 CPU cores per node and 100 Gb/s Intel OmniPath node interconnect. the integral file has been supplied by Ali Alavi. The data in a) has partly been published previously in Ref. [196].

to $\sim 3 - 4 \cdot 10^4$ cores, so further optimizations would be required for pushing towards even larger problems. To some extent, the increased cost at higher scales could also be attributed to the network topology of SuperMUC-NG, as the nodes are grouped into islands of 792 nodes, amounting for 38016 cores per island. Communication between islands is less efficient [351], which partly explains the drop in parallel performance once more than one island is used.

Appendix B

Specifications of the Hubbard lattices

The lattice models considered in section 4.1 for the Hubbard Hamiltonian (4.2) are uniquely defined by the set of momenta $\mathbf{k}$, i.e. the sampling of the Brillouin zone, the dispersion relation ϵ, and the lattice vectors defining the periodic boundary conditions. The corresponding data for the 16-site, 18-site and 24-site models used in section 4.1 are given in table B.1, an illustration of the clusters is depicted in Fig. B.1.

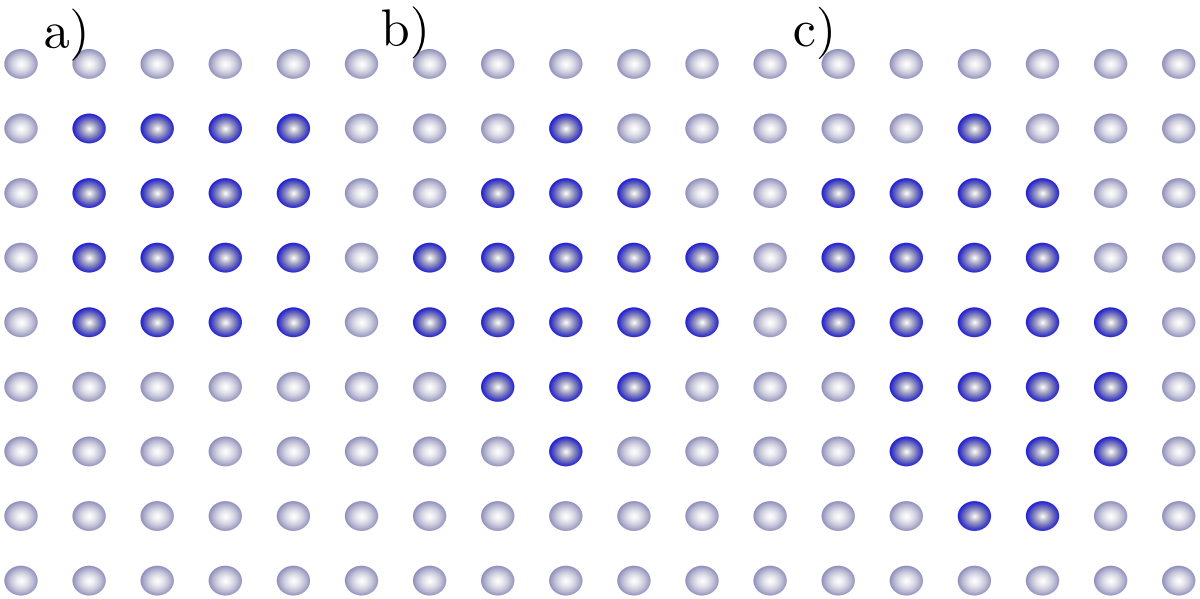


Figure B.1: Illustration of the square lattice clusters considered in section 4.1, a) 16-site cluster, b) tilted 18-site cluster, c) tilted 24-site cluster. The corresponding dispersion relations and lattice vectors are given in table B.1.

Table B.1: Dispersion relation $\epsilon_{\mathbf{k}}$ for the lattices considered in section 4.1. The corresponding lattice vectors are $\{(4,0),(0,4)\}$ (16-site), $\{(3,3),(3,-3)\}$ (18-site) and $\{(3,3),(-5,3)\}$(24-site). This data is taken from the `NECI` output of the corresponding calculations. The 24-site model has been implemented by Werner Dobrautz based on suggestions by Olle Gunnarsson specifically for this application.

16-site		18 site		24-site	
$\mathbf{k}$	$\epsilon_{\mathbf{k}}$	$\mathbf{k}$	$\epsilon_{\mathbf{k}}$	$\mathbf{k}$	$\epsilon_{\mathbf{k}}$
(0,0)	-4	(0,0)	-4	(0,0)	-4
(0,1)	-2	(0,1)	-2	(0,1)	$-2\sqrt{2}$
(0,-1)	-2	(0,-1)	-2	(0,-1)	$-2\sqrt{2}$
(1,0)	-2	(1,0)	-2	(1,0)	$-\sqrt{2}-2\cos\left(\frac{5\pi}{12}\right)$
(-1,0)	-2	(-1,0)	-2	(-1,0)	$-\sqrt{2}-2\cos\left(\frac{5\pi}{12}\right)$
(1,-1)	0	(1,-1)	-1	(1,-1)	$-\sqrt{3}$
(-1,1)	0	(-1,1)	-1	(-1,1)	$-\sqrt{3}$
(-1,-1)	0	(-1,-1)	-1	(-1,-1)	-1
(1,1)	0	(1,1)	-1	(1,1)	-1
(2,0)	0	(1,-2)	1	(1,-2)	$-2\cos\left(\frac{\pi}{12}\right)-2\cos\left(\frac{3\pi}{4}\right)$
(0,2)	0	(1,2)	1	(-1,2)	$-2\cos\left(\frac{\pi}{12}\right)-2\cos\left(\frac{3\pi}{4}\right)$
(-1,2)	2	(2,-1)	1	(0,2)	0
(1,2)	2	(2,1)	1	(0,-2)	0
(2,-1)	2	(2,0)	2	(-1,-2)	$2\cos\left(\frac{\pi}{12}\right)+2\cos\left(\frac{3\pi}{4}\right)$
(2,1)	2	(-2,0)	2	(1,2)	$2\cos\left(\frac{\pi}{12}\right)+2\cos\left(\frac{3\pi}{4}\right)$
(2,2)	4	(0,2)	2	(1,-3)	1
		(0,-2)	2	(2,-2)	1
		(3,0)	4	(-2,0)	$\sqrt{3}$
				(2,0)	$\sqrt{3}$
				(-2,1)	$\sqrt{2}+2\cos\left(\frac{5\pi}{12}\right)$
				(2,-1)	$\sqrt{2}+2\cos\left(\frac{5\pi}{12}\right)$
				(0,3)	$2\sqrt{2}$
				(0,-3)	$2\sqrt{2}$
				(3,-1)	4

List of Figures

List of Tables

Bibliography

[1] G. H. Booth and A. Alavi et. al. Standalone `NECI` codebase designed for FCIQMC and other stochastic quantum chemistry methods. `https://github.com/ghb24/NECI_STABLE`, 2013.

[2] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. McClain, S. Sharma, S. Wouters, and G. K.-L. Chan. PySCF: The python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):1340, 2018.

[3] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science Engineering*, 9(3):90–95, 2007.

[4] A. Einstein. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322(6):132–148, 1905.

[5] H. Rubens and F. Kurlbaum. Anwendung der methode der reststrahlen zur prüfung des strahlungsgesetzes. *Annalen der Physik*, 309(4):649–666, 1901.

[6] M. Planck. Über das Gesetz der Energieverteilung im Normalspektrum. *Annalen der Physik*, 4(553):1, 1901.

[7] W. Gerlach and O. Stern. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik*, 9(1):349–352, 1922.

[8] A. H. Compton. A quantum theory of the scattering of X-rays by light elements. *Physical Review*, 21(5):483–502, 1923.

[9] L. De Broglie. *Recherches sur la théorie des quanta*. PhD thesis, Migration-université en cours d'affectation, 1924.

[10] M. Le Bellac. *Quantum physics*. Cambridge University Press, Cambridge, 2011.

[11] W. Nolting. Viel-Teilchen-Modellsysteme. In *Grundkurs Theoretische Physik 7: Viel-Teilchen-Theorie*, chapter 2, pages 35–104. Springer-Verlag, Heidelberg, 2014.

[12] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049, 1926.

[13] H. Shull and G. Hall. Atomic units. *Nature*, 184(4698):1559, 1959.

[14] G. E. Uhlenbeck and S. Goudsmit. Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons. *Die Naturwissenschaften*, 13(47):953–954, 1925.

[15] P. A. M. Dirac and R. H. Fowler. The quantum theory of the electron. *Proceedings of the Royal Society of London A*, 117(778):610–624, 1928.

[16] M. Fierz. Über die relativistische Theorie kräftefreier Teilchen mit beliebigem spin. *Helvetica Physica Acta*, 12:3–37, 1939.

[17] W. Pauli. The connection between spin and statistics. *Physical Review*, 58(8):716–722, 1940.

[18] I. Duck, W. Pauli, and E. Sudarshan. *Pauli and the spin-statistics theorem*. World Scientific, Singapore, 1997.

[19] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar. Sign problem in the numerical simulation of many-electron systems. *Physical Review B*, 41(13):9301–9307, 1990.

[20] W. Pauli. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik*, 31:765–783, 1925.

[21] C. Kittel, J. M. Gress, and A. Lessard. *Einführung in die Festkörperphysik*. Oldenbourg München, München, 1969.

[22] A. M. Zagoskin. *Quantum theory of many-body systems*. Springer, Heidelberg, 1998.

[23] F. Schwabl. *Quantenmechanik (QM I): Eine Einführung*. Springer, Berlin, 2007.

[24] T. Helgaker, P. Jørgensen, and J. Olsen. The standard models. In *Molecular Electronic-Structure Theory*, chapter 5, pages 142–200. John Wiley & Sons, Ltd, Chichester, 2014.

[25] R. A. Friesner. Ab initio quantum chemistry: Methodology and applications. *Proceedings of the National Academy of Sciences*, 102(19):6648–6653, 2005.

[26] J. Hubbard. Electron Correlations in Narrow Energy Bands. *Proceedings of the Royal Society A*, 276(1365):238, 1963.

[27] W. Heisenberg. Zur theorie des Ferromagnetismus. *Zeitschrift für Physik*, 46:619–636, 1928.

[28] P. W. Anderson. Absence of diffusion in certain random lattices. *Physical Review*, 109(5):1492–1505, 1958.

[29] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Physical Review*, 108(5):1175, 1957.

[30] T. Helgaker, P. Jørgensen, and J. Olsen. Second quantization. In *Molecular Electronic-Structure Theory*, chapter 1, pages 1–33. John Wiley & Sons, Ltd, Chichester, 2014.

[31] U. Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96 – 192, 2011. January 2011 Special Issue.

[32] S. R. White and R. L. Martin. Ab initio quantum chemistry using the density matrix renormalization group. *The Journal of Chemical Physics*, 110(9):4127–4130, 1999.

[33] G. K.-L. Chan and D. Zgid. The density matrix renormalization group in quantum chemistry. *Annual Reports in Computational Chemistry*, 5:149–162, 2009.

[34] T. Helgaker, W. Klopper, and D. P. Tew. Quantitative quantum chemistry. *Molecular Physics*, 106(16-18):2107–2143, 2008.

[35] W. Kohn and L. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

[36] R. Parr. Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry*, pages 5–15. Springer, 1980.

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

[38] K. Burke. Perspective on density functional theory. *The Journal of Chemical Physics*, 136(15):150901, 2012.

[39] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169, 1996.

[40] Z. Zhao, M. Marsman, F. Wende, and J. Kim. Performance of hybrid MPI/OpenMP VASP on Cray XC40 based on Intel Knights landing many integrated core architecture. *CUG Proceedings*, 2017.

[41] A. D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140(18):18A301, 2014.

[42] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618–622, 1934.

[43] J. Pople, R. Krishnan, H. Schlegel, and J. S. Binkley. Derivative studies in Hartree-Fock and Møller-Plesset theories. *International Journal of Quantum Chemistry*, 16(13):225–241, 1979.

[44] K. Hirao. Multireference Møller—Plesset method. *Chemical Physics Letters*, 190(3-4):374–380, 1992.

[45] R. Bishop. An overview of coupled cluster theory and its applications in physics. *Theoretica chimica acta*, 80(2):95, 1991.

[46] P. Knowles, C. Hampel, and H.-J. Werner. Coupled cluster theory for high spin, open shell reference wave functions. *The Journal of Chemical Physics*, 99(7):5219, 1993.

[47] T. D. Crawford and H. F. Schaefer. An introduction to coupled cluster theory for computational chemists. *Reviews in computational Chemistry*, 14:33–136, 2000.

[48] R. J. Bartlett and M. Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291, 2007.

[49] S. R. White. Density matrix formulation for quantum renormalization groups. *Physical Review Letters*, 69(19):2863, 1992.

[50] S. White. Density-matrix algorithms for quantum renormalization groups. *Physical Review B*, 48(14):10345, 1993.

[51] U. Schollwöck. The density-matrix renormalization group. *Reviews of modern Physics*, 77(1):259, 2005.

[52] B. O. Roos, P. R. Taylor, and P. E. Sigbahn. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics*, 48(2):157–173, 1980.

[53] B. Roos. The complete active space SCF method in a fock-matrix-based super-CI formulation. *International Journal of Quantum Chemistry*, 18(S14):175, 1980.

[54] P. Siegbahn, J. Almlöf, A. Heiberg, and B. Roos. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *The Journal of Chemical Physics*, 74(4):2384, 1981.

[55] F. Illas, J. Rubio, J. Ricart, and P. Bagus. Selected versus complete configuration interaction expansions. *The Journal of Chemical Physics*, 95(3):1877–1883, 1991.

[56] R. J. Harrison. Approximating full configuration interaction with selected configuration interaction and perturbation theory. *The Journal of Chemical Physics*, 94(7):5021–5031, 1991.

[57] A. A. Holmes, N. M. Tubman, and C. Umrigar. Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. *Journal of chemical Theory and Computation*, 12(8):3674–3680, 2016.

[58] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *Journal of chemical Theory and Computation*, 13(4):1595–1604, 2017.

[59] G. H. Booth, A. J. W. Thom, and A. Alavi. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *The Journal of Chemical Physics*, 131(5):054106, 2009.

[60] D. Cleland, G. H. Booth, and A. Alavi. Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 132(4):041103, 2010.

[61] F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar. Semistochastic projector Monte Carlo method. *Physical Review Letters*, 109(23):230201, 2012.

[62] G. H. Booth, S. D. Smart, and A. Alavi. Linear-Scaling and Parallelisable Algorithms for Stochastic Quantum Chemistry. *Molecular Physics*, 112:1855–1869, 2014.

[63] S. R. White. Strongly correlated electron systems and the density matrix renormalization group. *Physics Reports*, 301(1-3):187–204, 1998.

[64] M. Imada and T. Miyake. Electronic structure calculation by first principles for strongly correlated electron systems. *Journal of the Physical Society of Japan*, 79(11):112001, 2010.

[65] G. Li Manni, S. D. Smart, and A. Alavi. Combining the complete active space self-consistent field method and the full configuration interaction quantum Monte Carlo within a super-CI framework, with application to challenging metal-Porphyrins. *Journal of Chemical Theory and Computation*, 12(3):1245–1258, 2016.

[66] G. Li Manni and A. Alavi. Understanding the mechanism stabilizing intermediate spin states in Fe(II)-Porphyrin. *Journal of Physical Chemistry A*, 122(22):4935–4947, 2018.

[67] G. Li Manni, D. Kats, D. P. Tew, and A. Alavi. Role of valence and semicore electron correlation on spin gaps in Fe(II)-Porphyrins. *Journal of Chemical Theory and Computation*, 15(3):1492–1497, 2019.

[68] D. Cleland, G. H. Booth, C. Overy, and A. Alavi. Taming the first-row diatomics: A full configuration interaction quantum Monte Carlo study. *Journal of Chemical Theory and Computation*, 8(11):4138–4152, 2012.

[69] K. Ghanem, A. Y. Lozovoi, and A. Alavi. Unbiasing the initiator approximation in full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 151(22):224108, 2019.

[70] N. A. Bogdanov, G. Li Manni, S. Sharma, O. Gunnarsson, and A. Alavi. New superexchange paths due to breathing-enhanced hopping in corner-sharing cuprates. `https://arxiv.org/abs/1803.07026`, 2018.

[71] J. J. Shepherd, G. E. Scuseria, and J. S. Spencer. Sign problem in full configuration interaction quantum Monte Carlo: Linear and sublinear representation regimes for the exact wave function. *Physical Review B*, 90(15):155130, 2014.

[72] W. Dobrautz, H. Luo, and A. Alavi. Compact numerical solutions to the two-dimensional repulsive hubbard model obtained via nonunitary similarity transformations. *Physical Review B*, 99(7):075119, 2019.

[73] M. Yang, E. Pahl, and J. Brand. A stochastic approach to the exact diagonalization study of bosonic quantum phase transition. In *11th Annual Dodd-Walls Centre Symposium University of Auckland 27-29 June 2018*, 2018.

[74] C. Homes. High-tc superconductors: Universal scaling relations. In K. J. Buschow, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer, S. Mahajan, and P. Veyssière, editors, *Encyclopedia of Materials: Science and Technology*, pages 1 – 4. Elsevier, Oxford, 2007.

[75] D. W. Turner. A discussion on photoelectron spectroscopy - molecular photoelectron spectroscopy. *Philosophical Transactions of the Royal Society of London A*, 268(1184):7–31, 1970.

[76] F. J. Himpsel and T. Fauster. Probing valence states with photoemission and inverse photoemission. *Journal of Vacuum Science & Technology A*, 2(2):815–821, 1984.

[77] J. Berkowitz. *Photoabsorption, photoionization, and photoelectron spectroscopy*. Academic Press, New York, 2012.

[78] J. J. Pireaux, S. Svensson, E. Basilier, P.-A. Malmqvist, U. Gelius, R. Caudano, and K. Siegbahn. Core-electron relaxation energies and valence-band formation of linear alkanes studied in the gas phase by means of electron spectroscopy. *Physical Review A*, 14(6):2133–2145, 1976.

[79] P. Mills and J. Sullivan. A study of the core level electrons in iron and its three oxides by means of X-ray photoelectron spectroscopy. *Journal of Physics D: Applied Physics*, 16(5):723, 1983.

[80] E. A. McCullough Jr and R. E. Wyatt. Quantum dynamics of the collinear (H, H_2) reaction. *The Journal of Chemical Physics*, 51(3):1253–1254, 1969.

[81] K. Kulander. Collision induced dissociation in collinear H+H_2: Quantum mechanical probabilities using the time-dependent wavepacket approach. *The Journal of Chemical Physics*, 69(11):5064–5072, 1978.

[82] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore. Colloquium: Nonequilibrium dynamics of closed interacting quantum systems. *Review of Modern Physics*, 83(3):863–883, 2011.

[83] W. Nolting. Green-Funktionen. In *Grundkurs Theoretische Physik 7: Viel-Teilchen-Theorie*, chapter 3, pages 105–194. Springer-Verlag, Heidelberg, 2014.

[84] D. Porras, F. Verstraete, and J. I. Cirac. Renormalization algorithm for the calculation of spectra of interacting quantum systems. *Physical Review B*, 73(1):014410, 2006.

[85] N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi. An excited-state approach within full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 143(13):134117, 2015.

[86] A. A. Holmes, C. J. Umrigar, and S. Sharma. Excited states using semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 147(16):164111, 2017.

[87] O. Gunnarsson and K. Schönhammer. Many-body formulation of spectra of mixed valence systems. In *High Energy Spectroscopy*, volume 10 of *Handbook on the Physics and Chemistry of Rare Earths*, chapter 64, pages 103 – 163. Elsevier, 1987.

[88] R. Kosloff. Time-dependent quantum-mechanical methods for molecular dynamics. *The Journal of Physical Chemistry*, 92(8):2087–2100, 1988.

[89] J. L. Krause, K. J. Schafer, and K. C. Kulander. High-order harmonic generation from atoms and ions in the high intensity regime. *Physical Review Letters*, 68(24):3535, 1992.

[90] M. Lewenstein, P. Balcou, M. Y. Ivanov, A. L'huillier, and P. B. Corkum. Theory of high-harmonic generation by low-frequency laser fields. *Physical Review A*, 49(3):2117, 1994.

[91] E. Runge and E. Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52(12):997, 1984.

[92] E. Gross, J. Dobson, and M. Petersilka. Density functional theory of time-dependent phenomena. In *Density functional theory II*, pages 81–172. Springer, 1996.

[93] M. Casida and M. Huix-Rotllant. Progress in time-dependent density-functional theory. *Annual review of physical chemistry*, 63:287, 2012.

[94] L. Hedin. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Physical Review*, 139(3A):A796, 1965.

[95] M. Hybertsen and S. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B*, 34(8):5390, 1986.

[96] F. Aryasetiawan and O. Gunnarsson. The GW method. *Reports on Progress in Physics*, 61(3):237, 1998.

[97] G. Onida, L. Reining, and A. Rubio. Electronic excitations: density-functional versus many-body Green's-function approaches. *Reviev of Modern Physics*, 74(2):601, 2002.

[98] T. J. Park and J. Light. Unitary quantum time evolution by iterative Lanczos reduction. *The Journal of Chemical Physics*, 85(10):5870–5876, 1986.

[99] X. Guan, C. J. Noble, O. Zatsarinny, K. Bartschat, and B. I. Schneider. ALTDSE: An Arnoldi-Lanczos program to solve the time-dependent Schrödinger equation. *Computer Physics Communications*, 180(12):2401–2409, 2009.

[100] T. Helgaker, P. Jørgensen, and J. Olsen. Configuration interaction theory. In *Molecular Electronic-Structure Theory*, chapter 11, pages 523–597. John Wiley & Sons, Ltd, Chichester, 2014.

[101] A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal. Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(04):04005, 2004.

[102] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig. Time-evolution methods for matrix-product states. *Annals of Physics*, 411:167998, 2019.

[103] E. Ronca, Z. Li, C. A. Jimenez-Hoyos, and G. K.-L. Chan. Time-step targeting time-dependent and dynamical density matrix renormalization group algorithms with ab initio hamiltonians. *Journal of chemical theory and computation*, 13(11):5560–5571, 2017.

[104] L.-H. Frahm and D. Pfannkuche. Ultrafast ab initio quantum chemistry using matrix product states. *Journal of Chemical Theory and Computation*, 15(4):2154–2165, 2019.

[105] S. Wouters and D. Van Neck. The density matrix renormalization group for ab initio quantum chemistry. *The European Physical Journal D*, 68(9):272, 2014.

[106] J. Geertsen, M. Rittby, and R. J. Bartlett. The equation-of-motion coupled-cluster method: Excitation energies of Be and CO. *Chemical Physics Letters*, 164(1):57–62, 1989.

[107] J. F. Stanton and R. J. Bartlett. The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *The Journal of Chemical Physics*, 98(9):7029–7039, 1993.

[108] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis. Taming the dynamical sign problem in real-time evolution of quantum many-body problems. *Physical Review Letters*, 115(26):266802, 2015.

[109] B. M. Austin, D. Y. Zubarev, and W. A. Lester. Quantum Monte Carlo and related approaches. *Chemical Reviews*, 112(1):263–288, 2012.

[110] R. N. Silver, D. S. Sivia, and J. E. Gubernatis. Maximum-entropy method for analytic continuation of quantum Monte Carlo data. *Physical Review B*, 41(4):2380–2389, 1990.

[111] J. Gubernatis, M. Jarrell, R. Silver, and D. Sivia. Quantum Monte Carlo simulations and maximum entropy: Dynamics from imaginary-time data. *Physical Review B*, 44(12):6011, 1991.

[112] T. Helgaker, P. Jørgensen, and J. Olsen. Gaussian basis sets. In *Molecular Electronic-Structure Theory*, chapter 8, pages 287–335. John Wiley & Sons, Ltd, Chichester, 2014.

[113] T. Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications in Pure and Applied Mathematics*, 10(2):151, 1957.

[114] C. Schwartz. Importance of angular correlations between atomic electrons. *Physical Review*, 126(3):1015, 1962.

[115] D. P. Carroll, H. J. Silverstone, and R. M. Metzger. Piecewise polynomial configuration interaction natural orbital study of $1s^2$ helium. *The Journal of Chemical Physics*, 71(10):4142–4163, 1979.

[116] W. Kutzelnigg. r12-dependent terms in the wave function as closed sums of partial wave amplitudes for large l. *Theoretica chimica acta*, 68(6):445–469, 1985.

[117] J. Longstaff and K. Singer. A survey of Gaussian two-electron functions. *Theoretica chimica acta*, 2(4):265–280, 1964.

[118] W. A. Lester Jr and M. Krauss. Gaussian correlation functions: Two-electron systems. *The Journal of Chemical Physics*, 41(5):1407–1413, 1964.

[119] H. F. King. Some theorems concerning symmetry, angular momentum, and completeness of atomic geminals with explicit r12 dependence. *The Journal of Chemical Physics*, 46(2):705–713, 1967.

[120] L. Kong, F. A. Bischoff, and E. F. Valeev. Explicitly correlated R12/F12 methods for electronic structure. *Chemical Reviews*, 112(1):75–107, 2012.

[121] T. Shiozaki and H.-J. Werner. Multireference explicitly correlated F12 theories. *Molecular Physics*, 111(5):607–630, 2013.

[122] G. H. Booth, D. Cleland, A. Alavi, and D. P. Tew. An explicitly correlated approach to basis set incompleteness in full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 137(16):164112, 2012.

[123] S. F. Boys and N. C. Handy. The determination of energies and wavefunctions with full electronic correlation. *Proceedings of the Royal Society of London A*, 310(1500):43–61, 1969.

[124] S. F. Boys and N. C. Handy. A condition to remove the indeterminacy in interelectronic correlation functions. *Proceedings of the Royal Society of London A*, 309(1497):209–220, 1969.

[125] S. F. Boys and N. C. Handy. A calculation for the energies and wavefunctions for states of neon with full electronic correlation accuracy. *Proceedings of the Royal Society of London A*, 310(1500):63–78, 1969.

[126] N. Handy. The transcorrelated method for accurate correlation energies using Gaussian-type functions: examples on He, H_2, LiH and H_2O. *Molecular Physics*, 23(1):1–27, 1972.

[127] N. Handy. Towards an understanding of the form of correlated wavefunctions for atoms. *The Journal of Chemical Physics*, 58(1):279–287, 1973.

[128] H. Luo and A. Alavi. Combining the transcorrelated method with full configuration interaction quantum Monte Carlo: Application to the homogeneous electron gas. *Journal of Chemical Theory and Computation*, 14(3):1403–1411, 2018.

[129] A. J. Cohen, H. Luo, K. Guther, W. Dobrautz, D. P. Tew, and A. Alavi. Similarity transformation of the electronic Schrödinger equation via Jastrow factorization. *The Journal of Chemical Physics*, 151(6):061101, 2019.

[130] C. Lanczos. 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*, 45:225–282, 1950.

[131] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[132] D. R. Hartree. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 89–110. Cambridge University Press, 1928.

[133] V. Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2):126–148, 1930.

[134] E. H. Lieb and B. Simon. The Hartree-Fock theory for coulomb systems. *Communications in Mathematical Physics*, 53(3):185–194, 1977.

[135] P. A. M. Dirac and R. H. Fowler. On the theory of quantum mechanics. *Proceedings of the Royal Society of London A*, 112(762):661–677, 1926.

[136] W. Heisenberg. Mehrkörperproblem und Resonanz in der Quantenmechanik. *Zeitschrift für Physik*, 38(6-7):411–426, 1926.

[137] T. Helgaker, P. Jørgensen, and J. Olsen. Atomic basis functions. In *Molecular Electronic-Structure Theory*, chapter 6, pages 201–255. John Wiley & Sons, Ltd, Chichester, 2014.

[138] T. Helgaker, P. Jørgensen, and J. Olsen. Molecular integral evaluation. In *Molecular Electronic-Structure Theory*, chapter 9, pages 336–432. John Wiley & Sons, Ltd, Chichester, 2014.

[139] E. R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *Journal of Computational Physics*, 17(1):87–94, 1975.

[140] G. L. G. Sleijpen and H. A. V. der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 17(2):401–425, 1996.

[141] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9(1):17–29, 1951.

[142] A. Weiße and H. Fehske. Exact diagonalization techniques. In H. Fehske, R. Schneider, and A. Weiße, editors, *Computational Many-Particle Physics*, pages 529–544. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[143] Y. Lu and M. W. Haverkort. Exact diagonalization as an impurity solver in dynamical mean field theory. *The European Physical Journal Special Topics*, 226(11):2549–2564, 2017.

[144] D. Medvedeva, S. Iskakov, F. Krien, V. V. Mazurenko, and A. I. Lichtenstein. Exact diagonalization solver for extended dynamical mean-field theory. *Physical Review B*, 96(23):235149, 2017.

[145] C. Forssén, B. Carlsson, H. Johansson, D. Sääf, A. Bansal, G. Hagen, and T. Papenbrock. Large-scale exact diagonalizations reveal low-momentum scales of nuclei. *Physical Review C*, 97(3):034328, 2018.

[146] A. Wietek and A. M. Läuchli. Sublattice coding algorithm and distributed memory parallelization for large-scale exact diagonalizations of quantum many-body systems. *Physical Review E*, 98(3):033309, 2018.

[147] A. Krylov. On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined. *Izvestija AN SSSR (News of Academy of Sciences of the USSR), Otdel. mat. i estest. nauk*, 7(4):491–539, 1931.

[148] R. V. Mises and H. Pollaczek-Geiringer. Praktische Verfahren der Gleichungsauflösung . *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 9(2):152–164, 1929.

[149] J. K. Cullum and R. A. Willoughby. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations: Vol. 1: Theory*, volume 41. Siam, Philadelphia, 2002.

[150] Y. Saad. On the rates of convergence of the Lanczos and the block-Lanczos methods. *SIAM Journal on Numerical Analysis*, 17(5):687–706, 1980.

[151] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU press, Baltimore, 2012.

[152] A. Meister. *Numerik Linearer Gleichungssysteme*, volume 5. Springer, Wiesbaden, 2011.

[153] A. M. Läuchli, J. Sudan, and E. S. Sørensen. Ground-state energy and spin gap of spin-$\frac{1}{2}$ Kagomé-Heisenberg antiferromagnetic clusters: Large-scale exact diagonalization results. *Physical Review B*, 83(21):212401, 2011.

[154] J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar. Fast semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 149(21):214110, 2018.

[155] A. A. Holmes, H. J. Changlani, and C. J. Umrigar. Efficient heat-bath sampling in Fock space. *Journal of Chemical Theory and Computation*, 12(4):1561–1571, 2016.

[156] W. L. McMillan. Ground state of liquid . *Physical Review*, 138(2A):A442–A451, 1965.

[157] D. Ceperley, G. V. Chester, and M. H. Kalos. Monte Carlo simulation of a many-fermion study. *Physical Review B*, 16(7):3081–3099, 1977.

[158] R. Grimm and R. Storer. Monte-carlo solution of Schrödinger's equation. *Journal of Computational Physics*, 7(1):134 – 156, 1971.

[159] J. B. Anderson. A random-walk simulation of the Schrödinger equation: H_3^+. *The Journal of Chemical Physics*, 63(4):1499–1503, 1975.

[160] J. B. Anderson. Quantum chemistry by random walk. *The Journal of Chemical Physics*, 65(10):4121–4127, 1976.

[161] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar. Monte Carlo calculations of coupled boson-fermion systems. I. *Physical Review D*, 24(8):2278–2286, 1981.

[162] G. Sugiyama and S. Koonin. Auxiliary field monte-carlo for quantum many-body ground states. *Annals of Physics*, 168(1):1 – 26, 1986.

[163] S. Zhang and M. H. Kalos. Bilinear quantum Monte Carlo: Expectations and energy differences. *Journal of Statistical Physics*, 70(3):515–533, 1993.

[164] S. Zhang and H. Krakauer. Quantum Monte Carlo method using phase-free random walks with Slater determinants. *Physical Review Letters*, 90(13):136401, 2003.

[165] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein. Continuous-time quantum Monte Carlo method for fermions. *Physical Review B*, 72(3):035122, 2005.

[166] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[167] R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta numerica*, 7:1–49, 1998.

[168] O. C. Ibe. Basic concepts in probability. In *Markov Processes for Stochastic Modeling*, chapter 1, pages 1 – 27. Elsevier, Oxford, Second edition edition, 2013.

[169] J. Bernoulli. *Ars Conjectandi*. 1713.

[170] W. Feller. *An introduction to probability theory and its applications. Vol. 1*. John Wiley & Sons, New York, 1968.

[171] S. M. Ross. Distributions of sampling statistics. In *Introductory Statistics*, chapter 7, pages 297 – 330. Academic Press, Boston, Third edition edition, 2010.

[172] H. Kahn and A. W. Marshall. Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America*, 1(5):263–278, 1953.

[173] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. 1970.

[174] I. Kosztin, B. Faber, and K. Schulten. Introduction to the diffusion Monte Carlo method. *American Journal of Physics*, 64(5):633–644, 1996.

[175] P. H. Acioli. Review of quantum Monte Carlo methods and their applications. *Journal of Molecular Structure: Theochem*, 394(2-3):75–85, 1997.

[176] R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos. Continuum variational and diffusion quantum Monte Carlo calculations. *Journal of Physics: Condensed Matter*, 22(2):023201, 2009.

[177] J. Feldt and C. Filippi. Excited states with quantum Monte Carlo. In L. Gonzales and R. Lindh, editors, *Quantum Chemistry and Dynamics of Excited States: Methods and Applications*. John Wiley & Sons Ltd., 2020. Forthcoming.

[178] D. Ceperley and B. Alder. Quantum Monte Carlo. *Science*, 231(4738):555–560, 1986.

[179] M. A. Suhm and R. O. Watts. Quantum Monte Carlo studies of vibrational states in molecules and clusters. *Physics Reports*, 204(4):293 – 329, 1991.

[180] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

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

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

[183] J. E. Hirsch and R. M. Fye. Monte Carlo method for Magnetic Impurities in Metals. *Physical Review Letters*, 56(23):2521, 1986.

[184] J. Carlson, S. Gandolfi, K. E. Schmidt, and S. Zhang. Auxiliary-field quantum Monte Carlo method for strongly paired fermions. *Physical Review A*, 84(6):061602, 2011.

[185] E. Gull, P. Werner, O. Parcollet, and M. Troyer. Continuous-time auxiliary-field Monte Carlo for quantum impurity models. *EPL (Europhysics Letters)*, 82(5):57003, 2008.

[186] E. Vitali, H. Shi, M. Qin, and S. Zhang. Computation of dynamical correlation functions for many-fermion systems with auxiliary-field quantum Monte Carlo. *Physical Review B*, 94(8):085140, 2016.

[187] S. Zhang, J. Carlson, and J. E. Gubernatis. Constrained path Monte Carlo method for fermion ground states. *Physical Review B*, 55(12):7464, 1997.

[188] M. Motta and S. Zhang. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *WIREs Computational Molecular Science*, 8(5):1364, 2018.

[189] H. Hao, B. M. Rubenstein, and H. Shi. Auxiliary field quantum Monte Carlo for multiband Hubbard models: Controlling the sign and phase problems to capture Hund's physics. *Physical Review B*, 99(23):235142, 2019.

[190] D. Thouless. Stability conditions and nuclear rotations in the Hartree-Fock theory. *Nuclear Physics*, 21:225 – 232, 1960.

[191] R. Stratonovich. *Soviet Physics Doklady*, 2:416, 1958.

[192] J. Hubbard. Calculation of partition functions. *Physical Review Letters*, 3(2):77, 1959.

[193] W. Al-Saidi, S. Zhang, and H. Krakauer. Auxiliary-field quantum Monte Carlo calculations of molecular systems with a Gaussian basis. *The Journal of Chemical Physics*, 124(22):224101, 2006.

[194] W. Al-Saidi, H. Krakauer, and S. Zhang. A study of $H+H_2$ and several H-bonded molecules by phaseless auxiliary-field quantum Monte Carlo with plane wave and Gaussian basis sets. *The Journal of Chemical Physics*, 126(19):194105, 2007.

[195] M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E. J. Walter. Phaseless auxiliary-field quantum Monte Carlo calculations with plane waves and pseudopotentials: Applications to atoms and molecules. *Physical Review B*, 75(24):245123, 2007.

[196] K. Guther, R. J. Anderson, N. S. Blunt, N. A. Bogdanov, D. Cleland, D. Dattani, W. Dobrautz, K. Ghanem, P. Jeszenski, N. Liebermann, G. Li Manni, A. Y. Lozovoi, H. Luo, D. Ma, F. Merz, C. Overy, M. Rampp, P. K. Samanta, L. R. Schwarz, J. J. Shepherd, S. D. Smart, E. Vitale, O. Weser, G. H. Booth, and A. Alavi. NECI: N-electron configuration interaction with an emphasis on state-of-the-art stochastic methods. *The Journal of Chemical Physics*, 153(3):034107, 2020.

[197] J. von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, chapter 13, pages 36–38. US Government Printing Office, Washington, DC, 1951.

[198] J. E. Gentle. *Random number generation and Monte Carlo methods*. Springer Science & Business Media, New York, 2006.

[199] Intel Corporation. *Intel 64 and IA-32 Architectures Software Developer Manuals*, 2016.

[200] Amd secure random number generator library. `http://developer.amd.com/wordpress/media/2013/12/56310-AMD-Secure-Random-Number-Generator-Library.pdf`, 2018.

[201] J.-S. Coron. On the security of random sources. In *International Workshop on Public Key Cryptography*, pages 29–42. Springer, 1999.

[202] K. Hongo, R. Maezono, and K. Miura. Random number generators tested on quantum Monte Carlo simulations. *Journal of computational Chemistry*, 31(11):2186–2194, 2010.

[203] D. P. Landau and K. Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, Cambridge, 2014.

[204] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

[205] J. M. Cardoso, J. G. F. Coutinho, and P. C. Diniz. High-performance embedded computing. In *Embedded Computing for High Performance*, chapter 2, pages 17 – 56. Morgan Kaufmann, Boston, 2017.

[206] I. Corporation. Intel instruction set extension technology. `https://www.intel.co.uk/content/www/uk/en/support/articles/000005779/processors.html`. Version from 07.01.2020.

[207] Advanced Micro Devices Inc. *AMD64 Architecture Programmers Manual, Volume 4: 128-Bit and 256-Bit Media Instructions*, 2020. Order Nr. 26568.

[208] M. Saito and M. Matsumoto. SIMD-oriented fast mersenne twister: a 128-bit pseudorandom number generator. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 607–622, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[209] M. Saito and M. Matsumoto. double precision SIMD oriented fast mersenne twister. `https://github.com/MersenneTwister-Lab/dSFMT`, 2008.

[210] F. James. A review of pseudorandom number generators. *Computer physics communications*, 60(3):329–344, 1990.

[211] G. Marsaglia and A. Zaman. A new class of random number generators. *The Annals of Applied Probability*, pages 462–480, 1991.

[212] M. Matsumoto and Y. Kurita. Twisted GFSR generators. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2(3):179–194, 1992.

[213] M. Matsumoto and Y. Kurita. Twisted GFSR generators II. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 4(3):254–266, 1994.

[214] D. E. Knuth. *The art of computer programming*, volume 3. Pearson Education, Boston, 1997.

[215] F. Panneton, P. L'Ecuyer, and M. Matsumoto. Improved long-period generators based on linear recurrences modulo 2. *ACM Trans. Math. Softw.*, 32(1):1–16, 2006.

[216] G. Marsaglia. Generating discrete random variables in a computer. *Commun. ACM*, 6(1):37–38, 1963.

[217] G. Marsaglia, W. W. Tsang, and J. Wang. Fast generation of discrete random variables. *Journal of Statistical Software, Articles*, 11(3):1–11, 2004.

[218] A. J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Transitions of Mathematical Software*, 3(3):253–256, 1977.

[219] M. D. Vose. A linear algorithm for generating random numbers with a given distribution. *IEEE Transactions on software engineering*, 17(9):972–975, 1991.

[220] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, 1986.

[221] G. Booth. Fciqmc: A practical perspective, lecture at molssi qmc summer school, pittsburgh, pa. `https://github.com/lkwagner/StochasticSchool/tree/master/Day3_FCIQMC`, 2019.

[222] S. M. Greene, R. J. Webber, J. Weare, and T. C. Berkelbach. Beyond walkers in stochastic quantum chemistry: Reducing error using fast randomized iteration. *Journal of chemical Theory and Computation*, 15(9):4834–4850, 2019.

[223] L.-H. Lim and J. Weare. Fast randomized iteration: Diffusion Monte Carlo through the lens of numerical linear algebra. *SIAM Review*, 59(3):547–587, 2017.

[224] G. Li Manni, K. Guther, D. Ma, and W. Dobrautz. Foundation of Multi-Configurational Quantum Chemistry. In L. Gonzales and R. Lindh, editors, *Quantum Chemistry and Dynamics of Excited States: Methods and Applications*, chapter 6. John Wiley & Sons Ltd., 2020. Forthcoming.

[225] J. S. Spencer, N. S. Blunt, S. Choi, J. Etrych, M.-A. Filip, W. M. C. Foulkes, R. S. T. Franklin, W. J. Handley, F. D. Malone, V. A. Neufeld, R. Di Remigio, T. W. Rogers, C. J. C. Scott, J. J. Shepherd, W. A. Vigor, J. Weston, R. Xu, and A. J. W. Thom. The HANDE-QMC project: Open-source stochastic quantum chemistry from the ground state up. *Journal of Chemical Theory and Computation*, 15(3):1728–1742, 2019.

[226] H. Flyvbjerg and H. G. Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466, 1989.

[227] N. S. Blunt, A. J. W. Thom, and C. J. C. Scott. Preconditioning and perturbative estimators in full configuration interaction quantum Monte Carlo. *Journal of Chemical Theory and Computation*, 15(6):3537, 2019.

[228] N. S. Blunt. A hybrid approach to extending selected configuration interaction and full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 151(17):174103, 2019.

[229] J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes. The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method. *The Journal of Chemical Physics*, 136:054110, 2012.

[230] D. M. Cleland, G. H. Booth, and A. Alavi. A study of electron affinities using the initiator approach to full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 134(2):024112, 2011.

[231] V. A. Neufeld and A. J. W. Thom. Exciting determinants in quantum Monte Carlo: Loading the dice with fast, low-memory weights. *Journal of Chemical Theory and Computation*, 15(1):127–140, 2019.

[232] S. D. Smart. *The use of spin-pure and non-orthogonal Hilbert spaces in full configuration interaction quantum monte-carlo*. PhD thesis, University of Cambridge, 2014.

[233] W. Dobrautz, S. D. Smart, and A. Alavi. Efficient formulation of full configuration interaction quantum Monte Carlo in a spin eigenbasis via the graphical unitary group approach. *The Journal of Chemical Physics*, 151(9):094104, 2019.

[234] N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi. Semi-stochastic full configuration interaction quantum Monte Carlo: Developments and application. *The Journal of Chemical Physics*, 142(18):184107, 2015.

[235] L. Veis, A. Antalík, Ö. Legeza, A. Alavi, and J. Pittner. The intricate case of Tetramethyleneethane: A full configuration interaction quantum Monte Carlo benchmark and multireference coupled cluster studies. *Journal of Chemical Theory and Computation*, 14(5):2439–2445, 2018.

[236] N. S. Blunt, A. Alavi, and G. H. Booth. Nonlinear biases, stochastically sampled effective Hamiltonians, and spectral functions in quantum Monte Carlo methods. *Physical Review B*, 98(8):085118, 2018.

[237] N. S. Blunt, G. H. Booth, and A. Alavi. Density matrices in full configuration interaction quantum Monte Carlo: Excited states, transition dipole moments, and parallel distribution. *The Journal of Chemical Physics*, 146(24):244105, 2017.

[238] R. J. Bartlett. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annual Review of Physical Chemistry*, 32(1):359–401, 1981.

[239] N. Bogoliubov. On the theory of superfluidity. *J. Phys*, 11(1):23, 1947.

[240] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski. *Methods of quantum field theory in statistical physics*. Courier Corporation, New York, 2012.

[241] E. Deumens, A. Diz, R. Longo, and Y. Öhrn. Time-dependent theoretical treatments of the dynamics of electrons and nuclei in molecular systems. *Reviews of Modern Physics*, 66(3):917, 1994.

[242] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Review of Modern Physics*, 20(2):367–387, 1948.

[243] R. P. Feynman. The theory of positrons. *Physical Review*, 76(6):749, 1949.

[244] J. Schwinger. On the Green's functions of quantized fields. II. *Proceedings of the National Academy of Sciences*, 37(7):455–459, 1951.

[245] M. Jarrell and J. E. Gubernatis. Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data. *Physics Reports*, 269(3):133–195, 1996.

[246] M. Born and P. Jordan. Zur Quantenmechanik. *Zeitschrift für Physik*, 34(1):858–888, 1925.

[247] M. Born, W. Heisenberg, and P. Jordan. Zur Quantenmechanik. II. *Zeitschrift für Physik*, 35(8-9):557–615, 1926.

[248] M. Feit, J. Fleck Jr, and A. Steiger. Solution of the Schrödinger equation by a spectral method. *Journal of Computational Physics*, 47(3):412–433, 1982.

[249] K. Guther, W. Dobrautz, O. Gunnarsson, and A. Alavi. Time propagation and spectroscopy of fermionic systems using a stochastic technique. *Physical Review Letters*, 121(5):056401, 2018.

[250] M. Schottenloher. *A mathematical introduction to conformal field theory*. Springer, Berlin, 2008.

[251] I. Bronstein, K. Semendjajew, G. Musiol, and H. Mülig. *Taschenbuch der Mathematik*. Wissenschaftlicher Verlag Harri Deutsch, 8 edition, 2012.

[252] H.-B. Schüttler and D. Scalapino. Monte Carlo studies of the dynamics of quantum many-body systems. *Physical Review Letters*, 55(11):1204, 1985.

[253] H.-B. Schüttler and D. Scalapino. Monte Carlo studies of the dynamical response of quantum many-body systems. *Physical Review B*, 34(7):4744, 1986.

[254] S. White, D. Scalapino, R. Sugar, and N. Bickers. Monte Carlo calculation of dynamical properties of the two-dimensional hubbard model. *Physical Review Letters*, 63(14):1523, 1989.

[255] A. W. Sandvik. Stochastic method for analytic continuation of quantum Monte Carlo data. *Physical Review B*, 57(17):10287–10290, 1998.

[256] K. Vafayi and O. Gunnarsson. Analytical continuation of spectral data from imaginary time axis to real frequency axis using statistical sampling. *Physical Review B*, 76(3):035115, 2007.

[257] K. Ghanem and E. Koch. Average spectrum method for analytic continuation: Efficient blocked-mode sampling and dependence on the discretization grid. *Physical Review B*, 101(8):085111, 2020.

[258] H. Vidberg and J. Serene. Solving the Eliashberg equations by means of N-point Padé approximants. *Journal of Low Temperature Physics*, 29(3-4):179–192, 1977.

[259] K. S. D. Beach, R. J. Gooding, and F. Marsiglio. Reliable Padé analytical continuation method based on a high-accuracy symbolic computation algorithm. *Physical Review B*, 61(8):5147–5157, 2000.

[260] A. Östlin, L. Chioncel, and L. Vitos. One-particle spectral function and analytic continuation for many-body implementation in the exact muffin-tin orbitals method. *Physical Review B*, 86(23):235107, 2012.

[261] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620, 1957.

[262] E. T. Jaynes. Information theory and statistical mechanics. II. *Physical Review*, 108(2):171, 1957.

[263] E. Jaynes. Maximum entropy formalism. In R. D. Levine and M. Tribus, editors, *Maximum Entropy Formalism Conference (1978 Massachusetts Institute of Technology)*. MIT Press, Cambridge, 1979.

[264] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni. Analytical continuation of imaginary axis data using maximum entropy. *Physical Review B*, 81(15):155107, 2010.

[265] J. Skilling and R. Bryan. Maximum entropy image reconstruction-general algorithm. *Monthly notices of the royal astronomical society*, 211:111, 1984.

[266] K. Nigam, J. Lafferty, and A. McCallum. Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering*, volume 1, pages 61–67. Stockholom, Sweden, 1999.

[267] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.

[268] J. Skilling. Classic maximum entropy. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*, volume 36, pages 45–52. Springer Science & Business Media, 1989.

[269] Y. Tikochinsky, N. Tishby, and R. Levine. Consistent inference of probabilities for reproducible experiments. *Physical Review Letters*, 52(16):1357, 1984.

[270] G. Erickson and C. R. Smith. *Maximum-Entropy and Bayesian Methods in Science and Engineering*, volume 1. Springer Science & Business Media, Dordrecht, 1988.

[271] R. Bryan. Maximum entropy analysis of oversampled data problems. *European Biophysics Journal*, 18(3):165–174, 1990.

[272] C. Runge. Ueber die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 46:167–178, 1895.

[273] W. Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Zeitschrift fü Mathematische Physik*, 46:435–53, 1901.

[274] E. Süli and D. F. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, Cambridge, 2003.

[275] L. Verlet. Computer experiments on classical fluids. i. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98, 1967.

[276] C. Overy, G. H. Booth, N. S. Blunt, J. J. Shepherd, D. Cleland, and A. Alavi. Unbiased reduced density matrices and electronic properties from full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 141(24):244117, 2014.

[277] N. S. Blunt, A. Alavi, and G. H. Booth. Krylov-projected quantum Monte Carlo method. *Physical Review Letters*, 115(5):050603, 2015.

[278] K. Wildberger, P. Lang, R. Zeller, and P. Dederichs. Fermi-Dirac distribution in ab initio Green's-function calculations. *Physical Review B*, 52(15):11502, 1995.

[279] M. Inc. Matlab r2017a, version 9.2.0.556344, 2017.

[280] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Review of Modern Physics*, 68(1):13, 1996.

[281] N. Bulut. $d_{x^2-y^2}$ superconductivity and the Hubbard model. *Advances in Physics*, 51(7):1587, 2002.

[282] A.-M. S. Tremblay, B. Kyung, and D. Sénéchal. Pseudogap and high-temperature superconductivity from weak to strong coupling. towards a quantitative theory. *Low Temperature Physics*, 32(4):424, 2006.

[283] A. Auerbach. *Interacting electrons and quantum magnetism*. Springer Science & Business Media, New York, 2012.

[284] E. Dagotto. Correlated electrons in high-temperature superconductors. *Review of Modern Physics*, 66(3):763, 1994.

[285] R. M. Noack and S. R. Manmana. Diagonalization- and numerical renormalization-group-based methods for interacting quantum systems. *AIP Conference Proceedings*, 789(1):93, 2005.

[286] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V.

Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull. Solutions of the two-dimensional hubbard model: Benchmarks and results from a wide range of numerical algorithms. *Physical Review X*, 5(4):041041, 2015.

[287] E. H. Lieb and F.-Y. Wu. Absence of Mott transition in an exact solution of the short-range, one-band model in one dimension. In *Exactly Solvable Models Of Strongly Correlated Electrons*, pages 9–12. World Scientific, 1994.

[288] J. von Delft and H. Schoeller. Bosonization for beginners — refermionization for experts. *Annalen der Physik*, 7(4):225–305, 1998.

[289] H. Haken and H. C. Wolf. *Molekülphysik und Quantenchemie: Einführung in die experimentellen und theoretischen Grundlagen.* Springer, Berlin, 2013.

[290] W. Nolting. Wechselwirkende Teilchensysteme. In *Grundkurs Theoretische Physik 7: Viel-Teilchen-Theorie*, chapter 4, pages 195–304. Springer-Verlag, Heidelberg, 2014.

[291] J. E. Hirsch. Two-dimensional Hubbard model: Numerical simulation study. *Physical Review B*, 31(7):4403, 1985.

[292] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz. `Molpro`: a general-purpose quantum chemistry program package. *WIREs Computational Molecular Science*, 2(2):242–253, 2012.

[293] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms Boron through Neon and Hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.

[294] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson. Gaussian basis sets for use in correlated molecular calculations. VII. valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theoretical Chemistry Accounts*, 128(1), 2011.

[295] A. Kramida and W. C. Martin. A compilation of energy levels and wavelengths for the spectrum of neutral Beryllium (Be i). *Journal of Physical and Chemical Reference Data*, 26(5):1185–1194, 1997.

[296] P. Z. et al. (Particle data group). 2020 review of particle physics. *Progress of Theoretical and Experimental Physics*, page 083C01, 2020. To be published.

[297] V. Kaufman and J. F. Ward. Measurement and calculation of Cu II, Ge II, Si II, and CI vacuum-ultraviolet lines. *Journal of the Optical Society of America*, 56:1591, 1966.

[298] G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi. Breaking the carbon dimer: The challenges of multiple bond dissociation with Full Configuration Interaction quantum Monte Carlo methods. *J. Chem. Phys.*, 135(8):084104, 2011.

[299] K. Huber and G. Herzberg. Molecular structure and molecular spectra. iv. constants of diatomic molecules. *Van Rostrand-Reinhold, New York*, 1979.

[300] M. Martin. C_2 spectroscopy and kinetics. *Journal of Photochemistry and Photobioligy A*, 66:263–289, 1992.

[301] G. H. Booth and A. Alavi. Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: A study of ionization potentials. *The Journal of Chemical Physics*, 132(17):174104, 2010.

[302] J. Schlappa, K. Wohlfeld, K. Zhou, M. Mourigal, M. Haverkort, V. Strocov, L. Hozoi, C. Monney, S. Nishimoto, S. Singh, et al. Spin–orbital separation in the quasi-one-dimensional Mott insulator Sr_2CuO_3. *Nature*, 485(7396):82–85, 2012.

[303] R. Neudert, M. Knupfer, M. Golden, J. Fink, W. Stephan, K. Penc, N. Motoyama, H. Eisaki, and S. Uchida. Manifestation of spin-charge separation in the dynamic dielectric response of one-dimensional Sr_2CuO_3. *Physical Review Letters*, 81(3):657, 1998.

[304] N. Bulut, D. Scalapino, and S. White. Quasiparticle dispersion in the cuprate superconductors and the two-dimensional Hubbard model. *Physical Review B*, 50(10):7215, 1994.

[305] A. Macridin, M. Jarrell, T. Maier, and G. Sawatzky. Physics of cuprates with the two-band Hubbard model: The validity of the one-band Hubbard model. *Physical Review B*, 71(13):134527, 2005.

[306] H. Rosner, H. Eschrig, R. Hayn, S.-L. Drechsler, and J. Málek. Electronic structure and magnetic properties of the linear chain cuprates Sr_2CuO_3 and ca_2CuO_3. *Physical Review B*, 56(6):3402, 1997.

[307] J. Lorenzana, J. Eroles, and S. Sorella. Does the Heisenberg model describe the multimagnon spin dynamics in antiferromagnetic CuO layers? *Physical Review Letters*, 83(24):5122, 1999.

[308] D. I. Khomskii. *Transition Metal Compounds*. Cambridge University Press, Cambridge, 2014.

[309] S. F. Boys. Some bilinear convergence characteristics of the solutions of dissymmetric secular equations. *Proceedings of the Royal Society of London A*, 309(1497):195–208, 1969.

[310] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics*, 96(9), 1992.

[311] E. A. Hylleraas. Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium. *Zeitschrift für Physik*, 54(5-6):347–366, 1929.

[312] R. T. Pack and W. B. Brown. Cusp conditions for molecular wavefunctions. *The Journal of Chemical Physics*, 45(2):556–559, 1966.

[313] R. Jastrow. Many-body problems with strong forces. *Physical Review*, 98(5):1479–1484, 1955.

[314] D. Ceperley. Ground state of the fermion one-component plasma: A Monte Carlo study in two and three dimensions. *Physical Review B*, 18(7):3126–3138, 1978.

[315] R. Needs, M. Towler, N. Drummond, P. Lopez Rios, and J. Trail. Variational and diffusion quantum Monte Carlo calculations with the CASINO code. *The Journal of Chemical Physics*, 152(15):154106, 2020.

[316] R. Gaudoin, M. Nekovee, W. M. C. Foulkes, R. J. Needs, and G. Rajagopal. Inhomogeneous random-phase approximation and many-electron trial wave functions. *Physical Review B*, 63(11):115115, 2001.

[317] S. Ten-no. A feasible transcorrelated method for treating electronic cusps using a frozen Gaussian geminal. *Chemical Physics Letters*, 330(1):169 – 174, 2000.

[318] H. Luo. Variational transcorrelated method. *The Journal of Chemical Physics*, 133(15):154109, 2010.

[319] J. O. Hirschfelder. Removal of electron—electron poles from many-electron hamiltonians. *The Journal of Chemical Physics*, 39(11):3145–3146, 1963.

[320] K. E. Schmidt and J. W. Moskowitz. Correlated Monte Carlo wave functions for the atoms He through Ne. *The Journal of Chemical Physics*, 93(6):4172–4178, 1990.

[321] S. Chakravorty, S. Gwaltney, E. Davidson, F. Parpia, and C. Fischer. Ground-state correlation energies for atomic ions with 3 to 18 electrons. *Physical Review A*, 47(5):3649, 1993.

[322] R. J. Le Roy, N. S. Dattani, J. A. Coxon, A. J. Ross, P. Crozet, and C. Linton. Accurate analytic potentials for $Li_2(x^1\sigma_g^+)$ and $Li_2(a^1\sigma_u^+)$ from 2 to 90 å, and the radiative lifetime of Li(2p). *The Journal of Chemical Physics*, 131(20):204309, 2009.

[323] J. M. Merritt, V. E. Bondybey, and M. C. Heaven. Beryllium dimer—caught in the act of bonding. *Science*, 324(5934):1548–1551, 2009.

[324] V. V. Meshkov, A. V. Stolyarov, M. C. Heaven, C. Haugen, and R. J. LeRoy. Direct-potential-fit analyses yield improved empirical potentials for the ground $x^1\sigma_g^+$ state of Be_2. *The Journal of Chemical Physics*, 140(6):064315, 2014.

[325] R. J. Anderson, T. Shiozaki, and G. H. Booth. Efficient and stochastic multireference perturbation theory for large active spaces within a full configuration interaction quantum Monte Carlo framework. *The Journal of Chemical Physics*, 152(5):054101, 2020.

[326] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, Cambridge, MA, 2009.

[327] I. Corporation. Intel vtune profiler. `https://software.intel.com/content/www/us/en/develop/tools/vtune-profiler.html`. Formerly known as Intel VTune Amplifier.

[328] The HDF Group. Hierarchical Data Format, version 5, 1997-2020. http://www.hdfgroup.org/HDF5/.

[329] P. J. Knowles and N. C. Handy. A determinant based full configuration interaction program. *Computer Physics Communications*, 54(1):75 – 83, 1989.

[330] K. Patkowski, R. Podeszwa, and K. Szalewicz. Interactions in diatomic dimers involving closed-shell metals. *The Journal of Physical Chemistry A*, 111(49):12822–12838, 2007.

[331] K. Patkowski, V. Špirko, and K. Szalewicz. On the elusive twelfth vibrational state of beryllium dimer. *Science*, 326(5958):1382–1384, 2009.

[332] J. Koput. The ground-state potential energy function of a beryllium dimer determined using the single-reference coupled-cluster approach. *Physical Chemistry Chemical Physics*, 13(45):20311–20317, 2011.

[333] S. Sharma, T. Yanai, G. H. Booth, C. J. Umrigar, and G. K.-L. Chan. Spectroscopic accuracy directly from quantum chemistry: Application to ground and excited states of Beryllium dimer. *The Journal of Chemical Physics*, 140(10):104112, 2014.

[334] M. Lesiuk, M. Przybytek, J. G. Balcerzak, M. Musiał, and R. Moszynski. Ab initio potential energy curve for the ground state of Beryllium dimer. *Journal of Chemical Theory and Computation*, 15(4):2470–2480, 2019.

[335] D. Kats and A. Köhn. On the distinguishable cluster approximation for triple excitations. *The Journal of Chemical Physics*, 150(15):151101, 2019.

[336] V. E. Bondybey. Laser-Induced Fluorescence and Bonding of Metal Dimers. *Science*, 227(4683):125–131, 1985.

[337] D. Bohm and D. Pines. A collective description of electron interactions: III. coulomb interactions in a degenerate electron gas. *Physical Review*, 92(3):609–625, 1953.

[338] N. D. Drummond, M. D. Towler, and R. J. Needs. Jastrow correlation factor for atoms, molecules, and solids. *Physical Review B*, 70(23):235119, 2004.

[339] P. López Ríos, P. Seth, N. D. Drummond, and R. J. Needs. Framework for constructing generic Jastrow correlation factors. *Physical Review E*, 86(3):036703, 2012.

[340] C. Filippi and C. J. Umrigar. Multiconfiguration wave functions for quantum monte carlo calculations of first-row diatomic molecules. *The Journal of Chemical Physics*, 105(1):213–226, 1996.

[341] S. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970.

[342] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe. State of the art in counterpoise theory. *Chemical Reviews*, 94(7):1873–1885, 1994.

[343] B. V. Noumerov. A Method of Extrapolation of Perturbations. *Monthly Notices of the Royal Astronomical Society*, 84(8):592–602, 1924.

[344] I. Fdez. Galván, M. Vacher, A. Alavi, C. Angeli, F. Aquilante, J. Autschbach, J. J. Bao, S. I. Bokarev, N. A. Bogdanov, R. K. Carlson, L. F. Chibotaru, J. Creutzberg, N. Dattani, M. G. Delcey, S. S. Dong, A. Dreuw, L. Freitag, L. M. Frutos, L. Gagliardi, F. Gendron, A. Giussani, L. Gonzàlez, G. Grell, M. Guo, C. E. Hoyer, M. Johansson, S. Keller, S. Knecht, G. Kovačević, E. Källman, G. Li Manni, M. Lundberg, Y. Ma, S. Mai, J. P. Malhado, P. Å. Malmqvist, P. Marquetand, S. A. Mewes, J. Norell, M. Olivucci, M. Oppel, Q. M. Phung, K. Pierloot, F. Plasser, M. Reiher, A. M. Sand, I. Schapiro, P. Sharma, C. J. Stein, L. K. Sørensen, D. G. Truhlar, M. Ugandi, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, O. Weser, T. A. Wesołowski, P.-O. Widmark, S. Wouters, A. Zech, J. P. Zobel, and R. Lindh. OpenMolcas: From source code to insight. *Journal of Chemical Theory and Computation*, 15(11):5925–5964, 2019.

[345] L. Stanisic and K. Reuter. MPCDF HPC Performance Monitoring System: Enabling insight via job-specific analysis. *arXiv preprint arXiv:1909.11704*, 2019.

[346] M. Harris. Mapping computational concepts to GPUs. In *ACM SIGGRAPH 2005 Courses*, pages 50–65. 2005.

[347] S. Williams, A. Waterman, and D. Patterson. Roofline: an insightful visual performance model for multicore architectures. *Communications of the ACM*, 52(4):65, 2009.

[348] J. Gustafson. Reevaluating Amdahl's law. *Communications of the ACM*, 31(5):532, 1988.

[349] R. Chamberlain, D. Chace, and A. Patil. How are we doing? An efficiency measure for shared, heterogeneous systems. In *Proc. of the ISCA 11th International*

Conference on Parallel and Distributed Computing Systems, pages 15–21. Citeseer, 1998.

[350] G. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the April 18-20, 1967, spring joint computer conference*, pages 483–485, 1967.

[351] Supermuc-ng specifications. `https://doku.lrz.de/display/PUBLIC/SuperMUC-NG`, accessed 05.06.2020.