The FCIQMC Sign Problem in the Real-Space Hubbard Model

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

vorgelegt von

Niklas Julian Liebermann

aus Tuttlingen

Hauptberichter:	Prof. Dr. Ali Alavi
Mitberichter:	Prof. Dr. Andreas Köhn
Prüfungsvorsitzender:	Prof. Dr. Blazej Grabowski

Tag der mündlichen Prüfung:

7. Februar 2023

MAX-PLANCK-INSTITUT FÜR FESTKÖRPERFORSCHUNG
UNIVERSITÄT STUTTGART

2023

Contents

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel „The FCIQMC Sign Problem in the Real-Space Hubbard Model“ 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 “The FCIQMC Sign Problem in the Real-Space Hubbard Model” is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Stuttgart, 7. Februar 2023

Niklas Julian Liebermann

List of Abbreviations

AFQMC	auxiliary-field quantum Monte Carlo
CI	configuration interaction
CPU	central processing unit
DMC	diffusion Monte Carlo
DMRG	density matrix renormalisation group
EN2	Epstein–Nesbet second-order perturbation theory
FCI	full configuration interaction
FCIQMC	full configuration interaction quantum Monte Carlo
GB	gigabyte
GFMC	Green's function Monte Carlo
hf	half-filling
MPO	matrix product operator
MPS	matrix product state
QMC	quantum Monte Carlo
SCI	selected configuration interaction
VMC	variational Monte Carlo
e.g.	for example *(exempli gratia)*
i.e.	that means *(id est)*
1-d	one-dimensional
2-d	two-dimensional

Abstract

Full configuration interaction quantum Monte Carlo (FCIQMC) is an electronic structure method that has been applied to a variety of *ab initio* molecular and solid-state systems as well as the Hubbard model in delocalised bases. In this thesis, the behaviour of FCIQMC in the Hubbard model in the real-space formulation is investigated. A special emphasis is put on the consequences of the fermionic sign problem.

Firstly, a classification of Hubbard lattice geometries based on their strength of the sign problem is performed. It is discovered that the commonly used ground state of the so-called stoquastic version of the Hamiltonian is not a good predictor for the difficulty to resolve the sign problem in FCIQMC in general. The notion of size-extensive and non-size-extensive behaviour of the sign problem is established. It is shown that although the vast majority of non-trivial fermionic systems suffer from the fermion sign problem when attempting to solve them using quantum Monte Carlo (QMC) methods, there are certain system configurations in the Hubbard model systems that are sign-problem-free.

In principle, this allows for the unbiased treatment of systems with very large Hilbert space sizes in FCIQMC. However, attempting to solve these systems uncovers a new systematic bias in the FCIQMC algorithm, the population control bias. This is a bias that has been observed previously in other QMC methods, like diffusion Monte Carlo. A method that allows for the removal of this bias entirely with negligible computational overhead, mainly through introducing importance sampling to FCIQMC, is presented. This allows for the calculation of ground-state energies of the one-dimensional Hubbard model with up to 150 sites at and close to half-filling in the difficult intermediate interaction regime. Also, the fundamental many-particle gaps between the ground states of the half-filled and the system with one hole are calculated for up to 102 sites.

Moving to sign-problematic systems, it is shown that the usual method of controlling the sign problem in FCIQMC, the initiator method, performs poorly in weakly sign-problematic Hubbard systems. Instead, it is demonstrated how applying the newly developed importance-sampled FCIQMC together with the exact non-initiator algorithm greatly reduces the minimum number of walkers necessary to obtain an unbiased ground-state energy in

real-space Hubbard models. This allows for the calculation of numerically exact ground-state energies for width-two Hubbard ladders – which exhibit a size-extensive yet very weak sign problem – in the intermediate interaction regime at half-filling and with one hole. Again, this makes the calculation of the fundamental many-particle gaps possible.

Finally, to deal with full two-dimensional Hubbard systems, a way to define fixed initiator subspaces in FCIQMC based on analytic wavefunction ansatzes is presented. This leads to far superior results compared to the usual population-based initiator criterion. Additionally, the newly developed two-shift method allows for the perturbative inclusion of the entire non-initiator space. This new scheme is shown to be compatible with importance sampling. Furthermore, an extrapolation scheme to the exact ground-state energy is presented. This allows for the estimation of the ground-state energy for systems up to 32 sites in the honeycomb lattice geometry.

Zusammenfassung

Full configuration interaction quantum Monte Carlo (FCIQMC) ist eine Methode der Elektronenstrukturtheorie, die bereits auf eine Vielzahl von *Ab-initio*-Quantensystemen (Molekül- als auch Festkörpersysteme) und das Hubbardmodell in delokalisierten Basen angewandt wurde. In dieser Arbeit wird das Verhalten von FCIQMC bei Behandlung des Hubbardmodells in seiner Realraum-Formulierung untersucht. Ein besonderer Schwerpunkt wird auf die Konsequenzen des Fermionen-Vorzeichenproblems gelegt.

Als erstes wird eine Klassifizierung von Gittergeometrien des Hubbardmodells aufgrund der jeweiligen Stärke des Vorzeichenproblems vorgenomen. Es wird festgestellt, dass der häufig genutzte Grundzustand der so genannten stoquastischen Version des Hamiltonoperators keine gute Vorhersage darüber erlaubt, wie aufwändig es ist, das Vorzeichenproblem in FCIQMC zu überwinden. Das Konzept des extensiven und nicht-extensiven Verhaltens von Vorzeichenproblemen wird eingeführt. Es wird gezeigt, dass obwohl die große Mehrheit von nicht-trivialen fermionischen Systemen ein Vorzeichenproblem bei Benutzung von Quanten-Monte-Carlo-Algorithmen (QMC-Algorithmen) aufweist, eine Reihe von Systemkonfigurationen unter Hubbardmodellen existieren, die nicht mit dem Vorzeichenproblem behaftet sind.

Dies erlaubt es im Prinzip, dass ein solches System ohne systematische Fehler auch in sehr großen Hilberträumen mittels FCIQMC gelöst werden kann. Bei der Behandlung von vorzeichenproblemfreien Systemen wird jedoch eine neue Quelle eines systematischen Fehlers in FCIQMC aufgedeckt. Dabei handelt es sich um einen Fehler aufgrund der Populationskontrolle im FCIQMC-Algorithmus. Ein ähnlicher Fehler wurde bereits in verwandten QMC-Methoden wie Diffusions-Monte-Carlo beobachtet und untersucht. Es wird eine Methode zur Entfernung dieses systematischen Fehlers mit vernachlässigbarem Rechenaufwand präsentiert, was hauptsächlich durch die Einführung von Stichprobennahme nach Wichtigkeit *(Importance Sampling)* möglich gemacht wird. Das erlaubt es, die Grundzustandsenergien von halb- oder nahezu halbgefüllten eindimensionalen Hubbardmodellen mit bis zu 150 Gitterplätzen im schwierig zu behandelnden Regime mittlerer Interaktionsstärke zu berechnen. Außerdem werden die fundamentalen Vielteilchen-Bandlücken zwischen dem Grundzustand des halbgefüllten Sy-

stems und dem des Systems mit einem Loch für bis zu 102 Gitterplätze berechnet.

Für Systeme mit Vorzeichenproblem wird gezeigt, dass die übliche Methode das Vorzeichenproblem in FCIQMC zu kontrollieren, die *Initiator*-Methode, nur unzureichend in den schwach vorzeichenproblembehafteten Hubbardsystemen funktioniert. Stattdessen wird gezeigt, wie das neu entwickelte *Importance Sampling* in Verbindung mit FCIQMC ohne *Initiator*-Näherung die Mindestanzahl der benötigten *Walker,* um die Grundzustandsenergie ohne systematischen Fehler zu berechnen, in Hubbardsystemen in der Realraum-Basis drastisch reduziert. Dies erlaubt es, die numerisch exakten Grundzustandenergien für halbgefüllte Hubbardsysteme und Hubbardsysteme mit einem Loch in der Leitergeometrie mit Breite zwei bei mittlerer Interaktionsstärke zu berechnen. Die Leitersysteme weisen ein extensives, jedoch sehr schwaches Vorzeichenproblem auf. Wiederum erlaubt dies die Berechnung der fundamentalen Bandlücken.

Schließlich wird ein Weg präsentiert, wie auf analytischen Ansatzen für die Vielteilchenwellenfunktion beruhende feste *Initiator*-Unterräume definiert werden können, um volle zweidimensionale Hubbardsysteme zu behandeln. Dies führt zu deutlich besseren Resultaten als die übliche populationsbasierte *Initiator*-Bedingung. Zusätzlich erlaubt die neu entwickelte Zwei-*Shift*-Methode die perturbative Miteinbeziehung des gesamten Nicht-*Initiator*-Raumes. Es wird gezeigt, dass diese Methode mit *Importance Sampling* vereinbar ist. Außerdem wird eine Methode zur Extrapolation zur exakten Grundzustandsenergie präsentiert. Dies erlaubt es, die Grundzustandsenergie für Systeme bis hin zu 32 Gitterplätzen in der Honigwabengittergeometrie zu berechnen.

Acknowledgements

Without the support of many people in many different forms, this thesis would not have been possible. Therefore, I would like to thank some of them here.

First and foremost, I want to thank my day-to-day supervisor *Prof. Ali Alavi* who gave me the possibility to do my PhD work in his department. He supported, challenged, and guided me with many different inspiring ideas but also gave me the freedom to pursue my own. Working with him made the last four years an enjoyable experience.

Also, I want to thank *Prof. Andreas Köhn* for taking over the position as the secondary examiner and *Prof. Blazej Grabowski* for being the head of my examination committee.

Furthermore, I would like to express my gratitude towards the entire group for making my PhD both a scientifically highly interesting and inspiring but also a fun time. Although everyone in the group contributed to my PhD in some shape or form, I want to mention some people in particular: I want to thank *Khaldoon Ghanem* with whom I had the pleasure to work together and discuss many of the topics in this thesis in great detail. I would also like to thank *Pablo López Ríos* for helping me with understanding other QMC methods apart from FCIQMC and providing me with a code to conduct VMC calculations for the Hubbard model. Then, I want to thank my office mates *Giovanni Li Manni* and *Oskar Weser* for the great working atmosphere in our office. I learned a lot about chemistry and a lot of fruitful ideas and side projects emerged just from casual conversations. But also, it was just a lot of fun merging Italian and German culture together. (Yes, it is definitely possible.) I also want to thank *Philip Haupt, Robert Anderson,* and *Michael Willatt* for discussing about the thesis and, apart from scientific advice, as native speakers and attentive proofreaders helped me a lot with writing good English. Finally, I also want to thank *Kai Guther* and *Werner Dobrautz*, who were in the middle of their PhDs when I started, and helped me a lot to get started, get to know the code, and just let me know what this is all about.

From the institute, I thank the *IT department* for the perfect administration of the computers and the cluster, our secretary *Brigitte Köber* for always knowing how to deal with all the administrative stuff, and *Christian Ast* for taking the role as secondary supervisor from the institute side.

Last but not least, I want to thank *my parents* for always being there for me. In the same way, I thank my girlfriend *Lisa* for warmly supporting me while working for and writing the thesis.

1 *Introduction*

"More is different" is the seemingly simple title of an article by later Nobel laureate P. W. Anderson published in *Science* in 1972 [1]. In this primarily philosophical paper, Anderson argues against the widespread perception that the more fundamental a scientific field and the more universal a law of nature is the more important it is. Instead, with every layer of complexity added to a system not only the scientific principles of the constituent parts need to be applied but also new laws emerge and are to be discovered. These emergent laws can be of equal significance as the most fundamental principles.

When Anderson wrote the article in the 1970s, computational methods to solve scientific questions were still in their infancy. Nowadays, with both powerful hardware and sophisticated algorithms available the familiarity with the emergent laws of complex systems is more important than ever.

This is especially true in the fields of electronic structure theory, quantum chemistry, and computational solid state physics, the topics this thesis will touch on. Even though the complexity of the calculations that can be performed on modern computers has increased tremendously over the past decades, there are countless systems that still evade a precise analysis with computational methods. Because of the exponential scaling of computational complexity with increasing system size, there is no other way forward than to improve our understanding of the laws of the interplay of the electrons and exploit them to come up with efficient algorithms.

Therefore, starting from a brief overview of the fundamental physical laws we are dealing with – the Schrödinger equation – I will introduce one of the many complications that arise from the interaction of electrons, the main topic of this thesis: the fermionic sign problem in quantum Monte Carlo methods.

1.1 *Schrödinger Equation and Many-Electron Problem*

As opposed to classical mechanics where the state of a dynamical system at time t is fully characterised by the positions $\boldsymbol{r}_i(t)$ and momenta $\boldsymbol{p}_i(t)$, the state of a quantum system is described by the many-particle wavefunction $\Psi(\{\boldsymbol{r}_i\}, t)$ [2]. In the Copenhagen interpretation of quantum mechanics, $|\Psi\rangle$ itself is not experimentally observable. $\left|\Psi(\{\boldsymbol{r}_i\}, t)\right|^2$ can be interpreted as

a probability density and therefore can be measured as an average over multiple measurements. The dynamics of $|\Psi\rangle$ is governed by the time-dependent Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle \tag{1.1}$$

where $\hbar$ is the reduced Planck constant and $\hat{H}$ is the Hamiltonian operator [3]. The Hamiltonian operator – mostly just called the *Hamiltonian* – can be derived from the classical Hamiltonian function and contains information about the involved particles and their interactions.

For time-independent Hamiltonians, the solutions of the Schrödinger equation can form standing waves, so-called stationary states. The stationary states are the eigenfunctions of $\hat{H}$ because

$$\hat{H}|\Psi_k\rangle = E_k|\Psi_k\rangle\ . \tag{1.2}$$

This equation is called the stationary Schrödinger equation. The corresponding eigenvalues E_k are the energies of the system. Additionally, it is possible to express the continuous Hamiltonian in a discrete and approximate fashion in a finite basis. In this way, the complicated partial differential equation (1.1) is reduced to a standard problem of linear algebra: the diagonalisation of a matrix. Still, despite the apparent simplicity of equation (1.2) and the well-understood form of the problem, the numerically exact solution of realistic systems – even medium-sized atoms, not to mention molecules and solids – poses a huge challenge, even on today's computer hardware. This is due to the fact that the dimension of the basis the Hamiltonian has to be represented in scales exponentially with system size. Therefore, accurate approximations and efficient algorithms are an inevitable necessity [4].

A very important and usually good approximation when attempting to solve the aforementioned atoms, molecules, and solid-state systems is the *Born–Oppenheimer approximation* [5, 6]. It separates the motion of the two constituents of these systems: the slow motion of the nuclei and the fast motion of electrons. The Hamiltonian in Born–Oppenheimer approximation in atomic units reads

$$\hat{H}_{\text{elec}} = -\sum_i \frac{1}{2}\hat{\nabla}_i^2 - \sum_{i,K}\frac{Z_K}{\hat{r}_{iK}} + \sum_{i>j}\frac{1}{\hat{r}_{ij}} + \sum_{K>L}\frac{Z_K Z_L}{\hat{r}_{KL}}\,. \tag{1.3}$$

Here, $\hat{r}_{ab} = |\hat{\boldsymbol{r}}_a - \hat{\boldsymbol{r}}_b|$ is the spatial distance of two particle wherein a lower-case index indicate electronic positions, upper-case indices stands for nuclear positions. Z_K is the electric charge of nucleus K.

The electronic part of the Schrödinger equation is crucially important since electrons play a key role in phenomena like electricity, magnetism, in thermal properties, and in all of chemistry. Developing efficient yet accurate approximations to the true solution of electronic structure problems and therefore expanding the scope of physical and chemical systems that can be predicted and analysed with computational methods is performed within the scientific field of electronic strucure theory which is a sub-field of quantum chemistry [7, 8]. Conversely, looking for ways to reduce the complexity of the electronic structure problem cannot only lead to quantitative improvements but also allows one to gain qualitative insights into the physics and chemistry of the systems at hand.

One important qualitative differentiation of electronic systems is the distinction between strongly and weakly correlated systems [9, 10]. If the electron correlation is weak, each electron can be treated to a good approximation as if it moves according to the mean field of all other electrons without introducing a large bias. These methods are comparatively easy to treat with mean-field methods like *Hartree–Fock theory* [11, 12]. On the other hand, systems in which the motion of electrons is highly correlated require much more expensive and sophisticated methods. One class among the many computational methods that try to tackle strongly correlated systems is *quantum Monte Carlo* (QMC) [13–15]. One particular QMC method, namely *full configuration interaction QMC* (FCIQMC), will be analysed and used as a tool throughout this thesis [16–18].

1.2 *Fermions and Pauli Exclusion Principle*

A fundamental fact of nature, that creates the structure of matter as we know it, is the fermionic nature of electrons. It acts on top of the Schrödinger equation (1.1) and puts an additional constraint on the many-particle wavefunction $|\Psi\rangle$. It is of particular importance for the analysis of electronic structure by QMC methods as it is the root cause for the infamous *fermionic sign problem* which will be the main topic of this thesis.

There are two fundamental classes of sub-atomic particles in the world. Every known particle is either a *boson* or a *fermion* [19]. The spin–statistics theorem provides a deep connection between the spin of a particle (a magnetic property that can only be correctly described in the context of quantum mechanics) and the particle statistics they obey – a concept that is only meaningful when dealing with multiple indistinguishable particles [20, 21].

While bosons have integer spin quantum numbers ($S = 0, 1, \dots$) and adhere to Bose–Einstein statistics, fermions have half-integer quantum numbers ($S = \frac{1}{2}, \frac{3}{2}, \dots$) and follow Fermi–Dirac statistics. This has major consequences for the structure of the physical world that we observe: While multiple bosons can occupy the same quantum state, the *Pauli exclusion principle* – that is a direct consequence of the Fermi–Dirac statistics – forbids this for fermions [22–24].

The Pauli exclusion principle can be formulated in many ways. For electrons in atoms, the exclusion principle states that no two electrons can have all equal quantum numbers (which are the principal quantum number n, the azimuthal quantum number ℓ, the z-projection the azimuthal quantum number m_ℓ, the spin quantum number s and the z-projection of the spin quantum number m_s) [25].

Another more formal way to define the Pauli exclusion principle is to put an additional sign constraint on the many-particle wavefunction $|\Psi\rangle$. While a many-particle wavefunction of bosons remains unchanged when two particles are exchanged, i.e.

$$\begin{aligned}\Psi_{\text{bos}}(\boldsymbol{r}_1, s_1; \dots; \boldsymbol{r}_i, s_i; \dots; \boldsymbol{r}_j, s_j; \dots; \boldsymbol{r}_n, s_n) \\ = \Psi_{\text{bos}}(\boldsymbol{r}_1, s_1; \dots; \boldsymbol{r}_j, s_j; \dots; \boldsymbol{r}_i, s_i; \dots; \boldsymbol{r}_n, s_n)\,,\end{aligned} \tag{1.4}$$

a fermionic wavefunction changes its sign in this case, i.e.

$$\begin{aligned}\left|\Psi_{\text{ferm}}(\boldsymbol{r}_1, s_1; \dots; \boldsymbol{r}_i, s_i; \dots; \boldsymbol{r}_j, s_j; \dots; \boldsymbol{r}_n, s_n)\right\rangle \\ = -\left|\Psi_{\text{ferm}}(\boldsymbol{r}_1, s_1; \dots; \boldsymbol{r}_j, s_j; \dots; \boldsymbol{r}_i, s_i; \dots; \boldsymbol{r}_n, s_n)\right\rangle\,.\end{aligned} \tag{1.5}$$

In other words, a fermionic wavefunction needs to be antisymmetric with respect to the exchange of particles. When trying to solve the stationary Schrödinger equation numerically, this constraint of the wavefunction can be built into a finite basis quite straightforwardly, e.g. by using *Slater determinants* [26–28]. However, the sign problem that is rooted in this antisymmetry property of fermionic wavefunctions lies much deeper and turns out to be NP-hard [29–32]. Being NP-hard however does not mean insoluble and that it cannot be remedied at least. In this thesis, the manifestations, implications, and possible mitigations of the sign problem in FCIQMC will discussed in detail.

1.3 *Lattice Models*

Although the Born–Oppenheimer approximation is already a significant simplification of an atomic or molecular Hamiltonian, it is still hard to solve when the number of electrons and orbitals becomes large. Especially in solids – which can be seen as macroscopic periodic molecules – solving the *ab initio* Hamiltonian from equation (1.3) is impossible when going beyond the mean-field approximation. Interesting physical effects like Mott insulating states, superconductivity etc. however occur precisely because of electron correlation effects.

It is therefore useful to further simplify the *ab initio* Hamiltonian and only let contributions remain that are necessary for a certain effect to occur. This not only makes numerical calculations easier, it also allows qualitative insights into which effective interactions are responsible for observed physical effects with more clarity, without the complication of irrelevant *ab initio* interaction terms.

Two examples of effective model Hamiltonians for solids are the *Heisenberg* [26] and the *Hubbard model* [33–35]. The Heisenberg model describes the magnetic interaction of localised spins on a lattice. Due to the localisation of the spins, the Heisenberg model can only describe insulating materials in a qualitatively correct fashion. For metals, kinetic movement of electrons across lattice sites has to be captured in the model. This is modeled in the Hubbard Hamiltonian where electrons can hop from one lattice sites to another while experiencing an on-site interaction on doubly occupied sites. The Heisenberg model is a limiting case of the Hubbard model for an infinite on-site interaction strength as the electrons become localised again.

1.4 *Context and Structure of the Thesis*

The focus of this thesis will be the application of FCIQMC to real-space lattice models. FCIQMC is a method that has been successfully applied to a variety of real systems, mostly in its initiator formulation [17, 36, 37]. While in principle the Hamiltonians of lattice models can be treated like the usual *ab initio* Hamiltonian, there are certain key differences. These differences are rooted in the fact that the structure of the wavefunction is different when using a real-space basis [38, 39]. Also, the strength of the sign problem is strongly affected. In this thesis, the FCIQMC method will be adapted and new features will be added in order to successfully tackle this class of problems, most of them dealing with overcoming the sign problem.

In part I *Theory & Foundations*, I will first introduce the foundations of QMC algorithms with a detailed introduction to FCIQMC, focusing on manifestations and existing ways of controlling the sign problem. I will also give a detailed overview of the real-space Hubbard lattice model.

In part II *Concepts & Results*, I will first present a classification of real-space Hubbard lattice geometries based on their respective strength of their sign problem in chapter 5. Ground-state eigenvalues of stoquastised Hamiltonians will be used for this purpose [38]. It will be found that there are certain non-trivial, yet sign-problem-free configurations. Also, the notion of non-size-extensive and size-extensive sign problems will be introduced.

I will then discuss the behaviour of FCIQMC in the special case of sign-problem-free systems. The absence of a sign problem allows for the calculation of very large lattices with an equally large number of electrons. However, these calculations uncover a new type of systematic bias in the FCIQMC algorithm that was previously masked by larger biases in the usual sign-problematic case. The bias is caused by walker population control. Similar biases have been shown to occur in diffusion Monte Carlo (DMC) [40–43] and Green's function Monte Carlo (GFMC) [44]. In chapter 6, large non-trivial sign-problem-free systems will be calculated. The results will be unbiased with respect to the population control bias by combining the newly introduced importance sampling with Gutzwiller-type guiding wavefunctions [34] and an a-posteriori reweighting scheme similar to one developed for DMC [40]. Importance sampling is widely used in other QMC methods like GFMC [45], DMC [46–49], and auxiliary-field QMC (AFQMC) [50–53] but has not been found useful in a systematic manner in FCIQMC thus far.

This will lead to chapter 7 where importance sampling with a Gutzwiller-like guiding wavefunction is applied to weakly sign-problematic systems like two-legged Hubbard ladders in the challenging intermediate interaction regime [54]. I will show how initiator-FCIQMC performs poorly in the treatment of these weakly sign-problematic systems. Instead, importance sampling has a significant and systematically beneficial effect onto the convergence of the ground-state energies. This allows for the calculation of the many-particle fundamental gaps which requires calculations at and close to half-filling in a numerically unbiased fashion. Unlike in other approaches where one attempts to reduce the gap between the ground-state energies between the stoquatised and the true Hamiltonian, respectively [55], through basis rotations, in case of importance sampling this gap is left un-

changed. Instead the efficiency of FCIQMC's discrete annihilation procedure is enhanced.

If one wants to tackle larger systems, new approximations are needed. In chapter 8, I will show how the usual population-based criterion can be effectively replaced by a criterion that is only based on the occupation structure of a Slater determinant itself. Unlike in a previously presented scheme where the initiator space has been defined based on prior selected configuration interaction (SCI) calculations [56], the selection of the fixed initiator space is based on analytical wavefunction ansatzes for the Hubbard model, feature-based, and therefore fast-to-evaluate. This way, a large portion of the correlation energy can be covered with only experiencing a very weak sign problem. This new method is subsequently combined with importance sampling. Lastly, I will introduce the two-shift expansion of FCIQMC, a way of perturbatively including the neglected space when using fixed initiator spaces. Unlike in previous approaches to correct for the bias caused by the initiator approximation – like the adaptive-shift method [57, 58] and a second-order Epstein–Nesbet (EN2) correction [59] – the two-shift method is specifically designed for the weak-sign-problem real-space Hubbard systems. Results for systems up to the 32-site Hubbard model in honeycomb geometry for intermediate interaction are presented.

PART I

THEORY & FOUNDATIONS

2 *Basic Concepts of Electronic Systems*

As already outlined in chapter 1, the fundamental goal in the field of electronic structure theory is to solve the electronic part of the *ab initio* Hamiltonian. In this chapter, I will present some basic concepts that are common to all approaches to the problem. I will also give an overview of two of these approaches that will be relevant in this thesis:

- the class of quantum Monte Carlo (QMC) methods to which full configuration interaction QMC (FCIQMC) belongs and
- density matrix renormalisation group (DMRG) which will be used as a benchmarking method at many points.

2.1 *The Second-Quantisation Formalism*

A very useful and insightful mathematical concept in the context of quantum many-body physics is the second quantisation formalism. In this formalism, bosonic many-particle states and operators are expressed in terms of creation operators $\hat{b}^\dagger_{i\sigma}$ and annihilation operators $\hat{b}_{i\sigma}$. Their fermionic counterparts are named $\hat{c}^\dagger_{i\sigma}$ and $\hat{c}_{i\sigma}$. One can interpret the action of $\hat{b}^\dagger_{i\sigma}$ ($\hat{c}^\dagger_{i\sigma}$) as creating a boson (fermion) in basis state i with spin σ. Similarly, one can interpret the adjoint $\sigma, \hat{b}_{i\sigma}$ ($\hat{c}_{i\sigma}$) as annihilating a boson (fermion) in basis state i with spin σ (if present). Because of this property, they are also called *ladder operators.*

2.1.1 *Commutation and Anticommutation Relations*

The crucial advantage of this formulation is the fact that the symmetry or antisymmetry of $|\Psi\rangle$ with respect to particle exchange can be readily included into the commutation relations of the respective creation and annihilation operators. The commutation relations for the bosonic operators are given by

$$\left[\hat{b}^\dagger_{i\sigma}, \hat{b}^\dagger_{i'\sigma'}\right] = \left[\hat{b}_{i\sigma}, \hat{b}_{i'\sigma'}\right] = 0\,, \tag{2.1a}$$

$$\left[\hat{b}_{i\sigma}, \hat{b}^\dagger_{i'\sigma'}\right] = \delta_{ii'}\delta_{\sigma\sigma'}\,. \tag{2.1b}$$

Here, δ_{ab} is the usual Kronecker delta with

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{else.} \end{cases} \tag{2.2}$$

$\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A}$ is the usual *commutator* of two operators $\hat{A}$ and $\hat{B}$. For fermions, one simply has to replace the commutator by the *anticommutator* defined as $\left\{\hat{A}, \hat{B}\right\} = \hat{A}\hat{B} + \hat{B}\hat{A}$. This means that

$$\left\{\hat{c}^\dagger_{i\sigma}, \hat{c}^\dagger_{i'\sigma'}\right\} = \left\{\hat{c}_{i\sigma}, \hat{c}_{i'\sigma'}\right\} = 0\,, \tag{2.3a}$$

$$\left\{\hat{c}_{i\sigma}, \hat{c}^\dagger_{i'\sigma'}\right\} = \delta_{ii'}\delta_{\sigma\sigma'}\,. \tag{2.3b}$$

So in reordering two fermionic ladder operators that create or annihilate particles in different single-particle basis states ($i\sigma \neq i'\sigma'$), a factor of -1 arises. As any fermionic many-body wavefunction can be written as

$$\prod_{n=1}^{N_\mathrm{el}} \hat{c}^\dagger_{i_n\sigma_n} |\rangle\,, \tag{2.4}$$

i.e. the action of fermionic creation operators onto the vacuum state, this exactly ensures the antisymmetry property. N_el denotes the number of electrons in the system.

2.1.2 *Ab Initio Hamiltonian in Second Quantisation*

With this knowledge, we can now construct the electronic part of the *ab initio* Hamiltonian from equation (1.3) in terms of the second-quantised operators:

$$\hat{H}_\mathrm{elec} = \sum_{ia\sigma} h^a_i \hat{c}^\dagger_{a\sigma}\hat{c}_{i\sigma} + \frac{1}{2}\sum_{ijab\sigma\sigma'} V^{ab}_{ij}\hat{c}^\dagger_{a\sigma}\hat{c}^\dagger_{b\sigma}\hat{c}_{j\sigma'}\hat{c}_{i\sigma'}\,. \tag{2.5}$$

Here, i, j, a, and b index the spatial single-particle basis functions, called *orbitals*. The coefficients h^a_i and V^{ab}_{ij} of the operator products are the integrals

$$h^a_i = \int \mathrm{d}\boldsymbol{r}\, \varphi^*_a(\boldsymbol{r})\left[-\frac{\nabla^2}{2} - \sum_K \frac{Z_K}{|\boldsymbol{r} - \boldsymbol{R}_K|}\right]\varphi_i(\boldsymbol{r})\,, \tag{2.6a}$$

$$V^{ab}_{ij} = \int \mathrm{d}\boldsymbol{r}_1\, \mathrm{d}\boldsymbol{r}_2\, \varphi^*_a(\boldsymbol{r}_1)\varphi^*_b(\boldsymbol{r}_2)\varphi_i(\boldsymbol{r}_1)\varphi_j(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \tag{2.6b}$$

where $\varphi_i(\boldsymbol{r})$ are the spatial orbitals chosen. $\boldsymbol{R}_K$ are the positions of the nuclei with charges Z_K. φ^* indicates the Hermitian conjugate of the orbital function φ.

Equation (2.5) also shows another distinct advantage of second quantisation: The information about the system's geometry and the orbital set is contained in the integrals from equation (2.6) entirely. The structure of the electronic Hamiltonian itself remains unchanged. This allows for the invention of generic solution algorithms.

2.1.3 *FCI Wavefunction and Slater Determinants*

Electronic wavefunctions can be represented in various ways. One of the most commonly used representations in quantum chemistry is the *full configuration interaction* (FCI) expansion

$$|\Psi_0\rangle = \sum_i C_i \, |D_i\rangle \, . \tag{2.7}$$

This is a full expansion of the ground-state wavefunction in terms of Slater determinants $|D_i\rangle$. A Slater determinant describes a wavefunction of multiple fermions in a way that the fermionic antisymmetry condition from equation (1.5) is automatically fulfilled [26–28]. As its name suggests, it is defined as the determinant of matrix. In a set of $2N_{\text{orb}}$ spin orbitals $\left\{\varphi_n(\boldsymbol{r})\right\}$, out of which the orbitals $n = n_1, n_2, \ldots, n_N$ are occupied by N_{el} electrons, it is defined as

$$\begin{aligned}
\left\langle \boldsymbol{r}_1 \boldsymbol{r}_2 \ldots \boldsymbol{r}_{n_N} \middle| D \right\rangle &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_{n_1}(\boldsymbol{r}_1) & \varphi_{n_2}(\boldsymbol{r}_1) & \ldots & \varphi_{n_N}(\boldsymbol{r}_1) \\ \varphi_{n_1}(\boldsymbol{r}_2) & \varphi_{n_2}(\boldsymbol{r}_2) & \ldots & \varphi_{n_N}(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{n_1}(\boldsymbol{r}_{N_{\text{el}}}) & \varphi_{n_2}(\boldsymbol{r}_{N_{\text{el}}}) & \ldots & \varphi_{n_N}(\boldsymbol{r}_{N_{\text{el}}}) \end{vmatrix} \\
&= \frac{1}{\sqrt{N!}} \left| \varphi_{n_1} \varphi_{n_2} \ldots \varphi_{n_{N_{\text{el}}}} \right| \\
&= \left\langle \boldsymbol{r}_1 \boldsymbol{r}_2 \ldots \boldsymbol{r}_{n_N} \middle| n_1, n_2, \ldots, n_{N_{\text{el}}} \right\rangle \, .
\end{aligned} \tag{2.8}$$

The expressions in the second and third lines are used as short forms if it is clear that wavefunction in question is a Slater determinant. One can see that the number of Slater determinants, i.e. the size of the Hilbert space $|\mathcal{H}|$, scales combinatorially as

$$|\mathcal{H}| = \binom{N_{\text{orb}}}{N_\uparrow} \binom{N_{\text{orb}}}{N_\downarrow} \tag{2.9}$$

where N_{orb} denotes the number of orbitals. $N_\uparrow$ ($N_\downarrow$) is the number of ↑-electrons (↓-electrons). This scaling behaviour makes determining the exact FCI expansion hard even for small system sizes and approximate methods have to be used.

Alternatively, totally antisymmetric states can also be expressed in terms of the second-quantisation formalism. In this case, a set of fermionic creation operators $\hat{c}_n^\dagger$ that each create an electron in spin orbital n act onto the vacuum state $|\rangle$. With this the definition of a fermionic many-particle state using second quantisation and using the above definition of a Slater determinant are equivalent:

$$\prod_n^{N_{\text{el}}} \hat{c}_n^\dagger |\rangle = |n_1, n_2, \dots, n_N\rangle \ . \tag{2.10}$$

The antisymmetry of the wavefunction in this formalism is already built in via the anticommutation relation. The determinant does not need to be evaluated explicitly.

Slater determinants are eigenfunctions of the $\hat{S}_z$ operator which is the observable that encodes the projection of the spin onto the z axis. However, not every Slater determinant is an eigenfunction of the total spin operator $\hat{S}^2$. Since the electronic *ab initio* Hamiltonian from equation (1.3) commutes with $\hat{S}^2$, $\left[\hat{H}, \hat{S}^2\right] = 0$, it already becomes clear that the solution of a general *ab initio* Hamiltonian cannot consist of a single Slater determinant. Truncations of the FCI expansion from equation (2.7) can have significant *spin contamination*, unless carefully constructed to take of this. That means that the truncated CI wavefunction is not an eigenstate of $\hat{S}^2$.

2.1.4 *Building the Many-Body Hamiltonian*

With the ingredients introduced in the previous sections, we can now construct the many-body Hamiltonian in Slater determinant space which will be diagonalised by methods like FCIQMC. There are two parts to building the many-body Hamiltonian:

- the interaction strength which will be calculated using the integrals from equation (2.6) using the *Slater–Condon rules* [60] and
- the Fermi phase which is a direct consequence of the antisymmetry constraint and depends on the chosen ordering of orbitals.

As the fermionic sign problem plays an important role in this work, the Fermi phase of the matrix elements merits further discussion.

When defining a many-particle state, one has to establish a convention for the ordering of the spin orbitals. In principle, one can choose any ordering but there are two principal ways of doing it:

- *orbital-first ordering* in which a Slater determinant in second quantisation is defined as

$$\hat{c}^\dagger_{i_1\sigma_1}\hat{c}^\dagger_{i_2\sigma_2}\dots\hat{c}^\dagger_{i_N\sigma_N}\,|\rangle \tag{2.11}$$

  and

- *spin-first ordering* in which a Slater determinant is defined as

$$\hat{c}^\dagger_{a_1\uparrow}\hat{c}^\dagger_{a_2\uparrow}\dots\hat{c}^\dagger_{a_n\uparrow}\hat{c}^\dagger_{b_1\downarrow}\hat{c}^\dagger_{b_2\downarrow}\dots\hat{c}^\dagger_{b_n\downarrow}\,|\rangle\,. \tag{2.12}$$

$i_1 \leq i_2 \leq \dots \leq i_N$ are the spatial orbitals that are occupied by the N_el electrons, each with their spin projection σ_i. $a_1 \leq a_2 \leq \dots \leq a_n$ in the second representation are the spatial orbitals occupied with ↑-electrons, $b_1 \leq b_2 \leq \dots \leq b_n$ are the ones occupied with ↓-electrons. ↑ and ↓ are the spin projections of the respective electrons. Using the fermionic commutation relations from equation (2.3), we can already see that by transforming equations (2.11) and (2.12) into each other the different representations do not have the same sign. Let us look at a minimal example with three electrons in spin orbitals $a_1\uparrow$, $a_2\uparrow$, and $b_1\downarrow$. In orbital-first ordering, the corresponding many-particle state is given by $\hat{c}^\dagger_{a_1\uparrow}\hat{c}^\dagger_{b_1\downarrow}\hat{c}^\dagger_{a_2\uparrow}\,|\rangle$. To obtain spin-first ordering, the operators $\hat{c}^\dagger_{b_1\downarrow}$ and $\hat{c}^\dagger_{a_1\uparrow}$ have to be exchanged. Thus, according to equation (2.3), a factor of -1 is picked up:

$$\hat{c}^\dagger_{a_1\uparrow}\hat{c}^\dagger_{b_1\downarrow}\hat{c}^\dagger_{a_2\uparrow}\,|\rangle = -\hat{c}^\dagger_{a_1\uparrow}\hat{c}^\dagger_{a_2\uparrow}\hat{c}^\dagger_{b_1\downarrow}\,|\rangle\,. \tag{2.13}$$

Let us now look at matrix elements in different orbital orderings. Suppose we are looking at a single-electron excitation of an ↑-electron from spatial orbital 2 to 3 starting from determinant $|D_i\rangle = |1\uparrow\ 2\uparrow\ 2\downarrow\rangle$ with amplitude t. This means that we are exciting to $|D_j\rangle = |1\uparrow\ 2\downarrow\ 3\uparrow\rangle$. The matrix element of this excitation in orbital-first ordering (of) is given by

$$\begin{aligned} H^\mathrm{of}_{ij} = \left\langle D^\mathrm{of}_j \middle| \hat{H} \middle| D^\mathrm{of}_i \right\rangle &= \left\langle \overleftarrow{\left|\hat{c}^\dagger_{3\uparrow}\hat{c}^\dagger_{2\downarrow}\hat{c}^\dagger_{1\uparrow}\right|} \overrightarrow{t\hat{c}^\dagger_{3\uparrow}\hat{c}_{2\uparrow}\hat{c}^\dagger_{1\uparrow}\hat{c}^\dagger_{2\uparrow}\hat{c}^\dagger_{2\downarrow}} \right\rangle \\ &= t\left\langle \left|\hat{c}_{3\uparrow}\hat{c}_{2\downarrow}\hat{c}_{1\uparrow}\underbrace{\hat{c}^\dagger_{3\uparrow}\hat{c}_{2\uparrow}}\hat{c}^\dagger_{1\uparrow}\hat{c}^\dagger_{2\uparrow}\hat{c}^\dagger_{2\downarrow}\right| \right\rangle \\ &= t\left\langle \left|\hat{c}_{3\uparrow}\hat{c}_{2\downarrow}\hat{c}_{1\uparrow}\hat{c}^\dagger_{1\uparrow}\underbrace{\hat{c}^\dagger_{3\uparrow}\hat{c}^\dagger_{2\downarrow}}\right| \right\rangle \\ &= -t\,. \end{aligned} \tag{2.14}$$

Conversely, the matrix element in spin-first ordering (sf) is

$$\begin{aligned} H_{ij}^{\mathrm{sf}} = \left\langle D_j^{\mathrm{sf}} \middle| \hat{H} \middle| D_i^{\mathrm{sf}} \right\rangle &= \left\langle \middle| \overleftarrow{\hat{c}_{2\downarrow}^\dagger \hat{c}_{3\uparrow}^\dagger \hat{c}_{1\uparrow}^\dagger} \middle| t \overrightarrow{\hat{c}_{3\uparrow}^\dagger \hat{c}_{2\uparrow} \hat{c}_{1\uparrow}^\dagger \hat{c}_{2\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger} \middle| \right\rangle \\ &= t \left\langle \middle| \hat{c}_{2\downarrow} \hat{c}_{3\uparrow} \hat{c}_{1\uparrow} \underset{\curvearrowright}{\hat{c}_{3\uparrow}^\dagger \hat{c}_{2\uparrow}} \underset{\curvearrowleft}{\hat{c}_{1\uparrow}^\dagger} \hat{c}_{2\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger \middle| \right\rangle \\ &= t \,. \end{aligned} \tag{2.15}$$

Arrows above operators indicate the direction they are applied in. No arrow means that the operators are applied to the right. Curved arrows below indicate possibly sign-changing commutations of operators. Therefore, these equations show that the sign of the matrix elements in Slater determinant space $\langle D_j|\hat{H}|D_i\rangle$ is not independent of the choice of the ordering.

The Hamiltonian matrices $\mathbf{H}^o$ and $\mathbf{H}^{o'}$ that are representated in Slater determinant bases with different orderings o and o' are connected via similarity transformations by a purely diagonal matrix $\mathbf{D}$. Since a change of ordering does not affect the magnitudes of the matrix elements, the diagonal elements of $\mathbf{D}$ are either $+1$ or -1. This can be written as

$$\mathbf{H}^{o'} = \mathbf{D}\mathbf{H}^o\mathbf{D} \,. \tag{2.16}$$

As we will see in section 3.2, this means that a change in ordering does not change the severity of the QMC sign problem in methods that diagonalise the Hamiltonian in Slater determinant space, like FCIQMC.

2.2 *Overview of QMC Methods*

The general idea behind Monte Carlo methods (MC methods) is the following: Instead of solving the problem either analytically or numerically by using deterministic algorithms, Monte Carlo methods use random samples of the defining equation of the respective problem [61, 62]. For a large number of samples, the numerical results will approach the true solution on average if certain conditions are met. A basic paradigmatic example of applying MC methods is the estimation of π by filling a 1×1 quadrant with random points with coordinates (x_i, y_i). The ratio of the number of random points that fulfill $\sqrt{x_i^2 + y_i^2} \leq 1$, i.e. that are contained in a quadrant of a unit circle, to the total number of points approaches $\frac{\pi}{4}$. Ultimately, what has been done here is the MC evaluation of a two-dimensional integral over a constant function.

In general, MC can be used to solve arbitrary integrals

$$I = \int_\Omega \mathrm{d}\boldsymbol{r}\, f(\boldsymbol{r}) \tag{2.17}$$

of arbitrary dimension. To do this, the integrand $f(\boldsymbol{r})$ is rewritten as $\frac{f(\boldsymbol{r})}{p(\boldsymbol{r})} p(\boldsymbol{r})$. Here, $p(\boldsymbol{r})$ is a probability distribution, i.e.

$$\int_\Omega \mathrm{d}\boldsymbol{r}\, p(\boldsymbol{r}) = 1 \quad \text{and} \quad p(\boldsymbol{r}) \geq 0 \text{ for all } \boldsymbol{r}. \tag{2.18}$$

The MC approximation to the integral is then given by

$$I_n \approx \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{r}_i) \tag{2.19}$$

where $\boldsymbol{r}_i$ are random samples drawn from $p(\boldsymbol{r})$ and n is the number of random samples. I_n converges to I for large n independent of the choice of $p(\boldsymbol{r})$ as long as it fulfills the criteria from equation (2.18). The expectation value $\overline{I_n} = I$ for all n. The variance of the estimate I_n for finite n however strongly depends on the choice of $p(\boldsymbol{r})$. In general, the variance of I_n can be determined using the *central limit theorem* (CLT). The CLT states that when summing up independent random variables that are not necessarily drawn from a normal distribution – like in equation (2.19) where $f(\boldsymbol{r}_i)$ can be distributed arbitrarily – the sample mean I_n is distributed according to a normal distribution with mean I and standard deviation

$$\sigma_n = \frac{\sigma_f}{\sqrt{n}}\,. \tag{2.20}$$

σ_f is the standard deviation of the function f. As a consequence of the CLT, the $1/\sqrt{n}$ scaling is universally observed in all MC methods. When $p(\boldsymbol{r})$ is chosen equal to $f(\boldsymbol{r})$, then $\sigma_f = 0$ and thus $\sigma_n = 0$ for all n. Selecting a sampling distribution as close as possible to the sampled function f is therefore desirable and is called *importance sampling* [63].[1] Selecting a sequence of configurations $\boldsymbol{r}_i$ for arbitrary probability distributions $p(\boldsymbol{r})$ is non-trivial, especially if $\boldsymbol{r}$ has a large dimension. This problem is solved in general by the *Metropolis–Hastings algorithm* which can be applied even if the normalisation of $p(\boldsymbol{r})$ is unknown [64, 65]. It is an algorithm that creates a *Markov chain* of $\boldsymbol{r}_i$ samples, i.e. $\boldsymbol{r}_{j+1}$ only depends on its predecessor $\boldsymbol{r}_j$. If one is dealing with a discrete and normalised distribution, the efficient *alias method* can be used instead [66].

[1] Applying importance sampling to FCIQMC will play an important role throughout part II.

As solving the Schrödinger equation (1.2) can be regarded as an integration problem, among the countless solution methods quantum Monte Carl (QMC) methods play an important role. In this section, I will give a brief overview of some of the most widely used QMC methods. This is to put FCIQMC – that will be discussed in detail in chapter 3 – into a more

general context and explore similarities and differences. I will touch on the following methods:

- *Variational Monte Carlo* (VMC) which optimises the parameters of a wavefunction ansatz with respect to the variational energy or its variance. This plays a role in part II when optimising wavefunction ansatzes for importance sampling.
- *Diffusion Monte Carlo* (DMC) which is closely related to FCIQMC as it is also a projector QMC technique and it also uses stochastic walkers to sample the wavefunction. Also, a similar bias to the population control in FCIQMC that will be discussed in chapter 6 has been known in DMC before.
- *Auxiliary-field QMC* (AFQMC) which solves the Schrödinger equation in a space of so-called auxiliary fields. Its special significance lies in the fact that the two-dimensional Hubbard model at half-filling is sign-problem-free which is not the case in FCIQMC.

2.2.1 *Variational Monte Carlo*

In variational Monte Carlo (VMC), a trial wavefunction $|\Psi_\mathrm{t}(\boldsymbol{a})\rangle$ is optimised according to the variational principle [15, 67, 68]. $\boldsymbol{a}$ is a vector of parameters the trial wavefunction depends on. The variational principle says that the variational energy

$$E(\boldsymbol{a}) = \frac{\left\langle \Psi_\mathrm{t}(\boldsymbol{a}) \middle| \hat{H} \middle| \Psi_\mathrm{t}(\boldsymbol{a}) \right\rangle}{\langle \Psi_\mathrm{t}(\boldsymbol{a}) \,|\, \Psi_\mathrm{t}(\boldsymbol{a}) \rangle} \tag{2.21}$$

is minimal if and only if $|\Psi_\mathrm{t}(\boldsymbol{a})\rangle \propto |\Psi_0\rangle$, the true ground-state wavefunction.

In VMC, the variational energy integral from equation (2.21) is evaluated using Monte Carlo sampling. To do this, it has to be expressed like the integral in equation (2.17). This leads to

$$E(\boldsymbol{a}) = \frac{\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \left|\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i\}, \boldsymbol{a}\big)\right|^2 \frac{\hat{H}\Psi_\mathrm{t}\,(\{\boldsymbol{r}_i\}, \boldsymbol{a})}{\Psi_\mathrm{t}\,(\{\boldsymbol{r}_i\}, \boldsymbol{a})}}{\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \left|\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i\}, \boldsymbol{a}\big)\right|^2} \,. \tag{2.22}$$

Here,

$$p\big(\{\boldsymbol{r}_i\}\big) = \frac{\left|\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i\}, \boldsymbol{a}\big)\right|^2}{\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \left|\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i\}, \boldsymbol{a}\big)\right|^2} \tag{2.23}$$

obviously has the properties of a probability density from equation (2.18). This means that the function

$$E_{\mathrm{loc}}\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big) = \frac{\hat{H}\Psi_{\mathrm{t}}\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big)}{\Psi_{\mathrm{t}}\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big)} \tag{2.24}$$

can be sampled by drawing positions $\boldsymbol{r}_i$ from the distribution $p\big(\{\boldsymbol{r}_i\}\big)$, e.g. by using the Metropolis–Hastings algorithm. As mentioned before, it is not even necessary to know the normalisation factor in this case. E_{loc} is called the *local energy*. The energy is thus given as the sum over the local energies:

$$E(\boldsymbol{a}) = \frac{1}{n}\sum_{j=1}^{n} E_{\mathrm{loc}}\big(\{\boldsymbol{r}_i\}_j,\boldsymbol{a}\big)\,. \tag{2.25}$$

$\{\boldsymbol{r}_i\}_j$ denotes the j-th sample of the electronic positions. Accordingly, the parameter vector $\boldsymbol{a}$ can be optimised by minimising $E(\boldsymbol{a})$ directly. It is also possible to not optimise $E(\boldsymbol{a})$ itself but rather minimise the variance of the local energy. The variance of E_{loc} with respect to the spatial coordinates $\{\boldsymbol{r}_i\}$ is given by

$$\mathrm{var}\big[E_{\mathrm{loc}}\big](\boldsymbol{a}) = \frac{\displaystyle\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \Big|\Psi_{\mathrm{t}}\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big)\Big|^2 E_{\mathrm{loc}}^2\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big)}{\displaystyle\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \Big|\Psi_{\mathrm{t}}\big(\{\boldsymbol{r}_i\},\boldsymbol{a}\big)\Big|^2} - E^2(\boldsymbol{a})\,. \tag{2.26}$$

Because the local energy is a constant when the trial wavefunction equals the exact ground-state wavefunction,

$$\frac{\hat{H}\Psi_0\big(\{\boldsymbol{r}_i\}\big)}{\Psi_0\big(\{\boldsymbol{r}_i\}\big)} = E_0\,, \tag{2.27}$$

the variance is not only minimal but zero in this case. Further, the minimisation of the variance can also be applied to excited states [69]. However, empirically it is found that in most cases the trial wavefunction is better able to estimate other properties when the energy is minimised instead of the variance. Alternatively, a linear combination of $E(\boldsymbol{a})$ and $\mathrm{var}\big[E_{\mathrm{loc}}\big](\boldsymbol{a})$ can be optimised. Finding the best optimisation procedure of VMC trial wavefunctions is still an active research topic.

A commonly used trial wavefunction for *ab initio* systems is the *Slater–Jastrow wavefunction* [70]. Its functional form is given by

$$\Psi_{\mathrm{t}}^{\mathrm{SJ}} = \exp(\mathcal{J})\sum_{i=1}^{N_{\mathrm{det}}} C_i\,|D_i\rangle\,. \tag{2.28}$$

The *Jastrow correlation factor* $\mathcal{J}$ is of the form

$$\mathcal{J} = \sum_{ijAB} u_1(R_{iA}) + u_2(r_{ij}) + u_3(r_{ij}, R_{iA}, R_{jB}) + \ldots \tag{2.29}$$

where r and R denote electron–electron and R the electron–nucleus distances in a molecule. Wavefunction ansatzes for the lattice models are discussed in detail in part II as they are used in a novel fashion in FCIQMC as well.

All VMC calculations in this thesis were conducted using a VMC code by P. López Ríos.

2.2.2 *Diffusion Monte Carlo*

Diffusion Monte Carlo (DMC) is the Monte Carlo that is conceptionally closest to FCIQMC as it is also a projector technique, i.e. it also uses the imaginary-time projection of the Schrödinger equation to project out the ground state [46–49]. Also DMC, like FCIQMC, has the concept of *stochastic walkers* propagating according to a master equation. Due to the similarities, problems and biases like the *population control bias* that are observed in DMC are also observed in FCIQMC, as we will see in part II. We will also see there how concepts like importance sampling that are used routinely in DMC can also be applied in FCIQMC and lead to algorithmic improvements. There are also considerable differences between the algorithms. As will be discussed in detail in chapter 3, FCIQMC works in a discrete set of Slater determinants to stochastically represent the FCI wavefunction from equation (2.7). Walkers in DMC propagate in a continuous space. In contrast, the discrete propagation of walkers in FCIQMC has significant advantages in mitigating the sign problem through *annihilations.* Based on the annihilation algorithm, we can further improve the mitigation of the sign problem in FCIQMC. Therefore, it is necessary to give a brief overview of the DMC algorithm and introduce the concepts mentioned above in this context.

Basic Algorithm

DMC attempts to solve the stationary Schrödinger equation (1.1) in imaginary time $\tau = it$ which is called a *Wick rotation* [71]. When writing $|\Psi(\tau)\rangle$ in a real-space basis, this leads to

$$-\frac{\partial \Psi\left(\{\boldsymbol{r}_i\}, \tau\right)}{\partial \tau} = \left(\hat{H} - E_\mathrm{t}\right)\Psi\left(\{\boldsymbol{r}_i\}, \tau\right). \tag{2.30}$$

E_t is the trial energy that will be defined later on. Inserting the Hamiltonian in the Born–Oppenheimer approximation from equation (1.3) and separating it into a kinetic part $\hat{T}$ and a potential part $\hat{V}$ leads to

$$\begin{aligned} -\frac{\partial\Psi\left(\{\boldsymbol{r}_i\},\tau\right)}{\partial\tau} &= \left(\hat{T}+\hat{V}-E_\mathrm{t}\right)\Psi\left(\{\boldsymbol{r}_i\},\tau\right) \\ &= \underbrace{-\frac{1}{2}\nabla_i^2\Psi\left(\{\boldsymbol{r}_i\},\tau\right)}_{\text{diffusion}} + \underbrace{\left[V\left(\{\boldsymbol{r}_i\}\right)-E_\mathrm{t}\right]\Psi\left(\{\boldsymbol{r}_i\},\tau\right)}_{\text{branching}} . \end{aligned} \tag{2.31}$$

The kinetic part $\hat{T}$, which equals the off-diagonal elements in a Slater determinant basis, leads to a differential equation that describes a diffusion process. The potential part $\hat{V}$, which equals the diagonal elements, describes a branching process. Integrating equation (2.31) and using the *Trotter decomposition*

$$\exp\left[\tau\left(\hat{T}+\hat{V}\right)\right] \approx \left[\exp\left(\Delta\tau\hat{T}\right)\exp\left(\Delta\tau\hat{V}\right)\right]^n \quad \text{with}\,\tau=\frac{\Delta\tau}{n}\,\text{and}\,\Delta\tau \ll \tau \tag{2.32}$$

allows for the stepwise solution of the differential equation. The wavefunction is propagated from the N_el electron positions $\{\boldsymbol{r}_i'\}$ at imaginary time τ to the positions $\{\boldsymbol{r}_i\}$ at time $\tau+\Delta\tau$ according to

$$\begin{aligned} &\Psi\left(\{\boldsymbol{r}_i\},\tau+\Delta\tau\right) \\ &= \int \mathrm{d}\boldsymbol{r}_1' \ldots \mathrm{d}\boldsymbol{r}_N' \left\langle \boldsymbol{r}_i \middle| \exp\left(\Delta\tau\hat{T}\right) \middle| \boldsymbol{r}_i' \right\rangle \exp\left[\Delta\tau\left(\hat{V}-E_\mathrm{t}\right)\right]\Psi\left(\{\boldsymbol{r}_i'\},\tau\right). \end{aligned} \tag{2.33}$$

Algorithmically, in each timestep the wavefunction is represented by a set of walkers that have spatial coordinates $\{\boldsymbol{r}_i\}$ assigned to them. Their propagation is implemented as follows:

- In the *diffusion step,* every walker is propagated according to

$$\boldsymbol{r}_i = \boldsymbol{r}_i' + \sqrt{\Delta\tau}\eta(\tau). \tag{2.34}$$

  η is a random number drawn from a Gaussian distribution. This is the solution to the diffusion part of the propagation. Significantly larger time steps can be used when adding a Metropolis-like accept/reject step [40].

- The *branching step* can be called a "birth and death" step and is accounted for by creating

$$N_\mathrm{w}\left(\{\boldsymbol{r}_i\}\right) = \exp\left[\Delta\tau\left(\hat{V}\left(\{\boldsymbol{r}_i\}\right)-E_\mathrm{t}\right)\right] \tag{2.35}$$

copies of a walker with positions $\{\boldsymbol{r}_i\}$.

It is easy to see that the algorithm is trivially parallelisable as the operations purely act on individual walkers. This means that each walker can be assigned to a specific parallel process with load-balancing every couple of iterations. Hashing and all-to-all communications of spawned particles like in FCIQMC are not required (see section 3.1.3).

Importance Sampling

With the scheme presented so far, the algorithm is exact. Usually when the algorithm is applied in this simple form, the total walker number N_{tot} fluctuates strongly. This is especially a problem when the wavefunction is sampled close to points where it has a cusp, e.g. at the nucleus. This can be improved by applying importance sampling. Instead of the wavefunction $\Psi\big(\{\boldsymbol{r}_i\}\big)$, the product of Ψ and a trial wavefunction Ψ_{t}

$$f\big(\{\boldsymbol{r}_i\},\tau\big) = \Psi\big(\{\boldsymbol{r}_i\},\tau\big)\Psi_{\text{t}}\big(\{\boldsymbol{r}_i\}\big) \tag{2.36}$$

is sampled. The master equation of the product function then reads

$$-\frac{\partial f\big(\{\boldsymbol{r}_i\},\tau\big)}{\partial\tau} = \underbrace{-\frac{1}{2}\nabla_i^2 f\big(\{\boldsymbol{r}_i\},\tau\big)}_{\text{diffusion}} \underbrace{-\nabla\Big[f\big(\{\boldsymbol{r}_i\},\tau\big)\nabla\ln\Psi_{\text{t}}\big(\{\boldsymbol{r}_i\}\big)\Big]}_{\text{drift}} + \underbrace{\Big[E_{\text{loc}}\big(\{\boldsymbol{r}_i\}\big) - E_{\text{t}}\Big]f\big(\{\boldsymbol{r}_i\},\tau\big)}_{\text{branching}}\,. \tag{2.37}$$

The local energy E_{loc} has the same definition as it has already been introduced for VMC in equation (2.24). In contrast to the master equation (2.31), importance sampling adds a drift to the diffusion step according to

$$\boldsymbol{r}_i = \boldsymbol{r}_i' + \Delta\tau\nabla\ln\Psi_{\text{t}}(\boldsymbol{r}_i') + \sqrt{\Delta\tau}\eta(\tau)\,. \tag{2.38}$$

Intuitively, the drift pushes the walkers to locations where the trial wavefunction is large. Evaluating the trial wavefunction at the walker positions takes up most of the computing time in DMC.

Population Control

With the algorithm as presented so far, the total number of walkers is not conserved. This means that the total walker number can grow beyond

available computational resources or die out completely, even when applying importance sampling. Therefore, population control is mandatory [72].

The most commonly used population control algorithm is the following: The trial energy E_t is adjusted according to

$$E_\mathrm{t}(\tau) = E_\mathrm{DMC}(\tau) + \frac{\gamma}{\Delta\tau}\frac{N_\mathrm{tot}}{N_\mathrm{target}}\,. \tag{2.39}$$

The total walker population is given by

$$N_\mathrm{tot} = \int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \Big| f\big(\{\boldsymbol{r}_i\}, \tau\big)\Big|\,. \tag{2.40}$$

N_target is the desired target walker population. E_DMC is the best estimate of the total energy and can be calculated by averaging the local energy according to

$$E_\mathrm{DMC}(\tau) = \frac{\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N\, f\big(\{\boldsymbol{r}_i\}, \tau\big) E_\mathrm{loc}\big(\{\boldsymbol{r}_i\}\big)}{\int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N\, f\big(\{\boldsymbol{r}_i\}, \tau\big)}\,. \tag{2.41}$$

γ is a damping parameter. The calculation of N_tot and E_DMC requires communication of information between walkers that, in a parallel implementation, might be located on different computational processes.

Population control introduces a bias in the sampled wavefunction [41–43]. In chapter 6, we will discover a similar bias in FCIQMC as well. There, it will also be compared to the population control bias in DMC.

Fixed-Node Approximation

Like any other QMC method, DMC also suffers from the sign problem when trying to solve a general fermionic system. The direct cause for this problem in DMC is the fact that the sampled function – f in the case of importance sampling – is not positive everywhere. Therefore, it is not a proper probability distribution. If no further measures are taken, Ψ would collapse to the solution of the *stoquastised Hamiltonian.*[2] To mitigate the problem, the *fixed-node approximation* has been introduced [69, 73, 74]. In this scheme, the nodes of the sampled wavefunction are not allowed to change. This is ensured by one additional algorithmic step after the diffusion and drift step that moves an electron from $\boldsymbol{r}_i'$ to $\boldsymbol{r}_i$: It is checked whether the sign of $\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i\}\big)$ agrees with the sign of $\Psi_\mathrm{t}\big(\{\boldsymbol{r}_i'\}\big)$. If it does not, i.e. a node is crossed, the move is rejected and the algorithms jumps to the next walker. By applying this constraint a systematic bias is introduced.

2 For a detailed discussion see section 3.2 and part II.

Generally, the fixed-node approximation becomes better when the nodes of the trial wavefunction are more accurate, e.g. by using a CI wavefunction with a large number of Slater determinants. They, however, are usually hard to obtain. The fixed-node approximation can be weakened at the price of reducing the size of the treatable systems: In the *released-node method,* the simulation is first equilibrated in the fixed-node approximation [75–78]. Subsequently, the walkers are allowed to cross the nodes of the trial wavefunction. If the fixed nodes are close to the true nodes, the simulation is stable for a short period of imaginary time and a released-node energy can be obtained.

2.2.3 *Auxiliary-Field QMC*

Auxiliary-Field QMC (AFQMC) is another QMC method that will be discussed in the context of this work to contrast it with FCIQMC. Like FCIQMC, it is a projector QMC method that uses the imaginary-time propagation according to equation (3.2) [50–53]. Another similarity lies in the fact that it operates in a space of Slater determinants (see section 2.1.3).

However, unlike FCIQMC which uses a linearised version of the exponential propagator which is stochastically applied as shown in equation (3.3), AFQMC sticks to the exponential propagator. Instead, AFQMC exploits the fact that for an arbitrary single-particle operator $\hat{O}_1 = \sum_{ij} O_{ij} \hat{c}_i^\dagger \hat{c}_j$ the exponential operator $\exp(\hat{O}_1)$ is simply an orbital rotation. It means that

$$\exp\left(\hat{O}_1\right)\left|\varphi_{n_1}\varphi_{n_2}\dots\varphi_{n_N}\right| = \left|\varphi'_{n_1}\varphi'_{n_2}\dots\varphi'_{n_N}\right|, \tag{2.42}$$

i.e. the exponential of a single-particle operator maps a single Slater determinant in an orbital basis $\left\{\varphi_n(\boldsymbol{r})\right\}$ onto another single Slater determinant in a different orbital basis $\left\{\varphi'_n(\boldsymbol{r})\right\}$. As we have already seen before, generic quantum chemistry as well as model Hamiltonians can be written as $\hat{H} = \hat{T} + \hat{V}$. $\hat{T}$ again is a kinetic one-body operator, $\hat{V}$ is a potential-energy two-body operator. In AFQMC, the two-body operator $\hat{V}$ is mapped onto one-body operators using a *Hubbard–Stratonovich transformation.*

Hubbard–Stratonovich Transformation

First, the exponential propagator $\exp\left(-\tau\hat{H}\right) = \exp\left[-\tau\left(\hat{T} + \hat{V}\right)\right]$ needs to be decomposed into a product. Like in DMC, this is done using a *Trotter decomposition* as shown in equation (2.32). To evaluate the effect of the prop-

agation of the two-body part $\hat{V}$ for a small time step $\Delta\tau$, i.e. of $\exp\left(-\Delta\tau\hat{V}\right)$, $\hat{V}$ is rewritten as

$$\hat{V} = \sum_i \lambda_i \hat{v}_i^2 \tag{2.43}$$

where $\hat{v}_i$ are one-body operators. This is possible for all two-body operators. The Hubbard–Stratonovich of one of these factors after the Trotter breakup is then given by

$$\exp\left(\frac{\Delta\tau}{2}\lambda_i\hat{v}_i^2\right) = \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \exp\left(\sqrt{-\Delta\tau\lambda_i}x\hat{v}_i\right). \tag{2.44}$$

This means that the exponential of a two-body operator is now expressed as an integral over infinitely many single-particle operators [79, 80]. x is called the auxiliary-field variable. There is one of these auxiliary-field integrals for each λ_i. The entire exponential propagator, including the one-body part and all auxiliary-field integrals from the two-body part, can then be rewritten as

$$\exp\left(-\tau\hat{H}\right) = \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x})\hat{B}(\boldsymbol{x}). \tag{2.45}$$

$\boldsymbol{x}$ is the vector of all auxiliary-field variables. $p(\boldsymbol{x}) = (2\pi)^{-\frac{1}{2}}\exp(-\boldsymbol{x}^2/2)$ is a multi-dimensional standard Gaussian probability distribution while $\hat{B}(\boldsymbol{x})$ is a product of single-particle operators. The imaginary-time projection of an initial state $|\Psi^0\rangle$, which is a superposition of Slater determinants $|D_i^0\rangle$,

$$|\Psi_0\rangle = \lim_{\tau\to\infty} \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x})\hat{B}(\boldsymbol{x})\,|\Psi^n\rangle \tag{2.46}$$

just like in FCIQMC ultimately leads to the true ground-state solution. Since the integrals are of the type that have already been introduced in equation (2.17), they can be evaluated using MC sampling. As the single Slater determinants perform random walks, they are often called walkers. They should not be mistaken with DMC walkers – that consist of positions in real-space as wavefunction coordinates – or FCIQMC walkers – that randomly sample coefficients of an FCI expansion and reside on Slater determinants. A walker i in iteration n, i.e. a specific Slater determinant $|D_i^n\rangle$ is propagated to another – possibly non-orthogonal – Slater determinant $|D_i^{n+1}\rangle$ by drawing an $\boldsymbol{x}$ according to the distribution $p(\boldsymbol{x})$ and applying $\hat{B}(\boldsymbol{x})$ to it. The wavefunction estimate in iteration n is then given by

$$|\Psi^n\rangle = \sum_i w_i \left|D_i^n\right\rangle \tag{2.47}$$

where w_i are scalar weights. The ground-state energy can simply be estimated by projecting $\hat{H}\,|\Psi^n\rangle$ onto a trial wavefunction $|\Psi_\mathrm{t}\rangle$ like

$$E^n_\mathrm{AFQMC} = \frac{\left\langle \Psi_\mathrm{t} \middle| \hat{H} \middle| \Psi^n \right\rangle}{\langle \Psi_\mathrm{t} \,|\, \Psi^n \rangle}\,, \tag{2.48}$$

just like the trial energy in FCIQMC as given in equation (3.15). Like any other MC method, AFQMC suffers from the sign problem which manifests as the *phase problem* in the most general case which will be discussed later.

Importance Sampling

The sampling distribution $p(\boldsymbol{x})$ so far does not incorporate any information about the sampled wavefunction. Even though the algorithm is exact in principle, sampling might be inefficient and unimportant determinants might be sampled often. So just like in DMC, to improve the sampling efficiency importance sampling is applied. This means that the sampling distribution is scaled to incorporate information about the ground state according to

$$\tilde{p}(\boldsymbol{x}) = \frac{\left\langle \Psi_\mathrm{t} \middle| D_i^{n+1} \right\rangle}{\left\langle \Psi_\mathrm{t} \middle| D_i^{n} \right\rangle} p(\boldsymbol{x})\,. \tag{2.49}$$

$|\Psi_\mathrm{t}\rangle$ again is a trial wavefunction that should have a large overlap with the true ground state $|\Psi_0\rangle$. Also, $|\Psi_\mathrm{t}\rangle$ should allow for an efficient calculation of the overlaps $\langle \Psi_\mathrm{t} | D_i^n \rangle$.

The importance-sampled imaginary-time projection is then given by

$$\left|\tilde{\Psi}_0\right\rangle = \lim_{\tau\to\infty} \int \mathrm{d}\boldsymbol{x}\; \tilde{p}(\boldsymbol{x}) \hat{B}(\boldsymbol{x})\,|\Psi^n\rangle \tag{2.50}$$

where instead of $|\Psi_0\rangle$ the rescaled ground-state wavefunction

$$|\Psi_0\rangle = \sum_i \langle \Psi_\mathrm{t} \,|\, D_i \rangle\, w_i\, |D_i\rangle \tag{2.51}$$

is approached in the infinite-τ limit. For $|\Psi_\mathrm{t}\rangle = |\Psi_0\rangle$, the weights w_i remain constant. Thus, all walkers $|D_i^n\rangle$ contribute equally to the calculation of the energy according to equation (2.48) after equilibration. Therefore, the variance approaches zero.

The Phase Problem

When looking at the Hubbard–Stratonovich transformation in equation (2.44), it is evident that the exponential one-body operator $\exp\left(\sqrt{-\Delta\tau\lambda_i}x\hat{v}_i\right)$ can be complex as the λ_i can either be positive or negative. This means that

Slater determinants will pick up complex phases when propagated along the auxiliary-field paths according to equation (2.46). As there is an infinite set of equally valid solutions to the Schrödinger equation that only differ by a complex phase $\exp(\mathrm{i}\varphi)$ with $\varphi \in [0, 2\pi)$, the phase of the AFQMC-sampled solution is not predefined. Therefore, the propagated determinants will all assume near-random phases and all walkers will be distributed almost uniformly in the complex plane with exponentially decaying signal-to-noise ratio in the large-τ limit.

As in other QMC methods, there are methods that mitigate the sign problem at the cost of introducing a systematic bias. To understand the mitigation procedures, one has to discriminate the different ways a sign problem can occur in AFQMC. This depends on the structure of the auxiliary field coefficients λ_i which are system-dependent:

- In Hubbard-like systems at half-filling, the calculations are entirely sign-problem-free and no mitigation procedure is required [54, 81, 82].

- In Hubbard-like systems at arbitrary fillings, $\lambda_i \leq 0$ for all i. Therefore, the auxiliary fields are purely real and there are no complex phases. However, there is still the coexistence of the solutions $|\Psi\rangle$ and $-|\Psi\rangle$. The general phase problem then reverts back to the usual sign problem where there is only a superposition of two instead of infinitely many solutions. In this case, the *constrained-path approximation* (CP approximation) can be applied [83–85]. In this approximation, the sign constraint

$$\left\langle \Psi_\mathrm{t} \middle| D_i^n \right\rangle > 0 \tag{2.52}$$

  is imposed for all walkers i, i.e. the overlap with the trial wavefunction cannot change sign throughout all iterations n. In the limit of small $\Delta\tau$, this prevents $|\Psi^n\rangle$ from crossing the nodal surface, i.e. the path of the random walk is constrained to either $|\Psi\rangle$ or $-|\Psi\rangle$. This comes at the cost of introducing a systematic bias and increasing the energy estimate above the exact energy. The approximation is improved if the trial wavefunction resembles the exact ground state more closely. It is exact for $|\Psi_\mathrm{t}\rangle = |\Psi_0\rangle$.

- For a general Hamiltonian, the λ_i can be positive and negative. Thus, the auxiliary fields can be complex which leads to the general phase problem. In this case, the *phaseless approximation* can be used to reduce the exponentially increasing signal-to-noise ratio. Roughly

speaking, the approximation consists of three steps: First, a generalised importance sampling method with complex overlaps between trial wavefunction and sampled Slater determinants is developed. This leads to walker weights that are rescaled by the local energy every iteration where the local energy has a definition like in DMC according to equation (2.24). Second, the complex local energy is approximated by its real part. Third, walkers that undergo phase rotations of more than $\pm\frac{\pi}{2}$ are discarded, analogous to the sign constraint from equation (2.52). The phaseless approximation is a true generalisation of the CP method as it reverts back to it for real auxiliary fields. Unlike importance sampling and CP-like constraints, the phaseless approximation will have no equivalent method developed for FCIQMC in this thesis. Therefore, the reader is referred to the corresponding literature for a more detailed presentation [53, 86–88].

2.3 *Density-Matrix Renormalisation Group*

In contrast to QMC, *density-matrix renormalisation group* (DMRG) is a deterministic method to perform quantum-chemical calculations. DMRG will be used as a benchmark method at multiple points in this thesis and both results and necessary computational resources will be compared. Therefore, I will give a brief overview of the method.

DMRG has its roots in the 1990s in pioneering work of S. R. White in condensed matter physics [89, 90]. White himself based the work on the renormalisation group approach by K. G. Wilson on the Kondo problem [91, 92] to which the modern formulation of DMRG is only loosely connected. A more wavefunction-based view has allowed for a wider range of application in quantum chemistry and electronic structure theory [93]. This brief overview will go along the lines of this wavefunction-based view of DMRG.

As in the QMC methods, the goal of DMRG is to find a good approximate solution to the electronic Hamiltonian from equation (1.3). Like FCIQMC (see chapter 3), DMRG is based on the FCI expansion of a many-particle quantum state that has been introduced in equation (2.7). To better suit the derivation of DMRG, the FCI wavefunction can be written as

$$|\Psi\rangle = \sum_{n_1 n_2 \dots n_S} C^{n_1 n_2 \dots n_S} |n_1 n_2 \dots n_S\rangle \ . \tag{2.53}$$

In contrast to equation (2.7), the coefficients will be indexed by the occupation numbers but have the same meaning. The n_i indicate how each orbital i

is occupied: empty ($|0\rangle$), with an ↑ electron ($|\uparrow\rangle$), with a ↓ electron ($|\downarrow\rangle$), or doubly occupied ($|\uparrow\downarrow\rangle$). S is the number of orbitals.

2.3.1 *Matrix Product States*

According to equation (2.9), the Hilbert space size grows combinatorially and as such, not all coefficients can be kept in memory for sufficiently large systems. While in FCIQMC only a stochastic representation of walker populations is stored at every instant, DMRG uses so-called *matrix product states* (MPS) to compactify and approximate the information contained in the wavefunction [94, 95]. The simplest MPS would be the full factorisation of the coefficients according to

$$C^{n_1 n_2 \dots n_S} \approx C^{n_1} C^{n_2} \dots C^{n_S} \,. \tag{2.54}$$

This is memory-efficient as only $4S$ coefficients need to be stored but it is also a crude approximation. A gradual improvement of the approximation with a controlled increase in memory can be achieved by moving from scalars C^{n_i} to matrices $C^{n_i}_{kk'}$. k and k' are then contracted over like

$$C^{n_1 n_2 \dots n_S} \approx \sum_{k_1 k_2 \dots k_{S-1}} C^{n_1}_{k_1} C^{n_2}_{k_1 k_2} C^{n_3}_{k_2 k_3} \dots C^{n_S}_{k_{S-1}} \quad \text{with } k_i = 1, \dots, M \text{ for all } i. \tag{2.55}$$

Inserting these coefficients into equation (2.53) then leads to the DMRG wavefunction ansatz

$$|\Psi_{\text{DMRG}}\rangle = \sum_{n_1 n_2 \dots n_S} \sum_{k_1 k_2 \dots k_{S-1}} C^{n_1}_{k_1} C^{n_2}_{k_1 k_2} \dots C^{n_S}_{k_{S-1}} \, |n_1 n_2 \dots n_S\rangle \,. \tag{2.56}$$

This definition explains the name MPS as a quantum state is approximated by a product of matrices. M is called the *bond dimension* of the MPS, a parameter which determines the accuracy and storage requirement. The dimension of the tensors $C^{n_i}_{kk'}$ is $4M^2$ so for the total wavefunction $4M^2S$ numbers have to be stored.

In the same way as MPS, operators can be represented as *matrix product operators* (MPO) [96]. A generic operator $\hat{O}$ in the occupation number basis can be written as

$$\hat{O} = \sum_{n_1 n_2 \dots n_S} \sum_{n'_1 n'_2 \dots n'_S} C^{n_1 n_2 \dots n_S, n'_1 n'_2 \dots n'_S} \, |n_1 n_2 \dots n_S\rangle \langle n'_1 n'_2 \dots n'_S| \,. \tag{2.57}$$

Again, the factor $C^{n_1 n_2 \dots n_S, n_1' n_2' \dots n_S'}$ can be approximated as a matrix product

$$C^{n_1 n_2 \dots n_S, n_1' n_2' \dots n_S'} \approx \sum_{k_1 k_2 \dots k_{S-1}} C_{k_1}^{n_1 n_1'} C_{k_1 k_2}^{n_2 n_2'} C_{k_2 k_3}^{n_3 n_3'} \dots C_{k_{S-1}}^{n_S n_S'} \tag{2.58}$$

of bond dimension M. In contrast to an MPS, the MPO matrices have two instead of one uncontracted indices.

With this, DMRG, like FCIQMC, keeps all the advantages of FCI-based methods. It allows for the variational evaluation of the energy and can express wavefunctions with arbitrarily strong multireference character as the reference Slater determinant does not have special significance. Additionally, DMRG wavefunctions are size-consistent in localised bases [93]. All important properties, like the energy, an be calculated without ever evaluating full $C^{n_1 n_2 \dots n_S}$ coefficients but rather from the $C_{kk'}^{n_i}$ tensors directly. This allows for a memory-efficient calculation of properties in cases where the true coefficients are approximated by matrix products with small M well. This is true in cases where the correlations between different orbitals are small and therefore the *entanglement entropy* is small [97, 98]. This is typically the case in one-dimensional lattice systems without periodic boundary conditions that will be studied in part II.

2.3.2 *Sweeping Algorithm*

MPS can be optimised in a stepwise fashion in so-called sweeps [99, 100]. Mostly, an MPS is optimsed such that the variational energy

$$E_{\mathrm{DMRG}} = \left\langle \Psi_{\mathrm{DMRG}} \middle| \hat{H} \middle| \Psi_{\mathrm{DMRG}} \right\rangle \tag{2.59}$$

is optimised. The representation of an MPS is not unique. The individual matrices C_{n_i} that make up the matrix product can be transformed with specific transformations and $|\Psi_{\mathrm{DMRG}}\rangle$ is invariant under those. To resolve this ambiguity, the *canonical representation* for a given orbital a is defined as

$$|\Psi_{\mathrm{DMRG}}^{a}\rangle = \sum_{n_1 \dots n_a \dots n_S} \sum_{\ell_1 \dots \ell_{a-1}} \sum_{r_a \dots r_{S-1}} L_{\ell_1}^{n_1} \dots L_{\ell_{a-2} \ell_{a-1}}^{n_{a-1}} A_{\ell_{a-1} r_a}^{n_a} R_{r_a r_{a+1}}^{n_{a+1}} \dots R_{r_{S-1}}^{n_S} |n_1 \dots n_a \dots n_S\rangle \ . \tag{2.60}$$

The matrices L are the site matrices to the left of site i, the matrices R are the ones to the right thereof. The L and R matrices are orthogonal, i.e.

$$\sum_{\ell n} L_{\ell n, \ell'}^{i} L_{\ell n, \ell''}^{i} = \delta_{\ell' \ell''} \tag{2.61a}$$

and

$$\sum_{rn} R^i_{r',rn} R^i_{r'',rn} = \delta_{r'r''} \,. \tag{2.61b}$$

Here, the occupation indices n_i are grouped with the lower indices to obtain a true two-dimensional matrix: $\mathbf{L}^i = L^i_{\ell n,\ell'} = L^{n_i}_{\ell\ell'}$ and $\mathbf{R}^i = R^i_{r',rn} = R^{n_i}_{r'r}$, respectively.

Since $|\Psi_{\mathrm{DMRG}}\rangle$ is defined to be invariant with respect to the site for which the canonical representation is defined, the following wavefunctions are equivalent:

$$\sum \mathbf{L}^1 \dots \mathbf{L}^{a-1} \mathbf{A}^a \mathbf{R}^{a+1} \mathbf{R}^{a+2} \dots \mathbf{R}^S |n_1 \dots n_a \dots n_S\rangle \tag{2.62a}$$

$$= \sum \mathbf{L}^1 \dots \mathbf{L}^{a-1} \mathbf{L}^a \mathbf{A}^{a+1} \mathbf{R}^{a+2} \dots \mathbf{R}^S |n_1 \dots n_a \dots n_S\rangle \,. \tag{2.62b}$$

The lower indices are omitted for the sake of notational simplicity. This means that

$$\mathbf{A}^a \mathbf{R}^{a+1} = \mathbf{L}^a \mathbf{A}^{a+1} \tag{2.63}$$

That leads to the fact that the canonical form of $|\Psi_{\mathrm{DMRG}}\rangle$ at site a can be transformed to the canonical representation at site $a+1$ via a singular value decomposition (SVD) according to

$$A^a_{\ell n,r} = \sum_{\ell'} L^a_{\ell n,\ell'} \sigma_{\ell'} V_{\ell' r} \tag{2.64a}$$

and

$$A^{a+1}_{\ell,rn} = \sum_{r'} \sigma_\ell V_{\ell r'} R^{a+1}_{r',rn} \,. \tag{2.64b}$$

With this knowledge about the transformations between canonical representations of adjacent sites, an MPS can be optimised in sweeps. This means that $|\Psi_{\mathrm{DMRG}}\rangle$ is first represented in the canconical representation for orbital $a = a_0$. In this representation the respective site matrix $\mathbf{A}^{a_0}$ is optimised for some criterion. Then the representation is changed to $a = a_0 + 1$ and so on until the last orbital is reached. Then, the MPS is optimised backwards until convergence is achieved. In most cases, the variational energy is minimised so that

$$\left\langle \Psi^a_{\mathrm{DMRG}} \middle| \hat{H} - E_0 \middle| \Psi^a_{\mathrm{DMRG}} \right\rangle = 0 \tag{2.65}$$

is the criterion for which the $\mathbf{A}^a$ is optimised.

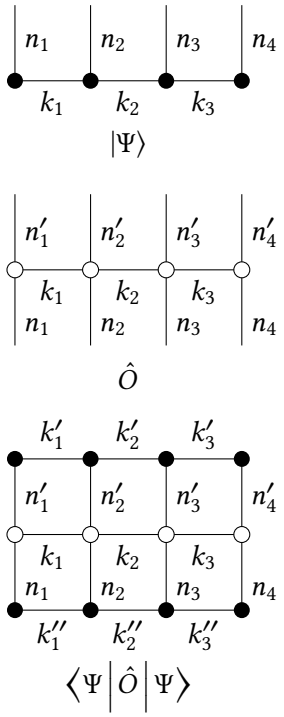


Figure 2.1. Graphical representation of a matrix product state $|\Psi\rangle$, a matrix product operator $\hat{O}$, and an expectation value $\langle\Psi|\hat{O}|\Psi\rangle$. Dots indicate orbital matrices, connected lines indicate contracted indices, and open-ended lines indicate non-contracted indices.

2.3.3 *Graphical Representations of MPS and MPO*

MPS and MPO can be represented in a visual way [101]. This makes it easier to understand contractions of different MPS and MPO which may get convoluted due to the many contracted and uncontracted indices involved. This representation will make is easy to calculate the memory demand of DMRG calculations which will be compared with FCIQMC in section 6.3.

The basic idea behind the graphical representation is that dots (●) represent general $C^{nn'}_{kk'}$ tensors. Lines that originate at a dot indicate indices. Lines that connect to another dot (●—●) are contracted over. Open-ended lines (●—) are uncontracted indices. Examples are shown in figure 2.1.

3 *Full Configuration Interaction Quantum Monte Carlo*

Full configuration interaction quantum Monte Carlo (FCIQMC) is a QMC method that was first published in 2009 [16] and has been applied and developed since then. In contrast to other QMC methods, FCIQMC is formulated in a finite basis set – originally in Slater determinants but recently also in configuration state functions (CSFs). FCIQMC has its roots in *Green's function Monte Carlo* (GFMC) [102–105]. There are two major implementations of FCIQMC available: `NECI` [18] and `HANDE` [106]. Also available is the `Dice` code, containing FCIQMC as part of a broader *ab initio* QMC framework [107], and the `Rimu` code, allowing to use FCIQMC in systems with bosons and fermions in model systems [108]. All numerical FCIQMC results in this thesis were obtained using `NECI`. After the general introduction into QMC methods in section 2.2, I will give a detailed overview of the FCIQMC algorithms with special emphasis on the manifestation of the fermion sign problem.

3.1 *The FCIQMC Algorithm*

FCIQMC is a stochastic method to solve the stationary Schrödinger equation (1.2) for the lowest eigenvalue E_0 with eigenfunction $|\Psi_0\rangle$. To do this, FCIQMC uses imaginary-time projection. The fundamental working equation of imaginary-time projector techniques can be derived from equation (1.1) by defining the imaginary time $\tau = \mathrm{i}t$. It is given by

$$\frac{\partial}{\partial\tau}|\Psi\rangle = -\left(\hat{H} - S\hat{\mathbb{1}}\right)|\Psi\rangle \,. \tag{3.1}$$

S is called the *shift* and will be used for walker population control later on. At this stage, it is a simple scalar diagonal shift of $\hat{H}$ as $\hat{\mathbb{1}}$ denotes an identity matrix of the same size as $\hat{H}$. Solving this differential equation leads to

$$\begin{aligned} |\Psi(\tau)\rangle &= \exp\left[-\tau(\hat{H} - S\hat{\mathbb{1}})\right]|\Psi(\tau=0)\rangle \\ &= \sum_n \exp\left[-\tau(E_n - S)\right]|\Psi_n\rangle\langle\Psi_n|\,\Psi(\tau=0)\rangle \end{aligned} \tag{3.2}$$

where in the second line the equation was expanded in the eigenbasis $\{|\Psi_n\rangle\}$ of $\hat{H}$ with respective eigenvalues E_n in ascending order. E_0 denotes the

ground-state energy. For $\tau \to \infty$, all basis states with $n \geq 1$ are projected out and $|\Psi(\tau)\rangle$ converges to the ground state $|\Psi_0\rangle$ if the starting guess $|\Psi(\tau = 0)\rangle$ has non-zero overlap with $|\Psi_0\rangle$.

Originally, FCIQMC was formulated with discrete integer walkers [16, 17]. Later, it has been generalised with walker populations that can assume any real number [109] which removes the stochasticity from some algorithmic steps but keeps it in others. In the following, I will only describe the latter, continuous-walker formulation of FCIQMC which is also used in the practical implementation in `NECI`.

3.1.1 *Stochastic Evaluation of the Imaginary-Time Propagation*

While the imaginary-time projection approach is common to all projector Monte Carlo methods, one of the unique features of FCIQMC is the representation of the wavefunction. In FCIQMC, the instantaneous wavefunction is represented using walkers that reside on Slater determinants $|D_i\rangle$. The number of walkers N_i on $|D_i\rangle$ is an estimate of the C_i coefficient in the FCI expansion from equation (2.7) such that the long-time average $\overline{N_i} = C_i$.[3]

[3] FCIQMC can also be used not in a basis of Slater determinants but in *configuration state functions* (CSF) that are eigenfunctions of $\hat{S}^2$ using the *graphical unitary group approach* (GUGA) [110–114]. This adds some complications to the algorithm and will not be used in this thesis.

The dynamics of these walkers in FCIQMC is given by a linearised version of equation (3.2). The continuous imaginary time is discretised into small time steps $\Delta\tau$. This leads to

$$\Delta N_i(\tau) = -\Delta\tau \left[\underbrace{\sum_{j \neq i} H_{ij} N_j(\tau)}_{\text{spawning}} + \underbrace{\big(H_{ii} - S\big) N_i(\tau)}_{\text{death/cloning}} \right]. \tag{3.3}$$

$\Delta N_i(\tau)$ is the number of walkers that is added to the population N_i in an iteration. The two terms are applied stochastically in the FCIQMC algorithm in two distinct steps called *spawning* and *death/cloning*. They govern the dynamics of the walkers and will be discussed in the following.

Spawning Step

The spawning step stochastically applies the off-diagonal term of the discrete master equation (3.3), $\Delta\tau \sum_{i \neq j} H_{ij}$, to the walker occupation vector $N_j(\tau)$ in a given iteration $p = \frac{\tau}{\Delta\tau}$. The stochastic application works as follows: The algorithm loops through all determinants in memory which have non-zero walker population. At a given determinant $|D_j\rangle$ in the loop, another determinant $|D_i\rangle$ that is connected to $|D_j\rangle$ via a non-zero Hamiltonian matrix element H_{ij} is selected at random for each integer walker on $|D_j\rangle$. For fractional walkers sitting on $|D_j\rangle$, i.e. the remaining part $N_i - \lfloor N_i \rfloor$, a random

number $R \in [0, 1]$ is drawn. A spawning attempt for the fractional walker is conducted only if $R \geq N_i - \lfloor N_i \rfloor$ such that the spawning procedure is unbiased on average.[4]

$$\Delta N_i = -\frac{\Delta\tau H_{ij} \operatorname{sgn} N_j}{p_{\text{gen}(i|j)}} \tag{3.4}$$

walkers are then added to a separate spawn vector which will be added to the main occupation vector at the end of the iteration. $\operatorname{sgn}(N_j)$ is the sign of the walkers on $|D_j\rangle$. Therefore, the sign of the spawned walkers created from $|D_j\rangle$ onto $|D_i\rangle$ is given by the product

$$\operatorname{sgn} \Delta N_i = -\operatorname{sgn} H_{ij} \operatorname{sgn} N_j \,. \tag{3.5}$$

$p_{\text{gen}}(i \mid j)$ is the probability with which $|D_i\rangle$ has been selected from $|D_j\rangle$. Any non-zero choice of generation probabilities leaves the algorithm unbiased. However, it is advantageous if $p_{\text{gen}}(i \mid j)$ is chosen approximately proportional to H_{ij} such that the ratio ΔN_i is constant. This reduces statistical fluctations in the walker occupations which is especially important when applying the initiator approximation (see section 3.2.3). It is however computationally prohibitive to choose $p_{\text{gen}}(i \mid j)$ exactly proportional to H_{ij} in general. This requires the evaluation of the column sum $\sum_k H_{ik}$ which is needed as a normalisation factor and scales as $\mathcal{O}(N^2M^2)$. Therefore, methods that approximate the optimal p_{gen} are available. These methods are called *excitation generators*. Examples are the heat-bath, Cauchy–Schwarz, and Power–Pitzer excitation generators which can also be combined with each other and can be used in either an on-the-fly and or precomputed fashion [115, 116]. For the model systems that will play the main role in this thesis however, a uniform excitation generator is sufficient. In the real-space Hubbard model all the off-diagonal elements have the same magnitude so the uniform excitation generator is already optimal.

Death/Cloning Step

The death/cloning step is the application of the diagonal part of equation (3.3). The death/cloning step happens before the contributions from the spawning step – saved in the spawn vector – are added into the main wavefunction vector. In this step, the algorithm again loops through all determinants with non-zero walker occupations. The walkers on every determinant are multiplied by $1 - \Delta\tau\left[H_{ii} - S(\tau)\right]$. S is a simulation parameter that is dynamically adjusted during the simulation so that the total walker population is held

4 This convention is a compromise between reducing stochastic noise and reducing computational effort. In principle, the algorithm would also be unbiased if only one connected determinant would be chosen for each occupied determinant or if more connected determinants would be drawn. All that is required is that the spawned walkers are scaled accordingly.

approximately constant. It will be discussed in the respective section. Generally speaking, this way energetically unfavourable determinants with large diagonal elements die off quicker.

Cloning of walkers occurs when $H_{ii} - S(\tau)$ is negative. Cloning events are rare during the simulation of realistic systems. They typically occur during the equilibration phase when the instantaneous wavefunction is different from the sought-for solution and S therefore can become large.

Annihilation Step

The annihilation step is an important part of the algorithm that controls the fermion sign problem in FCIQMC which will be discussed in section 3.2. Annihilation means that walker contributions from existing walker populations on determinants $N_i(\tau)$ and the newly spawned contributions $\Delta N_i(\tau)$ will be added with their respective sign. This means that walkers of opposite signs cancel out. In contrast to other QMC methods discussed in section 2.2, annihilations are straightforwardly possible in FCIQMC since it works in a finite basis of Slater determinants. For example, due to the continuous representation of the wavefunction in DMC, which is also a projector-QMC method, an annihilation step is not as easy to realise [117].

In a practical implementation, annihilations take part in two steps:

1. Newly spawned walkers of opposite signs in the spawn vector directly annihilate immediately there.

2. When the spawn vector is added to the main occupation vector, two contributions of opposite signs also annihilate.

Details on when and how these annihilations take place in a parallel computation on multiple CPU cores are given in section 3.1.3.

Stochastic Rounding

An additional stochastic step taking place after the conclusion of the annihilation step is the stochastic rounding of walker populations. Again, the algorithm loops through all occupied determinants which now also includes determinants that were previously unoccupied and have now been spawned upon. To this end, the occupation threshold t_{occ} which is chosen to be 1 in

most cases is defined. All determinants that have an occupation of $|N_i| < t_{\text{occ}}$ after annihilation are now rounded according to the following provision:

$$N_i' = \begin{cases} t_{\text{occ}} & \text{with probability } \frac{N_i}{t_{\text{occ}}}, \\ 0 & \text{with probability } 1 - \frac{N_i}{t_{\text{occ}}}. \end{cases} \tag{3.6}$$

In practice, this is achieved by drawing a random number r_i between 0 and 1 from a uniform distribution for each occupied determinant $|D_i\rangle$. If $r_i > \frac{N_i}{t_{\text{occ}}}$, the occupation is rounded up. If $r_i < \frac{N_i}{t_{\text{occ}}}$, the occupation is rounded to zero. Whenever a determinant's occupation is rounded to zero, it is removed from memory. It is only added again when it gets spawned upon in a later iteration once more. The stochastic rounding step is crucial as it ensures that in every iteration only a very small fraction of all determinants is stored in memory. This makes the FCIQMC algorithm highly memory-efficient at the expense that the instantaneous wavefunction only contains limited information about the wavefunction and its properties. Still, averaged quantities show high precision.

The Shift

As introduced before, the *shift* S is a global scalar parameter in the algorithm. S shifts the diagonal elements and enters the calculation in the death/cloning step according to equation (3.3). An FCIQMC run is usually started with a constant shift $S = S_0$. If S_0 is larger than the true ground-state energy, the total walker population

$$N_{\text{tot}}(\tau) = \sum_i |N_i(\tau)| \tag{3.7}$$

will grow on average. The larger S_0 is chosen, the faster N_{tot} will grow. This phase of the simulation is called the walker-growth phase.

As soon as a certain fraction of the desired total population N_{target} is reached, the algorithm will subsequently enter variable-shift mode. In this mode, the shift is dynamically updated every A iterations according to

$$S(\tau + A\Delta\tau) = S(\tau) - \frac{\gamma}{A\Delta\tau} \ln\left(\frac{N_{\text{tot}}(\tau + A\Delta\tau)}{N_{\text{tot}}(\tau)} \right). \tag{3.8}$$

The second term is responsible for keeping the shift at a constant value and counters the exponential growth of walkers. γ is a damping parameter.

With this way of updating the shift, one can only approximately converge to the desired population. If initial walker growth is rapid, then the walker number may severely overshoot the target number. As the computational effort roughly scales linearly with N_{tot}, this can lead to the problem that the assigned hardware capacity is exceeded and the simulation runs slowly or crashes when the available memory is exceeded. Especially in the model systems I wish to study, fast walker growth in the constant-shift phase and overshooting the total population are common. To remedy this, the shift-update formula (3.8) was extended by an additional term [118]. The shift is then updated using

$$S(\tau + A\Delta\tau) = S(\tau) - \frac{\gamma}{A\Delta\tau} \ln\left(\frac{N_{\text{tot}}(\tau + A\Delta\tau)}{N_{\text{tot}}(\tau)}\right) - \frac{\chi}{A\Delta\tau} \ln\left(\frac{N_{\text{tot}}(\tau + A\Delta\tau)}{N_{\text{target}}}\right). \quad (3.9)$$

The additional term in the equation takes care that the shift converges to N_{target} in the large-τ limit. In this scheme, no initial walker-growth phase is necessary as the additional term is large for low walker numbers that differ much from N_{target}. This way, the initial walker growth is still fast but slows down as soon as N_{tot} approaches N_{target}. χ is an additional free damping parameter but it can be determined as a function of γ when assuming a scalar model of walker population dynamics. In this model, it is assumed that equation (3.3) is applied deterministically instead of stochastically and that the initial walker distribition is already proportional to the true solution – which is approximately true when the simulation is already equilibrated. The master equation then reduces to the differential equation of the damped harmonic oscillator in N_{tot}. The solution of this differential equation has three fundamental regimes: overdamped, underdamped, and critical. The convergence of N_{tot} to N_{target} is fastest in the critical-damping case. This is the case when

$$\chi = \frac{\gamma^2}{4}. \quad (3.10)$$

Therefore, this choice of χ will be used whenever the improved shift-update formula (3.9) will be used throughout this thesis.

A third way of controlling the population and adapting the shift is the *fixed-N_0 method.* In this scheme, not the overall population N_{tot} is fixed but rather the population N_0 on a prespecified *reference determinant* $|D_0\rangle$. The reference determinant is usually chosen to be the determinant that is

expected to have the largest weight in the FCI expansion. A constant N_0 can be achieved by updating the shift according to

$$S(\tau) = \frac{\sum_j H_{0j} N_j(\tau)}{N_0(\tau)} \,. \tag{3.11}$$

Here, again an initial growth phase with $S = S_0$ like in the original shift-update scheme is required until N_0 reaches its target value. H_{0j} are all the matrix elements connecting the reference determinant. When inserting equation (3.11) into equation (3.3), it is easy to see that this way of updating S keeps N_0 constant:

$$\Delta N_0(\tau) = -\Delta\tau \left[\left(H_{00} - \frac{\sum_j H_{0j} N_j(\tau)}{N_0(\tau)} \right) N_0(\tau) + \sum_{j \neq 0} H_{0j} N_j(\tau) \right] = 0 \,. \tag{3.12}$$

The shift update schemes according to equations (3.8) and (3.9) only lead to a constant N_{tot}. In contrast, equation (3.12) indicates that N_0 is held exactly constant in the fixed-N_0 scheme. N_{tot} however still fluctuates after equilibration. The value to which N_{tot} will eventually converge cannot be predicted a priori. This constitutes a disadvantage of the fixed-N_0 method as the total memory requirement cannot be determined beforehand. The fixed-N_0 method is one possibility to determine the minimum walker number that still ensures a sign-coherent wavefunction. Sign coherence means that only one of the possible solutions $|\Psi\rangle$ and $-|\Psi\rangle$ is sampled and the sign problem is overcome. This will be discussed in section 3.2 in more detail.

3.1.2 *Energy Estimators and Properties*

As mentioned before, the shift can be used as an estimator for the ground-state energy of a Hamiltonian in FCIQMC. Regardless on how the shift is calculated, in equilibrium it will always converge to and then fluctuate around the sought-for ground-state energy. Still, there are other independent ways of estimating the ground-state energy. In the context of this thesis, different energy estimators may be required to judge whether an FCIQMC run has converged properly in an unbiased manner, especially with respect to the sign problem (see section 3.2). The different energy estimators are affected by different systematic biases in different ways, as we will see in part II. Statistical uncertainties are estimated using the *blocking analysis* which is described in appendix A.2.

The Projected Energy

The most commonly used energy estimator in FCIQMC is the *projected energy*. The projected energy in iteration p is given by

$$E_{\text{proj}} = \frac{\left\langle D_0 \middle| \hat{H} \middle| \Psi(p\Delta\tau) \right\rangle}{\langle D_0 \,|\, \Psi(p\Delta\tau)\rangle} \,. \tag{3.13}$$

If $|\Psi(\tau)\rangle$ equals the ground state $|\Psi_0\rangle$ exactly, i.e. is an eigenstate of $\hat{H}$, E_{proj} equals the exact ground-state energy E_0. If $|\Psi(\tau)\rangle$ only approximately equals $|\Psi_0\rangle$, as it is the case in every FCIQMC simulation due to the stochastic nature of the algorithm, E_{proj} is a non-variational estimator of E_0. The advantage of the projected energy is that it can be obtained with minimal overhead. When writing the projected energy in terms of instantaneous walker populations,

$$E_{\text{proj}} = \frac{\sum_j H_{0j} N_j(p\Delta\tau)}{N_0(p\Delta\tau)} \,, \tag{3.14}$$

and comparing it with the FCIQMC master equation (3.3), we can see that the numerator is simply given by the spawns onto the reference determinant $|D_0\rangle$. This makes the calculation of the projected energy possible with almost no overhead.

The calculation of E_{proj} is only possible when there is at least one permanently occupied determinant. When a system has a sign problem, which is usually the case, this is a necessity for an unbiased calculation. For sign-problem-free system however, it is possible that the simulation is unbiased without a permanently occupied determinant. In these cases, the only usable energy estimator is the shift (also see section 3.2).

When comparing equations (3.11) and (3.14), one can see that they are equivalent. So when using the fixed-N_0 way of updating the shift, there is also only a single energy estimator.

The Trial Energy

In systems where the population N_0 on the reference determinant is low, it can be useful to project the vector $\hat{H}\,|\Psi(\tau)\rangle$ not only onto a single determinant but onto a linear combination of many. In this case, one is dealing with the trial energy

$$E_{\text{t}} = \frac{\left\langle \Psi^{\mathcal{T}} \middle| \hat{H} \middle| \Psi(p\Delta\tau) \right\rangle}{\left\langle \Psi^{\mathcal{T}} \middle| \Psi(p\Delta\tau) \right\rangle} \,. \tag{3.15}$$

$\mathcal{T}$ denotes a subspace of the full Hilbert space $\mathcal{H}$ called the trial space [119]. It is advantageous be $\mathcal{T}$ be chosen such that it contains the $N^{\mathcal{T}}$ determinants with the largest coefficients in the FCI expansion. $\mathcal{T}$ can be determined approximately by simply taking the $N^{\mathcal{T}}$ determinants with the largest walker occupations N_i in an iteration after the initial FCIQMC equilibration phase.

After the trial space has been defined, the Hamiltonian constrained to $\mathcal{T}$ is constructed which will be called $\hat{H}^{\mathcal{T}}$. $\hat{H}^{\mathcal{T}}$ is then diagonalised exactly once after $\mathcal{T}$ has been determined to acquire

$$\left|\Psi^{\mathcal{T}}\right\rangle = \sum_{i\in\mathcal{T}} C_i^{\mathcal{T}} \left|D_i\right\rangle \tag{3.16}$$

which is the ground state in the subspace with a ground-state energy $E^{\mathcal{T}}$.

Like for the projected energy, we can write E_t in terms of instantaneous FCIQMC walker populations as

$$E_\mathrm{t} = E^{\mathcal{T}} + \frac{\sum_{j\in\mathcal{C}} C_j V_j}{\sum_{i\in\mathcal{T}} C_i C_i^{\mathcal{T}}} \tag{3.17a}$$

with

$$V_j = \sum_{i\in\mathcal{T}} \left\langle D_j \middle| \hat{H} \middle| D_i \right\rangle C_i^{\mathcal{T}}, \quad \left|D_i\right\rangle \in \mathcal{T} \text{ and } \left|D_j\right\rangle \in \mathcal{C}. \tag{3.17b}$$

$\mathcal{C}$ is the space of determinants connected to $\mathcal{T}$ via $\hat{H}$, not including $\mathcal{T}$ itself and containing $N^{\mathcal{C}}$ determinants. The $C_i^{\mathcal{T}}$ coefficient (an array of length $N^{\mathcal{T}}$) and the V_j (an array of length $N^{\mathcal{C}}$) are kept in memory during the entire simulation. Since typically $N^{\mathcal{C}} \gg N^{\mathcal{T}}$, the storage of the vector $\boldsymbol{V}$ is the bottleneck in the calculation of E_t.

3.1.3 *Parallel Implementation*

A huge advantage of the FCIQMC algorithm lies in the fact that it can be implemented in parallel on a large number of CPU cores in a distributed-memory architecture [120]. Near-linear scaling with the number of CPU cores has been shown in up to 24 000 cores using the `NECI` code [18].

However, FCIQMC is not what is typically called "embarrassingly parallel". "Embarrassingly parallel" is a term used when an algorithm can be easily divided into subtasks that can be executed largely independently without depending on results from other subtasks, i.e. no or little communication between processes is required and there is no computational overhead caused

by a parallel implementation [121]. The aforementioned estimation of π using MC in section 2.2 is an example of an embarrassingly parallel algorithm.

FCIQMC can be most efficiently parallelised using a paradigm that resembles *domain decomposition* [122, 123]. In domain decomposition, an entire boundary value problem is subdivided into local subdomains with each subdomain also being a boundary value problem. The algorithm is iterated once on the subdomains in parallel and then the results are communicated between adjacent subdomains. The result of the adjacent subdomains enter as boundary values for the next iteration. A typical example for this procedure is the solution of a (partial) differential equation on a grid [124, 125].

To implement parallelism, `NECI` uses the *message passing interface* (MPI) computational standard [126]. MPI implements distributed-memory parallelism, i.e. each process, each of which runs on exactly one physical CPU core, has its own share of memory. Information that is needed by another process needs to be sent there using MPI routines first.[5]

5 There exist limited read-only ways of sharing memory between different processes in the MPI standard. In `NECI`, this is used for the storage of the integrals for example.

In FCIQMC, the basis functions, which are Slater determinants in the cases considered in this thesis, can be regarded as the domains. FCIQMC differs from domain decomposition in the following ways:

- Only a small fraction of Slater determinants are kept in memory at any given iteration and that the occupied Slater determinants can change from iteration to iteration.
- The annihilation step requires a communication from all processes to all other processes which cannot be circumvented.

In this section, I will describe how these differences are handled in a memory- and CPU-efficient manner.

3.1.4 *Hashing, Local Spawning, and Death/Cloning Step*

For FCIQMC to work efficiently, the determinants need to be distributed as equally as possible amongst the MPI processes. In the computational implementation, the Slater determinants are represented as bit strings. Since the sampled wavefunction typically lives in only a small part of the Hilbert space, the instantaneously occupied determinants are all similar. To ensure this, the occupied determinants are assigned hash values using a hashing function. The crucial property of hash functions exploited here is the following: It assigns Slater determinants that have similar bit strings uniformly distributed hash values $h(i)$ between 0 and 1. The processor p on which the

i-th determinant is stored and where it performs its algorithmic steps is then determined by

$$p(i) = \left\lfloor h(i) n_{\text{proc}} \right\rfloor \tag{3.18}$$

with n_{proc} being the total number of processes. Together with the determinant's bit string, all other determinant-related information is stored in the memory assigned to $p(i)$. This includes the instantaneous walker population and flags which are boolean variables that contain information about the determinants status in the calculation. These flags will become crucial when adaptations to the algorithm will be made to control the sign problem.

The spawning step is then performed on every process fully independently in parallel by looping through all determinants on a respective process. The information about new spawns that have been generated according to the first term of equation (3.3) is stored locally first. Subsequently, the death/-cloning step according to the second term of equation (3.3) is also performed locally and in parallel.

3.1.5 *All-to-All Communication and Annihilation*

The annihilation step is the most computationally expensive step in parallel FCIQMC because it requires communication of the spawn arrays from each process to all others. In MPI language, this is called an all-to-all communication. The parallel annihilation step is performed as follows: The hash value and thus the respective process is looked up for each determinant in the spawn array according to equation (3.18). Then, the send spawn table is partitioned by destination MPI process index.[6] The newly spawned contribution is written into the partition given by its hash value. All new contributions and the already existing walker number are then added locally. This way, walkers of opposite signs annihilate. Global simulation variables like the shift are computed once on the head process and then communicated across after the conclusion of the iteration. One is again left with the situation that each determinant is located on a unique process which allows for parallel spawning in the next iteration.

6 Local annihilation already on the parent process is possible when using a hash table only for the spawn array. This is not done routinely, however.

3.2 *The Sign Problem in FCIQMC*

When trying to solve for the ground or any excited state of a general Hamiltonian, FCIQMC, like any other QMC method, suffers from the infamous sign problem. The underlying reason of the problem is always the same: when diagonalising a Hamiltonian with positive and negative off-diagonal

matrix elements, in most cases large positive and large negative contributions need to be summed up with their respective sign. The result of this summation, i.e. the true weight of the contribution, might be comparatively small, however. This is not a problem in deterministic methods as long as no numerical overflows occur in a practical implementation because then all occuring signed sums are computed numerically exactly. However, in stochastic methods it is a problem because sums are evaluated stochastically. If large positive and large negative cancelling weights are summed by using infrequent sampling, i.e. using a small number of stochastic walkers, the signal-to-noise ratio becomes very small. This then leads to long integration times which for sizeable systems can become computationally prohibitive.

3.2.1 *Stoquastic and Stoquastised Matrices*

A more quantitative approach to understanding the sign problem is based on the notion of *stoquastic matrices.* A *stoquastic matrix* $\mathbf{S}$ is defined as a matrix that only contains off-diagonal matrix elements of one sign. If $S_{ij} > 0$ for all $i \neq j$, then the largest eigenvalue can be found without a sign problem. Since in electronic-structure problems, one is mostly interested in low-lying eigenvalues, the case $S_{ij} < 0$ for all $i \neq j$ is the relevant one. In this case of a stoquastic matrix, the solution for the smallest eigenvalue is sign-problem-free.[7]

[7] These are only the trivial cases where a sign problem is absent. There are also matrices that do not have a sign problem that are not stoquastic. They will be discussed in part II.

Based on this definition, the concept of *stoquastised matrices* can be introduced. It applies to any real matrix $\mathbf{M}$. The matrix elements of the two stoquastised versions $\mathbf{M}^{\text{stoq},\pm}$ of $\mathbf{M}$ are defined as

$$M_{ij}^{\text{stoq},\pm} = \begin{cases} \pm\left|M_{ij}\right| & \text{for } i \neq j, \\ M_{ij} & \text{for } i = j. \end{cases} \tag{3.19}$$

Since we are only interested in cases where the solution is sign-problem-free for the lowest eigenvalue, for the remainder of the thesis we will only refer to the minus-sign version $\mathbf{M}^{\text{stoq},-} =: \mathbf{M}^{\text{stoq}}$ as the stoquastised version of $\mathbf{M}$. In words, building the stoquastised version of a matrix means flipping the sign of all positive off-diagonal elements such that the resulting matrix is stoquastic.

For the quantitative understanding of the sign problem, it is crucial to note that $\lambda_0^{\text{stoq}} \leq \lambda_0$, i.e. that the lowest eigenvalue of the stoquastised version $\mathbf{M}^{\text{stoq}}$ is always lower than or equal to the lowest eigenvalue of the original matrix $\mathbf{M}$. This is easy to see with the following argument [38]:

Suppose, $\boldsymbol{v}_0 = c_0^i \boldsymbol{b}^i$ is a normalised eigenvector with eigenvalue λ_0, the lowest eigenvalue of $\mathbf{M}$. c_0^i are the real coefficients and $\boldsymbol{b}^i$ is a set of basis vectors. When $\mathbf{M}$ is expressed in the same basis, we can then write the corresponding eigenvalue as

$$\lambda_0 = \sum_{ij} c_0^i M_{ij} c_0^j \,. \tag{3.20}$$

Let us now consider the vector $-|\boldsymbol{v}_0| = -|c_0^i| \boldsymbol{b}^i$ which can be considered as the stoquastised version of the ground state of the non-stoquastised matrix $\mathbf{M}$. First, we note that for the expectation value of $\mathbf{M}^{\text{stoq}}$ with respect to this vector we find

$$\sum_{ij} |c_0^i| M_{ij}^{\text{stoq}} |c_0^j| \leq \sum_{ij} c_0^i M_{ij} c_0^j \tag{3.21}$$

using the definition of a stoquastised matrix from equation (3.19). By applying the variational principle, we also find for the lowest eigenvalue of $\mathbf{M}^{\text{stoq}}$ that

$$\lambda_0^{\text{stoq}} \leq |c_0^i| M_{ij}^{\text{stoq}} |c_0^j| \,. \tag{3.22}$$

Combining equations (3.20) to (3.22), we conclude that

$$\lambda_0^{\text{stoq}} \leq \lambda_0 \tag{3.23}$$

which is what we wanted to prove.

3.2.2 *The Role of Annihilations*

To see what is the significance of the stoquastised Hamiltonian $\hat{H}^{\text{stoq}}$ and its relationship with the annihilation step in the FCIQMC algorithm, let us imagine a simulation with an extremely low walker population (close to the limit of a single walker). This is such that the Hilbert space is only populated by so few walkers such that annihilation events never occur for all τ. It is apparent that in this edge case of zero annihilations – even though the propagation of the imaginary-time projection is governed by $-\hat{H}$, the signed, true Hamiltonian – the sign information loses importance when calculating the shift estimator. The algorithm cannot distinguish between ground-state solutions of $\hat{H}$ and of the stoquastised version $\hat{H}^{\text{stoq}}$. The stoquastised solution is given by $|\Psi_0^{\text{stoq}}\rangle = \left(N_i^+ + N_i^-\right)|D_i\rangle$ instead of the true fermionic solution $|\Psi_0\rangle = \left(N_i^+ - N_i^-\right)|D_i\rangle$. N_i^+ are the positive walker contributions, N_i^- are the negative walker contributions. In the former case, it is obvious that the sign information is disregarded since the two types of walker contributions are simply summed up. Since the ground state of $\hat{H}^{\text{stoq}}$ is

always lower than the ground state of $\hat{H}$ and FCIQMC is a projector method that projects out all but the state with the lowest energy, the algorithm evolves according to the incorrect master equation

$$\Delta\left(N_i^+ + N_i^-\right)(\tau) = -\Delta\tau\left[\sum_{i\neq j} H_{ij}^{\text{stoq}}\left(N_j^+ + N_j^-\right) + \left(H_{ii}^{\text{stoq}} - S\right)\left(N_i^+ + N_i^-\right)\right]. \quad (3.24)$$

The shift energy estimator S converges to the ground-state energy of $\hat{H}^{\text{stoq}}$ and the signal of true fermionic ground state is lost in noise.

When adding more walkers that simultaneously occupy the Hilbert space, the likelihood of annihilation events increases. This means that the ground state of $\hat{H}^{\text{stoq}}$ is now penalised through annihilation events. This penalty procedure can be regarded as a damping term in the evolution of the undesired stoquastised solution according to

$$\Delta\left(N_i^+ + N_i^-\right)(\tau) = \Delta\tau\left[\sum_{i\neq j} H_{ij}^{\text{stoq}}\left(N_i^+ + N_i^-\right)N_j(\tau) + \left(H_{ii}^{\text{stoq}} - S\right)\left(N_i^+ + N_i^-\right)\right] + \kappa N_i^+ N_i^- \quad (3.25)$$

where κ is an average annihilation rate of positive and negative walkers [38]. The emergence of the true solution $N_i^+ - N_i^-$ according to the desired master equation driven by $\hat{H}$ is not affected by annihilations as walkers of opposite signs do not contribute here anyway. This way, annihilation events stabilise the true fermionic solution.

It is important to note that there is a system-dependent minimum number of walkers N_{min} above which the different energy estimators converge to the exact energy. It is a crucial observation that not the entire Hilbert space needs to be occupied with walkers such that the fermionic instead of the stoquastised solution emerges. In some cases, less than 1 % of the Hilbert space needs to be occupied such that enough annihilation events occur on average to ensure that the simulation converges to the fermionic solution. However, there are also other cases where almost the entire Hilbert space needs to be filled with walkers. In these cases, an additional way to control the sign problem is necessary which will be discussed in the next section. Empirically, in most *ab initio* systems and some model systems – especially when they are formulated in delocalised orbitals – one can find that during

the walker-growth phase, during which a constant shift $S(\tau) = S_0$ is kept, an *annihilation plateau* occurs. The walker-growth rate $N_{\text{tot}}(\tau + \Delta\tau)/N_{\text{tot}}(\tau)$ is slowed down during the plateau phase which can be observed in the simulation dynamics. During the plateau phase, more annihilations take place and thus the correct sign structure emerges [38]. For a walker number below the annihilation plateau, the sign problem is still unresolved and only biased results can be obtained. For any walker number above the plateau, any sign-problem-related bias is removed. Often in systems that are formulated in terms of localised orbitals, the annihilation plateau cannot be observed and therefore cannot be used to determine N_{min}. When using the shift update equation equation (3.9) without a walker-growth phase with constant shift, an alternative method to determine the height of the annihilation plateau has been presented [118].

3.2.3 *The Initiator Approximation*

The initiator method is the standard way of controlling the sign problem in FCIQMC [17, 36, 37]. It is used when reaching N_{min} is computationally impossible which is the case for most systems of interest. Unlike the discrete annihilation mechanism, the initiator approximation – as the name suggests – introduces a systematic bias into the sampled wavefunction and thus the extracted properties. The bias is called the *initiator bias.*

When using the standard initiator method, the notion of *initiators* is introduced. A Slater determinant $|D_i\rangle$ is called an initiator in iteration p if its absolute population $\left|N_i(p\Delta\tau)\right| > t_{\text{init}}$. t_{init} is called the *initiator threshold.* It is a real number that is specified at the beginning of the simulation. Additionally, a determinant is defined as an initiator if it has been spawned upon by two or more different determinants, the so-called multiple-spawns rule. The space of initiators can change in every iteration depending on the instantaneous occupations. The initiator approximation modifies the spawning step of the FCIQMC algorithm depending on whether a determinant is an intiator or not:

- *Initiators* can spawn freely to any other determinant.
- *Non-initiators* can only spawn to already occupied determinants.

The initiator method typically allows for convergence of the simulation at walker numbers well below N_{min}. This is because determinants that have a well-established population of one sign are sign-coherent with respect to the average signs of all other determinant populations. Therefore, a correct

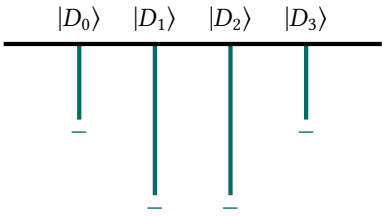
global sign structure has been formed inside the intiator subspace. The global t_{init}, for which the number of sign-coherent determinants is large enough such that the sign problem is resolved, in principle depends on the system and the total population used. Also, the initiator bias increases for larger t_{init}. In practice however, the value for which the sign problem is overcome and the magnitude of the initiator bias are rather insensitive with respect to t_{init}. Typically, t_{init} is chosen between 1 and 10.

Effectively, the initiator method is similar to a truncation of the Hilbert space. There are, however, two key differences to a strict truncation:

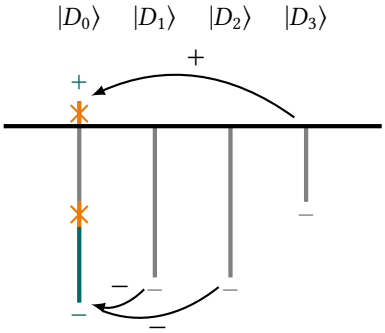
- Non-initiators can passively acquire walker population. This walker population contributes to all estimators of wavefunction properties and spawns onto occupied determinants are possible.
- The initiator space is dynamically changing and depends on the total population N_{tot}.

In the limit $N_{\text{tot}} \to \infty$, the initiator bias approaches zero as the population on all determinants exceeds the initiator threshold. Due to the similarity to a strict truncation, the initiator bias in most cases introduces a positive shift to the energy estimators shift, projected energy, and trial energy. The energy estimates then converge monotonically to the exact ground state energy as a function of N_{tot}. In some rare cases however, the initiator bias is negative and the energy estimates converge from below. Typically, an extrapolation to the exact ground-state energy is not possible due to an irregular convergence of the initiator bias with respect to the total walker number.

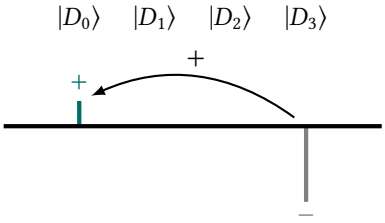
An extension of the initiator method is available with the *adaptive-shift algorithm*. It includes spawning events that are rejected due to the initiator criterion into the diagonal elements of non-initiators [57, 58, 127, 128]. Furthermore, a second-order Epstein–Nesbet correction has been used to correct for the initiator bias [59]. Also, selected configuration interaction (SCI) calculations have previously been used to preselect initiator spaces [56] where the SCI space has been constructed using the heat-bath technique [129–131]. A related approach will be used in chapter 8 where fixed initiator spaces will be built based on analytical fast-to-evaluate wavefunction ansatzes instead.



(a) Eigenvector corresponding to the lowest eigenvalue. All coefficients are negative which, however, does not indicate that there is no sign problem.



(b) Deterministic imaginary-time propagation with $\Delta\tau = 1$. Only spawns to $|D_0\rangle$ from all other basis functions are shown. Due to opposite-sign contributions, an annihilation takes place (indicated by the orange crosses).



(c) Stochastic imaginary-time propagation with $\Delta\tau = 1$. Here, for low N_{tot} it can happen that $|D_0\rangle$, $|D_1\rangle$, and $|D_2\rangle$ are unoccupied in a given stochastic snapshot. Still, $|D_3\rangle$ can spawn to $|D_0\rangle$, leading to a wrong sign in the next iteration.

Figure 3.1. Model of imaginary-time propagation according to the toy Hamiltonian from equation (3.26). If the Hilbert space is saturated with walkers and all possible spawns are realised, which effectively corresponds to a deterministic propagation, the sign problem is resolved in every iteration due to annihilations. However, if the wavefunction is represented in a stochastic manner in a low-walker regime, spawning can lead to the propagation of wrong signs.

3.2.4 *Toy Example*

Let us make a small practical example of how the sign problem emerges in FCIQMC. For this, I define the matrix

$$\hat{H} = \begin{array}{c} \begin{matrix} |D_0\rangle & |D_1\rangle & |D_2\rangle & |D_3\rangle \end{matrix} \\ \begin{pmatrix} d & -1 & -1 & +1 \\ -1 & d & -1 & -1 \\ -1 & -1 & d & -1 \\ +1 & -1 & -1 & d \end{pmatrix} \end{array} \tag{3.26}$$

with the four basis vectors $|D_0\rangle$, $|D_1\rangle$, $|D_2\rangle$, and $|D_3\rangle$ which would correspond to Slater determinants in actual FCIQMC calculations of fermionic systems. d is some diagonal element that is not relevant for the consideration here as the sign problem is mainly a feature of the off-diagonal part of a matrix.

The eigenvector corresponding to the lowest eigenvalue $E_0 = -2.2361 + d$ is given by

$$|\Psi_0\rangle = -0.372\,|D_0\rangle - 0.602\,|D_1\rangle - 0.602\,|D_2\rangle - 0.372\,|D_3\rangle\ . \tag{3.27}$$

It is depicted in figure 3.1a. Even though there are coefficients of only one sign, this does not mean there is no sign problem.

To see this, let us look at one application of the off-diagonal part of $-\hat{H}$, corresponding to one step in the imaginary-time propagation with $\Delta\tau = 1$. For an equilibrated walker-saturated simulation, which means that all basis functions are already occupied with a walker weight proportional to their C_i weight, the part of this application that involves spawning events to $|D_0\rangle$ is shown in figure 3.1b. This corresponds to a deterministic (i.e. non-stochastic) evolution. It becomes evident that there are opposite-sign walker contributions that are added with their respective sign. The walkers therefore annihilate.

Figure 3.1c shows the situation in which the Hilbert space is only occupied by a small number of walkers. In the instantaneous snapshot of the wavefunction depicted, $|D_0\rangle$, $|D_1\rangle$, and $|D_2\rangle$ are unoccupied. A spawning event from $|D_3\rangle$ to $|D_0\rangle$ leads to a walker on $|D_0\rangle$ with a wrong positive sign which will be present in the next iteration (unless removed due to the diagonal death/cloning step). This is a manifestation of the sign problem in the low-walker limit and causes noise that makes trial energy estimators unusable. Also, the shift energy estimator will be biased as the walker growth

rate $N_{\text{tot}}(\tau + \Delta\tau)/N_{\text{tot}}(\tau)$, which is measured by the shift, is too large due to non-occuring annihilations. This leads to a too low shift value.

3.2.5 *Practical Implications of the Sign Problem*

Now that the sign problem in FCIQMC has been discussed fundamentally, I will look at how the sign problem affects actual FCIQMC calculations. Figure 3.2 shows the behaviour of important simulation variables that are influenced by the sign problem for a simple, yet non-trivial example system: the Hubbard model at interaction strength $U/t = 8$ with a 4×4 square lattice geometry with periodic boundary conditions. The Hubbard model will be introduced in chapter 4, however here it is just used as a placeholder for any kind of FCIQMC-simulable system. The simulations are conducted at two different walker numbers that each are targeted by using the improved population control according to equation (3.9), respectively: one at $N_{\text{tot}} = 1 \times 10^6 < N_{\text{min}}$ and the other at $N_{\text{tot}} = 5 \times 10^6 > N_{\text{min}}$. N_{min}, as defined above, is the minimum number of walkers above which the sign problem is overcome and the correct wavefunction is sampled.

One can now judge from the behaviour of

(a) the population on the determinant with the highest population, the reference determinant,

(b) the projected energy, and

(c) the shift

whether the sign problem is overcome or not when N_{min} is unknown. As visualised in figure 3.2, for $N_{\text{tot}} < N_{\text{min}}$ the simulation variables show the following behaviour:

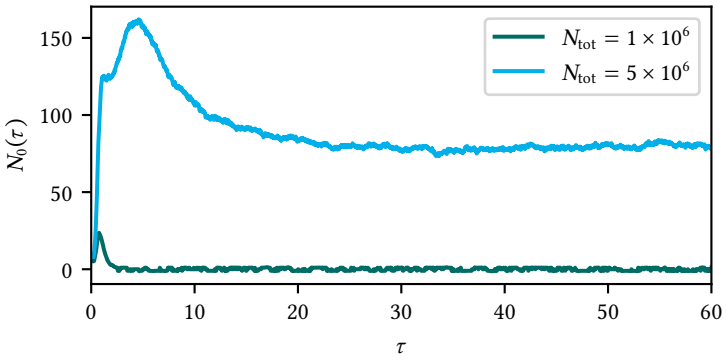
(a) The population on the reference drops to and then fluctates around 0, with no properly established consistent sign.[8]

(b) The projected energy fluctuates wildly and is undefined when the reference population is exactly zero.

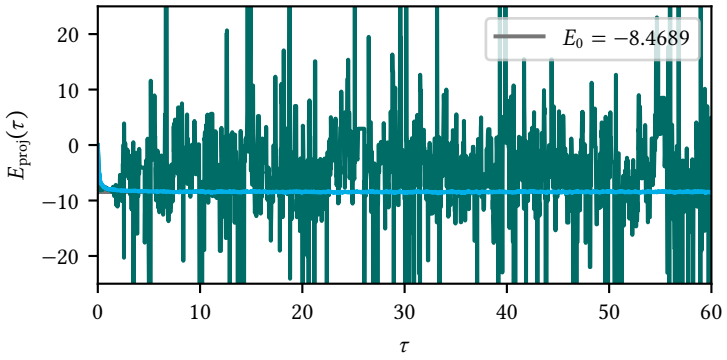
(c) The shift is below the exact energy E_0 and does not agree with the average projected energy. It lies above the stoquastised energy, however. This is due to the fact that some but not all necessary annihilation events take place during the run.

For $N_{\text{tot}} \geq N_{\text{min}}$, the variables show the following behaviour:

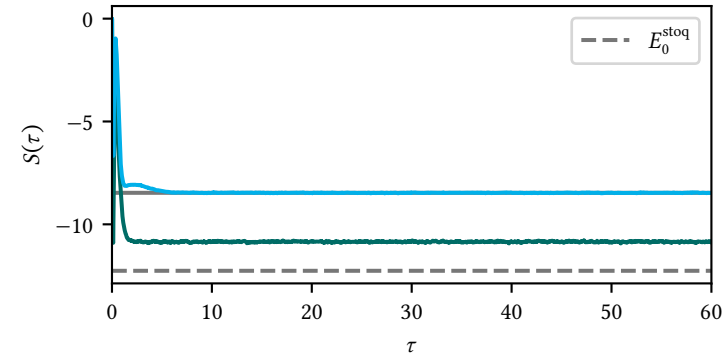
8 In principle, the overlap with the reference determinant and the sampled wavefunction $\langle D_0|\Psi\rangle$ can vanish and the simulation can still be sign-coherent. This occurs when there is a better trial wavefunction $|\Psi_t\rangle$ such that $\langle\Psi_t|\Psi\rangle$, the denominator of the trial energy, still has a well-established sign and is non-zero. The sign problem is unresolved if no trial wavefunction can be found (except for the exact ground-state wavefunction) such that the overlap shows this behaviour. However, the reference population is usually a good approximation and can be determined with no computational overhead.



(a) Reference population N_0 as a function of imaginary time τ. For $N_{\text{tot}} \geq N_{\text{min}}$, N_0 approaches a constant value of one sign. For $N_{\text{tot}} < N_{\text{min}}$, N_0 vanishes and fluctuates around zero with both signs being present.



(b) Projected energy E_{proj} as a function of τ. For $N_{\text{tot}} \geq N_{\text{min}}$, E_{proj} approaches the correct ground-state energy E_0 on average. For $N_{\text{tot}} < N_{\text{min}}$, E_{proj} fluctuates strongly due to a vanishing N_0 and its average does not approach E_0.



(c) Shift S as a function of τ. For $N_{\text{tot}} \geq N_{\text{min}}$, S like E_{proj} converges to E_0 on average. For $N_{\text{tot}} < N_{\text{min}}$, the τ-averaged S converges to a value lower than E_0 due to the unresolved sign problem.

Figure 3.2. Simulation variables (reference population, projected energy, and shift) as a function of imaginary time in an actual FCIQMC simulation of a 4×4 Hubbard model at $U/t = 8$ with periodic boundary conditions. Two total walker populations are looked at: $N_{\text{tot}} = 1 \times 10^6$ which is below the minimum walker population N_{min}, where the sign problem is still unresolved, and $N_{\text{tot}} = 5 \times 10^6$ which is above N_{min}.

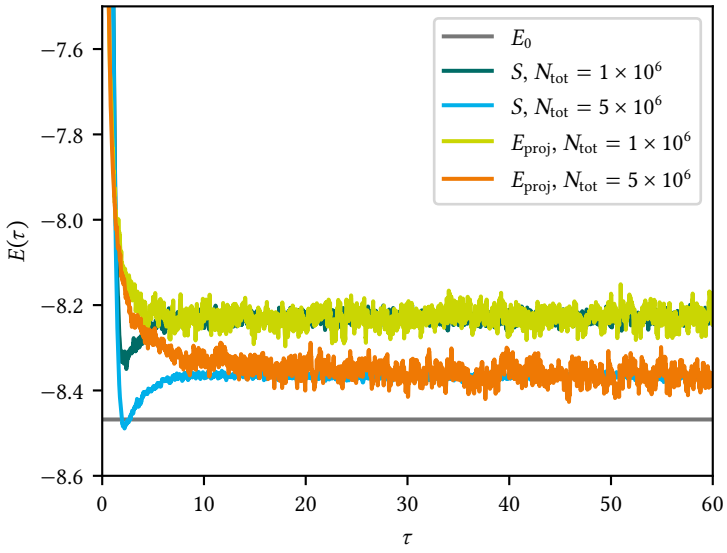


Figure 3.3. FCIQMC simulation runs when applying the initiator approximation with $t_{\text{init}} = 3$. When using the initiator method, one observes convergent behaviour also for $N_{\text{tot}} = 1 \times 10^6 < N_{\text{min}}$, unlike in the full scheme shown in figure 3.2. This comes at the expense of introducing a bias in the energy estimators. The biases vanish in the limit of infinite total population but are still non-zero for $N_{\text{tot}} = 5 \times 10^6 > N_{\text{min}}$.

(a) The population on the reference fluctuates around a non-zero value and does not change sign during the entire simulation.

(b) The projected energy fluctuates around E_0.

(c) The shift also fluctuates around E_0 and on average agrees with the projected energy. When increasing N_{tot}, the average shift does not change.

Examples for FCIQMC runs on the same system with the initiator method are shown in figure 3.3. Unlike the full unbiased algorithm with annihilations only, the algorithm shows converging behaviour not only for $N_{\text{tot}} \geq N_{\text{min}}$ but also for $N_{\text{tot}} < N_{\text{min}}$: Also for $N_{\text{tot}} = 1 \times 10^6$, the reference population has a well-established sign and the projected energy and the shift converge to the same value on average. As discussed before, this comes at the expense of a systematic bias in the wavefunction and especially the energy estimates. It is important to note that the energy estimators converge to the exact energy for $N_{\text{tot}} \to \infty$. However, the energy is still biased for $N_{\text{tot}} \geq N_{\text{min}}$ alone when applying the initiator approximation. In this case, it is advantageous to revert back to the full method. Especially in the lattice model systems that will be introduced in chapter 4, the convergence behaviour with respect to N_{tot} can be problematic and too slow. This is the main justification for the introduction of improved approximations in chapter 8.

4 *The Hubbard Model*

Realistic physical systems are hard to treat numerically, regardless of the numerical method and approximations applied. This is true for any sizable molecule but it is especially true for macroscopic solids. However, macroscopic solids have a key feature that can be exploited when attempting to study them: their periodicity. Macroscopic solids consist of an almost infinite number of ever-repeating primitive cells. This limit of infinite number of primitive cells is called the *thermodynamic limit.*

Even when limiting a calculation to a finite number of primitive cells, the problem can still be hard to treat if all constituent particles in a large number of orbital basis functions have to be treated. This is one of the reasons for the success of lattice model systems throughout theoretical sciences. The foundation of the lattice models is again the Born–Oppenheimer approximation. Solids can be well approximated by negatively-charged electrons moving in a field of positively-charged ions that form a static periodic lattice which is the first defining feature of lattice models. The geometrical shape of this lattice emerges from the chemical constituents of the solids but is simply imposed upon the lattice model by construction. The second feature that defines lattice models is the fact that they are model systems. This means that – aside from the Born–Oppenheimer approximation – also some degrees of freedom of the electronic part of the *ab initio* Schrödinger equation are removed. This has three reasons:

1. Removing degrees of freedom typically reduces the computational cost in simulating the system with high accuracy. Phenomenona observed in experiments can often be observed in simulations, even when some interactions are not simulated.

2. Inversely, models reduce the complexity and therefore enhance fundamental understanding. When a realistic system is simplified and a specific phenomenological effect is still observed, the underlying mechanisms can be pinpointed more easily. Complex interactions can possibly be quantified and classified by introducing effective interaction parameters as we will see in the Hubbard and Heisenberg model.

3. Due to their universality, model systems are ideal when used as benchmarking systems.

In this chapter, I will introduce the Hubbard model in more detail. This is the system in which the numerical studies in part II will be conducted.

The Hubbard model is one of the most well-known and well-studied lattice model systems. It was independently introduced by J. Hubbard, M. Gutzwiller, and H. Kanamori [33–35] in 1963 and was subsequently refined by J. Hubbard in a series of articles [132–136].

It is a simplified model of a real periodic solid, consisting of only two energetic contributions in the Hamiltonian: the *kinetic part* which describes hopping processes of electrons between ionic sites and the *potential part* which describes the on-site interaction between electrons. The kinetic part is parametrised by the *hopping parameter* t. The potential part is parametrised by the *correlation parameter* U.

4.1 *Derivation of the Hubbard Hamiltonian*

To introduce the approximations of the general *ab initio* Hamiltonian from equation (1.3) that lead us to the simplified Hubbard Hamiltonian, let us start from its second quantised version introduced in section 2.1.2 [137]. Since we are dealing with a periodic system, the electronic Hamiltonian from equation (2.5) is modified slightly to become

$$\hat{H}_{\text{lattice}} = \sum_{ia} \sum_{\mu} \sum_{\sigma} t_{i\mu}^{a\mu} \hat{c}_{a\mu\sigma}^{\dagger} \hat{c}_{i\mu\sigma} + \frac{1}{2} \sum_{ijab} \sum_{\mu\nu\alpha\beta} \sum_{\sigma\sigma'} U_{i\mu,j\nu}^{a\alpha,b\beta} \hat{c}_{a\alpha\sigma}^{\dagger} \hat{c}_{b\beta\sigma'}^{\dagger} \hat{c}_{j\nu\sigma'} \hat{c}_{i\mu\sigma} \,. \tag{4.1}$$

Every creation and annihilation operator $\hat{c}_{i\mu}^{(\dagger)}$ now not only carries a *site index* i but also a *band index* μ. The site index numbers the ionic sites as positions in real space, the band index numbers the occupied energy band. Thus the transition amplitudes are – analogously to equation (2.6) – given by

$$t_{i\mu}^{a\mu} = \int \mathrm{d}\boldsymbol{r}\, \varphi_{\mu}^{*}(\boldsymbol{r} - \boldsymbol{R}_a) \hat{h}_1 \varphi_{\mu}(\boldsymbol{r} - \boldsymbol{R}_i) \,, \tag{4.2a}$$

$$U_{i\mu,j\nu}^{a\alpha,b\beta} = \int \mathrm{d}\boldsymbol{r}_1\, \mathrm{d}\boldsymbol{r}_2\, \varphi_{\alpha}^{*}(\boldsymbol{r}_1 - \boldsymbol{R}_a) \varphi_{\beta}^{*}(\boldsymbol{r}_2 - \boldsymbol{R}_b) \hat{h}_2 \varphi_{\mu}(\boldsymbol{r}_1 - \boldsymbol{R}_i) \varphi_{j}(\boldsymbol{r}_2 - \boldsymbol{R}_j) \,. \tag{4.2b}$$

$\hat{h}_1$ and $\hat{h}_2$ are again the one- and two-body parts of the Hamiltonian, respectively. The orbital functions $\varphi_\mu(\boldsymbol{r})$ in the case of a periodic lattice are given by the *Wannier functions* that are defined as

$$\varphi_\mu(\boldsymbol{r}-\boldsymbol{R}) = L^{-\frac{1}{2}} \sum_{\boldsymbol{k}} \exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R})\psi_{\mu\boldsymbol{k}}(\boldsymbol{r}) \tag{4.3}$$

where L is the number of primitive cells considered. They form an orthogonal set of one-body basis functions and they are strongly localised around $\boldsymbol{r} = \boldsymbol{R}$, similar to localised molecular orbitals. The operators $\hat{c}^\dagger_{i\mu\sigma}$ and $\hat{c}_{i\mu\sigma}$ therefore create and annihilate an electron in the Wannier orbital $\varphi_\mu\big(\boldsymbol{r}-\boldsymbol{R}_i\big)$ with spin σ, respectively. $\psi_{\mu\boldsymbol{k}}$ in turn are a set of *Bloch functions. Bloch's theorem* states that for a periodic potential – as it is obviously present in a periodic lattice – the eigensolutions of the Schrödinger equation are of the form

$$\psi_{\mu\boldsymbol{k}}(\boldsymbol{r}) = \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})u_{\mu\boldsymbol{k}}(\boldsymbol{r})\,. \tag{4.4}$$

$u_{\mu\boldsymbol{k}}$ are functions that have the periodicity of the lattice. μ numbers the different solution to the eigenvalue equation

$$\hat{h}_1\psi_{\mu\boldsymbol{k}}(\boldsymbol{r}) = \varepsilon_{\mu\boldsymbol{k}}\psi_{\mu\boldsymbol{k}}(\boldsymbol{r})\,. \tag{4.5}$$

$\varepsilon_{\mu\boldsymbol{k}}$ thus define the energy bands, i.e. they are the dispersion relations and define the relation between the single-particle energy and the quasi momentum $\boldsymbol{k}$ in energy band μ.

4.1.1 *The Hubbard Hamiltonian in the Real-Space Basis*

Now that we have established the basic form of the electronic Hamiltonian in a periodic system and defined the Wannier functions as basis states, we can now introduce the Hubbard approximations to equation (4.1). When all electron–electron interactions $U^{a\alpha,b\beta}_{i\mu,j\nu} = 0$ are small compared to all $t^{a\mu}_{i\mu}$, it is a good approximation to ignore them. This corresponds to the band picture of a solid. In the Hubbard model however, it is assumed that the electron–electron interactions are not negligible but they are local. This means that the on-site interactions dominate all other interactions. Therefore, only the $U^{i\alpha,i\beta}_{i\mu,i\nu}$ are chosen to be non-zero. Furthermore, the Hubbard model assumes that only one band is close to the Fermi level. Thus, we are only left with the interaction terms $U^{i\mu,i\mu}_{i\mu,i\mu}$. The effect of the higher-energy bands can be included in the hopping and interaction parameters of the one conduction band considered, making them effective parameters. Since all ions are of the

same type, the single effective interaction parameter can simply be written as U. Therefore, we are left with the general Hubbard Hamiltonian

$$\hat{H}_{\text{Hubbard, general}} = \sum_{ia\sigma} t_i^a \hat{c}_{a\sigma}^\dagger \hat{c}_{i\sigma} + U \sum_i \hat{c}_{i\uparrow}^\dagger \hat{c}_{i\downarrow}^\dagger \hat{c}_{i\downarrow} \hat{c}_{i\uparrow} \,. \tag{4.6}$$

Here, also the factor $1/2$ has been absorbed into U.

Additional constraints can also be put on the kinetic part of the Hamiltonian. Accounting for the the fact that the Wannier basis functions are typically localised around $\boldsymbol{r} = \boldsymbol{R}$, hopping processes between nearest-neighbour lattice sites usually dominate. Therefore, in the so-called *tight-binding approximaton* the general amplitudes t_i^a are restricted to be non-zero only if i and a are the indices of neighbouring lattice sites. These combinations are denoted as $\langle ia \rangle$ and the hopping amplitude is chosen to be $-t$ for all of those. With this, we finally arrive at

$$\hat{H}_{\text{Hubbard}} = -t \sum_{\langle ia \rangle \sigma} \hat{c}_{a\sigma}^\dagger \hat{c}_{i\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \tag{4.7}$$

where also the occupation number operators $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^\dagger \hat{c}_{i\sigma}$ are used to simplify the notation. This is the form of the Hubbard model that will be used in the remainder of the thesis. Despite its apparent simplicity and the many approximations that have been made, the Hubbard model still exhibits a very rich spectrum of physical phenomena, as will be discussed in section 4.2. It is also numerically hard or even impossible to solve for a general lattice geometry and for arbitrary hopping and interaction parameters. Even though the Hubbard model in tight-binding approximation only has one effective parameter U/t, both the physical behaviour and the most efficient ways to solve it numerically differ vastly for different U/t. For $t > 0$ and $U > 0$, the system is called repulsive. For $U < 0$, it is attractive.

When expanding the Hubbard Hamiltonian in a many-particle basis of Slater determinants, the potential part will become the diagonal and the kinetic part will become the off-diagonal contributions. It is therefore apparent that $\hat{H}_{\text{Hubbard}}$ becomes purely diagonal for $U/t \to \infty$. It is purely off-diagonal for $U/t = 0$. This case however can be solved by a basis rotation analytically which will be discussed in the next section.

4.1.2 *The Hubbard Hamiltonian in the Reciprocal-Space Basis*

To diagonalise the Hamiltonian for $U/t = 0$, a basis transformation into *reciprocal space* needs to be performed. This leads to a different representation

of the Hubbard Hamiltonian, also for $U/t \neq 0$. To do this, we will use the Bloch functions defined in equation (4.4) directly instead of the localised Wannier functions from equation (4.3). The Bloch functions can be obtained from the Wannier functions by inverse Fourier transformation according to

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = L^{-\frac{1}{2}} \sum_{\boldsymbol{R}} \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R})\varphi(\boldsymbol{r}-\boldsymbol{R})\,. \tag{4.8}$$

The band index is neglected here. Accordingly, also the creation and annihilation operators in Wannier space and in Bloch space are connected via

$$\hat{c}^{(\dagger)}_{\boldsymbol{k}\sigma} = L^{-\frac{1}{2}} \sum_{i} \exp\big(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}_i\big)\hat{c}^{(\dagger)}_{i\sigma}\,. \tag{4.9}$$

$\hat{c}^{(\dagger)}_{\boldsymbol{k}\sigma}$ now annihilates (creates) an electron with quasi momentum $\boldsymbol{k}$ and spin σ in the delocalised Bloch orbital $\psi_{\boldsymbol{k}}$. Inserting this definition into equation (4.7) and using

$$L^{-1} \sum_{\boldsymbol{R}} \exp\Big[\big(\boldsymbol{k}-\boldsymbol{k}'\big)\cdot\boldsymbol{R}_i\Big] = \delta_{\boldsymbol{k}\boldsymbol{k}'} \quad \text{and} \tag{4.10a}$$

$$L^{-1} \sum_{\boldsymbol{k}} \exp\Big[\boldsymbol{k}\cdot\big(\boldsymbol{R}-\boldsymbol{R}'\big)\Big] = \delta_{\boldsymbol{R}\boldsymbol{R}'}\,, \tag{4.10b}$$

we arrive at

$$\hat{H}^{k\text{-space}}_{\text{Hubbard}} = \sum_{\boldsymbol{k}\sigma} \varepsilon_{\boldsymbol{k}}\hat{n}_{\boldsymbol{k}\sigma} + \frac{U}{L} \sum_{\boldsymbol{k}\boldsymbol{k}'\boldsymbol{q}\sigma\sigma'} \hat{c}^{\dagger}_{(\boldsymbol{k}-\boldsymbol{q})\sigma}\hat{c}^{\dagger}_{(\boldsymbol{k}'+\boldsymbol{q})\sigma'}\hat{c}_{\boldsymbol{k}'\sigma'}\hat{c}_{\boldsymbol{k}\sigma}\,. \tag{4.11}$$

We can see that for a non-interacting system with $U/t = 0$, the Hamiltonian is already diagonal in this basis. $\varepsilon_{\boldsymbol{k}}$ again is the single-particle dispersion relation. For a lattice with lattice vectors $\boldsymbol{a}_i$, it is given by

$$\varepsilon_{\boldsymbol{k}} = -2t \sum_{i} \Big[\cos\big(\boldsymbol{k}\cdot\boldsymbol{a}_i\big)\Big]\,. \tag{4.12}$$

The interacting part of the Hamiltonian leads to off-diagonal elements in the many-particle basis. It describes a scattering process of two electrons with opposite spins.

4.2 *Physical Features of the Hubbard Model*

Despite the many approximation steps taken from the original multiband, multiorbital *ab initio* Hamiltonian in equation (4.1) to the Hubbard Hamiltonians in equations (4.7) and (4.11), the Hubbard model exhibits many striking

features of more complex solid-state systems. Solely by introducing a finite repulsive on-site interaction $U > 0$, the system's collective behaviour can change fundamentally depending on the system's characteristics. A Hubbard system in the tight-binding approximation can be characterised by

- the number of sites N_s and the lattice geometry,
- the ratio of the on-site interaction to the hopping amplitude U/t,
- the filling given as the ratio between number of electrons and number of sites $n_\mathrm{f} = N_\mathrm{el}/N_\mathrm{s}$, and
- the inverse temperature β in a canonical ensemble.

It is unquestionably impossible to give an exhaustive overview of the phenomenology and applications of the Hubbard model in this thesis. I will therefore just give a brief overview to explain why it is highly relevant to the field of solid state physics and chemistry. The systems considered in this thesis are all at $1/\beta = 0$, do have repulsive interaction ($U/t > 0$), and are close to half-filling.

One of the most crucial features that is correctly described by the Hubbard model is the existence of a *Mott insulator transition* [138, 139]. A Mott insulator is an electronic system that would be expected to be a conductor in a mean-field picture purely judging by the band structure but turns out to be a insulator due to the electron correlation. This is a feature of Hubbard systems: The kinetic part of the tight-binding Hubbard Hamiltonian clearly is conducting. In a one-dimensional lattice geometry however, one finds that despite this fact the system is insulating for all $U/t > 0$. For two-dimensional systems, the phase diagram is more involved.

Apart from mottism, unconventional superconductivity [140, 141] with d-wave pairing away from half-filling [142, 143], striped states [144, 145], charge and spin density waves [146], and pseudogaps have been observed in the Hubbard model phase diagram.

4.3 *Numerical Solution*

As mentioned before, the solution of the Hubbard model in general is very challenging numerically. In the special case of one-dimensional systems, an analytical solution using the *Bethe ansatz* exists. It allows for the derivation of a set of algebraic equations, the *Lieb–Wu equations,* that can be solved analytically in the thermodynamic limit. In three and more dimensions,

a mean-field description is typically a sufficiently accurate approximation. Therefore, methods like dynamical mean-field theory (DMFT) can be applied successfully in these cases [147]. The two-dimensional case turns out to be by far the most challenging one. This is especially true in the intermediate interaction regime for $4 \lesssim U/t \lesssim 12$ for which both the non-interacting case – where the k-space description becomes exact – as well as the infinite-interaction case – where the system becomes the Heisenberg or t–J model (see section 4.4), respectively – are bad approximations.

There are countless computational methods to solve the Hubbard model, all with their own strengths and weaknesses. Although there are overlaps, roughly speaking they can be organised into three categories [54]:

- *Embedding methods* approximate properties of a desired infinite system by solving the problem for a finite embedded system that is self-consistently optimised. Examples are the density matrix embedding theory (DMET) [148], the dynamical cluster approximation (DCA) [149], and the dual fermion method (DF) [150].

- *Green's function-based methods* stochastically evaluate the so-called many-body perturbation series and provide the self energies and Green's functions. Properties on the real-frequency axis can be evaluated subsequently but limit the system size the method can be applied to. Diagrammatic Monte Carlo (DiagMC) [151, 152] is an example, but also DCA and DF make use of Green's function techniques.

- *Wavefunction-based methods* calculate an approximation of the ground-state wavefunction of the system. Examples are unrestricted coupled cluster theory including singles and doubles (UCCSD) [153–155], diffusion Monte Carlo with fixed nodes (DMC) [45], multireference projected Hartree–Fock (MRPHF) [156], density matrix renormalisation group (DMRG) [157], auxiliary-field QMC (AFQMC) [82, 158, 159], and also FCIQMC [112, 160, 161].

DMC, AFQMC, and DMRG have already been presented in more detail in chapter 2. Apart from the analytical Bethe ansatz, DMRG performs well in one-dimensional (1-d) and close-to-1-d systems due to its reliance on locality in the MPS description. Monte Carlo methods typically exhibit a sign problem because the standard Hubbard model deals with fermions. AFQMC however has the special feature that it is sign-problem-free for $n_\mathrm{f} = 1$ (half-filling) on bipartite lattices. This is due to a special symmetry that makes all auxiliary fields positive (see section 2.2.3).

In chapter 5, it will be shown that FCIQMC can also treat very large one-dimensional systems with more than 100 sites because of a non-extensive sign problem. I will also demonstrate how the FCIQMC algorithm can be adapted to deal with the weak sign roblem in Hubbard systems close to half-filling which exhibit a highly spread-out wavefunction.

4.4 *Large-Interaction Limit*

For $U/t \to \infty$, the Hubbard model transforms into the Heisenberg model. Even though it was described in the 1920s already, it can be seen as a special case of the Hubbard model [26, 28, 162–164]. It is derived considering the the kinetic term as a perturbation to the on-site interaction term and then doing second-order perturbation theory [165]. The resulting Hamiltonian is given by

$$\hat{H}_{\text{Heisenberg}} = J \sum_{\langle ij \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \,. \tag{4.13}$$

$\langle ij \rangle$ again defines the sum over nearest-neighbouring sites i and j. J is the isotropic interaction parameter. $J > 0$ leads to antiferromagnetic coupling, the case mostly studied. $\hat{\mathbf{S}}_i = \left(\hat{S}_i^x, \hat{S}_i^y, \hat{S}_i^z\right)$ is the spin operator on site i. $\hat{\mathbf{S}}_i^2$ has the well-known eigenvalues $S(S+1)$ with $S \in \left\{\frac{1}{2}, 1, \frac{3}{2}, \dots\right\}$. Here, the focus will on the case $S = \frac{1}{2}$, i.e. there is a single electron on each site.[9]

9 There are also anisotropic generalisations of the model: If there are separate J_x, J_y, and J_z for each direction of the spin, the model is called the *XYZ Heisenberg model.* If $J_x = J_y$, it is called the *XXZ Heisenberg model.*

The Heisenberg Hamiltonian describes magnetic interactions of quantum-mechanical spins that are localised at their respective lattice site. Like the Hubbard model, it is an effective model neglecting most of the degress of freedom but still carrying the most fundamental parts. There is a variety of one-dimensional [166–168], two-dimensional [169–171], and three-dimensional systems [172, 173] that show a behaviour that can be well described by the Heisenberg model. Like the Hubbard model, the Heisenberg model can be solved analytically in one dimension using the Bethe ansatz [174–178].

With respect to the success of QMC methods in treating the Heisenberg model, it is important to note that the Heisenberg model on a bipartite lattice is sign-problem-free in any number of dimensions. A proof for this will be given in section 5.1.2. It explains heuristically why the sign problem of the Hubbard model in the real-space basis decreases with increasing U/t.

4.5 *Energy Units*

All energies given for the Hubbard model throughout this thesis will be given as total energies in units of t.

PART II

CONCEPTS & RESULTS

5 *Classification of the Sign Problem in Model Systems*

With its annihilation step in a discrete basis, as described in sections 3.1.1 and 3.2.2, FCIQMC possesses the ability to mitigate the sign problem in a non-biasing fashion. This enables results to be obtained with a walker number N_{min} that is only a fraction of the Hilbert space size. Additionally, there is the initiator approximation, as discussed in section 3.2.3, which allows stable sign-coherent sampling at even lower walker numbers at the expense of introducing a bias in the sampled wavefunction, the initiator bias.

Parts of the results presented in this chapter are also contained in ref. 179. Collaborators: K. Ghanem and A. Alavi.

The strength of the sign problem is typically quantified in terms of the gap between the true ground-state energy of the full Hamiltonian E_0 and the ground-state energy of the stoquastised version of the Hamiltonian E_0^{stoq} as defined in section 3.2.1. The minimum walker number N_{min} strongly depends on this gap, even though the compactness of the wavefunction also has an influence. This will be studied and exploited in chapter 7. The initiator approximation works best for compact wavefunctions where the majority of the wavefunction's ℓ_1 norm is contained within few excitations of the reference Slater determinant.

Typically, molecular *ab initio* systems exhibit strong sign problems with large stoquastised gaps and relatively compact wavefunctions [39, 56]. The same is true for the Hubbard model in a reciprocal-space basis [38, 160, 180]. In this chapter, I will describe how the sign problem behaves in real-space lattice models. The stoquastised gap in systems of this kind strongly depends on the lattice geometry and how much the excitations are constrained compared to the *ab initio* Hamiltonian from wich the model systems are derived:

- For a one-dimensional (1-d) Hubbard model with nearest-neighbour hopping, there is even the possibility that the system can be solved without a sign problem which will be proven in section 5.1. The same is true for the 2-d Heisenberg model on a square lattice. These systems will be used as a paradigm to study the FCIQMC population control bias in chapter 6.

- Hubbard chains that are not sign-problem-free show another interesting feature: Their sign problem is not size-extensive, i.e. larger systems

closer to the thermodynamic limit can be easier to solve than smaller ones. This counterintuitve behaviour will be explained in section 5.2.1.

- All other Hubbard systems show an extensive sign problem, yet it is typically weak compared to the *ab initio* case. The strength of the sign problem strongly depends on the lattice geometry in these cases, especially on the length of the innermost cycle. This will be discussed in section 5.2.2.

5.1 *Sign-Problem-Free Systems in FCIQMC*

In this section, I will provide a proof about how some 1-d Hubbard systems and the 2-d Heisenberg model is sign-problem-free. I will also describe a rule that can be applied to determine which 1-d Hubbard systems are sign-problem-free.

5.1.1 *One-Dimensional Hubbard Model*

Whether the 1-d Hubbard model is sign-problem-free or not depends on the boundary conditions imposed and on the filling. It is sign-problem-free

- for open boundary conditions (i.e. the first and the last site of the chain are not connected via hopping terms) for any filling,
- for periodic boundary conditions (i.e. the first and the last site are connected via a hopping term with amplitude $-t$) only for an odd number of ↑-electrons ($N_\uparrow$) and an odd number of ↓-electrons ($N_\downarrow$), and
- for antiperiodic boundary conditions (i.e. the first and the last site are connected with amplitude $+t$) only for an even $N_\uparrow$ and an even $N_\downarrow$.

To prove this, I will use the representation of a Slater determinant as a product of creation operators according to equation (2.10). In spin-first ordering, a Slater determinant written as

$$|D_i\rangle = \prod_{p=1}^{N_\uparrow} c^\dagger_{\alpha_p\uparrow} \prod_{q=1}^{N_\downarrow} c^\dagger_{\beta_q\downarrow} |\rangle \tag{5.1}$$

is defined with a positive sign. α_p and β_q are spatial sites in ascending order. Since we are dealing with a Hubbard Hamiltonian with nearest-neighbour

interactions only, without loss of generality we can write an excitation of an ↑-electron from site j to one of its neighbours $j \pm 1$ as

$$c^\dagger_{j\pm1\uparrow} c_{j\uparrow} c^\dagger_{\alpha_1\uparrow} \dots c^\dagger_{j\uparrow} \dots c^\dagger_{\alpha_{N_\uparrow}\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{N_\downarrow}\downarrow} |\rangle = c^\dagger_{\alpha_1\uparrow} \dots c^\dagger_{j\pm1\uparrow} \underbrace{c_{j\uparrow} c^\dagger_{j\uparrow}}_{=1} \dots c^\dagger_{\alpha_{N_\uparrow}\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{N_\downarrow}\downarrow} |\rangle \ . \quad (5.2)$$

Of course, site $j \pm 1$ needs to be unoccupied in this case. As long as the excitation does not move an electron from the end of the chain to the beginning or vice versa, both the creation operator $c^\dagger_{j\pm1\uparrow}$ and the annihilation operator $c_{j\uparrow}$ have been anticommuted with the same number of creation operators. Therefore, there is no sign change due to fermionic antisymmetry. Trivially, the same is true for a ↓-electron. Therefore all Hubbard chains with open boundary conditions are sign-problem-free.

When using periodic boundary conditions, exciting an ↑-electron from site ℓ to site 1 leads to

$$c^\dagger_{1\uparrow} c_{\ell\uparrow} c^\dagger_{\alpha_1\uparrow} \dots c^\dagger_{\ell\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{n_\downarrow}\downarrow} |\rangle = (-1)^{N_\uparrow - 1} c^\dagger_{1\uparrow} c^\dagger_{\alpha_1\uparrow} \dots \underbrace{c_{\ell\uparrow} c^\dagger_{\ell\uparrow}}_{=1} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{N_\downarrow}\downarrow} |\rangle \ . \quad (5.3)$$

The factor $(-1)^{N_\uparrow - 1}$ occurs because the creation operator $c_{\ell\uparrow}$ has been commuted with $N_\uparrow - 1$ creation operators. The creation operator $c^\dagger_{1\uparrow}$ is already at its correct position and does not need to be commuted. This means that there is a sign change for even $N_\uparrow$. Equally, moving a ↓-electron from site ℓ to site 1 leads to

$$c^\dagger_{1\downarrow} c_{\ell\downarrow} c^\dagger_{\alpha_1\uparrow} \dots c^\dagger_{\alpha_{n_\uparrow}\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\ell\downarrow} |\rangle = (-1)^{2N_\uparrow + N_\downarrow - 1} c^\dagger_{\alpha_1\uparrow} \dots c^\dagger_{\alpha_{n_\uparrow}\uparrow} c^\dagger_{1\downarrow} c^\dagger_{\beta_1\downarrow} \dots \underbrace{c_{\ell\downarrow} c^\dagger_{\ell\downarrow}}_{=1} |\rangle \ . \quad (5.4)$$

The same rule follows.

5.1.2 *Two-Dimensional Heisenberg Model*

It is easy to see that the Heisenberg model on a square-lattice geometry is sign-problem-free when taking the linearised imaginary-time Schrödinger equation (3.3), that e.g. FCIQMC is based on, as a foundation. Qualitatively speaking, in this way the ground state is reached by subsequent applications of $-\hat{H}$. Suppose, the lattice is divided in two sublattices A and B such that a site in sublattice A only has neighbours of sublattice B and vice versa. This

way, in the Neél state one sublattice, say A, is only occupied by ↑-electrons and the other sublattice, say B, is only occupied by ↓-electrons. This can be written as $N_\downarrow^A = N_\uparrow^B = 0$ where N_σ^s is the number of σ-spin electrons on sublattice s. Since only exchange interactions are present, one application of $-\hat{H}_{\text{Heisenberg}}$ will change these numbers to $N_\downarrow^A = N_\uparrow^B = 1$. Since $J > 0$, every application of $-\hat{H}_{\text{Heisenberg}}$ will introduce a sign change. If $-\hat{H}_{\text{Heisenberg}}$ is applied again, this leads to either $N_\downarrow^A = N_\uparrow^B = 0$ or $N_\downarrow^A = N_\uparrow^B = 2$. This way, it is possible to identify whether an even or an odd number of applications p of the Hamiltonian have been necessary to reach a configuration merely by counting $N_\downarrow^A$ or $N_\uparrow^B$, respectively. Since every application of the Hamiltonian is sign-changing, the contributions onto a configuration can only ever have the sign of $(-1)^p$. Therefore, the Hilbert space can be subdivided into two disjoint subspaces, one subspace only getting contributions with positive and the other only getting contributions with negative sign. This implies the absence of a sign problem.

Because the Heisenberg model describes the $U/t \to \infty$ case of the Hubbard model (see section 4.4), this gives a hint for why the sign problem decreases with increasing U/t. Amongst other observations, a more quantitative discussion of this fact will be given in section 5.2.1.

5.2 *Size-Extensivity of the Hubbard Sign Problem*

To understand how the sign problem emerges in the Hubbard model and how it behaves as a function of system size, let us first make a couple of basic considerations using a path-integral-type argument. According to equation (3.3), the amplitude of a single walker propagated from a Slater determinant $|D_i\rangle$ to another Slater determinant $|D_j\rangle$ in a single time step $\Delta\tau$ is given by

$$P_{ji} = \left\langle D_j \middle| 1 - \Delta\tau \hat{H} \middle| D_i \right\rangle . \tag{5.5}$$

The amplitude of a set of walkers is then given by the sum over all connected Feynman paths

$$P_{ji}^{(K)} = \underbrace{\sum_{k,\ell,\ldots,y,z}}_{K} P_{jz} P_{zy} \ldots P_{\ell k} P_{ki} . \tag{5.6}$$

In the large-τ limit, one also approaches the large-K limit. Therefore, I define the total transition amplitude as

$$C_{ji} = \lim_{K\to\infty} P_{ji}^{(K)} . \tag{5.7}$$

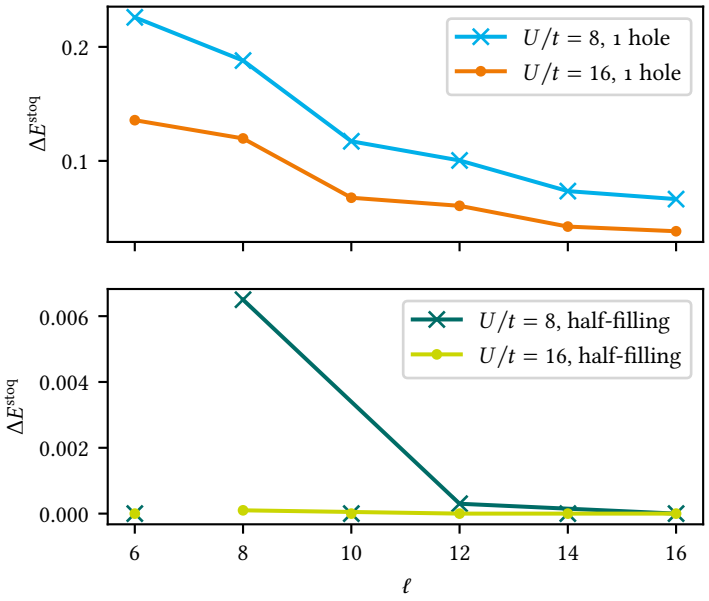


Figure 5.1. Stoquastised gaps ΔE^{stoq} of periodic 1-d Hubbard systems with increasing chain lengths ℓ with one hole and at half-filling [179]. While there are certain chain lengths in the half-filled system where the gap is exactly zero, i.e. the system is sign-problem-free, all other systems show a non-size-extensive sign problem. The stoquastised gaps close with increasing system size. The gaps in the one-hole case are generally larger compared to the half-filled case. Gaps for $U/t = 8$ are generally larger compared to $U/t = 16$.

Analogously, the amplitude of a set of walkers for the stoquastised Hamiltonian is given by

$$P_{ji}^{(K),\,\text{stoq}} = \underbrace{\sum_{k,\ell,\ldots,y,z}}_{K} \left| P_{jz} P_{zy} \ldots P_{\ell k} P_{ki} \right| \tag{5.8}$$

and the total transition amplitude thus is

$$C_{ji}^{\text{stoq}} = \lim_{K\to\infty} P_{ji}^{(K),\,\text{stoq}} \,. \tag{5.9}$$

A sign problem that opens up the gap between fermionic and stoquastised ground state emerges when there is at least one K for which

$$P_{ji}^{(K),\,\text{stoq}} - \left| P_{ji}^{(K)} \right| > 0 \,, \tag{5.10}$$

i.e.

$$C_{ji}^{\text{stoq}} > \left| C_{ji} \right| . \tag{5.11}$$

5.2.1 *Systems with Non-size-extensive Sign Problems*

For one-dimensional systems that are not sign-problem-free according to the rules from section 5.1.1, let us first collect data from some numerical experiments. Figure 5.1 shows the stoquastised gaps $\Delta E^{\text{stoq}} = E_0 - E_0^{\text{stoq}}$ of Hubbard chains of increasing length ℓ, each with one hole and at half-filling.

The half-filled systems with $\ell = 4n + 2,\ n \in \mathbb{N}^0$ are sign-problem-free as they contain an odd number of electrons of each spin species. As expected, chains of other lengths show a non-zero stoquastised gap. It is decreasing with increasing ℓ, almost going to zero for $\ell \geq 16$. An additional observation is that the gaps are much smaller for $U/t = 16$ compared to $U/t = 8$.

In the one-hole case, ΔE^{stoq} is roughly an order of magnitude larger than in the half-filled case. Also, as expected the gaps are no longer zero for $\ell = 4n + 2$. However, there still seems to be a spurious influence of the sign-problem-free half-filled chains as in these cases the gap is smaller than a simple extrapolation would predict. Also here, the gaps decrease with increasing ℓ. A larger U/t seems to imply smaller gaps.

How can these observations be explained? To illustrate this, let us study pathways from $|D_i\rangle$ to $|D_j\rangle$ via the 1-d Hubbard Hamiltonian. Without loss of generality, I consider a system with an odd number of ↓-electrons and an even number of ↑-electrons with periodic boundary conditions. Let us also say that the determinants are given by

$$|D_i\rangle = c^\dagger_{\alpha_1\uparrow} \dots c_{p\uparrow} \dots c^\dagger_{\alpha_{N_\uparrow}\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{N_\downarrow}\downarrow} |\rangle \quad \text{and} \tag{5.12a}$$

$$|D_j\rangle = c^\dagger_{\alpha_1\uparrow} \dots c_{q\uparrow} \dots c^\dagger_{\alpha_{N_\uparrow}\uparrow} c^\dagger_{\beta_1\downarrow} \dots c^\dagger_{\beta_{N_\downarrow}\downarrow} |\rangle\ , \tag{5.12b}$$

i.e. $|D_i\rangle$ and $|D_j\rangle$ only differ in the position of a single ↑-electron which is moved from site p to one of its neighbouring sites q. Following the argument for when a system is sign-problem-free, a sign problem emerges when there are multiple distinct pathways from $|D_i\rangle$ to $|D_j\rangle$, leading to opposite-sign contributions at $|D_j\rangle$. In a periodic 1-d situation, there are only two such pathways: one exploiting the periodic boundary conditions and one direct pathway. By using the FCIQMC master equation equation (3.3), it is possible to quantify the magnitudes of the different contributions as a function of the system parameters.

When not exploiting the boundary conditions and according to the definition in equation (5.6), the contribution from $|D_i\rangle$ onto $|D_j\rangle$ is given by

$$P^{ji}_{\text{non-periodic}} = (\Delta\tau t)^{|p-q|} \Big[1 - \Delta\tau \big(N_{\text{docc}} + 1\big) U\Big]^{N_\downarrow^{pq}} \Big[1 - \Delta\tau N_{\text{docc}} U\Big]^{|p-q| - N_\downarrow^{pq}}. \tag{5.13}$$

N_s is the number of lattice sites. In this case, I am only selecting the dominant pathway out of the K pathways described above to simplify the illustration. This is because each application of the off-diagonal part of $-\hat{H}$ contributes $\Delta\tau t$ and each application of the diagonal part contributes $1 - \Delta\tau(N_{\text{docc}} + 1)U$

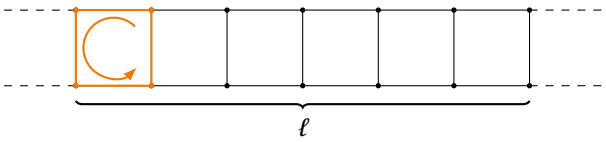


Figure 5.2. Ladder-like lattice geometry with length ℓ. This lattice shape is used as a paradigm for a Hubbard system with weak yet extensive sign problems. The sign problem is weak because the geometry is still close to 1-d. It is extensive because the number of sign-problematic four-site plaquettes (indicated in orange with an arrow) increases proportionally with respect to ℓ.

if the moving electron's position is already occupied by an opposite-spin electron. Otherwise, the diagonal part contributes $1 - \Delta\tau N_{\text{docc}} U$. N_{docc} is the number of doubly occupied sites between sites p and q in $|D_i\rangle$. $N_\downarrow^{pq}$ is the number of sites between p and q that are singly occupied by a $\downarrow$-electron. The product of all these contributions leads to the total contribution onto $|D_j\rangle$.

When exploiting periodic boundary conditions and after a similar consideration, the contribution from $|D_i\rangle$ onto $|D_j\rangle$ amounts to

$$P^{ji}_{\text{periodic}} = -(\Delta\tau t)^{N_s - |p-q|} \left[1 - \Delta\tau\big(N_{\text{docc}} + 1\big)U\right]^{N_\downarrow - N_\downarrow^{pq}} \left[1 - \Delta\tau N_{\text{docc}} U\right]^{(N_s - |j-i|) - (N_\downarrow - N_\downarrow^{pq})} . \quad (5.14)$$

According to equations (5.3) and (5.4), the periodic and the non-periodic contributions have opposite signs in the one-hole system.

The total contribution of these two paths onto $|D_j\rangle$ is thus given by

$$P^{ji} = P^{ji}_{\text{periodic}} + P^{ji}_{\text{non-periodic}} . \quad (5.15)$$

The stoquastised contribution is given by

$$P^{ji,\,\text{stoq}} = \left|P^{ji}_{\text{periodic}}\right| + \left|P^{ji}_{\text{non-periodic}}\right| . \quad (5.16)$$

The larger the difference $P^{ji,\,\text{stoq}} - P^{ji}$, the stronger is the contribution of these pathways to the opening of the stoquastised gap and therefore the sign problem. From this, one can conclude the following:

- The opposite-sign contribution of P^{ji}_{periodic} decreases with increasing N_s so $P^{ji,\,\text{stoq}} - P^{ji}$ shrinks and the sign problem becomes weaker.
- The fewer opposite-spin electrons there are between sites p and q compared to the total number of opposite-spin electrons, the smaller is $P^{ji,\,\text{stoq}} - P^{ji}$.
- The same is true for increasing U.

Table 5.1. Stoquastised gaps ΔE^{stoq} of the 4×4 Hubbard lattice at $U/t = 4$ and 8 both in the real- and the reciprocal-space basis. Clearly, when using ΔE^{stoq} as a measure for the strength of the sign problem, the sign problem in the real-space basis is much weaker than in the reciprocal-space basis. This is the case even for $U/t = 4$ where the reciprocal space is the basis representation that leads to significantly more compact wavefunctions.

U/t	real	reciprocal
4	8.401	40.720
8	3.792	97.201

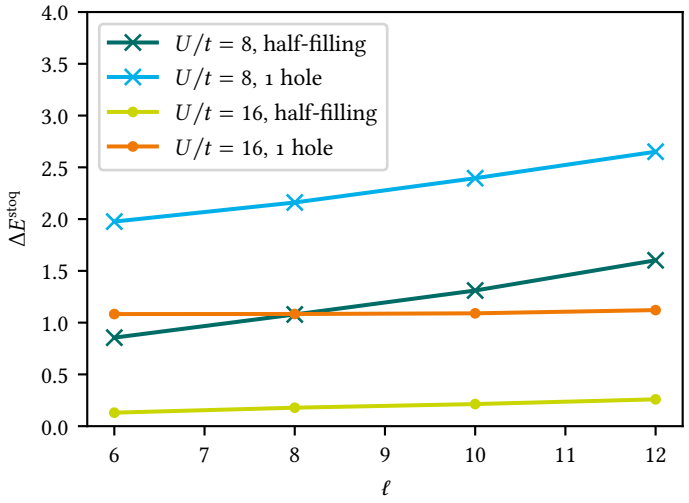


Figure 5.3. Stoquastised gaps ΔE^{stoq} of periodic Hubbard ladder systems with increasing chain lengths ℓ with one hole and at half-filling [179]. For all system parameters, ΔE^{stoq} increases with ℓ, making the sign problem size-extensive. Again, the gaps in the one-hole case are larger compared to the half-filled case. Gaps for $U/t = 8$ are larger compared to $U/t = 16$.

5.2.2 *Systems with Extensive Sign Problems*

In proper 2-d lattices, the sign problem is extensive, i.e. it grows with system size. First, it has to be noted that the Hubbard model in a real-space basis has a comparatively weak sign problem. To show this, I compare the stoquastised gaps of 4×4 Hubbard square lattices at $U/t = 4$ and 8 both in a real- and a reciprocal-space formulation in table 5.1. For $U/t = 8$, ΔE^{stoq} is almost two orders of magnitude smaller for the real-space basis compared to the reciprocal-space basis. Even for $U/t = 4$ where the reciprocal-space representation leads to a much more compact ground-state wavefunction representation, the sign problem, measured by ΔE^{stoq}, is much weaker in the real-space case.

I will illustrate this empirically by looking at a two-dimensional system that is still close to a 1-d system: the ladder-like geometry. A sketch of the lattice is shown in figure 5.2.[10] The ladder systems consist of $2 \times \ell$ sites consisting of four-site square plaquettes.

The stoquastised gaps for Hubbard ladders with increasing ℓ are shown in figure 5.3. Again, I look at systems at half-filling and with one hole at $U/t = 8$ and 16, respectively. Unlike in the 1-d case, the gaps increase with system size in all four cases.

Qualitatively, the main contribution to the sign problem comes from the shortest-possible sign-problematic loops. The contribution of each pathway in equations (5.6) and (5.8) decreases with each necessary application of the Hamiltonian due to the diagonal death step. In 1-d lattices, this shortest cycle has the length of the entire chain. In a square 2-d lattice, the shortest cycles are always the four-site plaquettes the lattice is made up of, as

[10] A synopsis of all lattice geometries used throughout the thesis is given in appendix A.1.

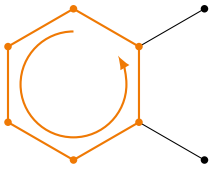


Figure 5.4. Honeycomb lattice geometry. Physically, it is a system of great interest because it resembles the lattice structure of graphene. From a sign-problem perspective, the honeycomb lattice differs from the square lattice in the fact that the innermost loop consists of six instead of four sites. Both the increased length but also the fact that the six-site half-filled periodic chain is sign-problem-free according to section 5.1.1 make the sign problem of honeycomb systems weaker for an equal number of lattice sites.

indicated in figure 5.3. Increasing the system size means that more of the four-site plaquettes are added while obviously not increasing the length of the innermost loop. In $\ell \times \ell$ square lattices, this problem is emphasised as the number of four-site plaquettes does not grow linearly but quadratically in ℓ. The observation that the stoquastised gap is larger for the one-hole system and for larger U/t still holds also in the case of extensive sign problems.

This insight leads us to the conclusion that increasing the length of the innermost plaquettes in a 2-d lattice will lead to a weaker sign problem compared to a system with an equal number of sites but made up of four-site plaquettes. A physically relevant lattice type in this regard is the *honeycomb lattice structure*. The primitive cell of the honeycomb lattice is depicted in figure 5.4. The honeycomb structure has gained significant interest in solid-state research because it is the lattice structure of graphene [181, 182]. It consists of two-dimensional sheets of sp^3-hybridised carbon. Among other special features in the phase diagram, the Hubbard model in honeycomb geometry shows signs of spin liquid behaviour between the antiferromagnetic and semimetallic phase [183–185].

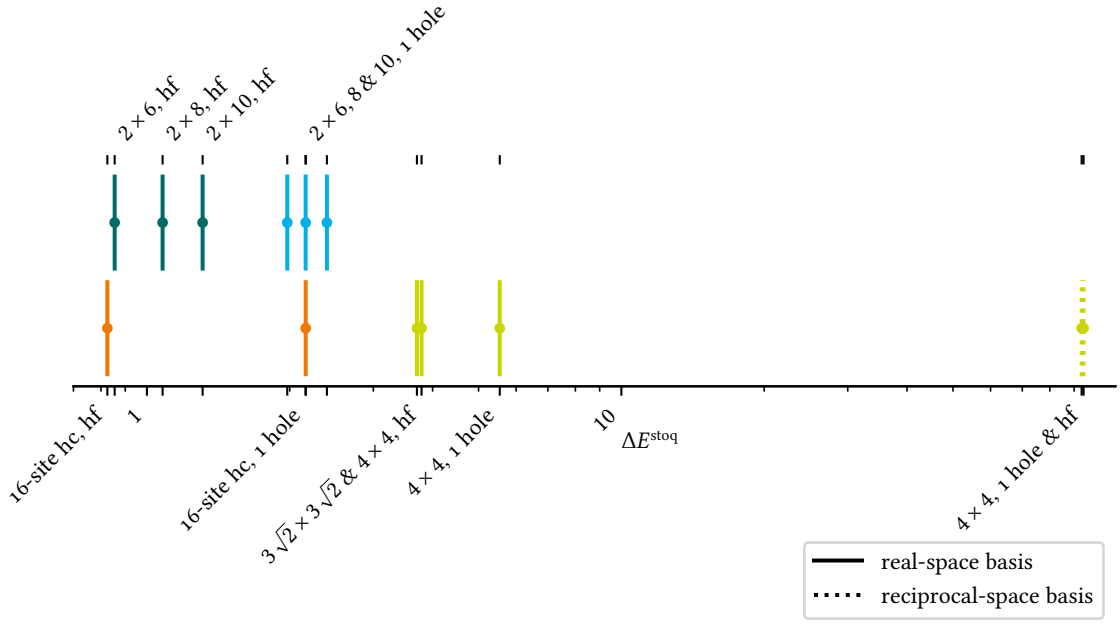


Figure 5.5. Synopsis of stoquastised gaps for ladder systems and 16- and 18-site 2-d lattices with different lattice geometries and orbital representations at $U/t = 8$ [179]. The ladder geometry but also the honeycomb geometry show weaker sign problems than lattices made up of four-site plaquettes. This shows that the sign problem in Hubbard systems is mainly influenced by the number and size of the innermost loop as defined in the main text.

Figure 5.5 summarises the stoquastised gaps for the ladder systems for $\ell = 6, 8$, and 10 and for 16- and 18-site lattices at $U/t = 8$ at half-filling and with one hole. ΔE^{stoq} for 16-site lattices is given both for a square

and a honeycomb geometry. One can clearly observe the influence of the lattice geometry on the strength of the sign problem as discussed above. For comparison, also the stoquastised gap of the 16-site lattice with square plaquettes in a reciprocal-space basis is shown. It is about an order of magnitude larger as already shown in table 5.1.

6 *Population Control Bias and Importance-Sampled FCIQMC*

As established in chapter 5, the overwhelming majority of electronic structure problems of interest, be it *ab initio* or model systems, show a sign problem. This depends on the gap between the ground-state energy of the stoquastised Hamiltonian $\hat{H}^{\text{stoq}}$, as defined in equation (3.19), and the ground-state energy of the fermionic Hamiltonian $\hat{H}$ and the structure of the wavefunction itself.[11] The sign problems can be either strong or weak and they can behave in a size-extensive and non-size-extensive manner.

[11] The dependency of the FCIQMC-related sign problem on the shape of the wavefunction will be discussed in detail in chapter 7.

In this chapter however, the focus will be on the calculation of the one-dimensional (1-d) Hubbard model with nearest-neighbour hopping which is sign-problem-free at half-filling for special configurations according to section 5.1.1. Furthermore, large 1-d Hubbard systems with one hole can be calculated due to the non-size-extensive sign problem.

Parts of the results presented in this chapter are also contained in ref. 186. Collaborators: K. Ghanem and A. Alavi. In this chapter, the analytical derivations and the implementation of the *a-posteriori* correction weights in a Python script for `NECI` postprocessing were done by K. Ghanem.

A system is sign-problem-free in FCIQMC if the stoquastised ground-state energy E_0^{stoq} equals the true ground-state energy E_0. This implies that there are strictly zero annihilations throughout an FCIQMC run when starting the simulation from a single Slater determinant or from a superposition of Slater determinants that already has the correct sign structure. This is because there is no possibility that opposite-sign contributions meet on any Slater determinant. Naively, one would expect that therefore convergence to the correct ground state can be achieved with an arbitrarily low number of walkers with a stochastic error only (that scales as $n^{-\frac{1}{2}}$ with the number of samples n) but with no systematic error. In practice however, this is not the case. Calculations of the aforementioned sign-problem-free systems show that there is a systematic positive bias in the energy estimators for low walker numbers. This bias will be called population control bias which will be justified during the analysis. This bias scales with system size which inhibits calculations of systems with more than approximately 50 sites with affordable hardware.

The goal of this chapter is to show

- that the population control can be corrected for by the application of importance sampling to FCIQMC and the development of an a-posteriori correction method and

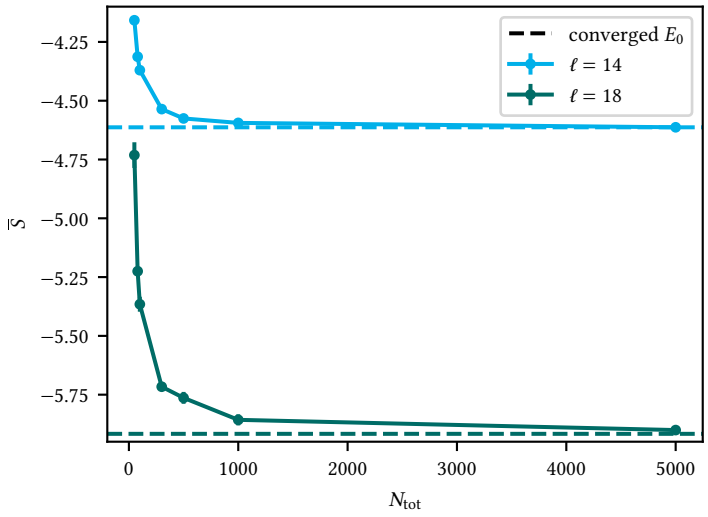


Figure 6.1. Biased average shifts $\overline{S}$ as a function of the total walker number N_{tot} for two Hubbard chains with lengths $\ell = 14$ and 18, respectively, at $U/t = 8$. This is not expected as the systems are sign-problem-free. Unlike the bias caused due to a sign problem, $\overline{S}$ is larger than the respective true ground-state energies E_0.

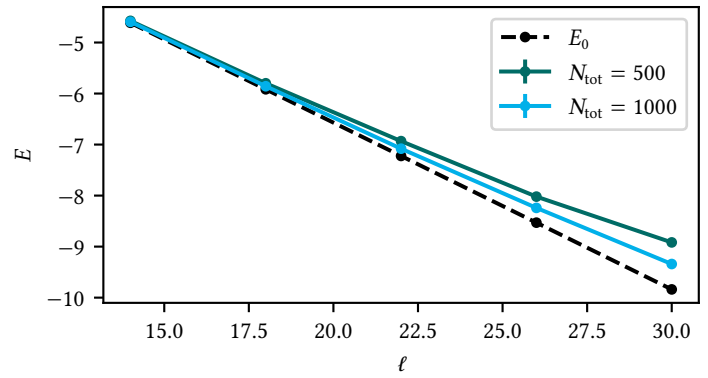


Figure 6.2. Uncorrected ground-state energies obtained with FCIQMC for 1-d Hubbard systems with increasing length ℓ at $U/t = 8$ for $N_{\text{tot}} = 500$ and 1000, respectively. They are compared with converged ground-state energies E_0. Clearly, the bias in the total energies is not scaling linearly.

- that with this correction systems up to 150 sites in the 1-d Hubbard case can be calculated to high numerical accuracy with a highly memory-efficient stochastic description of the wavefunction.

6.1 *Understanding the Bias*

After assessing that there are certain sign-problem-free lattice geometries in the Hubbard and Heisenberg model in chapter 5, it is a natural question to ask whether these kinds of systems can in principle be solved using FCIQMC with arbitrarily low walker numbers. So far, the only known systematic bias to FCIQMC is caused by the sign problem due to unresolved annihilations and ambiguous global sign structure. This however is not the case: Figure 6.1 shows the convergence of the shift and projected energy estimators with respect to walker number for the sign-problem-free 14- and 18-site 1-d Hubbard models at $U/t = 8$ at half-filling. As this chain contains $N_\uparrow = N_\downarrow = 7$ or 9 electrons of each spin species, i.e. both are odd, according to the rules established in section 5.1.1 the systems are sign-problem-free.

Still, the energy estimates for low walker numbers show a positive bias. Figure 6.2 shows the scaling of the bias with increasing system size for sign-problem-free chains for walker numbers $N_{\text{tot}} = 500$ and 1000.

Let us make some empirical observations about this bias:

- There are zero annihilations occuring during the run which confirms the fact that the system is sign-problem-free.
- The bias is independent of τ so it is not a time discretisation issue.
- The bias is positive and larger for the shift energy estimator compared to the projected energy.

How can these empiric observations be understood? Is there a possibility to quantify the error? And finally, is there even a way to correct for it?

To analyse this, let us go back to the FCIQMC master equation (3.2). The master equation governs the evolution of walkers and is correct for a constant shift $S = E_0$ with E_0 being the exact ground-state energy. When calculating ground-state energies and properties however, one is interested in averaging the energy estimators as a single shift or projected energy value does not have significance. Therefore, let us look at the averaged master equation

$$-\frac{\mathrm{d}}{\mathrm{d}\tau}\overline{|\Psi(\tau)\rangle} = \hat{H}\overline{|\Psi(\tau)\rangle} - \overline{S(\tau)\,|\Psi(\tau)\rangle} \tag{6.1}$$

where $\overline{x(\tau)}$ indicates ensemble-averaged properties in imaginary time. Since both $S(\tau)$ and $\Psi(\tau)$ depend on the instantaneous walker populations $N_i(\tau)$, they are correlated and cannot simply be factorised. The non-vanishing covariance needs to be added according to

$$-\frac{\mathrm{d}}{\mathrm{d}\tau}\overline{|\Psi(\tau)\rangle} = \hat{H}\overline{\Psi(\tau)} - \overline{S(\tau)\,|\Psi(\tau)\rangle}$$
$$\text{with } \overline{S(\tau)\,|\Psi(\tau)\rangle} = \overline{S(\tau)}\,\overline{|\Psi(\tau)\rangle} + \mathrm{Cov}\big(S(\tau), |\Psi(\tau)\rangle\big)\,. \tag{6.2}$$

The additional covariance acts like a walker source term and explains the origin of the aforementioned bias in FCIQMC. As the bias is associated with the shift that is necessary to control the walker population, this bias will be called population control bias.

6.1.1 *Population Control Bias in DMC*

Population control biases are not unknown in other QMC methods. This is especially true in DMC where, like in FCIQMC, an effective population control is essential (see section 2.2.2). This population control has been shown

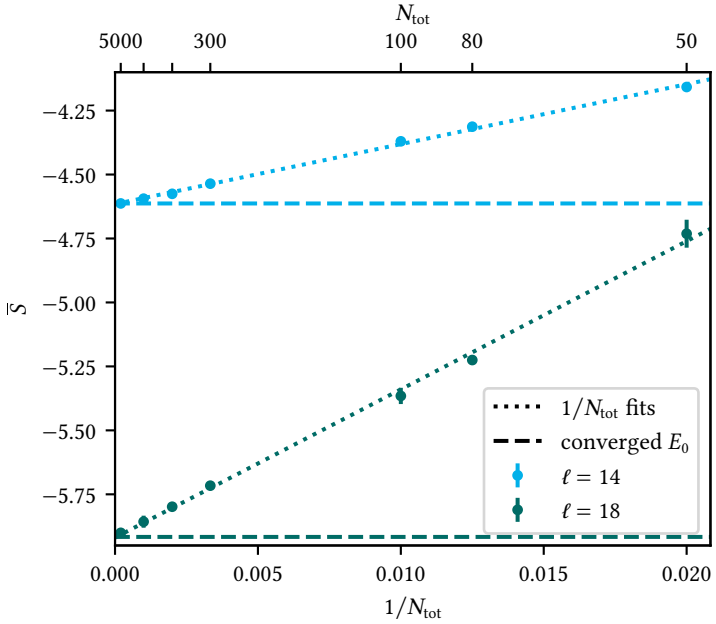


Figure 6.3. Data from figure 6.1 (average shifts $\overline{S}$ of Hubbard chains with lengths $\ell = 14$ and 18 at $U/t = 8$) plotted as a function of the inverse total walker number $1/N_{\text{tot}}$ and fitted accordingly. This confirms that the population control bias in the Hubbard chains follows the estimated $1/N_{\text{tot}}$ scaling from equation (6.7).

to introduce a systematic positive $1/N_{\text{tot}}$ population control bias in the energy estimate (with N_{tot} being the DMC walker population in this section) [187]. This problem becomes especially prominent when trying to tackle a bosonic system where, in the absence of a sign problem, no fixed-node approximation is required [43]. Furthermore, even in model systems like the d-dimensional harmonic oscillator, where the exact form of the ground-state wavefunction is known and therefore no population control is required, a positive bias to the energy estimate has been observed [41]. It can be mitigated but not entirely removed by employing a good-quality trial wavefunction. Also, a reweighting scheme to correct for the population control bias has been developed previously [40]. Unlike the former population control bias, this systematic error is introduced by non-negligible correlation times of walkers caused by the DMC branching process [42].

6.1.2 *Effects and Scaling Behaviour of the Population Control Bias*

When considering the dynamics of the individual walker populations instead of the entire wavefunction like in equation (3.3), one arrives at

$$-\frac{\mathrm{d}}{\mathrm{d}\tau}\overline{N_i(\tau)} = \left[H_{ii} - \frac{\operatorname{Cov}\big(S(\tau), N_i(\tau)\big)}{\overline{N_i(\tau)}} - \overline{S(\tau)}\right]\overline{N_i(\tau)} + \sum_{j\neq i} H_{ij}\overline{N_j(\tau)}\,. \tag{6.3}$$

Clearly, the additional covariance term acts in the diagonal part of the master equation, i.e. the population control bias is rooted in an effective modification of the diagonal matrix elements of the Hamiltonian.

To analyse the scaling behaviour of the bias, the covariance term is expressed more explicitly by inputting the shift update equation (3.8). It directly follows that

$$S(\tau + A\Delta\tau) + \frac{\gamma}{A\Delta\tau} \ln\big(N_{\text{tot}}(\tau + A\Delta\tau)\big) = S(\tau) + \frac{\gamma}{A\Delta\tau} \ln\big(N_{\text{tot}}(\tau)\big) \tag{6.4}$$

and therefore

$$\operatorname{Cov}\left[N_i, S(\tau) + \frac{\gamma}{A\Delta\tau} \ln\big(N_{\text{tot}}(\tau)\big)\right] = 0\,. \tag{6.5}$$

When expanding $\ln N_{\text{tot}}$ around its mean, this leads to

$$\begin{aligned}
\operatorname{Cov}\big(S(\tau), N_i(\tau)\big) &= -\frac{\gamma}{A\Delta\tau} \operatorname{Cov}\Big[\ln\big(N_{\text{tot}}(\tau)\big), N_i(\tau)\Big] \\
&\approx -\frac{\gamma}{A\Delta\tau} \operatorname{Cov}\left[\ln\big(\overline{N_{\text{tot}}}\big) + \frac{N_{\text{tot}}(\tau) - \overline{N_{\text{tot}}}}{\overline{N_{\text{tot}}}}, N_i(\tau)\right] \\
&= -\frac{\gamma}{A\Delta\tau} \frac{\operatorname{Cov}\big(N_{\text{tot}}(\tau), N_i(\tau)\big)}{\overline{N_{\text{tot}}}}\,.
\end{aligned} \tag{6.6}$$

As the diagonal elements are modified according to

$$\begin{aligned}
H'_{ii} &= H_{ii} - \underbrace{\frac{1}{\overline{N_i}} \operatorname{Cov}\big(S(\tau), N_i(\tau)\big)}_{=: b_i} \\
&\approx H_{ii} + \frac{\gamma}{A\Delta\tau} \frac{1}{\overline{N_i}} \frac{\operatorname{Cov}\big(N_{\text{tot}}(\tau), N_i(\tau)\big)}{\overline{N_{\text{tot}}}}\,,
\end{aligned} \tag{6.7}$$

one can see that the shift of the diagonal elements scales as $\overline{N_{\text{tot}}}^{-1}$. If $\operatorname{Cov}\big(N_{\text{tot}}(\tau), N_i(\tau)\big)$ is independent of $\overline{N_{\text{tot}}}$, it means that also the bias of the energy estimators scales as $\overline{N_{\text{tot}}}^{-1}$. In the Hubbard chains considered here, the $\overline{N_{\text{tot}}}^{-1}$ scaling of the shift estimator can be observed in actual simulations in good agreement. This is shown in figure 6.3 for $\ell = 14$ and 18.

Let us now analyse how the implicit modification of the diagonal elements due to population control affects the energy estimators. For this, assume the edge case that the biasing term b_i is the same for all determinants $|D_i\rangle$ ($b_i \equiv b$). In this case, the bias would just be a constant diagonal shift of the Hamiltonian. This means that the sampled wavefunction itself is unbiased, implying that all its associated energy estimators, like the projected energy E_{proj} as defined in equation (3.14), are unbiased. However, the shift energy estimator is still biased according to

$$\overline{S} - \overline{E_{\text{proj}}} = b\,. \tag{6.8}$$

This hints that in general $\overline{S}$ is especially susceptible to the population control bias.

In the case of determinant-dependent covariances, the difference between the average shift and the trial energy according to equation (3.15) can be expressed by projecting equation (6.2) onto a trial wavefunction $|\Psi_\mathrm{t}\rangle$ according to

$$\overline{S} - E_\mathrm{t} = -\frac{\mathrm{Cov}\big(S(\tau), \langle\Psi_\mathrm{t}|\Psi(\tau)\rangle\big)}{\langle\Psi_\mathrm{t}|\overline{\Psi(\tau)}\rangle}\,. \tag{6.9}$$

Since we are mainly dealing with sign-problem-free problems in this chapter, as the systematic biases due to the sign problem are typically masking the much weaker population control bias (see chapter 7), it is possible to define a trial wavefunction $|\Psi^\pm\rangle$ with $\langle D_i|\Psi^\pm\rangle = \pm 1$. The signs are defined according to the predictable signs in the true solution $|\Psi_0\rangle$ which is possible if there is no sign problem. Thus, $\langle\Psi^\pm|\Psi(\tau)\rangle = N_\mathrm{tot}(\tau)$ is easy to calculate and the bias of the shift with respect to $E^\pm = \langle\Psi^\pm|\hat{H}|\Psi\rangle$ is given by

$$\overline{S} - E^\pm = -\frac{\mathrm{Cov}\big(S(\tau), N_\mathrm{tot}(\tau)\big)}{\overline{N_\mathrm{tot}}}\,. \tag{6.10}$$

This is an easy way to correct the maximally biased $\overline{S}$ partly. In the next section, ways to improve the algorithm and correct for the population control bias completely will be presented.

6.2 *Correcting the Bias*

In this section, the two means of correcting the population control bias will be discussed. It consists of

- amending the FCIQMC algorithm by introducing importance sampling using the *Gutzwiller ansatz* and its easy-to-calculate approximation, the *Gutzwiller-like guiding wavefunction,* and
- introducing an a-posteriori way of reweighting the biased wavefunction and correct it and the energy estimators.

The introduction of importance-sampled FCIQMC in this chapter will also lay a foundation for its application to sign-problematic systems in chapters 7 and 8.

6.2.1 *Importance Sampling*

Importance sampling – which can be pictured as increasing MC sampling frequency in regions of importance – is widely used in QMC methods, as its

principles have been introduced in section 2.2. Only recently has importance sampling been applied in FCIQMC in *ab initio* simulations [39]. Also it has been used in Green's Function Monte Carlo (GFMC) which can be seen as a predecessor of FCIQMC [45]. The main ingredient of importance sampling is a guiding wavefunction $|\Psi_g\rangle$ that has to be similar to the true solution. Similarity of $|\Psi_g\rangle$ with $|\Psi_0\rangle$ can be defined in multiple ways. The simplest way is to maximise the overlap $\langle\Psi_g|\Psi_0\rangle$, although this is not possible without the knowledge of $|\Psi_0\rangle$. Another more indirect way would be to choose a wavefunction ansatz that is energy- or variance-optimised with respect to the same Hamiltonian. As discussed in section 2.2.1 however, a well-optimised wavefunction ansatz with respect to energy or variance does not necessarily yield similar wavefunctions with large overlap with the true solution. If $|\Psi_g\rangle$ exactly equals the true solution $|\Psi_0\rangle$, MC sampling of an integral typically leads to a variance of zero (see section 2.2).

Introducing Importance Sampling to FCIQMC

Introducing importance sampling to FCIQMC is conceptually simple. Instead of diagonalising the Hamiltonian matrix $\mathbf{H}$ directly, a similarity-transformed version $\mathbf{H}'$ is treated which is given by

$$\mathbf{H}' = \mathbf{D}^{-1}\mathbf{H}\mathbf{D}\,. \tag{6.11}$$

Similarity transformations leave the spectrum unchanged. Similarity transformations have been studied in FCIQMC both in the Hubbard model in a reciprocal-space basis for $U \leq 4$ [160], in atomic and molecular systems [188, 189], in periodic *ab initio* systems [190], and in the uniform electron gas [191]. In those cases, the similarity transformation is used to factor out correlations in the wavefunction explicitly and treat them analytically. This ansatz is called *transcorrelation* [192–194]. This leads to significantly more compact solutions that reduce the initiator bias in initiator-FCIQMC (see section 3.2.3) at the expense of having to negotiate three-body excitations.

Importance sampling alone, as it will be introduced in the following and will be studied both in the context of the population control bias here as well as with respect to weak-sign-problem systems in chapter 7, does not have

this complication. In these cases, the matrix $\mathbf{D}$ and therefore also $\mathbf{D}^{-1}$ are simple purely diagonal matrices with diagonal entries

$$D_{ii} = \left\langle D_i \middle| \Psi_\mathrm{g} \right\rangle \quad \text{and} \tag{6.12a}$$

$$D_{ii}^{-1} = \frac{1}{\left\langle D_i \middle| \Psi_\mathrm{g} \right\rangle} . \tag{6.12b}$$

This means that the entries of the similarity-transformed Hamiltonian are given by

$$H'_{ij} = \frac{\left\langle D_i \middle| \Psi_\mathrm{g} \right\rangle}{\left\langle D_j \middle| \Psi_\mathrm{g} \right\rangle} H_{ij} \tag{6.13}$$

where H'_{ij} is now the matrix element that is used for spawns from $|D_j\rangle$ to $|D_i\rangle$, i.e.

$$H'_{ij} = \left\langle D_i \middle| \hat{H}' \middle| D_j \right\rangle . \tag{6.14}$$

The right eigenfunctions of $\mathbf{H}$ and $\mathbf{H}'$, which is now an important distinction as $\mathbf{H}'$ is no longer Hermitian, are also related in a simple manner. The coefficients in an FCI expansion of $\mathbf{H}'$ are transformed to

$$C'_i = \left\langle D_i \middle| \Psi_\mathrm{g} \right\rangle C_i . \tag{6.15}$$

Algorithmically, the only change that is necessary in FCIQMC is in the spawning step. The number of walkers ΔN_j spawned onto $|D_j\rangle$ by a walker sitting on $|D_i\rangle$, is scaled by the weight

$$w_{ij} = \frac{\left\langle D_i \middle| \Psi_\mathrm{g} \right\rangle}{\left\langle D_j \middle| \Psi_\mathrm{g} \right\rangle} . \tag{6.16}$$

Because this ratio has to be evaluated at every spawning attempt, it is crucial that the evaluation of the guiding wavefunction is fast to keep the algorithm efficient.

Before considering specific choices for $|\Psi_\mathrm{g}\rangle$ and practical implementations, let us study the effect of importance sampling in FCIQMC by looking at the special case $|\Psi_\mathrm{g}\rangle = |\Psi_0\rangle$. In this case, the transformed matrix elements are given by

$$\begin{aligned} H'_{ij} &= \frac{\sum_k \langle D_i | \Psi_0 \rangle}{\sum_{k'} \left\langle D_j \middle| \Psi_0 \right\rangle} H_{ij} \\ &= \frac{\sum_k C_k \langle D_i | D_k \rangle}{\sum_{k'} C_{k'} \left\langle D_j \middle| D_{k'} \right\rangle} H_{ij} \\ &= \frac{C_i}{C_j} H_{ij} . \end{aligned} \tag{6.17}$$

This leads to the fact that the column sums of $\mathbf{H}'$ are

$$\sum_i H'_{ij} = \frac{\sum_i C_i H_{ij}}{C_j} = E_0 \tag{6.18}$$

which is exactly the expression for the projected energy with $|D_j\rangle$ as reference determinant according to equation (3.14). That makes $\mathbf{H}'$ a *column-stochastic matrix*. Also, according to equation (6.15) the FCI coefficients in this case are given by

$$C'_i = C_i^2\,, \tag{6.19}$$

i.e. the wavefunction is compactified. The compactification leads to the fact that important determinants are more permanently occupied for low walker numbers and are less often stochastically rounded. Also, according to the FCIQMC master equation equation (3.3) the average total contribution of $|D_i\rangle$ is given by

$$\sum_i 1 + \Delta\tau\left(H'_{ij} - S\right) = 1 + \Delta\tau\left(E_0 - S\right) \tag{6.20}$$

which is a constant. This removes the need for population control for permanently occupied determinants, therefore leading to $\mathrm{Cov}\left(S(\tau), N_i(\tau)\right) = 0$ and removing the population control bias. In actual simulations, not all determinants will be permanently occupied due to the stochastic rounding step. Therefore, there is still the need for population control and and a non-zero covariance, even if the exact solution would be used as a guiding wavefunction. Therefore, an a-posteriori reweighting procedure will be presented in the next section. However, the numerical results for large systems, which will be presented in section 6.3, will show that in practice the bias due to the remaining covariance is small.

Gutzwiller and Gutzwiller-like Guiding Wavefunction

Especially in the strongly correlated case at around $U/t = 8$, only approximations to the true solution are known and can be used. This is particularly true when considering the practical constraint that the guiding wavefunction has to be evaluated often and therefore efficiently.

One of the simplest approaches to the true ground-state solution in the Hubbard model is the Gutzwiller wavefunction

$$|\Psi_\mathrm{G}\rangle = \prod_{i=1}^{N_s} \exp\left(-gU\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}\right)|\Psi_0(U/t=0)\rangle \tag{6.21}$$

where $|\Psi_0(U/t = 0)\rangle$ is the ground-state solution of the non-interacting Hubbard model [34]. In words, the Gutzwiller ansatz equals the non-interacting solution where real-space configurations with double occupancies are exponentially suppressed. As we have seen in section 4.1, the non-interacting Hubbard Hamiltonian at $U/t = 0$ can be diagonalised analytically using the transformation from site orbitals in real space to delocalised orbitals in reciprocal space according to equation (4.9). To calculate the overlaps $\langle D_i|\Psi_\mathrm{G}\rangle$ with real-space basis functions $|D_i\rangle$, which is needed to calculate the weights w_{ij}, the non-interacting solution in the reciprocal-space basis

$$|\Psi_0(U/t = 0)\rangle = \prod_{\sigma=\{\uparrow,\downarrow\}} \prod_{m=1}^{N_\sigma} \hat{c}^\dagger_{\boldsymbol{k}_{m\sigma}\sigma} \, |\rangle \tag{6.22}$$

needs to be expressed in the real-space basis [70]. Here, the $\boldsymbol{k}_{m\sigma}$ vectors number the energetically lowest reciprocal-space spatial orbitals. N_σ denotes the number of σ-spin electrons. Thus, we write the non-interacting ground state as

$$|\Psi_0(U/t = 0)\rangle = \prod_{\sigma=\{\uparrow,\downarrow\}} \prod_{m=1}^{N_\sigma} \left(\sum_{i=1}^{N_\mathrm{s}} U^\sigma_{im} \hat{c}^\dagger_{i\sigma} \right) |\rangle \tag{6.23}$$

where, according to equation (4.9),

$$U^\sigma_{im} = L^{-\frac{1}{2}} \exp\left(\mathrm{i}\boldsymbol{k}_{m\sigma} \cdot \boldsymbol{R}_i\right) \tag{6.24}$$

are the matrix elements of two matrices $\mathbf{U}^\sigma$ of size $N_\mathrm{s} \times N_\sigma$. $\boldsymbol{R}_i$ are the lattice site positions in real space. The overlap of a real-space Slater determinant $|D\rangle$ with the non-interacting ground-state solution is then given by

$$\begin{aligned} \langle D \,|\, \Psi_0(U/t = 0)\rangle &= \prod_{\sigma=\{\uparrow,\downarrow\}} \langle | \prod_{n=1}^{N_\sigma} \hat{c}_{j_{N\sigma}\sigma} \prod_{m=1}^{n_\sigma} \left(\sum_{i=1}^{N_\mathrm{s}} U_{im} \hat{c}^\dagger_{i\sigma} \right) |\rangle \\ &= \prod_{\sigma=\{\uparrow,\downarrow\}} \det \tilde{\mathbf{U}}^\sigma \, . \end{aligned} \tag{6.25}$$

$j_{n\sigma}$ are the lattice sites of $|D\rangle$ occupied by an electron with spin σ. $\tilde{\mathbf{U}}^\sigma$ are two $N_\sigma \times N_\sigma$ submatrices of $\mathbf{U}$ where only the rows of the matrix corresponding to the lattice sites occupied with σ-spin electrons are selected. The matrix elements are given by

$$\tilde{U}^\sigma_{nm} = U^\sigma_{j_{n\sigma}m} \, . \tag{6.26}$$

The alternating sign of the determinant sign is caused by the commutation relations of fermionic operators.

Computationally, when evaluating the importance-sampling weights w_{ij}, the overlap $\langle D_j|\Psi_G\rangle$ for every occupied determinant, i.e. the denominator of w_{ij}, only needs to be calculated once and can be stored for the entire lifetime of $|D_j\rangle$. The numerator $\langle D_i|\Psi_G\rangle$ however needs to evaluted at every spawning attempt. The evaluation of a determinant of an $N\times N$-sized matrix via LU decomposition through Gaussian elimination scales as $\mathcal{O}\left(N^3\right)$ [195].[12] Since at every spawning attempt an $N_\uparrow \times N_\uparrow$ and an $N_\downarrow \times N_\downarrow$ matrix needs to be evaluated, the algorithm now formally scales as $\mathcal{O}\left[\max\left(N_\uparrow^3, N_\downarrow^3\right)\right]$, in addtition to the linear scaling with total walker number.

To avoid this computational overhead, that can be especially problematic for large systems with many electrons, I will introduce the Gutzwiller-like wavefunction. It has the simple form

$$|\Psi_{GL}\rangle = \sum_i \exp\left(-gH_{ii}\right)|D_i\rangle \ . \tag{6.27}$$

It is similar to the full Gutzwiller wavefunction $|\Psi_G\rangle$ but instead projecting out the double occupancies in $|\Psi_{HF}\rangle$, the uniform wavefunction

$$|\Psi_{\text{uniform}}\rangle = \sum_i |D_i\rangle \tag{6.28}$$

is used. It is easy to see that the overlaps

$$\langle D_i\,|\,\Psi_{GL}\rangle = \exp\left(-gH_{ii}\right) \tag{6.29}$$

solely depend on the diagonal Hamiltonian matrix element of $|D_i\rangle$. The weights are thus easily calculated via

$$\begin{aligned} w_{ij} &= \frac{\langle D_i|\Psi_{GL}\rangle}{\langle D_j|\Psi_{GL}\rangle} \\ &= \exp\left[-g\left(H_{ii}-H_{jj}\right)\right] \\ &= \exp\left[-\frac{gU}{t}\left(d_i-d_j\right)\right] \end{aligned} \tag{6.30}$$

where d_k is the number of doubly occupied sites in determinant $|D_k\rangle$. Since the calculation of diagonal elements of occupied determinants is anyway necessary for the death/cloning step, the computational complexity only increases for excitation attempts to determinants $|D_i\rangle$ that are rejected in the spawning step. Also, the calculation of diagonal elements in the Hubbard model is inexpensive since it only requires the number of double occupancies,

12 There are algorithms to calculate determinants of size-N square matrices that scale as efficiently as $\mathcal{O}\left(N^{2.376}\right)$ [196]. They are impractical to implement however. Therefore, for the purpose of importance sampling in `NECI`, the usual procedure using LU decomposition is used.)

so applying importance sampling with the Gutzwiller-like guiding wavefunction can be implemented with virtually no overhead.

Numerical Experiment

Figure 6.4 shows the significantly improved convergence both when using the easy-to-evaluate Gutzwiller-like guiding wavefunction $|\Psi_{\mathrm{GL}}\rangle$ and the full Gutzwiller wavefunction $|\Psi_{\mathrm{G}}\rangle$, both with Gutzwiller paramter $g = 0.15$. For example, for $N_{\mathrm{tot}} = 100$ the deviation of the average shift $\overline{S}$ is reduced from $\Delta E = 0.58(3)$ to $\Delta E^{\mathrm{GL}} = 0.18(1)$ with Gutzwiller-like importance sampling, so the bias is more than halved. For the full Gutzwiller wavefunction it is even reduced to $\Delta E^{\mathrm{G}} = 0.078(7)$. Still, for very low walker numbers there is still some bias left with both considered wavefunctions. That is why there is still the need for an additional way to remove the remaining bias. This will be derived in section 6.2.2.

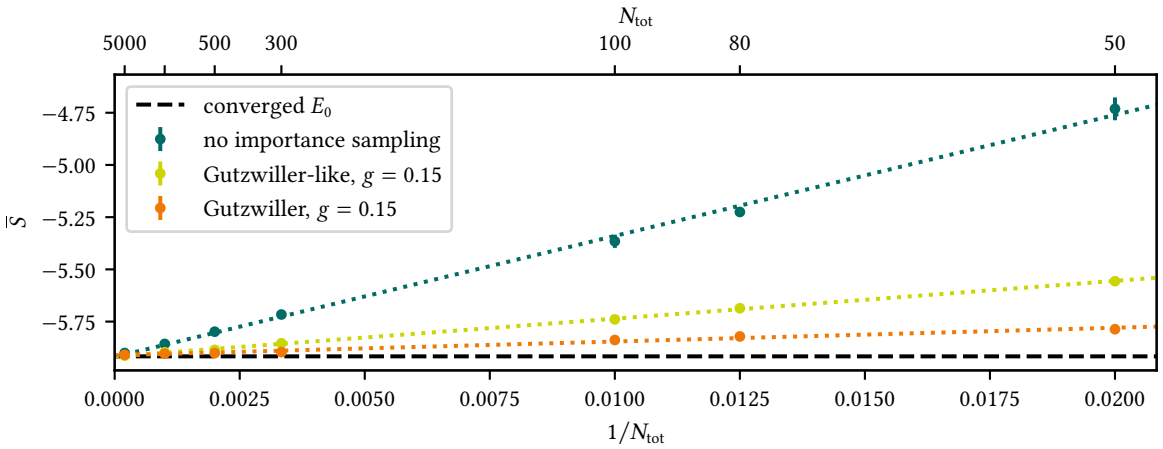


Figure 6.4. Improved convergence of the average shift estimator $\overline{S}$ to the exact ground-state energy E_0 with respect to total walker number N_{tot} for the 18-site Hubbard chain at $U/t = 8$ with importance sampling. Both the full Gutzwiller and the Gutzwiller-like guiding wavefunction reduce the population control bias significantly. The full Gutzwiller performs even better than the Gutzwiller-like, is more expensive to evaluate however.

Figure 6.5 compares the computational scaling of importance sampling with the full Gutzwiller and the Gutzwiller-like guiding wavefunction. It shows the actual time per iteration T of a `NECI` FCIQMC calculation of half-filled Hubbard chains with increasing chain length ℓ. The T for the Gutzwiller-like calcuation are fitted with a linear function. Due to the additional $\mathcal{O}\left(N^3\right)$ scaling, the Gutzwiller iteration times at half-filling are fitted with the polynomial

$$T(\ell) = a\ell^3 + b\ell + c\,. \tag{6.31}$$

guiding wavefunction	T [s]	E_0/N_{sites}
Gutzwiller-like, $g = 0.15$, $N_{\text{tot}} = 3 \times 10^7$	1.3492(2)	−0.327 54(1)
Gutzwiller, $g = 0.15$, $N_{\text{tot}} = 4 \times 10^5$	0.9282(1)	−0.327 52(2)

Table 6.1. Comparison of ground-state energies per site E_0/N_{sites} obtained with importance-sampled FCIQMC using the Gutzwiller-like and the full Gutzwiller guiding wavefunction, respectively, for roughly equal iteration times T on 16 computing nodes with 320 cores in total. The amount the population control bias is corrected by is almost the same.

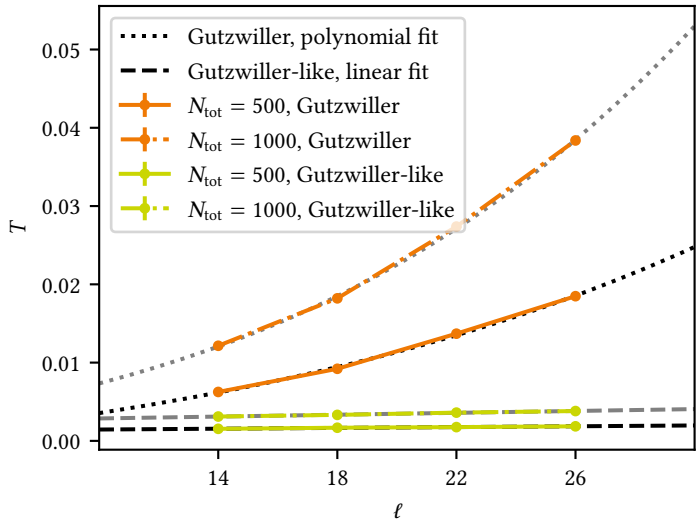


Figure 6.5. Time per iteration T on a fixed computational resource (single processor) as a function of the Hubbard chain length ℓ at $U/t = 8$ for the full Gutzwiller and the Gutzwiller-like guiding wavefunction in comparison for half-filled systems for $N_{\text{tot}} = 500$ and 1000. For the Gutzwiller-like guiding wavefunction, T only grows linearly with a small prefactor when increasing ℓ. This is the expected behaviour as there are some operations in the `NECI` implementation of FCIQMC that linearly depend on the length of the bit strings that encode the Slater determinants. The full Gutzwiller guiding wavefunction requires the calculation of the determinants of two $N_{\text{el}}/2 \times N_{\text{el}}/2$ matrices for each spawning attempt, thus it was fitted with a third-order polynomial in ℓ with vanishing second-order term.

For $\ell = 102$, one of the system sizes we want to study in more detail later on, the fits roughly predict a 74-fold increase when moving from the Gutzwiller-like to the full Gutzwiller guiding wavefunction.

To test which of the guiding wavefunctions to choose for the large-scale calculations, a 3×10^7-walker calculation with the Gutzwiller-like guiding wavefunction is conducted. It is compared to a 4×10^5-walker calculation with the full Gutzwiller guiding wavefunction. 4×10^5 walkers roughly equals a 74-fold reduction of the original 3×10^7 walkers such that the expected iteration times are approximately equal. The results of the test calculations are shown in table 6.1. The full Gutzwiller wavefunction has a slightly lower T as one would expect from the third-order polynomial fit in figure 6.5. The obtained ground-state energies agree within statistical errorbars. Therefore, one can conclude that the two guiding wavefunctions roughly perform equally. For simplicity, in all other large-scale calculations the Gutzwiller-like guiding wavefunction will be used.

Optimisation of the Gutzwiller Parameter

There are various ways to optimise the parameter g in the wavefunction ansatzes to achieve this. Two examples are the following:

- VMC, as introduced in section 2.2.1, can be used to optimise the wavefunction ansatz, either with respect to the variational energy or its variance.

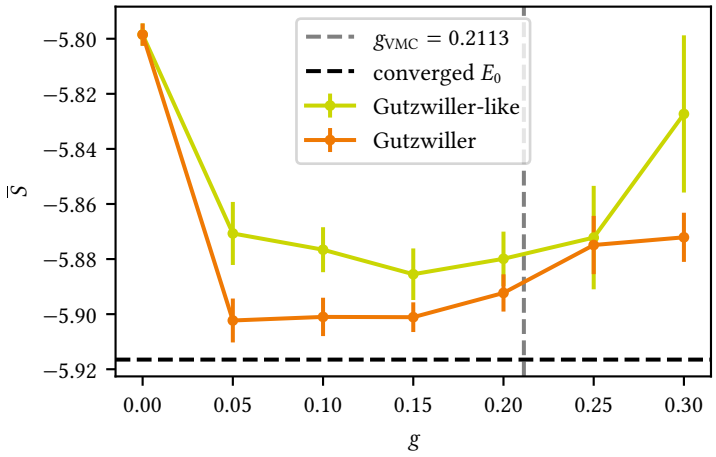


Figure 6.6. Average shift $\overline{S}$ as a function of the Gutzwiller parameter g using the Gutzwiller-like guiding wavefunction $|\Psi_{\mathrm{GL}}\rangle$ and the full Gutzwiller wavefunction $|\Psi_{\mathrm{G}}\rangle$ for an 18-site Hubbard chain at $U/t = 8$ with $N_{\mathrm{tot}} = 500$. The population control bias as well as the standard error of the estimates (shown as errorbars) is minimal roughly between $g = 0.05$ and 0.25 for the Gutzwiller-like and roughly between $g = 0.05$ and $g = 0.20$ for the full Gutzwiller wavefunction. This indicates that the correction is rather insensitive to the choice of the g parameter. Also shown is a VMC-optimised value for g for the full Gutzwiller wavefunction $|\Psi_{\mathrm{G}}\rangle$ which is located at the top end of the plateau region, slightly overestimating the optimal value for g. This could be due to the fact that the energy-optimised ansatz does not necessarily yield the best accuracy in the wavefunction.

- A simple bootstrapping procedure can also be used, i.e. multiple low-resource calculations at low walker numbers are performed for various values of g. The value that yields the lowest energy, i.e. reduces the population control bias most, is then chosen to perform a larger-scale calculation with.

As shown paradigmatically in figure 6.6 for the 18-site Hubbard chain for $N_{\mathrm{tot}} = 500$, there is an extended plateau region where both the Gutzwiller-like and the full Gutzwiller guiding wavefunction minimise the population control bias by roughly the same amount. Therefore, there is no necessity for precisely optimising g. The VMC energy optimisation of the true Gutzwiller wavefunction yields an optimised value that is located at the top end of the plateau region. The optimal g is slightly overestimated. As outlined already in section 2.2.1, this could be caused by the fact that a wavefunction ansatz optimised with respect to the variational energy does not necessarily imply maximal similarity between the optimised and the true ground-state wavefunction.

6.2.2 *A-Posteriori reweighting Procedure*

As shown previously, the population control bias can be significantly reduced by using importance sampling. Since only approximate wavefunctions can be used efficiently, there is still a remaining bias however which can make converging large sign-problem-free systems with affordable walker numbers impossible. Following the argument from section 6.2.1, the need for population control is entirely removed only if no stochastic rounding step is necessary. This means that the population control bias is not entirely removed if not the entire Hilbert space is occupied with walkers, even when using the exact solution as the guiding wavefunction.

Equation (6.10) has shown a simple way to correct the maximally biased shift with respect to $E^{\pm}$ by calculating the covariance between shift and total walker number. This will be extended to a correction not only to $E^{\pm}$ but to the exact ground-state energy E_0. It is possible to do this in an a-posteriori way, i.e. finish the (importance-sampled) calculation and then use the logged progression of $S(\tau)$ and N_{tot} as a function of τ. A similar reweighting scheme has been applied to DMC previously [40].

Correcting the Wavefunction

The starting point of the reweighting procedure is the imaginary-time propagation. Instead of the actually used linear propagator, the derivation will be based on the exponential propagation

$$|\Psi(\tau + \Delta\tau)\rangle = \exp\Big[-\Delta\tau\big(\hat{H} - S(\tau)\big)\Big]\,|\Psi(\tau)\rangle \tag{6.32}$$

Although this is the exact propagator, this makes the correction equations a small approximation. The propagator can be straightforwardly split into a constant part and a fluctuating part according to

$$\exp\Big[-\Delta\tau\big(\hat{H} - S(\tau)\big)\Big] = \exp\Big[-\Delta\tau\big(\hat{H} - C\big)\Big]\exp\Big[-\Delta\tau\big(C - S(\tau)\big)\Big] \tag{6.33}$$

Inserting this into equation (6.32) leads to

$$\exp\Big[-\Delta\tau\big(C - S(\tau)\big)\Big]\,|\Psi(\tau + \Delta\tau)\rangle = \exp\Big[-\Delta\tau\big(\hat{H} - C\big)\Big]\,|\Psi(\tau)\rangle\ . \tag{6.34}$$

Averaging both sides leads to

$$\overline{\underbrace{\exp\Big[-\Delta\tau\big(C - S(\tau)\big)\Big]}_{=:X_C(\tau)}\,|\Psi(\tau + \Delta\tau)\rangle} = \exp\Big[-\Delta\tau\big(\hat{H} - C\big)\Big]\overline{|\Psi(\tau)\rangle} \tag{6.35}$$

since the exponential term on the right-hand side is a constant term by construction. The left-hand side contains the reweighting factor $X_C(\tau)$ such that $X_C(\tau)\,|\Psi(\tau + \Delta\tau)\rangle$ is the unbiased evolution of $|\Psi(\tau)\rangle$. The ground-state

solution is obtained by repeated application of the propagator. This leads us to

$$\begin{aligned}\overline{\exp\left[-\Delta\tau\sum_{p=1}^{n}\big(C-S(\tau-p\Delta\tau)\big)\right]|\Psi(\tau)\rangle} &= \exp\Big[-\Delta\tau\big(\hat{H}-C\big)\Big]^n\overline{|\psi(\tau-n\Delta\tau)\rangle}\,,\\ \underbrace{\overline{\prod_{p=1}^{n}X_C(\tau-p\Delta\tau)\,|\Psi(\tau)\rangle}}_{=:W_{C,n}(\tau)} &= \exp\Big[-\Delta\tau\big(\hat{H}-C\big)\Big]^n\overline{|\psi(\tau-n\Delta\tau)\rangle}\,.\end{aligned} \tag{6.36}$$

That means that the true ground state can be obtained by repeatedly applying the reweighting factors to the sampled wavefunction according to

$$\lim_{n\to\infty}\overline{W_{C,n}\,|\Psi(\tau)\rangle} = |\Psi_0\rangle\ . \tag{6.37}$$

In practice, of course it is not possible to go to infinite correction order n so a large but finite n will be used as an approximation.

Correcting the Energy Estimators

Trial energies according to equation (3.15) can be obtained in a simple manner when the reweighted wavefunction is known. One simply projects onto the corrected wavefunction obtained using equation (6.37) according to

$$E_{\mathrm{t}}^{\mathrm{corr}} = \frac{\sum_{p=1}^{L}W_{C,n}(\tau_p)\left\langle\Psi^{\mathrm{t}}\middle|\hat{H}\middle|\Psi(\tau_p)\right\rangle}{\sum_{p=1}^{L}W_{C,n}(\tau_p)\left\langle\Psi^{\mathrm{t}}\middle|\Psi(\tau_p)\right\rangle}\,. \tag{6.38}$$

As mentioned before, trial energies and especially the projected energy are problematic to calculate in large systems with low walker populations. For large systems, the Hilbert space cannot be populated by enough walkers such that there is a permanently occupied reference determinant or trial space which would allow for the calculation of a projected or trial energy, respectively. However, the shift can still be used in these cases. Therefore, it is necessary to find a way to derive a corrected shift S^{corr} from the weights $W_{C,n}$.

The shift is a special case of what is called a growth estimator because it measures the tendency of walkers to grow in order to use it for population control. A general growth energy estimator projected onto a trial wavefunction $|\Psi_{\mathrm{t}}\rangle$ is given by

$$E_{\mathrm{growth}} = C - \frac{1}{\Delta\tau}\ln\left[\frac{\left\langle\Psi_{\mathrm{t}}\middle|\exp\Big[-\Delta\tau\big(\hat{H}-C\big)\Big]\middle|\Psi_0\right\rangle}{\langle\Psi_{\mathrm{t}}|\Psi_0\rangle}\right] \tag{6.39}$$

It can be derived from the exponential operator applied onto the exact solution like

$$\exp\Big[-\Delta\tau\big(\hat{H}-C\big)\Big]\,|\Psi_0\rangle = \exp\Big[-\Delta\tau\big(E_0-C\big)\Big]\,|\Psi_0\rangle \tag{6.40}$$

and subsequently projecting the equation onto $|\Psi_\mathrm{t}\rangle$. Inserting the reweighted FCIQMC-sampled wavefunction from equation (6.37) leads to

$$E_\mathrm{growth} = C - \frac{1}{\Delta\tau}\log\left[\frac{\sum_{p=1}^{L} W_{C,n+1}(\tau_{p+1})\left\langle\Psi_\mathrm{t}\middle|\Psi(\tau_{p+1})\right\rangle}{\sum_{p=1}^{L} W_{C,n}(\tau_p)\left\langle\Psi_\mathrm{t}\middle|\Psi(\tau_p)\right\rangle}\right]. \tag{6.41}$$

To make this general growth estimator usable to correct the average shift $\overline{S}$, two choices need to be made: the value of C and the trial wavefunction $|\Psi_\mathrm{t}\rangle$.

If C is chosen to be $\overline{S}$, the logarithmic term in equation (6.41) directly yields the corrections to the shift estimator. Choosing $C = \overline{S}$ also minimises the variance of the weight factors as their squared distance to the sampled shift values $S(\tau_p)$ are reduced.

About the choice of the trial wavefunction, the same argument as in the correction of the shift to $E^\pm$ from equation (6.10) can be used: In sign-problem-free systems, it is possible to predict the sign structure of the solution. In this case, one can project onto $|\Psi^\pm\rangle$. Then, $\langle\Psi_\mathrm{t}|\Psi(\tau_p)\rangle$ simply becomes $N_\mathrm{tot}(\tau_p)$.

Inserting both into equation (6.41) ultimately leads to

$$S^\mathrm{corr} = \overline{S} - \frac{1}{\Delta\tau}\left[\frac{\sum_{p=1}^{L} W_{\overline{S},n+1}(\tau_{p+1})N_\mathrm{tot}(\tau_{p+1})}{\sum_{p=1}^{L} W_{\overline{S},n}(\tau_p)N_\mathrm{tot}(\tau_p)}\right]. \tag{6.42}$$

Clearly, to calculate S^corr, only the trajectory of the total walker number N_tot as the weights $W_{C,n}$ as a function of τ needs to be known. $W_{C,n}$ itself, as defined in equation (6.36), only depends on the trajectory of $S(\tau)$. This allows for the correction of the shift by analysing global simulation parameters alone which does not increase the amount of data that needs to be written out during an FCIQMC run.

The correction procedure has also introduced an additional technical parameter, the correction order n. n introduces a tradeoff between how much the population control bias is corrected (smaller bias) and the statistical noise of S^corr (larger variance). Small n limit the correction as, according to equation (6.37), the true ground-state only emerges in the infinite-n limit. Large n effectively undoes the effects of population control, thus increasing the fluctuations of the reweighted wavefunction. A larger total integration

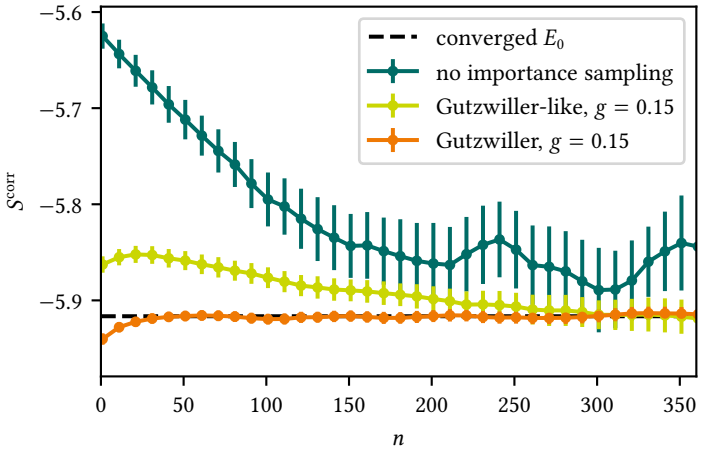


Figure 6.7. Corrected average shifts S^{corr} as a function of the correction order n for the $N_{\mathrm{tot}} = 100$ calculations of the half-filled 18-site Hubbard chain at $U/t = 8$ (for convergence of $\overline{S}$ see figure 6.4). The three cases no, Gutzwiller-like, and full Gutzwiller importance sampling are shown. With no importance sampling, it is not possible to reach an accurate estimate of the ground-state energy as the fluctuations in the weights become large. With Gutzwiller-like importance sampling, the exact ground-state energy is obtained for $n \gtrsim 300$. With the more exact full Gutzwiller importance sampling, the exact ground-state energy with even smaller standard deviation is reached for $n \gtrsim 40$ already.

time $\tau_{\mathrm{tot}} = L\Delta\tau$ also improves the ability to remove the bias as more terms in the summations in equation (6.42) can be taken into account.

6.2.3 *Numerical Example*

Figure 6.7 shows an example of the a-posterori correction applied to a half-filled 18-site 1-d Hubbard model in simulations with $N_{\mathrm{tot}} = 100$ walkers. The corrected shift estimator is plotted as a function of the correction order n. Figure 6.8 shows the convergence of the corrected shift as a function of the inverse walker number for different correction orders for the same system. Despite the fact that in theory the a-posteriori correction is able to remove the population control bias entirely for arbitrarily low walker numbers and large shift–wavefunction covariances, in practice this is not possible without attaining strongly fluctuating weights and slow convergence with respect to n. With no importance sampling at low walker numbers, one would require very long total integration times to obtain a correct estimate of the ground-state energy with acceptable errorbars. This underlines the necessity for importance sampling. Using a Gutzwiller-like guiding wavefunction, with $N_{\mathrm{tot}} = 100$ convergence to the exact ground-state energy can be obtained for $n \gtrsim 300$. With the higher-quality full Gutzwiller guiding wavefunction, the exact ground-state energy can be obtained for $n \gtrsim 40$ already.

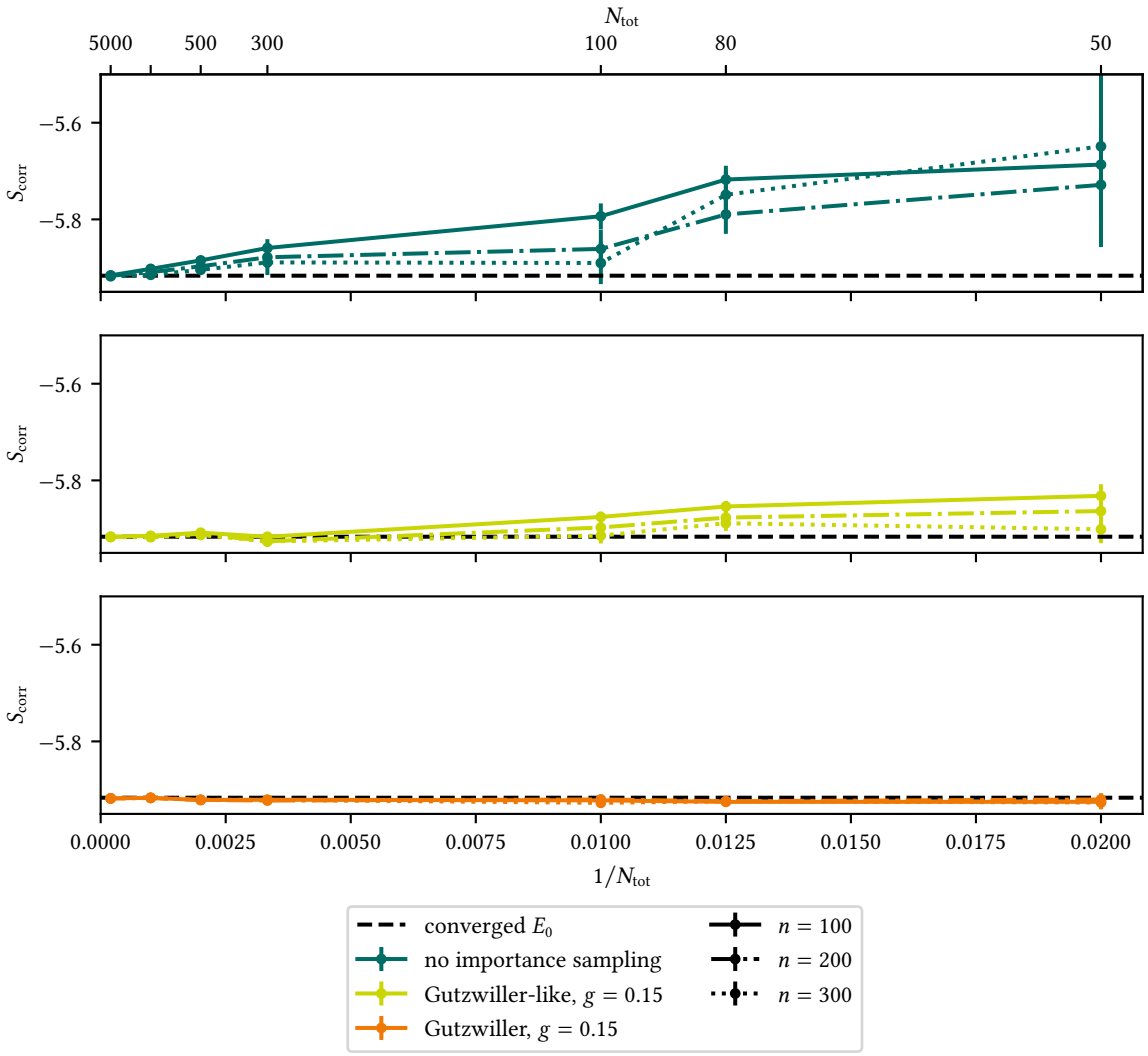


Figure 6.8. Convergence of the corrected shift S_{corr} after the a-posteriori correction with respect to the inverse total walker number $1/N_{\text{tot}}$ for different correction orders n with no, Gutzwiller-like, and full Gutzwiller importance sampling for the half-filled 18-site Hubbard chain at $U/t = 8$. The performance of the a-posteriori correction is clearly improved when using more accurate guiding wavefunctions.

Table 6.2. Results showing the correction of the population control bias in large-scale one-dimensional Hubbard models without a sign problem [186]. "FCIQMC (original)" indicates a plain non-initiator calculation without any corrections where the results strongly deviate from the reference DMRG and Bethe ansatz results. "FCIQMC (guided)" implies that importance sampling with the Gutzwiller-like factor has been used which brings the energy estimates already within two to three standard deviations of the reference results. "FCIQMC (corrected)" indicates that an a-posteriori correction on top of the importance sampling has been performed, yielding highly accurate results. Technical details for the FCIQMC calculations are given in table 6.4. Again, DMRG calculations were conducted with up to $M = 6000$ until there was convergence within $E_0^{\mathrm{DMRG}}/N_{\mathrm{sites}} = 1 \times 10^{-4}$. Chain lengths of ∞ indicate calculations in the thermodynamic limit. M_s is the spin-projection quantum number.

system	sites	method	energy/site (pbc)	energy/site (obc)
$U/t = 4$ half-filling	102	DMRG	−0.573 79	−0.570 13
		FCIQMC (original)	−0.570 76(40)	−0.566 80(50)
		FCIQMC (guided)	−0.573 71(8)	−0.570 11(7)
		FCIQMC (corrected)	−0.573 75(8)	−0.570 17(9)
	∞	Bethe ansatz	−0.573 73	
$U/t = 8$ half-filling	102	DMRG	−0.327 57	−0.325 50
		FCIQMC (original)	−0.322 99(49)	−0.321 19(43)
		FCIQMC (guided)	−0.327 54(1)	−0.325 48(2)
		FCIQMC (corrected)	−0.327 55(3)	−0.325 49(3)
	150	DMRG	−0.327 55	
		FCIQMC (original)	−0.302 54(154)	
		FCIQMC (guided)	−0.327 39(6)	
		FCIQMC (corrected)	−0.327 54(9)	
	∞	Bethe ansatz	−0.327 53	
$U/t = 8$ 4 holes ($M_s = 0$)	102	DMRG	−0.392 29	−0.390 04
		FCIQMC (original)	−0.388 52(32)	−0.387 09(39)
		FCIQMC (guided)	−0.392 28(3)	−0.390 02(2)
		FCIQMC (corrected)	−0.392 29(3)	−0.390 03(2)

6.3 *Applications*

In the following, I will demonstrate the effectiveness of both importance sampling and the a-posteriori correction on top in correcting the population control bias in large sign-problem-free systems and large sign-problematic systems with non-size-extensive sign problems. The results will be compared with DMRG and the analytical Bethe ansatz in the thermodynamic limit.

6.3.1 *One-Dimensional Half-Filled Hubbard Model*

Table 6.2 shows numerical results for some large sign-problem-free paradigmatic systems. The ground-state energies for half-filled Hubbard chains with 102 and 150 sites, respectively, as well the 102-site system with four holes are given. The results are both given for open and periodic boundary conditions. These half-filled systems have $N_\uparrow = N_\downarrow = 51$ and 75 electrons of each spin species, respectively. For the four-hole system, there are

Table 6.3. Results showing the correction of the population control bias in large-scale one-dimensional Hubbard models with one hole. The systems with open boundary conditions (obc) are sign-problem-free. The systems with periodic boundary conditions (pbc) in principle show a sign problem, however the sign problem is not size-extensive. Therefore, also these systems are calculable with importance-sampled FCIQMC and the results agree well with the DMRG benchmarks. "FCIQMC (guided)" again indicates that importance sampling with the Gutzwiller-like factor has been used which brings the energy estimates already within two to three standard deviations of the reference results. "FCIQMC (corrected)" indicates the a posteriori correction on top of the importance sampling, yielding highly accurate results. Technical details for the FCIQMC calculations are given in table 6.4. Again, DMRG calculations were conducted with up to $M = 6000$ until there was convergence within $E_0^{\mathrm{DMRG}}/N_{\mathrm{sites}} = 1 \times 10^{-4}$. M_s is the spin-projection quantum number.

system	sites	method	energy/site (pbc)	energy/site (obc)
$U/t = 8$ 1 hole ($M_s = 1/2$)	102	DMRG	−0.343 73	−0.351 52
		FCIQMC (guided)	−0.343 76(2)	−0.351 68(2)
		FCIQMC (corrected)	−0.343 77(1)	−0.351 68(3)
$U/t = 16$ 1 hole ($M_s = 1/2$)	102	DMRG	−0.188 59	−0.187 54
		FCIQMC (guided)	−0.188 60(1)	−0.187 57(2)
		FCIQMC (corrected)	−0.188 60(1)	−0.187 57(1)

$N_\uparrow = N_\downarrow = 49$ electrons. According to the rules established in chapter 5, these are sign-problem-free in a real-space basis for all U/t for both open and periodic boundary conditions. Results are given for Hubbard on-site interaction parameters $U/t = 4$ and 8 which both lie in the intermediate interaction regime. This regime is challenging because neither the solutions in the $U/t \to \infty$ limit – where the real-space basis diagonalises the Hamiltonian – nor the ones in the $U/t = 0$ limit – where the reciprocal-space basis diagonalises the Hamiltonian – are good approximations and the wavefunctions are highly spread-out in both bases.

Table 6.3 shows the results for 102-site Hubbard chains at $U/t = 8$ and 16, this time each with one hole. While according to section 5.1.1 the chains with open boundary conditions do not have a sign problem in all configurations, according to section 5.2.1 the periodic one-hole systems have one but it is not size-extensive. Therefore, for systems close to the thermodynamic limit both boundary conditions can be converged in good agreement with the DMRG benchmarking results.

The results for the half-filled are benchmarked with the analytical solution of the Bethe ansatz in the thermodynamic limit. The ground-state energy per site for the 1-d Hubbard model can be obtained via

$$\frac{E_0^{\mathrm{Bethe}}}{N_{\mathrm{sites}}} = -4t \int_0^\infty \mathrm{d}\omega \frac{J_0(\omega)J_1(\omega)}{1 + \exp\left(\frac{\omega U}{2t}\right)} \tag{6.43}$$

where J_n are the n-th order Bessel functions of first kind [137, 197].

system	sites	$N_{\text{tot}}/10^6$	g	n
$U/t = 4$, half-filling	102	50	0.17	2560
$U/t = 8$, half-filling	102	30	0.15	5120
	150	50	0.15	5120
$U/t = 8$, 4 holes ($M_s = 0$)	102	30	0.15	2560
$U/t = 8$, 1 hole ($M_s = 1/2$)	102	50	0.17	2560
$U/t = 16$, 1 hole ($M_s = 1/2$)	102	50	0.17	2560

Table 6.4. Technical details for the FCIQMC calculations of the large-scale one-dimensional Hubbard model from table 6.2 [186]. N_{tot} is the number of walkers, g is the Gutzwiller parameter, and n is the number of terms in the population control bias correction. Listed are the number of walkers N_{tot}, the Gutzwiller parameter g and the number of terms n in the expansion of the a-posteriori correction. g is chosen within the plateau region obtained in the bootstrapping optimisation.

All systems were also benchmarked using DMRG. As described in section 2.3, DMRG is an ideal solver for one-dimensional systems as there is a unambiguous site ordering – neighbouring sites are placed next to one another – such that the locality can be exploited. This leads to low entanglement entropies which allows for an accurate description with an MPS already for low bond dimensions M.

Analysis

When analysing the results that are all obtained using the same N_{tot} and equal total integration time τ_{tot}, it becomes clear that for systems this large – with Hilbert spaces up to 8.62×10^{87} Slater determinants for the half-filled 150-site system – the plain FCIQMC results are strongly biased compared to the reference results. Also the statistical errorbars are on the order of 10^{-4} per site and therefore quite large. Applying importance sampling using the Gutzwiller-like guiding wavefunction from equation (6.27), the population control bias is removed almost entirely and tames down the statistical errorbars to the order of 10^{-6} per site. Additional a-posteriori correction leads to FCIQMC results that lie within statistical errorbars of the DMRG reference results. Also, it becomes clear that calculations of systems with 102 or more sites already approximate the thermodynamic limit well.

Technical details for the FCIQMC calculations – i.e. the total walker number, the optimised Gutzwiller correlation factor, and the correction order in the a-posteriori correction – are given in table 6.4. Calculations of the half-filled and four-hole 102-site systems at $U/t = 8$ were conducted with $N_{\text{tot}} = 3 \times 10^7$ walkers. The one-hole 102-site calculations were conducted with 5×10^7 walkers to remove any potential bias due to the small remaining sign problem. The 150-site problem at $U/t = 8$ is also solved with 5×10^7 walkers because of the significantly larger Hilbert space which increases the population control bias. For $U/t = 4$, the N_{tot} was increased to 5×10^7 walkers as well. Since all calculations in table 6.2 are integrated over (roughly) the same imaginary time τ_{tot}, the statistical errors are increased

slightly. This is due to the fact that the real-space basis is less suitable in the low-U/t case, leading to more spread-out wavefunctions. This in turn leads to more stochastic rounding processes that lead to an increased need for population control which increases the population control bias.

Comparison of Memory Usage

The accuracy and precision of the ground-state energies are not the only things that can and should be compared when benchmarking FCIQMC against DMRG. DMRG is a close-to-optimal solver of 1-d systems because the representation of the DMRG wavefunction as an MPS exploits the locality and therefore reduces the required memory so much that a deterministic optimisation of the ground state is readily possible. FCIQMC uses a very different approach as it only stores a low-memory instantaneous stochastic snapshot. Let us therefore estimate the memory requirements of both algorithms and compare them.

Since the wavefunction representations are closely linked to the algorithms they are used within, it is not reasonable to only consider the storage requirements of the wavefunctions themselves. Rather, the additional essential data that are necessary to perform the respective FCIQMC or DMRG run will also be added.

In FCIQMC, memory is required to store

- a list of the instantaneously occupied Slater determinants (which is done in a binary representation which scales linearly in the number of sites ℓ),
- the walker occupation of each occupied Slater determinant (which is a floating point number),
- the spawn array (which in a conservative estimate requires an additional $\frac{1}{10}$ of the storage of the main walker list), and
- the hash table without which the algorithm would be inefficient (requiring two integers per hash table entry and an additional integer to store empty spots; all roughly scaling with the number of occupied determinants).

This amounts to the formula

$$n_{\text{int64}}^{\text{FCIQMC}} = \left[\left(1+\frac{1}{10}\right)\left(\left\lceil\frac{2\ell}{64}\right\rceil+1\right)+3\right]\times N_{\text{dets}}^{\text{max}} \tag{6.44}$$

where $n_{\text{int64}}^{\text{FCIQMC}}$ is the number of 64-bit integers that need to be stored maximally during an FCIQMC run. $N_{\text{dets}}^{\text{max}}$ is the maximum number of occupied Slater determinants during a particular simulation of a system.

In turn, for a DMRG run of an ℓ-site 1-d Hubbard model one needs to store

- the MPS $|\Psi^{\text{DMRG}}\rangle$ itself (which consists of 4ℓ matrices of dimension $M \times M$),
- the derivatives at each lattice site (which consist of $4M^2$ floating point numbers), and
- a contraction of the expectation value $\langle\Psi^{\text{DMRG}}|\hat{H}|\Psi^{\text{DMRG}}\rangle$ over right and left neighbours for each site (which leads to additional $2\ell M^2 D_{\hat{H}}$ floating point numbers where $D_{\hat{H}}$ is the bond dimension of the 1-d Hubbard Hamiltonian when conjugate terms are evaluated on the fly and not stored).

The entire formula for the DMRG memory requirement is thus given by

$$n_{\text{doubles}}^{\text{DMRG}} = M^2 \left[4(\ell + 1) + 2\ell D_{\hat{H}} \right] \tag{6.45}$$

with $n_{\text{doubles}}^{\text{DMRG}}$ being the number of double-precision floating point numbers to be stored.

With these formulas, let us now explicitly calculate the memory requirements for the calculation of the half-filled periodic 102-site system at $U/t = 8$ with the technical simulation parameters taken from table 6.4:

- In the corresponding FCIQMC calculation, $N_{\text{tot}} = 3 \times 10^7$ walkers were used. Due to the low walker number compared to the huge Hilbert space, this amounts to $N_{\text{dets}}^{\text{max}} \approx N_{\text{min}}$. Therefore, according to equation (6.44), this amounts to $n_{\text{int64}}^{\text{FCIQMC}} = 2.55 \times 10^8$ 64-bit integers which requires 2.04 GB of storage.
- In the DMRG benchmarking results, $M_{\text{min}} = 2000$ was necessary to achieve convergence. Inserting this into equation (6.45) adds up to 4.91×10^9 double-precision floating point numbers which require 39.3 GB of storage.

This leads to the conclusion that an instantaneous snapshot of a stochastically evoluting wavefunction whose sampling ultimately leads to correct ground-state energies can be more memory-efficient than an MPS, even in

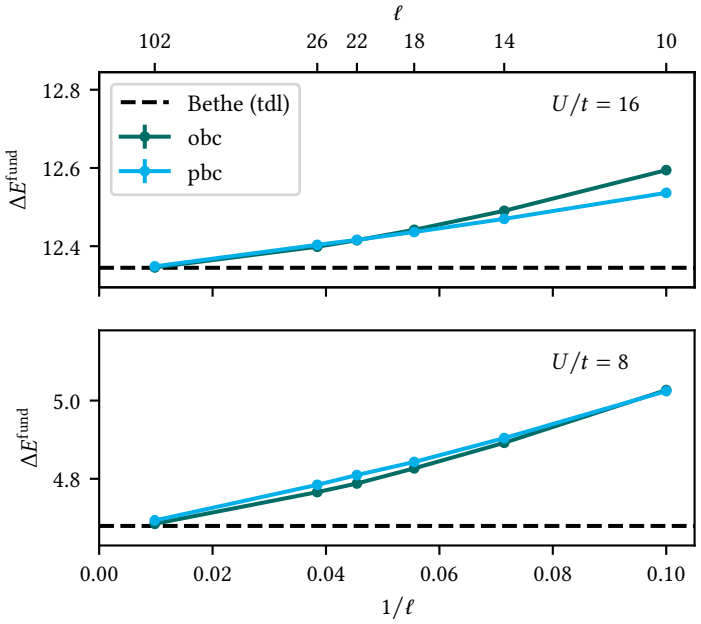


Figure 6.9. Fundamental many-particle gaps ΔE^{fund} for Hubbard chains at $U/t = 8$ and 16 with open (obc) and periodic boundary conditions (pbc). The dashed line indicates the Bethe ansatz result in the thermodynamic limit (tdl) using equation (6.48). While the obc case is sign-problem-free at any length, the periodic chains even up to more than 100 sites can only be converged due to the vanishing sign problem for $\ell \to \infty$.

cases where this representation is close-to-optimal due to low entanglement entropies.

6.3.2 *Fundamental Many-Particle Gaps of Hubbard Chains*

The fundamental many-particle gap is defined as

$$\begin{aligned} \Delta E^{\text{fund}} &= E_0(+1) + E_0(-1) - 2E_0(0) \\ &= 2\big[E_0(-1) - E_0(0)\big] + U \end{aligned} \tag{6.46}$$

where $E_0(0)$, $E_0(-1)$, and $E_0(+1)$ are the ground-state energies at half-filling, with one hole, and with one excess electron, respectively. It is the many-particle equivalent to the band gap in a mean-field picture. The second equivalence is true because for the Hubbard model

$$E_0(+1) = E_0(-1) + U\,. \tag{6.47}$$

The calculation of ΔE^{fund} is challenging because the ground-state energies need to be resolved with high precision as their difference is usually very small. Also, they cannot be calculated in a sign-problem-free manner even with AFQMC as it involves the calculation of the ground-state energy of a system off-half-filling.

In figure 6.9, ΔE^{fund} is shown for 1-d Hubbard systems at $U/t = 8$ and 16 with open (obc) and periodic boundary conditions (pbc) for up to 102 sites. The results were taken from tables 6.2 and 6.3, respectively, i.e. they were

obtained using importance sampling and a-posteriori correction combined. Also shown is an analytical reference for ΔE^{fund} in the thermodynamic limit (tdl) obtained via the Bethe ansatz. It is given by

$$\Delta E^{\text{fund, Bethe}} = U - 4t + 8t \int_0^\infty \mathrm{d}\omega \frac{J_1(\omega)}{\omega\left[1 + \exp\left(\frac{\omega U}{2t}\right)\right]} \tag{6.48}$$

where again $J_1(\omega)$ is the first-order Bessel function of first kind [198].

7 *Importance Sampling in Sign-Problematic Cases*

In chapter 6, it was shown how importance sampling with a fast-to-evaluate Gutzwiller-like guiding wavefunction is highly effective in reducing the population control bias in FCIQMC. A natural question to ask is what effect importance sampling has when dealing with sign-problematic systems, as it is known to increase the sampling effectiveness in other QMC methods.

Parts of the results presented in this chapter are also contained in ref. 179. Collaborators: K. Ghanem and A. Alavi.

However, the first thing to note is that the importance-sampled Hamiltonian as defined in equation (6.11) – which, as we know, leaves the spectrum unchanged – also does not change the stoquastised gap. This is easy to see: By definition, the matrix elements of the stoquastised version of the similarity-transformed Hamiltonian $\hat{H}'$ are given by

$$\left[H'_{ij}\right]^{\text{stoq}} = \left[\exp\left[g\left(H_{jj} - H_{ii}\right)\right] H_{ij}\right]^{\text{stoq}} \tag{7.1}$$

Since the exponential prefactor in front of each matrix element is positive only, the stoquastised version of the similarity-transformed Hamiltonian and the similarity-transformed version of the stoquastised Hamiltonian are equal:

$$\exp\left[g\left(H_{jj} - H_{ii}\right)\right] H_{ij}^{\text{stoq}} = \left[\exp\left[g\left(H_{jj} - H_{ii}\right)\right] H_{ij}\right]^{\text{stoq}}. \tag{7.2}$$

Thus, the stoquastised spectrum is the same as in the non-stoquastised version.[13]

The stoquastised gap has been established as the main quantifiable indicator of the strength of the sign problem as it both describes sign-problem-free situations correctly and allows us to classify sign-problematic systems based on their lattice geometry as shown in chapter 5. In the literature, it has been shown that the stoquastised gap in Hubbard systems can be reduced by basis rotations [55]. These facts may lead to the conclusion that it is not possible to reduce the computational effort to solve sign-problematic systems merely by introducting importance sampling. This does not capture the complete picture however, as I will demonstrate in this chapter.

13 It is noted that, even for signed guiding wavefunctions, ΔE^{stoq} is left unchanged. Applying importance sampling with a guiding wavefunction $|\Psi_{\text{g}}\rangle$ with $\langle D_i|\Psi_{\text{g}}\rangle < 0$ is equivalent to multiplying all elements of the i-th column and row of $\mathbf{H}$ with -1. This simply flips the sign of the C_i coefficient without affecting its magnitude, leaving the difference $\left|C'_i - |C_i|\right|$ and therefore ΔE^{stoq} unchanged.

7.1 *FCIQMC-Related Strength of the Sign Problem*

Apart from calculating the stoquastised gap, empirically one can determine the actual computational effort required to resolve the sign problem and determine the correct ground-state energy. As explained in section 3.2.2, in many systems the minimum walker number N_{min} to overcome the sign problem can be determined by inspecting the rate of walker growth during the constant-shift phase at the beginning of the simulation. The annihilation plateau however does typically not occur in the weak-sign-problem real-space model systems considered in this thesis.

Alternatively, as hinted in section 3.1.1, the fixed-N_0 method can be used. With this method, the shift is adapted such that the reference population N_0 is held constant after it has grown to its target value. With this, of course also the sign of N_0 is fixed. This then also fixes the signs of all connected determinants which again fixes the signs of their connections and so on. Therefore, the global sign of the wavefunction is fixed, leading to a resolved sign problem. The algorithm will therefore converge to an average total walker number $\overline{N_{\text{tot}}} \geq N_{\text{min}}$ for any non-zero N_0. Theoretically, running a fixed-N_0 calculation with $N_0 = t_{\text{occ}}$ would exactly yield $\overline{N_{\text{tot}}} = N_{\text{min}}$. In practice however, fixed-N_0 calculations with very low N_0 lead to very large flucuations in all simulation parameters. Therefore, in this chapter I will use $N_0 = 50$ in all determinations of N_{min} which leads to slight overestimations. This way, fluctuations are limited while still allowing comparisons between similar systems and giving correct orders of magnitude.

With this definition of N_{min}, it is then possible to define the *FCIQMC-related relative strength of the sign problem* by relating N_{min} and the Hilbert space size $|\mathcal{H}|$:

$$s = \frac{N_{\text{min}}}{|\mathcal{H}|}\,. \tag{7.3}$$

The Hilbert space size is given by

$$|\mathcal{H}| = \binom{N_{\text{sites}}}{N_\uparrow}\binom{N_{\text{sites}}}{N_\downarrow}\,. \tag{7.4}$$

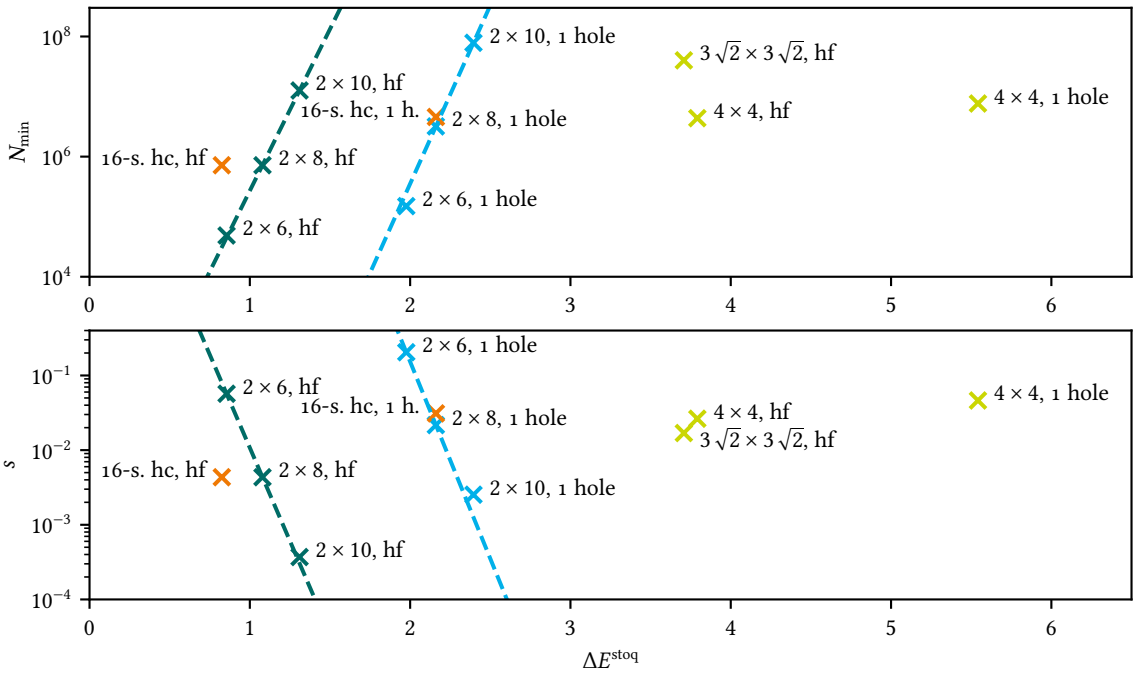


Figure 7.1. Minimum number of walkers $N_{\min}$ and FCIQMC-related relative strength of the sign problem s as a function of the stoquastised gap ΔE^{stoq} on a logarithmic scale [179]. Both the relation between N_{stoq} and ΔE^{stoq} and between s and ΔE^{stoq} are not monotonic across the systems considered here, as one would assume from a simple picture that is based on average walker dynamics. Instead, the structure of the sampled wavefunction and its relation to the stoquastised solution play a crucial role. The dashed lines are exponential fits used as a guide to the eye.

7.1.1 *Two-Dimensional Systems*

Figure 7.1 shows $N_{\min}$ and s determined using the fixed-N_0 method using $N_0 = 50$ as a function of ΔE^{stoq} for the systems already shown in figure 5.5. Clearly, both $N_{\min}$ and s do not monotonically depend on ΔE^{stoq} as one would naively assume from the general considerations about the emergence of the FCIQMC sign problem in section 3.2.2. For example, $N_{\min}$ for the larger 2×10 system is larger than for the smaller 16-site system, even though the stoquastised gap is smaller. Instead, as the length of the ladder systems is increased, the relative strength of the sign problem even decreases. This means that for an increase of the Hilbert space by some factor, the required walker number grows by less than this factor, even though the stoquastised gap grows. On the other hand, when removing an electron from the system, both s and ΔE^{stoq} are increased for a given lattice in all systems considered.

This leads to the conclusion that the required computational effort to solve a system with FCIQMC does not solely depend on ΔE^{stoq} which can be considered an averaged quantity. It also depends on the structures of the sampled wavefunction and its stoquastised counterpart which determine the effectiveness of the annihilation step. Thus, it is justified to study the effect of importance sampling also in sign-problematic systems.

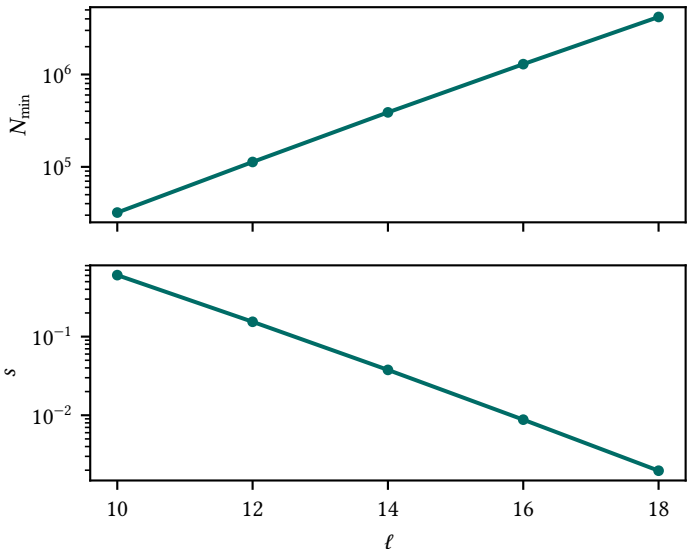


Figure 7.2. Minimum number of walkers N_{min} and FCIQMC-related relative strength of the sign problem s as a function of chain length ℓ for 1-d Hubbard systems at $U/t = 8$, each with one hole, on a logarithmic scale. Despite the non-size-extensivity of the sign problem for the 1-d Hubbard model as demonstrated in section 5.2.1 and a monotonically decreasing s, N_{min} grows exponentially. This indicates that also the computational effort to resolve the sign problem might grow exponentially. However, the N_{min} obtained with fixed-$N_0 > 1$ is only an approximation that overestimates the true minimum number of walkers, especially for very spread-out wavefunctions. Furthermore, for very large Hubbard chains ($\ell \gtrsim 100$), the stoquastised gap has closed to practically zero, meaning that a constant reference population is no longer necessary to achieve convergence.

7.1.2 *One-Dimensional Systems*

In section 5.2.1, it is shown that 1-d Hubbard chains show a non-size-extensive sign problem which is defined as a decreasing gap ΔE^{stoq} with increasing chain length ℓ. But, as just discussed, the monotonically decreasing stoquastised gap does not automatically translate into a decreasing difficulty to resolve the sign problem with FCIQMC. Even a small ΔE^{stoq} can pose a problem in large Hilbert spaces. This problem is illustrated in figure 7.2 where both N_{min} and s are plotted for Hubbard chains with $\ell = 10, 12, 14, 16,$ and 18, each with one hole which are sign-problematic configurations. Even though the relative strength of the sign problem decreases, like in the 2-d ladders, the required population N_{min} to achieve a constant reference population of $N_0 = 50$ increases exponentially.

ℓ	E_0 (FCIQMC)	E_0^{stoq} (FCIQMC)	E_0 (DMRG)
50	$-18.0338(4)$	$-18.0339(5)$	-18.0188
	$N_{\text{tot}} = 5 \times 10^7$	$N_{\text{tot}} = 1 \times 10^6$	$M = 2000$
102	$-35.0637(46)$	$-35.0639(49)$	-35.0601
	$N_{\text{tot}} = 5 \times 10^7$	$N_{\text{tot}} = 5 \times 10^7$	$M = 6000$

Table 7.1. Comparison of FCIQMC and DMRG results for the ground-state energy E_0 of periodic 1-d Hubbard chains at $U/t = 8$ with one hole. Also given is the stoquastised ground-state energy E_0^{stoq}, also calculated with FCIQMC. Clearly, the FCIQMC result for E_0 of the 50-site system is still biased, even for a fairly large walker number. For the same walker number, the result for the 102-site system is already converged as the non-size-extensive stoquastised gap has already closed to within statistical errorbars. Importance sampling with a Gutzwiller-like guiding wavefunction with $g = 0.17$ has been applied to correct for the population control bias.

Does this mean that the correct energy for large 1-d systems close to the thermodynamic limit cannot be obtained? The answer is no because the simple picture shown figure 7.2 is lacking some detail. Firstly, any N_{min} obtained with $N_0 > t_{\text{occ}}$ is only an approximation to the true walker number above which convergence can be achieved (also see figure 7.4). $N_0 = t_{\text{occ}}$ however is typically not calculable due to large fluctuations. Secondly, for a

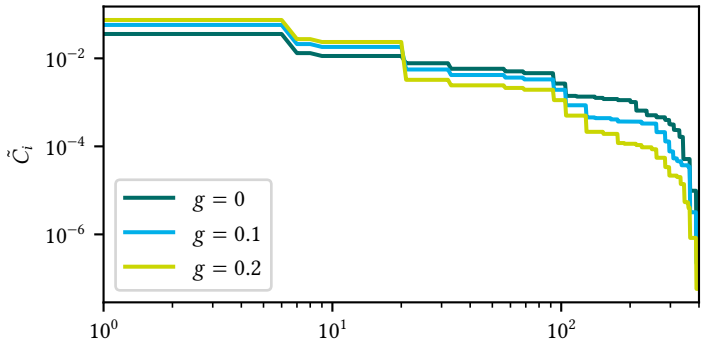


Figure 7.3. C_i coefficients of a 2×3 Hubbard model at $U/t = 8$, both without and with similarity transformation using $|\Psi_{\mathrm{GL}}\rangle$ with $g = 0.1$ and 0.2 [179]. The wavefunctions were obtained with exact diagonalisation and normalised according to their ℓ_1 norm. Applying the similarity transformation increases the compactness.

large enough ℓ the stoquastised gap becomes so small that E_0^{stoq} equals E_0 within statistical errors. Therefore, in practice no sign problem needs to be overcome. No constant reference population of one sign is required anymore and the definition of N_{min} loses its significance. However, this means that chains of intermediate lengths can be problematic or even impossible to solve as the Hilbert space is already too large. Not enough annihilations can be achieved to resolve the sign problem (even with importance sampling; see next section). The stoquastised gap however still has not fully closed so there is an unresolved bias in all calculations with an achievable total walker number. This is the case for 50-site system as can be seen in table 7.1.

7.2 *Weakly Sign-Problematic Systems*

Let us now see in a numerical experiment what is the effect of using importance sampling with a Gutzwiller-like guiding wavefunction $|\Psi_{\mathrm{GL}}\rangle$ – as defined in equation (6.27) – in non-initiator FCIQMC calculations of weakly sign-problematic systems.

To get a better understanding of the effects of importance sampling, let us look how the diagonal similarity transformation affects the FCI expansion of an exactly diagonalisable wavefunction. Figure 7.3 shows the C_i coefficients of a 2×3 Hubbard ladder system with $|\mathcal{H}| = 400$ at $U/t = 8$. The wavefunction – normalised according to the ℓ_1 norm to resemble the C_i coefficients obtained with FCIQMC at different N_{tot} – becomes more compact as g is increased.

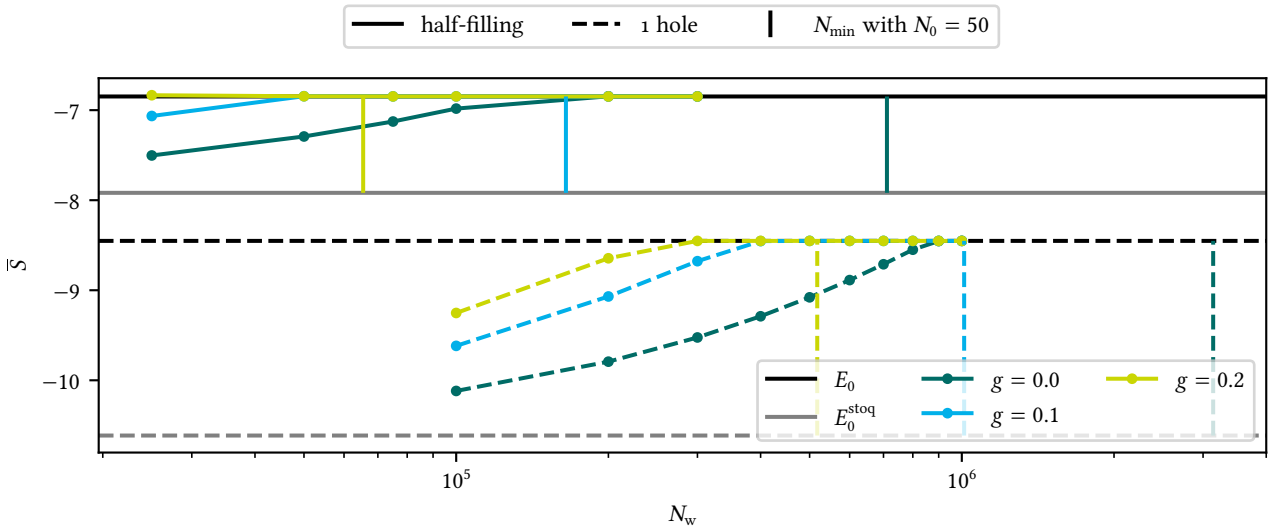


Figure 7.4. Convergence of the average shift $\overline{S}$ with respect to the total number of walkers N_{tot} for a half-filled 2×8 Hubbard ladder at $U/t = 8$ and with one hole [179]. The different curves show results obtained with no importance sampling and with importance sampling using $|\Psi_{\text{GL}}\rangle$ as a guiding wavefunction with $g = 0.1$ and 0.2. Importance sampling improves convergence significantly. The vertical lines show the N_{min}, obtained using the fixed-$N_0 = 50$ method. The true N_{min} is slightly overestimated as expected. Yet, they allow for good comparability of N_{min} among different systems of similar type.

7.2.1 *Improved Convergence of the Shift Estimator*

Figure 7.4 shows the convergence of the average shift $\overline{S}$ for a 2×8 weakly sign-problematic Hubbard ladder at $U/t = 8$ both at half-filling (top) and with one hole (bottom) with respect to the total number of walkers N_{tot}. The convergence is studied for standard non-initiator FCIQMC as well as importance-sampled FCIQMC using $|\Psi_{\text{GL}}\rangle$ as a guiding wavefunction with correlation parameters $g = 0.1$ and 0.2. As explained in section 3.2, $\overline{S}$ converges from below from somewhere close to E^{stoq} to the exact energy at N_{min}. $\overline{S} = E^{\text{stoq}}$ would be reached in the low-walker limit which cannot be performed in practice due to large stochastic fluctuations. In this case, no annihilations would take place and the shift estimator would not be able to distinguish between $\hat{H}$ and $\hat{H}^{\text{stoq}}$.

Clearly, applying importance sampling improves the convergence in both examples with $g = 0.2$, performing even better than $g = 0.1$. In the half-filled case, for $g = 0.2$ the exact energy is obtained already with less than 3×10^4 walkers. Without importance sampling, approximately 2×10^5 walkers are necessary. In the one-hole case, the required number of walkers decreases from roughly 1×10^6 without importance sampling to approximately 3×10^5 walkers. This goes in line with the values for N_{min} determined using fixed-$N_0 = 50$ which are indicated in the plot with vertical lines. As expected, the fixed-N_0 method leads to estimations of N_{min} that are above the true minimum number of walkers but allow for good comparability among different system setups.

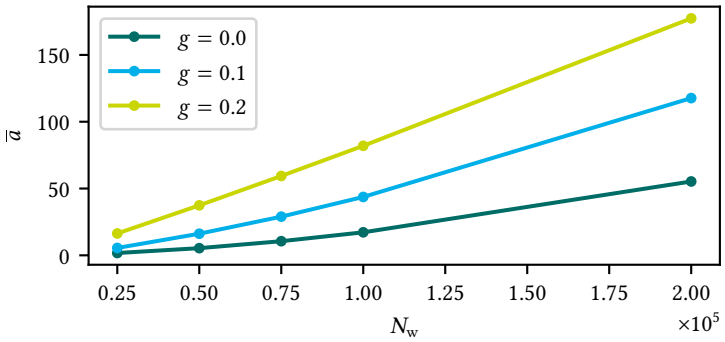


Figure 7.5. Annihilation rates $\overline{a}$ for the half-filled 2×8 Hubbard ladder system at $U/t = 8$ with no importance sampling and importance sampling using $|\Psi_{GL}\rangle$ with $g = 0.1$ and 0.2 as a function of the total walker number N_{tot} [179]. For a larger g value the annihilation rate is increased. This leads to an improved resolution of the sign problem for low N_{tot}.

How the increased compactness reduces the number of walkers required to resolve the sign problem can partly be understood from inspecting the average number of annihilations per iteration. For the calculations of the half-filled 2×8 ladder from figure 7.4, the corresponding average annihilation rates $\overline{a(\tau)}$ are shown in figure 7.5. The annihilation rate in an iteration at imaginary time τ is given by

$$a(\tau) = \sum_{i=1}^{N_{\text{dets}}} \left(\left|N_i(\tau - \Delta\tau)\right| + \sum_{j\neq i} \left|\Delta N_{ij}(\tau)\right| \right) - \left| N_i(\tau - \Delta\tau) + \sum_{j\neq i} \Delta N_{ij}(\tau) \right| \quad (7.5)$$

where $\Delta N_{ij}(p\Delta\tau)$ is the number of spawned walkers from $|D_j\rangle$ to $|D_i\rangle$ in iteration p. As before, $N_i(p\Delta\tau)$ is the number of walkers residing on $|D_i\rangle$ in iteration p. For a larger value of g and given N_{tot}, the annihilation rate is increased. Already for $N_{tot} = 2.5 \times 10^4$ walkers, the annihilation rate in the $g = 0.2$ calculation surpasses the one for $N_{tot} = 1 \times 10^5$ in the no-importance-sampling case. This is mirrored also in the values for $\overline{S}$ where in the former case, the estimator is already converged to the true solution whereas the latter case is still biased. This indicates that the annihilation rate is boosted due to the increased compactness of the sampled wavefunction and therefore the fermionic solution can emerge for lower walker numbers.

7.2.2 *Tradeoff between Compactness and Noise*

As established in the previous section by considering the annihilation rates, N_{min} is lowered due to the increased compactness while not affecting the stoquastised gap. This is confirmed by the data given in figure 7.6 where N_{min} is given as a function of g. These FCIQMC calculations were again conducted in the small 2×3 toy system. N_{min} was obtained using fixed-N_0 $= 10\,000$. The small size of the system and the large fixed-N_0 value are necessary for this demonstration to obtain data for values of g up to 0.5. For $g = 0.5$, large fluctuations are introduced due to a highly unbalanced Hamiltonian. The

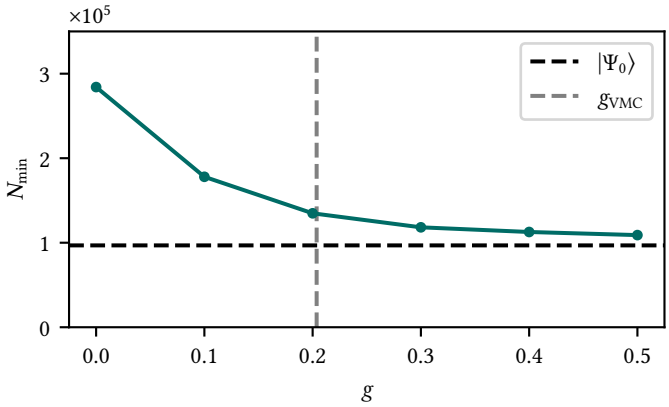


Figure 7.6. Minimum required walker number N_{min} to obtain $N_0 = 10\,000$ in a 2×3 Hubbard toy system at $U/t = 8$ as a function of the Gutzwiller correlation parameter g [179]. The horitzontal dashed line indicates N_{min} for when the exact ground-state solution $|\Psi_0\rangle$ is used as a guiding wavefunction. The vertical line indicates the optimal value for g in a full Gutzwiller ansatz $|\Psi_{\text{G}}\rangle$ energy-optimised by VMC. Although the exact guiding wavefunction presents a lower bound for N_{min} in this case, the increasing compactness for increasing g further lowers N_{min}, even far away from g_{VMC}.

small system size also allows one to use the exact solution $|\Psi_0\rangle$ as a guiding wavefunction.

Clearly, N_{min} decreases with increasing g. The dashed horizontal line indicates N_{min} for when $|\Psi_0\rangle$ is used as a guiding wavefunction. This is counter-intuitive as a VMC optimisation of the Gutzwiller wavefunction for this system yields $g_{\text{VMC}} = 0.2037$. Although a VMC energy optimisation does not necessarily lead to the optimal wavefunction, this cannot explain why the value for which the N_{min} approach each other are vastly different. So unlike for the population control bias where the bias was maximally reduced for $g \approx g_{\text{VMC}}$, here the compactification itself plays a crucial role.

To better understand the tradeoff between similarity of the guiding wavefunction and the true ground state – which influences the fluctuations according to the considerations in section 6.2.1 – and compactness, let us look at the actual FCIQMC dynamics of the simulations for the values of g from figure 7.6. These are shown in figure 7.7. It is obvious that for a Gutzwiller-like guiding wavefunction, $\overline{N_{\text{tot}}}$ decreases at the cost of introducing significant fluctuations because of correlations between the shift and the wavefunction as expected. Using the exact wavefunction as a guiding wavefunction leads to both the lowest $\overline{N_{\text{tot}}}$ as well as low fluctuations.

This implies that at some point one is suffering from diminishing returns when further increasing g because the total wall-clock time needs to be increased quadratically according to equation (2.20) to reach a certain accuracy for increased fluctuations.

This tradeoff is illustrated for the larger 2×8 Hubbard ladder in figure 7.8. The top part again illustrates that N_{min} decreases monotonically for increasing g. The bottom part shows the standard error σ of the shift energy estimator for equal wall-clock times T. Lower walker numbers allow for averaging over more samples for same T (and also longer τ_{tot} for equal $\Delta\tau$)

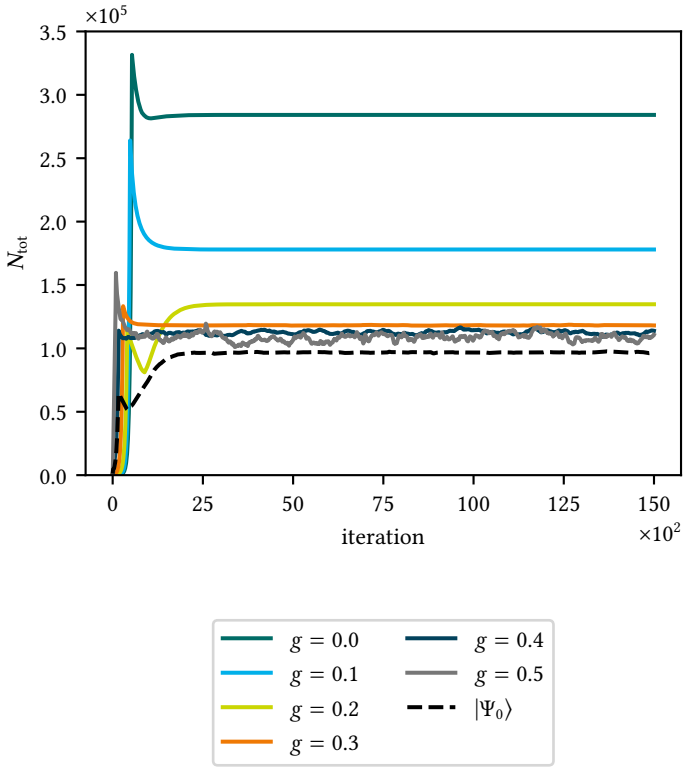


Figure 7.7. Dynamics of the total walker population N_{tot} in actual FCIQMC dynamics of the small 2×3 Hubbard system at $U/t = 8$ with importance sampling with $g = 0.1 \ldots 0.5$ and using the exact solution $|\Psi_0\rangle$, respectively [179]. $\overline{N_{\text{tot}}}$ decreases with increasing g, although fluctuations are introduced as the Gutzwiller-like guiding wavefunction deviates more and more from the true solution. Using $|\Psi_0\rangle$ leads to both the lowest $\overline{N_{\text{tot}}}$ and low fluctuations.

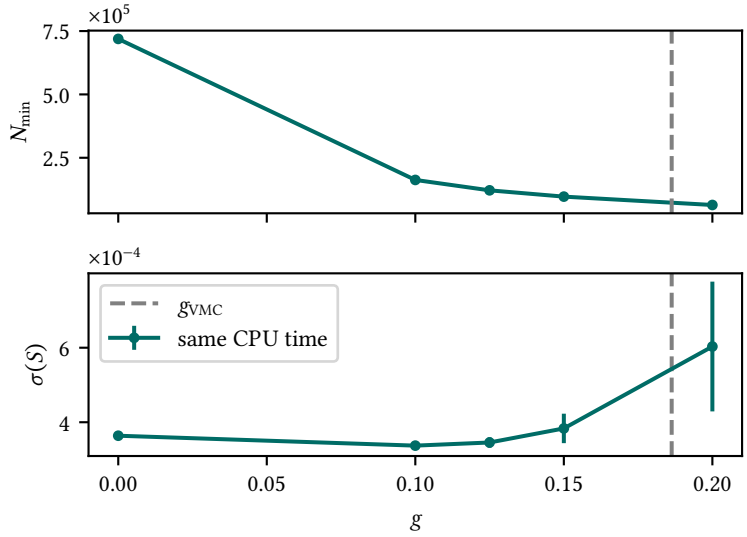


Figure 7.8. N_{min} obtained via fixed-$N_0 = 50$ (top) and standard error for fixed wall-clock time (bottom) for the 2×8 Hubbard ladder at $U/t = 8$ [179]. As expected from the 2×3 toy system, N_{min} decreases with increasing g. However, the lowest standard error and therefore the optimal tradeoff between lowering N_{min} and noise is given for $g \approx 0.125$. Energy-optimised VMC overestimates the optimal g with $g_{\text{VMC}} = 0.1863$.

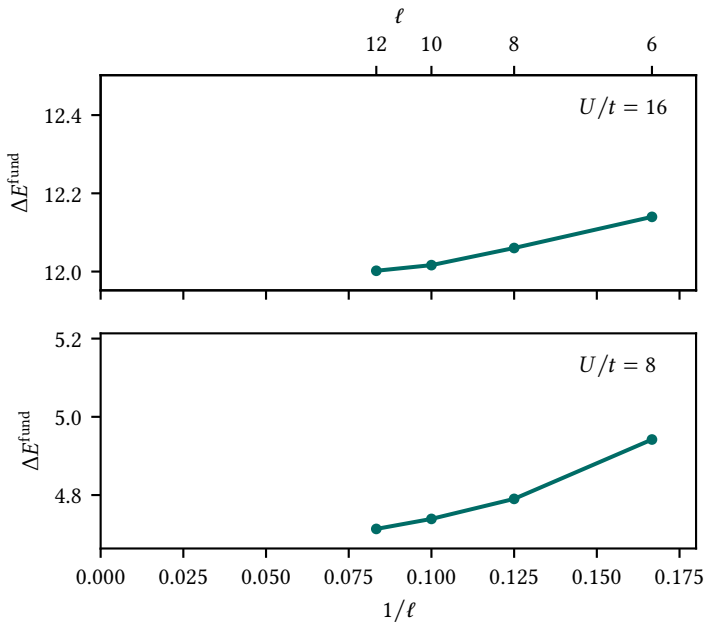


Figure 7.9. Fundamental many-particle gaps ΔE^{fund} for $2 \times \ell$ ($\ell = 6, 8, 10, 12$) Hubbard ladders at $U/t = 8$ and 16 with periodic boundary conditions as a function of the inverse length $1/\ell$ [179]. Since these systems show an inevitable but weak sign problem, the calculations are conducted in a non-initiator fashion using importance-sampled FCIQMC with the Gutzwiller-like guiding wavefunction $|\Psi_{\text{GL}}\rangle$. Raw numbers and technical details for the $U/t = 8$ calculations are given in table 7.2.

but above a certain g the fluctuations become dominant. The optimal tradeoff is reached for $g \approx 0.125$. When one can accept a larger uncertainty in the energy estimate however, it can be advised to use a larger-than-optimal g because it allows to reach the required precision for lower T. This is used in the next sections where I will calculate large ladder systems using the importance-sampling method. Again, VMC with energy optimisation overestimates the optimal point with $g_{\text{VMC}} = 0.1863$, not returning the optimal wavefunction.

7.3 *Applications*

In this section, I will apply the previous findings to find the *fundamental many-particle gaps* of Hubbard ladder systems in the intermediate ($U/t = 8$) and strong ($U/t = 16$) interaction regime. I will also calculate the ground-state energy of the 32-site honeycomb system in the difficult intermediate interaction regime $U/t = 8$, a system well beyond the capabilites of exact diagonalisation and out of reach with plain FCIQMC with currently available hardware capabilities.

7.3.1 *Fundamental Many-Particle Gaps of Hubbard Ladders*

By using importance sampling in sign-problematic systems, we can now calculate the fundamental many-particle gaps ΔE^{fund}, like in section 6.3.2, also for the ladder systems according to equation (6.46).

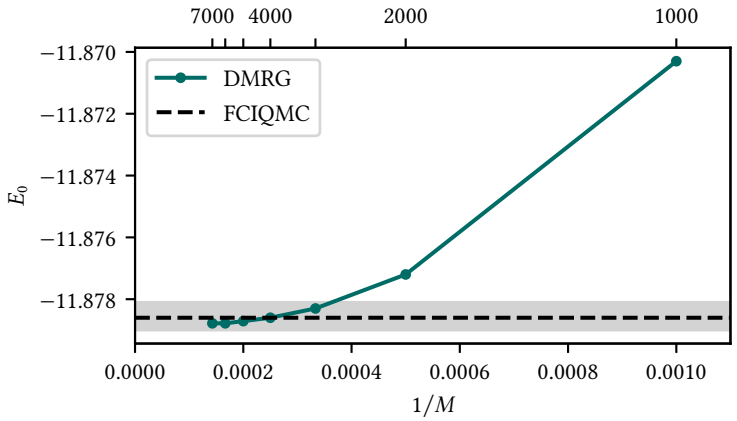


Figure 7.10. Convergence of the DMRG benchmark for the 2×12 Hubbard ladder with one hole at $U/t = 8$ with respect to the inverse bond dimension $1/M$ [179]. Fiedler-type ordering is used. Agreement with FCIQMC within statistical errorbars is reached for $M \gtrsim 4000$. As the ladder systems are geometrically close to a 1-d system, DMRG still benefits from a small entanglement entropy.

Figure 7.9 shows ΔE^{fund} for $2 \times \ell$ ladders for $\ell = 6, 8, 10$, and 12 at $U/t = 8$ and 16 for periodic boundary conditions. Importance sampling with the Gutzwiller-like guiding wavefunction $|\Psi_{\text{GL}}\rangle$ is used to reduce N_{min} significantly. The ground-state energies for the $U/t = 8$ case are given in table 7.2, also comparing the respective N_{min} for fixed-$N_0 = 50$ with and without importance sampling (where accessible) and the value for g used. Especially the 2×12 system at $U/t = 8$ with one hole, which is beyond the scope of exact diagonalisation [199], could not be be calculated using plain FCIQMC because well beyond 1×10^9 walkers would be required. For comparison, the convergence of the ground-state energy using DMRG with respect to the bond dimension M is shown in figure 7.10. The DMRG calculations were again conducted using the `BLOCK` code and Fiedler-type ordering of the lattice sites was used [200, 201]. Agreement between FCIQMC and DMRG within statistical errorbars is reached for $M \gtrsim 4000$. DMRG still performs relatively well as the two-legged ladders' geometries are still somewhat close to 1-d. Therefore, the entanglement entropy is still relatively small.

filling	lattice	E_0	N_{min} $[\times 10^3]$ using $\lvert\Psi_{\text{GL}}\rangle$	N_{min} $[\times 10^3]$	g
half-filling	2×6	−5.1600	48	11	0.15
	2×8	−6.8469	718	97	0.15
	2×10	−8.5382	12 638(2)	1085(1)	0.15
	2×12	−10.2353	n.a.	21 275(10)	0.15
one hole	2×6	−6.6890	149	55	0.15
	2×8	−8.4518	3168(1)	627	0.15
	2×10	−10.1687	80 494(10)	13 410(5)	0.15
	2×12	−11.8794	n.a.	297 850(9774)	0.25

Table 7.2. Raw numbers for the ground-state energy estimate E_0 and the minimum walker numbers N_{min} with and without importance sampling [179]. Also given is the Gutzwiller correlation parameter g used for $2 \times \ell$ Hubbard ladders at $U/t = 8$. The N_{min} are obtained with fixed-$N_0 = 50$. N_{min} numbers are not available for 2×12 systems as this exceeds available hardware capacity. $g = 0.15$ is close to the value where there is an optimal tradeoff between noise and reduction of N_{min} (see section 7.2.2). For the most difficult system, 2×12 with one hole, a larger g is chosen to further reduce N_{min} while slightly increasing the variance of the energy estimate.

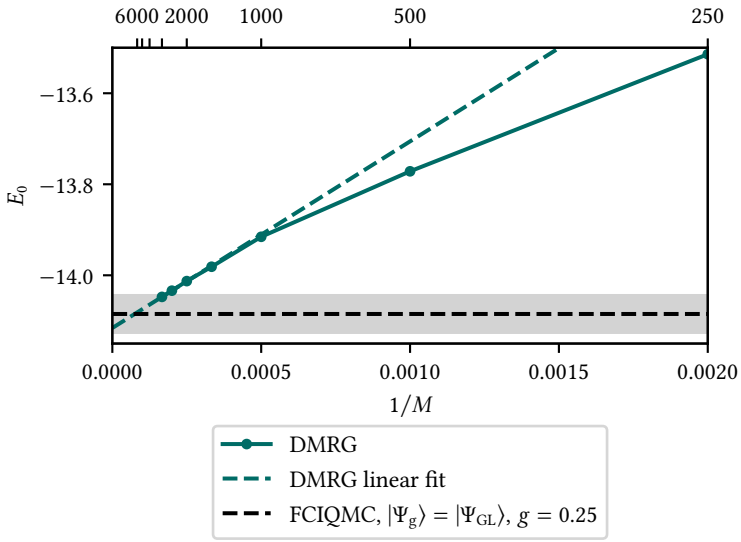


Figure 7.12. Convergence of the DMRG ground-state energy of the half-filled 32-site honeycomb Hubbard system at $U/t = 8$. Also shown is the result $E_0 = -14.085(41)$ and the respective errorbars obtained with importance-sampled FCIQMC using the Gutzwiller-like guiding wavefunction with $g = 0.25$. The linearly extrapolated DMRG for $1/M \to 0$ agrees with the FCIQMC result within statistical errors.

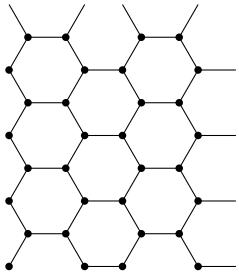


Figure 7.11. Lattice structure of the 32-site honeycomb system. Due to the six-site innermost loops, the system has a very weak sign problem. Thus, it can be effectively solved for the ground state using importance-sampled FCIQMC.

7.3.2 *Half-Filled 32-Site Honeycomb System*

The half-filled 32-site honeycomb lattice has a very weak sign problem due to the fact that its innermost loop consists of six instead of four sites. A more detailed discussion about the effect of this is given in section 5.2.2. The lattice structure is depicted in figure 7.11.

The 32-site honeycomb lattice at $U/t = 8$ with 32 electrons cannot be converged with $N_{\text{tot}} \lesssim 1 \times 10^9$ walkers. When applying importance sampling with $g = 0.25$ however, N_{min} is reduced to 1.3432×10^8 walkers. A ground-state energy of $E_0^{\text{FCIQMC}} = -14.085(41)$ is obtained. Some precision had to be sacrificed, i.e. larger errorbars had to be accepted, to lower N_{min} to a manageable value. The value obtained with FCIQMC agrees well with the DMRG benchmark which is shown in figure 7.12: For $M = 6000$ one obtains an unconverged ground-state energy $E_0^{\text{DMRG}} = -14.047$. Linear extrapolation for $1/M \to 0$ results in $E_0^{\text{DMRG, extrapol}} = -14.115(30)$. With this, the half-filled 32-site honeycomb Hubbard model is the largest 2-d model system solvable with FCIQMC in an unbiased fashion so far.

8 *Fixed Initiator Spaces and Two-Shift Method*

In systems with compact ground-state solutions, the initiator method is a good approximation, mostly insensitive to the strength of the sign problem [17, 36, 37]. However, in real-space Hubbard systems we find that the initiator energies only converge very poorly compared to reciprocal or *ab initio* systems. Figure 8.1 shows the convergence of the energies obtained by the initiator method with respect to the total walker number of the systems for which the stoquastised gaps were already shown in figure 5.5.

Real-space calculations allow for an initiator threshold of $t_{\text{init}} = 1$ which, in the case of the standard occupation threshold of $t_{\text{occ}} = 1$, is the smallest possible t_{init}. This means that every determinant that is occupied by slightly more than one walker is marked as an initiator. But even then, it is apparent that convergence is poor compared to the reciprocal basis problem even though, when looking at the N_{min}, the situation is actually the other way around.

Along this line, one can conjecture that compactness of a ground-state wavefunction and the strength of the sign problem are inversely related in many realistic cases. On top of the numbers presented in this thesis for lattice models, a similar observation has been made in *ab initio* systems when comparing localised with delocalised orbitals [39]. On the other hand, we have seen in chapter 7 that increasing the compactness by a diagonal similarity transformation leaves the stoquastised gap of the respective Hamiltonian unchanged and even lowers the FCIQMC-related strength of the sign problem as defined there.

Both observations can be explained using the following argument: Let us consider an equilibrated FCIQMC simulation where the total walker number is chosen large enough such that the entire Hilbert space is occupied with walker numbers N_i that are proportional to the C_i coefficients of the exact FCI expansion. Therefore, the net change of walkers on determinant $|D_i\rangle$ is $\Delta N_i = 0$. According to equation (3.3), the equilibrated walker number is then given by

$$N_i = -\frac{\sum_{i \neq j} H_{ij} N_j}{H_{ii} - S} \,. \tag{8.1}$$

A wavefunction is compact if a lot of N_i are small. N_i is small in the following cases:

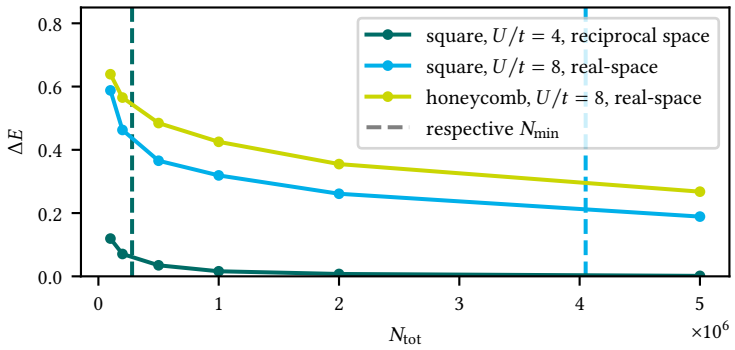


Figure 8.1. Convergence of the initiator biases ΔE with respect to the total walker number N_{tot} in FCIQMC initiator calculations of the 4×4 Hubbard model. Initiator thresholds of $t_{\text{init}} = 1$ for the real-space and $t_{\text{init}} = 1.3$ for the reciprocal space problems are chosen, respectively. Also shown are the walker numbers above which a full calculation without initiator approximation of the real-space systems would be exact. The annihilation plateau for the $U/t = 4$ system in the reciprocal basis lies well outside of the scope of this plot, indicating that the initiator method is very useful. On the other hand, initiator calculations in the real-space basis show very slow convergence. Here, it is much more useful to use the full method.

1. The diagonal element H_{ii} is large.
2. The connecting matrix elements H_{ij} are all small.
3. There are a lot of opposite-sign contributions $H_{ij}N_j$ that lead to annihilations on $|D_i\rangle$.

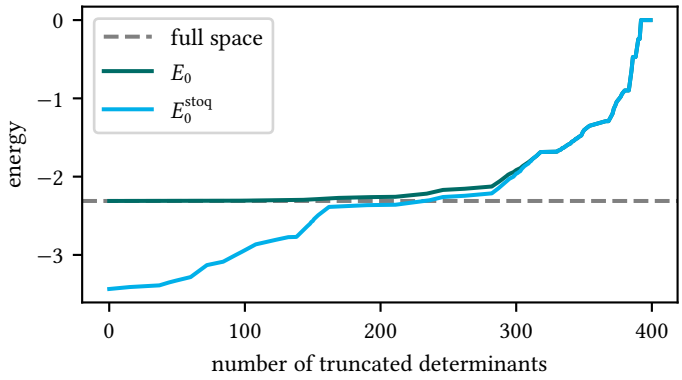
In the case of the diagonal similarity transformation, only the denominator is modified so there is no increased necessity for sign cancellations. When changing the basis however, compactness in the wavefunction can be introduced because of the third case, causing a more severe sign problem.

Therefore, in case of weak-sign-problem systems like the real-space Hubbard model, it is much more feasible to use the full FCIQMC method. However, for larger systems it might be computationally unaffordable to reach N_{min} and below N_{min} the simulation is uncontrolled and does not return a viable approximation. In chapter 7, I have shown how importance sampling can be used to reduce the mimumum number of walkers to achieve convergence in FCIQMC. Here, I will introduce an approximate method for when on the one hand the ground-state wavefunction is spread-out and thus the initiator approximation performs poorly but on the other hand the sign problem is relatively weak. It can be readily combined with importance sampling.

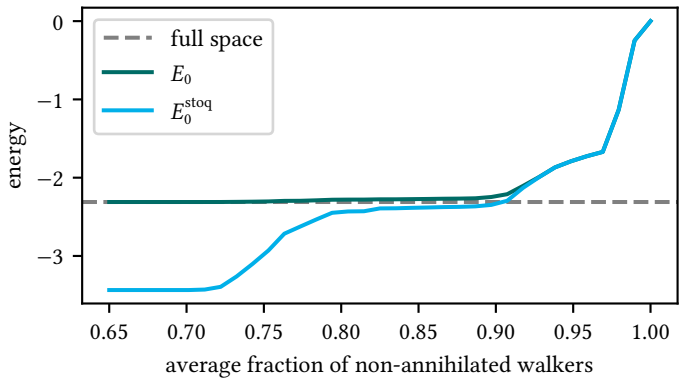
8.1 *Finding Subspaces with Very Weak Sign Problems*

The criterion to distinguish strongly and weakly sign-problematic systems using the gap between stoquastised and fermionic ground-state energy does not only hold on the global level. Also individual determinants $|D_i\rangle$ can be characterised as strongly or weakly sign-problematic by looking at the ratio

$$r_i^{\text{stoq}} = \left| \frac{C_i}{C_i^{\text{stoq}}} \right| \tag{8.2}$$



(a) Truncation based on r_i^{stoq} in ascending order.



(b) Truncation based on FCIQMC-acquired average flux of non-annihilated walkers $\overline{f_i}$. In the plot, determinants with $\overline{f_i}$ smaller than the given value on the abscissa are truncated.

Figure 8.2. Gap between E_0 and E_0^{stoq} for different truncations of the Hamiltonian of the 3×2 real-space Hubbard rectangle at $U/t = 8$. It is apparent that both the truncations based on r_i^{stoq} and on $\overline{f_i}$ can reduce the gap significantly by only biasing the true fermionic energy E_0 negligibly. The non-annihilated flux f_i is defined in equation (8.3).

between the ground-state wavefunction coefficients C_i of $\hat{H}$ and C_i^{stoq} of $\hat{H}^{\text{stoq}}$, respectively. If r_i^{stoq} is small, it means that a lot of annihilations have to take place to sample the correct amplitude of the fermionic ground-state wavefunction instead of the stoquastised one. These determinants are the reason why the Hilbert space has to be occupied by a larger number of walkers simultaneously. When the Hilbert space is truncated by the $|D_i\rangle$ with small r_i^{stoq}, this significantly reduces the sign problem and the global stoquastised gap. On the other hand, determinants with low r_i^{stoq} have comparably lower N_i because of the exact same reason. This means that removing them not only severely pushes up E_0^{stoq} but might also only affect E_0 slightly. This is desirable because a good approximate ground-state energy could be obtained with only a small number of walkers.

To illustrate this, let us look again at a very small paradigmatic system – the 2×3 Hubbard rectangle with 400 Slater determinants – where exact diagonalisation is easily possible. In this system, we now look at both the fermionic and the stoquastised ground-state energies for different truncations of the Hamiltonian. Figure 8.2a shows E_0 and E_0^{stoq} for a truncated

Hamiltonian based on r_i^{stoq} in ascending order. This means that the Hamiltonian of the respective system is diagonalised only in the subspace where r_i^{stoq} exceeds a certain threshold. It is apparent that by roughly truncating the 150 determinants with lowest r_i^{stoq}, the true fermionic E_0 is only biased by a small amount but the sign problem, measured by $E_0 - E_0^{\text{stoq}}$, is significantly reduced.

Figure 8.2b shows that similar truncations can also be achieved by using an FCIQMC-inherent parameter. Additional to the normal global statistics, determinant-specific statistics has been acquired, namely the average flux of non-annihilated walkers onto a respective determinant $|D_i\rangle$. It is defined as

$$f_i(\tau) = 1 - \frac{a_i(\tau)}{\Delta N_i(\tau)} \tag{8.3}$$

$$\text{with } a_i(\tau) = \left(\left|N_i(\tau - \Delta\tau)\right| + \sum_{j \neq i} \left|\Delta N_{ij}(\tau)\right| \right) - \left| N_i(\tau - \Delta\tau) + \sum_{j \neq i} \Delta N_{ij}(\tau) \right|$$

with $\Delta N_i(p\Delta\tau)$ being the total spawned walkers onto $|D_i\rangle$ in iteration p and $\Delta N_{ij}(p\Delta\tau)$ the spawns from $|D_j\rangle$ to $|D_i\rangle$. $N_i(p\Delta\tau)$ is the number of walkers already sitting on $|D_i\rangle$ in iteration p. With this, $A_i(p\Delta\tau)$ is the number of annihilations on $|D_i\rangle$ in iteration p.

The truncation of the Hamiltonian in this plot is done based on the average quantity $\overline{f_i}$ in ascending order. By truncating all determinants with $\overline{f_i} \lesssim 0.8$, the gap between E_0 and E_0^{stoq} is reduced to 1/10 of the size when looking at the full space. However, E_0 alone is only biased negligibly.

The existence of such truncations in small paradigmatic systems indicates that there might also be truncations of Hamiltonians of larger Hubbard lattices that remove determinants that only have a minor effect on the sought-for energy but reduce the sign problem and therefore the annihilation plateau significantly. r_i^{stoq} cannot be known a priori of course. Also sampling $\overline{f_i}$ is not feasible for large systems since empty determinants and all their determinant-related data have to be removed from memory in the FCIQMC algorithm to keep it memory-efficient.

To predict low-amplitude determinants, known analytical wavefunction ansatzes of the Hubbard model can be used instead. A possible simple ansatz was already presented in equation (6.21): the Gutzwiller ansatz. Other ansatzes include that the system favours configurations where doubly occupied sites (doublons) are spatially close to empty sites (holons) [202, 203]. The addition of this term is important for the correct description of Mott-insulator

transitions in two-dimensional half-filled square Hubbard models [204, 205]. This can be expressed as

$$|\Psi_{\mathrm{dh}}\rangle = \exp\left[\sum_{k>0}^{k_{\max}} c_k \sum_{ij} \delta_{r_{ij}k} \hat{n}_{i\uparrow}\hat{n}_{i\downarrow}\left(1-\hat{n}_{j\uparrow}\right)\left(1-\hat{n}_{j\downarrow}\right)\right]|\Psi_{\mathrm{HF}}\rangle \tag{8.4}$$

where r_{ij} is the spatial distance between site i and site j. c_k are real coefficients for the contributions of different distances k between doublons $\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$ at site i and holons $\left(1-\hat{n}_{j\uparrow}\right)\left(1-\hat{n}_{j\downarrow}\right)$ at site j up to a certain distance $k_{\max}$. VMC optimisations of the parameters c_k indicate that they decrease with increasing k. Another known wavefunction ansatz is one that enforces antiferromagnetic order. Configurations with adjacent singly occupied sites are favoured when they have opposite spin and suppressed when they are same spins. This can be written as

$$|\Psi_{\mathrm{af}}\rangle = \exp\left[\left(a\sum_{\langle ij\rangle}\left(\hat{n}_{i\uparrow}\hat{n}_{j\uparrow}+\hat{n}_{i\downarrow}\hat{n}_{j\downarrow}\right)\right)+\left(b\sum_{\langle ij\rangle}\left(\hat{n}_{i\uparrow}\hat{n}_{j\downarrow}+\hat{n}_{i\downarrow}\hat{n}_{j\uparrow}\right)\right)\right]|\Psi_{\mathrm{HF}}\rangle \tag{8.5}$$

with $a < b$.

Table 8.1 shows a combined VMC optimisation with respect to the variational energy using the Gutzwiller ansatz $|\Psi_{\mathrm{G}}\rangle$, the doublon–holon ansatz $|\Psi_{\mathrm{dh}}\rangle$, and the antiferromagnetic ansatz $|\Psi_{\mathrm{af}}\rangle$ for the 2×8 Hubbard ladder at $U/t = 8$. Configurations with larger doublon–holon distances are less important than the ones with small distances. Antiferromagnetic pairing is favoured over ferromagnetic pairing.

Table 8.1. Wavefunction parameters for a paradigmatic VMC energy optimisation of the wavefunction ansatzes $|\Psi_{\mathrm{G}}\rangle$, $|\Psi_{\mathrm{dh}}\rangle$, and $|\Psi_{\mathrm{af}}\rangle$ combined of a half-filled 2×8 Hubbard ladder at $U/t = 8$. Small doublon–holon distances as well as antiferromagnetic pairing are favoured.

$\lvert\Psi_{\mathrm{G}}\rangle$	g	0.2388
$\lvert\Psi_{\mathrm{dh}}\rangle$	c_1	5.0989
	c_2	1.5951
	c_3	−3.5017
$\lvert\Psi_{\mathrm{af}}\rangle$	a	9.8641
	b	1.4307

8.1.1 *Truncation Based on Doublon–Holon Distances*

Let us first consider ansatz $|\Psi_{\mathrm{dh}}\rangle$. The wavefunction is truncated based on the doublon–holon criterion above a certain k. In other words, I only consider a subspace $\mathcal{S}_{\mathrm{dh}}\left(k_{\max}\right)$ of $\hat{H}$ in a basis of Slater determinants in which a holon and the respective doublon only moved away from each other by a spatial distance up to $k_{\max}$. From an algorithmic perspective, this truncation only creates a small computational overhead. Although every determinant that is created as a possible excitation in the spawning step has to be checked for whether it is still contained in the selected subspace, this is typically fast. Only the one electron that is moved has to be checked for whether its move violates the truncation criterion. This makes it an $\mathcal{O}(1)$ operation since the check is independent of system size.

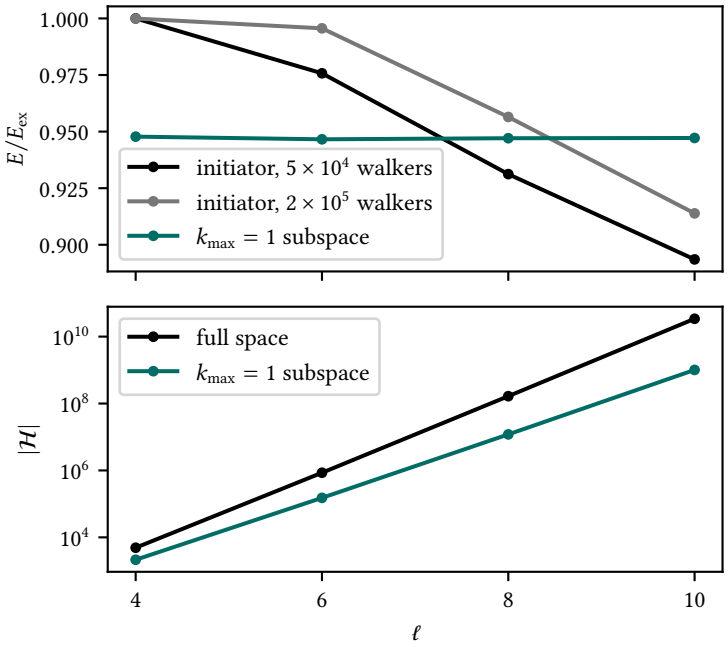
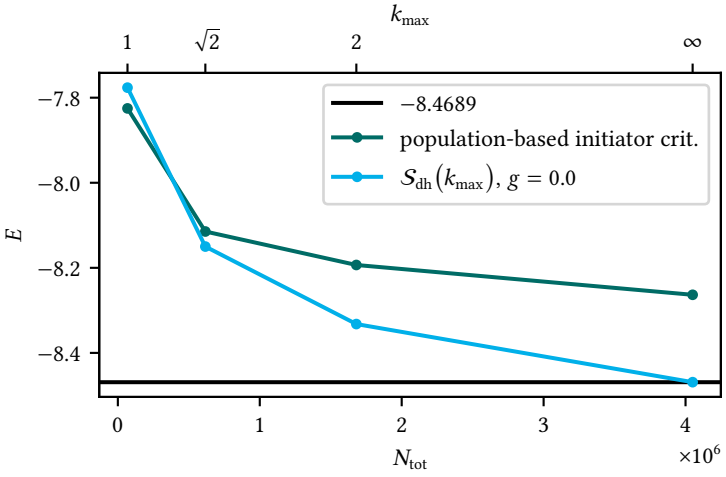


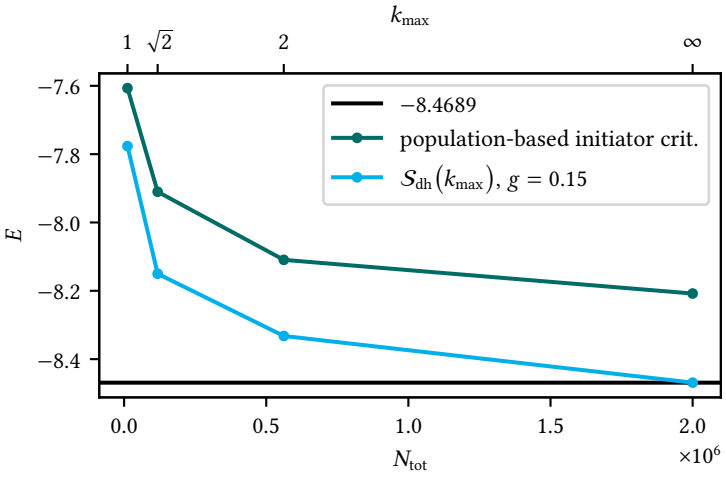
Figure 8.3. Hilbert space sizes and energies for increasing lengths of Hubbard ladder systems with $2 \times \ell$ sites for the full space and for a subspace based on the doublon–holon criterion with $k_{max} = 1$. Also the $k_{max} = 1$ subspace scales exponentially so the truncation is extensive. This leads to the fact that roughly a constant ratio of the ground-state energy is recovered in the $k_{max} = 1$ subspace. This is not true for the automatically sampled initiator subspace, regardless of the number of walkers.

It is also noted that this truncation is applied in an initiator-like fashion. This means that it is not applied as a strict truncation, i.e. all spawns to determinants that are not contained in $\mathcal{S}$ are strictly rejected, but rather occupation of every determinant in the full space is allowed. However, only determinants that are contained in $\mathcal{S}$ are allowed to spawn onto empty determinants. Determinants not contained in $\mathcal{S}$ can only spawn onto determinants in $\mathcal{S}$. Thus, only these determinants can propagate their sign through the Hilbert space whilst the non-spawning occupation of determinants not contained in $\mathcal{S}$ still gives a small energy correction, just like in the usual population-based initiator criterion. Furthermore, the population-based initiator criterion with $t_{init} = 1.3$ is applied on top. Every determinant on which the instantaneous population exceeds this threshold is treated like it would be contained in $\mathcal{S}$. This also gives a small energy correction without affecting N_{min} because these determinants can be deemed sign-coherent with a high probability.

As briefly described in section 3.2.3, a related preselection of initiator subspaces based on selected configuration interaction (SCI) in *ab initio* system has been presented previously [56]. While some observations, like the reintroduction of a weak sign problem in the fixed initiator space, nicely agree, there are key differences: The scheme presented in this thesis relies on analytical wavefunction ansatzes to determine the fixed space while the SCI-based truncation is founded on a preceding heat-bath SCI calculation.



(a) Without importance sampling.



(b) With importance sampling using the Gutzwiller-like guiding wavefunction.

Figure 8.4. Ground-state energies of the 4×4 real-space Hubbard lattice at $U/t = 8$ calculated with full FCIQMC for different extensive truncations based on the doublon–holon criterion and with initiator-FCIQMC, respectively. The points for the extensive truncations are put at the N_{min} of the respective subspace. Population-based initiator calculations are performed for the same walker numbers. For the extensive spaces, numbers with and without importance sampling using $|\Psi_{\mathrm{GL}}\rangle$ are compared. Importance sampling is not effective in initiator-FCIQMC.

The analytical wavefunction ansatzes can be evaluated in a discrete feature-based manner which allows for fast on-the-fly calculations. This also means that extensive, exponentially scaling subspaces can be selected which is necessary in the real-space Hubbard model with its spread-out ground-state wavefunctions.

Figure 8.3 shows that subspaces truncated based on this criterion are indeed extensive. To demonstrate this, we again use Hubbard lattices in ladder geometry. Due to the exponential scaling, the ground-state energy E_0^S of the subspace in this case roughly recovers a constant amount of the energy E_0. Dynamically sampled initiator subspaces based on the usual population-based initiator criterion do not scale exponentially. Therefore, E_0/E_0^S always falls off at some ℓ, regardless of the number of walkers.

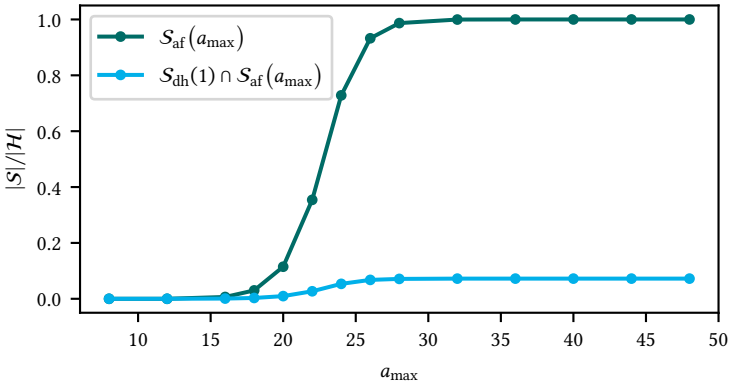


Figure 8.5. Size of the subspace $|\mathcal{S}|$ as a fraction of the complete Hilbert space size $|\mathcal{H}|$ for a 2×8 ladder-type Hubbard lattice as a function of $a_{\max}$. For the blue curve, $\mathcal{S}_{\mathrm{af}}$ is the only constraint. For the orange curve, there is the additional $\mathcal{S}_{\mathrm{dh}}$ constraint with $k_{\max} = 1$ which restricts $|\mathcal{S}|$ even for maximum $a_{\max}$ which is 48 for this lattice.

This then leads to the observation that an extensive truncation based on $|\Psi_{\mathrm{dh}}\rangle$ leads to improved results also for larger $k_{\max}$ when comparing them for the same number of walkers. This is shown in figure 8.4a for a 4×4 lattice at $U/t = 8$. The energy in the subspace is calculated and $N_{\min}$ in the subspace is determined. Subsequently, an initiator calculation for that number of walkers is performed and the obtained energies are compared. One can see that for $k_{\max} \geq \sqrt{2}$, the energy of the extensive subspace is consistently better than the result of initiator-FCIQMC. Furthermore, it is possible to lower $N_{\min}$ in the extensive subspaces by using importance sampling with $|\Psi_{\mathrm{GL}}\rangle$ which is shown in figure 8.4b. The extensive spaces now give a better energy estimate for all $k_{\max}$.

8.1.2 *Truncation Based on Antiferromagnetic Wavefunction Ansatz*

The truncation inspired by $|\Psi_{\mathrm{af}}\rangle$ works in a similar fashion as the doublon–holon criterion. To this end, we define a penalty score a. It is defined as

$$a = \sum_{\langle ij \rangle} s_{ij} + N_{\mathrm{links}} \tag{8.6}$$

$$\text{with} \quad s_{ij} = \begin{cases} -1 & \text{if } i = |\uparrow\rangle\,,\ j = |\downarrow\rangle \text{ or } i = |\downarrow\rangle\,,\ j = |\uparrow\rangle\,, \\ 0 & \text{if } i = |\cdot\rangle\,,\ i = |\uparrow\downarrow\rangle\,,\ j = |\cdot\rangle\,, \text{or } j = |\uparrow\downarrow\rangle\,, \\ 1 & \text{if } i = j = |\uparrow\rangle \text{ or } i = j = |\downarrow\rangle\,. \end{cases}$$

N_{links} is the total number of links between sites. For a genuine 2-d square lattice, it is $2N_{\mathrm{s}}$; for a honeycomb lattice, it is $1.5N_{\mathrm{s}}$. Again, $\langle ij \rangle$ is the sum over nearest neighbours. a is small if there are many $\uparrow$-electrons surrounded by many $\downarrow$-electrons in a given determinant or vice versa. For the half-filled case, $a = 0$ is only true for the two Néel determinants; the maximum

possible value is $2N_{\text{links}}$. The fixed initiator subspace $\mathcal{S}_{\text{af}}\left(a_{\max}\right)$ only includes determinants for which $a \leq a_{\max}$.

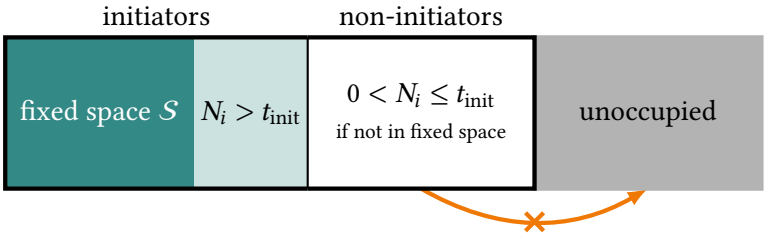
Again, $\mathcal{S}_{\text{af}}$ is extensive, i.e. scales exponentially with system size. It has to be noted that the Hilbert space size does not scale linearly with $a_{\max}$ but is rather a sigmoid function. The change in Hilbert space size from $a_{\max}$ to $a_{\max} + 1$ is largest for intermediate $a_{\max}$. The sigmoid function is not symmetric around the inflection point due to the existence of doubly occupied sites that do not contribute to a. The size $\left|\mathcal{S}_{\text{af}}\right|$ – both with and without the additional $\mathcal{S}_{\text{dh}}(1)$ constraint – as a fraction of the size of the whole space $|\mathcal{H}|$ is shown in figure 8.5 for a 2×8 ladder geometry.

8.2 *The Two-Shift Method*

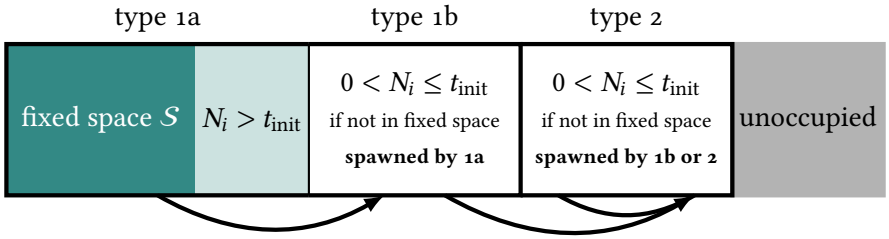
Although we have shown that applying FCIQMC in subspaces based on truncation criteria from analytical wavefunction ansatzes lead to improved results compared to initiator-FCIQMC, so far these results are not systematically improvable by adding more walkers. If the wavefunction ansatz is too coarse, adding more determinants into the subspace may be too expensive in terms of the sign problem but necessary to get a satisfactory result. To this end, I will introduce a scheme where one can still benefit from knowing important subspaces with a very weak sign problem but the outside space can be included in a perturbative manner. This inclusion will happen in an approximate way to not let the sign problem of the remaining part dominate the simulation.

The method will be based on the initiator-like method with a fixed initiator subspace $\mathcal{S}$ that was presented in section 8.1. The original initiator scheme as described in section 3.2.3 is depicted in figure 8.6a. It strictly truncates all spawns from non-initiators (all determinants not contained in $\mathcal{S}$ or fulfilling the occupation criterion $0 < N_i \leq t_{\text{init}}$) to empty determinants. This constraint will now be relaxed in a controlled manner.

In a first step, the spawning constraints are released all together. With a too small number of walkers, there would not be enough annihilations to stabilise the fermionic solution. As described in section 3.2, one would get a meaningless superposition of $|\Psi\rangle$ and $-|\Psi\rangle$. Like in chapter 7, also here the fixed-N_0 method, that employs a shift update according to equation (3.11), will be used to ensure a sign-coherent solution in the subspace $\mathcal{S}$. If not specified otherwise, fixed-$N_0 = 50$ will be used for all calculations as a compromise



(a) Original initiator-like method with fixed $\mathcal{S}$. No spawns are allowed from non-initiators onto empty determinants (with the exception of the multiple-spawns rule). The global shift S is applied to every occupied determinant.



(b) Two-shift method. All spawns are allowed. Shift S_1 is applied to type-1 determinants to target a minimum population on the reference determinant to ensure sign coherence. Shift S_2 is applied to type-2 determinants to target a total overall population N_{tot}.

Figure 8.6. Comparison of the initiator-like method with a fixed very-weak-sign-problem subspace $\mathcal{S}$ and its extension by the two-shift method.

between minimising the necessary total population and controlling stochastic noise.

8.2.1 *Application of Different Shifts Depending on Determinant History*

As mentioned before, N_{min} of the overall space may be too large. $\mathcal{S}$ however is specifically designed such that N_{min} is reachable with available computational resources. Therefore, a shift S_1 is applied only to the determinants in the extensive subspace $\mathcal{S}$ and the dynamically changing additional determinants that fulfill $N_i > t_{\text{init}}$. These determinants will be called *type-1a determinants.* Additionally, S_1 is applied to determinants that have been spawned from initiators. In the original scheme, they are called non-initiators. Here, they will be called *type-1b determinants.*

By construction, the outside space $\mathcal{S}^{\text{c}} := \mathcal{H} \setminus \mathcal{S}$ with $N_i \leq t_{\text{init}}$ has a comparatively strong sign problem but relatively weak contributions into the ground-state wavefunction. We therefore apply a dynamically adjusted second shift S_2 to determinants that are neither type-1a nor type-1b determinants. They will be called *type-2 determinants.* In line with the above definitions, these are determinants that have either been spawned by type-1b or other type-2 determinants. If an already occupied type-2 determinant is subsequently spawned upon by a type-1a determinant, it is marked as a type-1b determinant. If a type-1b or a type-2 determinant exceeds t_{init} at the

end of an iteration, it is marked as a type-1a determinant for the subsequent iteration.

It is noted that for $S_2 \to -\infty$ the two-shift method reverts back to the original initiator-like method with fixed $\mathcal{S}$ with the type-1 determinants being initiators, type-1b determinants being non-initiators and type-2 determinants being constantly unoccupied. On the other hand, for $S_2 = S_1$ we recover the full method without approximation. This can be used for extrapolation purposes as we will see later.

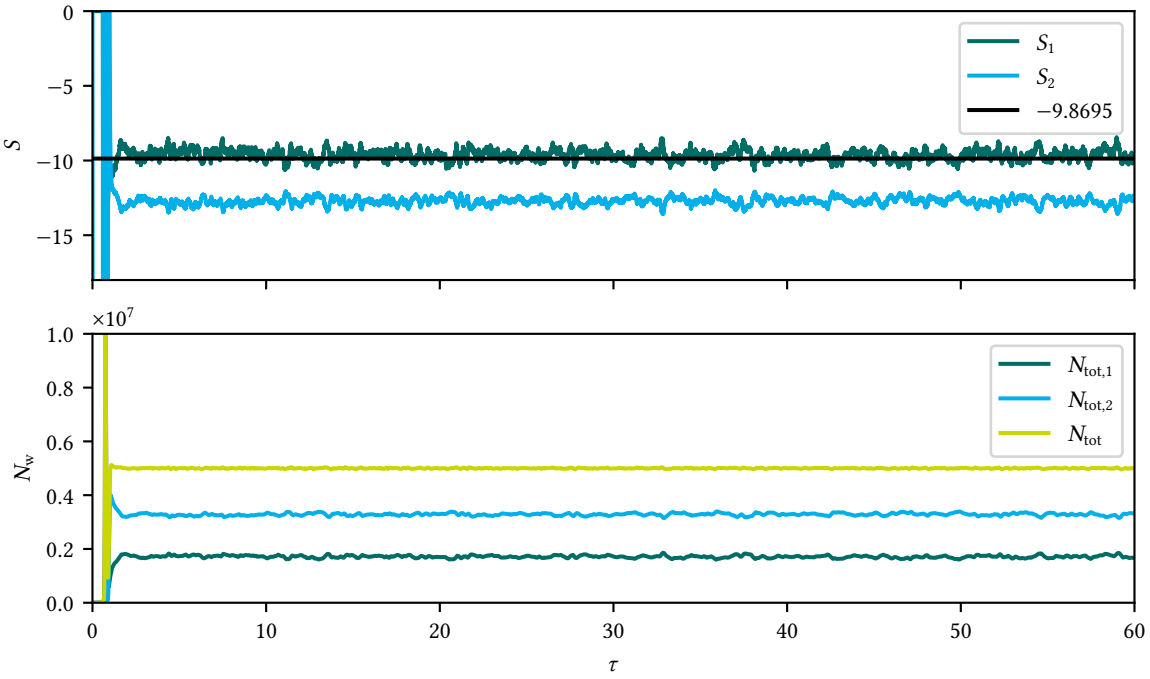


Figure 8.7. Simulation dynamics of the shifts S_1 and S_2 and the walker numbers $N_{\text{tot},1}$, $N_{\text{tot},2}$, and N_{tot} in a two-shift calculation as a function of imaginary time τ. The system is the half-filled 18-site 2-d Hubbard model at $U/t = 8$. $N_{\text{tot}} = 5 \times 10^6$ walkers is chosen. Since $N_{\text{tot}} < N_{\text{min}}$, $\overline{S_1} > E_0$ and $\overline{S_2} < E_0$. $N_{\text{tot},1}$ and $N_{\text{tot},2}$ add up to the desired N_{tot}.

We will now discuss how the shifts S_1 and S_2 are adjusted during a simulation. S_1 is adjusted according to equation (3.11). It fixes the population on the reference determinant which has to be contained in $\mathcal{S}$. It ensures sign coherence in the type-1 space and therefore will lead to a population $N_{\text{tot},1}(\tau)$. After the equilibration period, it will fluctuate around an average population of $\overline{N_{\text{tot},1}}$. Also defined is a total number of walkers N_{tot} that can be handled by the computational resources at hand. It has to be larger than the maximum value of $N_{\text{tot},1}$ at (almost) every given τ during the simulation, otherwise the run is not stable and the simulation variables cannot be averaged.

The excess walkers can be used in the type-2 space to correct the residual bias that comes from the truncation of the type-1 space. S_2 is adjusted in

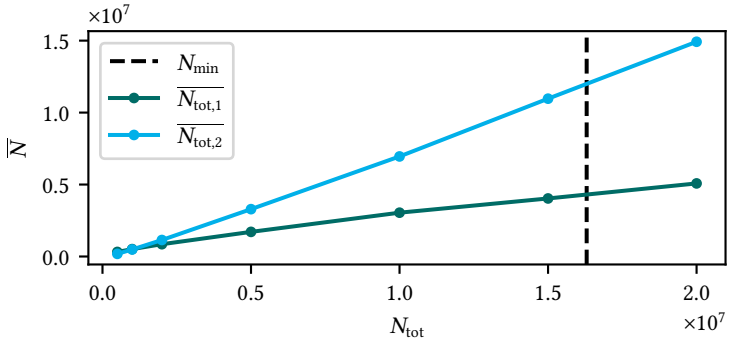


Figure 8.8. Average number of walkers on type-1 determinants ($\overline{N_{\text{tot},1}}$) and on type-2 determinants ($\overline{N_{\text{tot},2}}$), respectively, as a function of the total number of walkers N_{tot} for a half-filled 18-site tilted real-space Hubbard system at $U/t = 8$. $|\Psi_{\text{g}}\rangle$ is applied with $g = 0.15$. $|\Psi_{\text{dh}}\rangle$ with $k_{\max} = 1$ is used to define the subspace $\mathcal{S}$. $\overline{N_{\text{tot},1}}$ increases because more walkers are required to compensate for the non-sign-coherent influx from the type-2 space when increasing N_{tot}. Still, the largest fraction of the increase in N_{tot} goes into the increase of $\overline{N_{\text{tot},2}}$ which corrects the energy estimator $\overline{S_1}$.

every iteration such that the population in the type-2 space is given by $N_{\text{tot},2}(\tau) = N_{\text{tot}} - N_{\text{tot},1}(\tau)$. This can be achieved by updating S_2 according to

$$S_2(\tau) = S_2(\tau - \Delta\tau) - \frac{\gamma}{\Delta\tau} \ln\left(\frac{N_{\text{tot},2}(\tau)}{N_{\text{tot},2}(\tau - \Delta\tau)}\right) - \frac{\gamma^2}{4\Delta\tau} \ln\left(\frac{N_{\text{tot},2}(\tau)}{N_{\text{tot}} - N_{\text{tot},1}(\tau)}\right). \tag{8.7}$$

Instead of just leading to a constant population with only the first two terms, the shift update now takes the total population to the target population specified in the denominator of the logarithm $N_{\text{tot}} - N_{\text{tot},1}(\tau)$. As introduced in section 3.1.1, the prefactor $\gamma^2/4$ is chosen to achieve critical damping. In our case, the denominator is not constant but rather changes due to stochastic fluctuations of $N_{\text{tot},1}$. Yet, after the equilibration period also $N_{\text{tot},2}$ fluctuates around a constant average number $\overline{N_{\text{tot},2}}$.

The sign problem is not fully resolved in the type-2 space if N_{tot} is below N_{min} of the entire system. Therefore, $S_2 < S_1$ in that regime because there are excess walkers in the type-2 space that are not annihilated due to underpopulation. If $N_{\text{tot}} = N_{\text{min}}$, $S_2 = S_1$. For $N_{\text{tot}} > N_{\text{min}}$, consequently $S_2 > S_1$.

Figure 8.7 shows the simulation dynamics of both shifts S_1 and S_2 and the walker numbers $N_{\text{tot},1}$, $N_{\text{tot},2}$, and N_{tot} for a target population $N_{\text{tot}} = 5 \times 10^6 < N_{\text{min}}$. As expected, S_1 fluctuates around a value larger than E_0, S_2 fluctuates around a value smaller than E_0. Due to the interdependent population control, $N_{\text{tot},1}$ and $N_{\text{tot},2}$ add up to the desired N_{tot}.

Below N_{min}, S_1 is used as an approximate estimator to the exact energy. As the type-1 space now gets additional contributions from the type-2 space that were not there in the original method, $\overline{S_1}$ is always lowered compared to the ground-state energy $E_0^{\mathcal{S}}$ of $\mathcal{S}$ alone. It has to be noted that there can be spawns from the type-2 into the type-1 space that are not sign-coherent. Thus, increasing N_{tot} will also increase $\overline{N_{\text{tot},1}}$ starting from $N_{\text{min}}^{\mathcal{S}}$ of $\mathcal{S}$ because more walkers are needed in the type-1 space to guarantee sign coherence

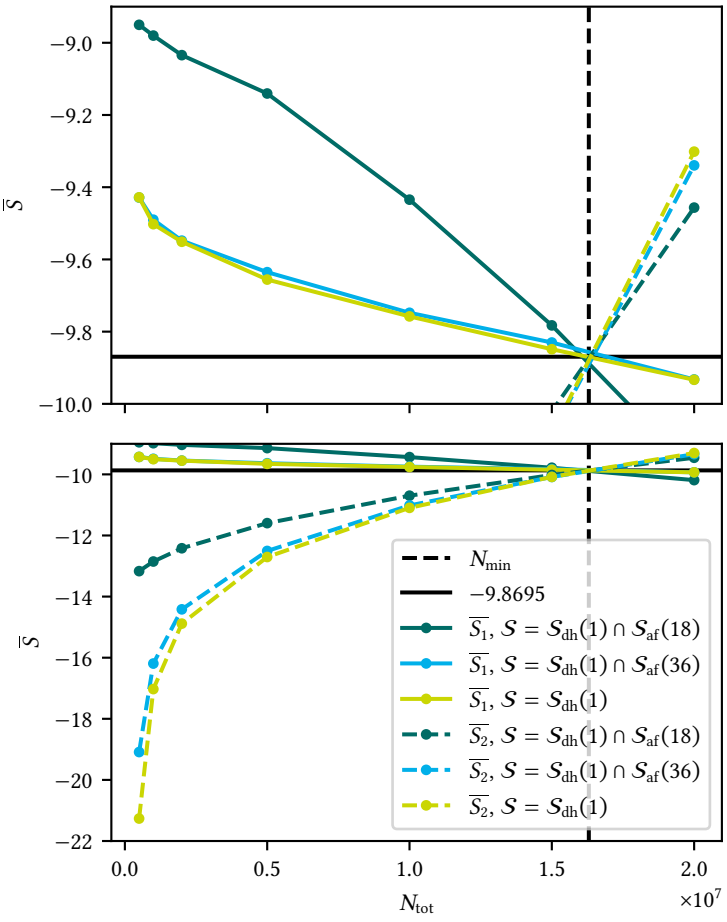


Figure 8.9. Average shifts $\overline{S_1}$ and $\overline{S_2}$ for a half-filled 18-site tilted real-space Hubbard system at $U/t = 8$ calculated with the two-shift method with respect to the total number of walkers. The green and blue curves are based on calculations with subspaces based on the doublon–holon ($k_{\text{max}} = 1$) and the antiferromagnetic ($a_{\text{max}} = 18$ and $a_{\text{max}} = 36$, respectively) truncation criterion, respectively. The yellow curve is based on the doublon–holon criterion with $k_{\text{max}} = 1$ only. A smaller subspace leads to a worse energy estimate via $\overline{S_1}$ for the same N_{tot}. All curves intersect at the exact energy for a number of walkers at the system's N_{min} with the Gutzwiller-like guiding wavefunction $|\Psi_{\text{GL}}\rangle$ with $g = 0.15$ applied.

there. Figure 8.8 shows $\overline{N_{\text{tot},1}}$ and $\overline{N_{\text{tot},2}}$ for increasing N_{tot} for a half-filled 18-site real-space Hubbard lattice.

Figure 8.9 shows the convergence of the average shifts $\overline{S_1}$ and $\overline{S_2}$ with respect to N_{tot} for different choices of the subspace $\mathcal{S}$. For a larger $\mathcal{S}$, the energy estimates via $\overline{S_1}$ are consistently lower for all walkers numbers below N_{min}. The shifts S_2 acting in the type-2 space are also consistently lower for larger $\mathcal{S}$. Regardless of $\mathcal{S}$, $\overline{S_1}$ and $\overline{S_2}$ intersect at $N_{\text{tot}} = N_{\text{min}}$ (for a given guiding wavefunction) with $\overline{S_1}$ and $\overline{S_2}$ being on top of the exact energy.

The two-shift method is related to the adaptive-shift method [57, 58] and the initiator-FCIQMC method corrected by second-order Epstein–Nesbet perturbation theory (EN2) [59]. The key difference to the adaptive-shift algorithm lies in the fact that the two-shift method allows for free spawning in the entire Hilbert space. Thus, it provides a correction accounting for contributions from all of the outside space. The adaptive-shift algorithm only corrects for contributions in the direct vicinity of the initiator space. Due to the spread-out nature of the sampled wavefunctions, this is not very effective in the real-space Hubbard model. The EN2-corrected initiator-FCIQMC method relies on the replica method [206] because two statistically independent samples of the wavefunction are required to evaluate the second-order energy contribution. This poses a problem for the real-space Hubbard model as the overlap between two independent samples of the wavefunction at a given instance typically goes to zero when using extensive initiator spaces. The two-shift method presented here does not rely on the replica method.

Similar in spirit is the usage of a sign-flip potential to remove the sign problem [39]. In this scheme, sign-violating off-diagonal contributions are zeroed and folded into the diagonal matrix elements. This however requires the knowledge of a highly accurate trial wavefunction, both with respect to the sign and the amplitude structure. Furthermore, the trial wavefunction needs to be evaluated many times to determine the sign-flip potential as sums over all connected determinants need to be evaluated for every occupied determinant. This prohibits large-scale calculations. In contrast, the wavefunction ansatzes used to determine $\mathcal{S}$ in the two-shift method are feature-based and can be evaluated highly efficiently.

8.2.2 *Extrapolation to the Unbiased Ground-State Energy*

So far, fixed initiator subspaces, guiding wavefunctions, and the two-shift method have greatly improved the best approximate ground-state energy

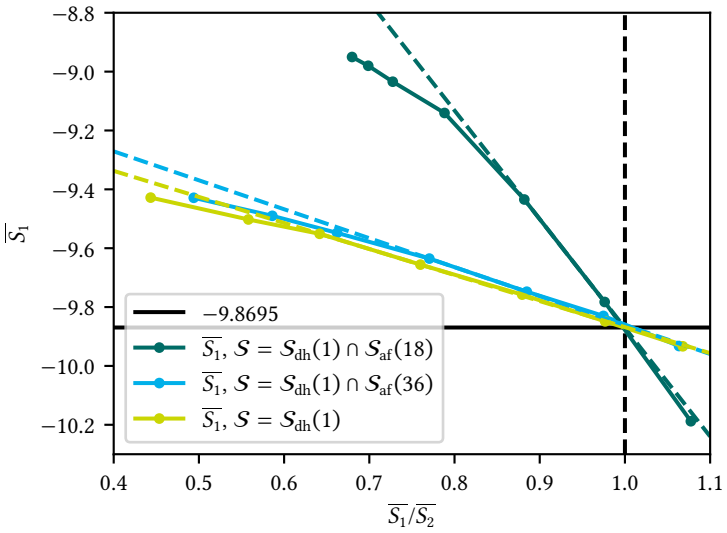


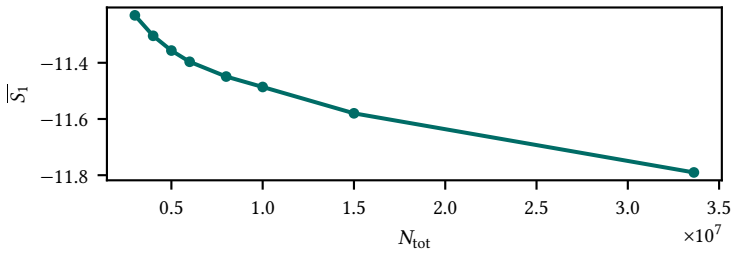
Figure 8.10. Extrapolation to the exact ground-state energy for the half-filled 18-site system at $U/t = 8$ with $g = 0.15$ for different subspace choices. The points in this plot are taken from the data shown in figure 8.9 at the respective total walker numbers. For $\overline{S_1}/\overline{S_2} = 1$, both shifts equal the exact energy but a linear extrapolation is possible for lower walker numbers. The dashed lines show extrapolations from 1×10^7 to 1.5×10^7 walkers for the green, from 5×10^6 to 1×10^7 walkers for the blue and from 2×10^6 to 5×10^6 walkers for the yellow curve.

	extrapolation			
$\mathcal{S}$	from …	to … walkers	E_0^{extrapol}	$\Delta E/E_0$
$\mathcal{S}_{\mathrm{dh}}(1) \cap \mathcal{S}_{\mathrm{af}}(18)$	2×10^6	5×10^6	−9.5107	3.64 %
	5×10^6	1×10^7	−9.8051	0.65 %
	1×10^7	1.5×10^7	−9.8705	0.01 %
$\mathcal{S}_{\mathrm{dh}}(1) \cap \mathcal{S}_{\mathrm{af}}(36)$	2×10^6	5×10^6	−9.8205	0.50 %
	5×10^6	1×10^7	−9.8611	0.08 %
	1×10^7	1.5×10^7	−9.8532	0.17 %
$\mathcal{S}_{\mathrm{dh}}(1)$	2×10^6	5×10^6	−9.8681	0.01 %
	5×10^6	1×10^7	−9.8608	0.08 %
	1×10^7	1.5×10^7	−9.8703	0.01 %

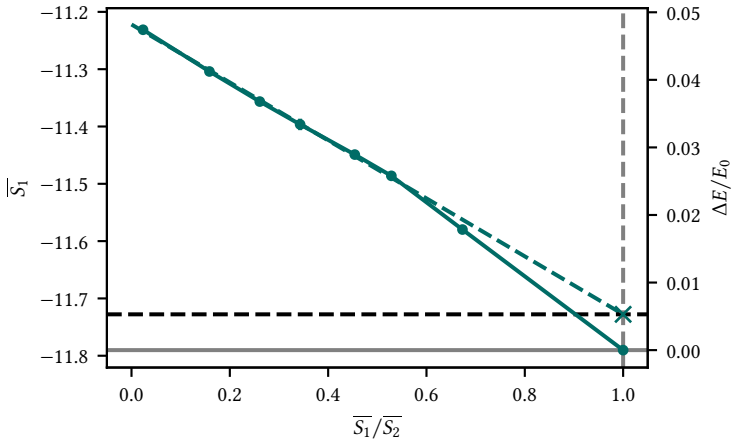
Table 8.2. Extrapolated energies E_0^{extrapol} for different subspace choices for the half-filled 18-site Hubbard model at $U/t = 8$. The linear extrapolation allows to determine the ground-state energy within a relative error $\Delta E/E_0$ of less than 0.1 % with walker numbers below N_{min}. For the largest subspace, an extrapolation from 2×10^6 to 5×10^6 walkers is already sufficient. For the smallest subspace $\mathcal{S}_{\mathrm{dh}}(1) \cap \mathcal{S}_{\mathrm{af}}(18)$, only extrapolations close to N_{min} (at 1.62×10^7 walkers for $g = 0.15$) return accurate results.

FCIQMC can return for systems with weak sign problems but very spread-out wavefunctions. Still, our solutions still have a systematic bias. Now, the fact that the shifts S_1 and S_2 intersect at the exact energy will be exploited for extrapolation.

An extrapolation to the point of equality is either possible using the difference $S_1 - S_2$ or the ratio S_1/S_2. We observe that the ratio actually enters a linear regime well below N_{min} as can be seen in figure 8.10 for the half-filled 18-site problem. This allows for a linear extrapolation to $S_1/S_2 = 1$. It is apparent that the linear regime is reached for a higher number of walkers the smaller $\mathcal{S}$ is chosen. Therefore, it is always beneficial to use the largest possible subspace in which N_{min} is still reached with the requested total number of walkers N_{tot}. The numerical extrapolation results are listed in table 8.2.



(a) Total walker numbers used.



(b) Extrapolation.

Figure 8.11. Extrapolation to the exact ground-state energy for the 18-site system with one hole at $U/t = 8$ with $g = 0.15$ for a subspace choice $S_{\text{dh}}(1)$. The extrapolation using calculations with only up to 5×10^6 walkers allows one to calculate a ground-state energy of $-11.728(2)$ which corresponds to a relative error of 0.5 %.

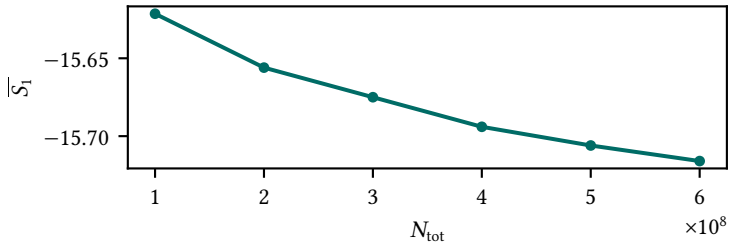
8.3 *Results for Systems with One Hole*

In this section, I will present results on the 18-site tilted square lattice and 32-site honeycomb lattice systems in the challenging intermediate interaction regime at $U/t = 8$, each with one hole. These systems show a sign problem even in AFQMC (see section 2.2.3). The 32-site system is well beyond the scope of exact diagonalisation.

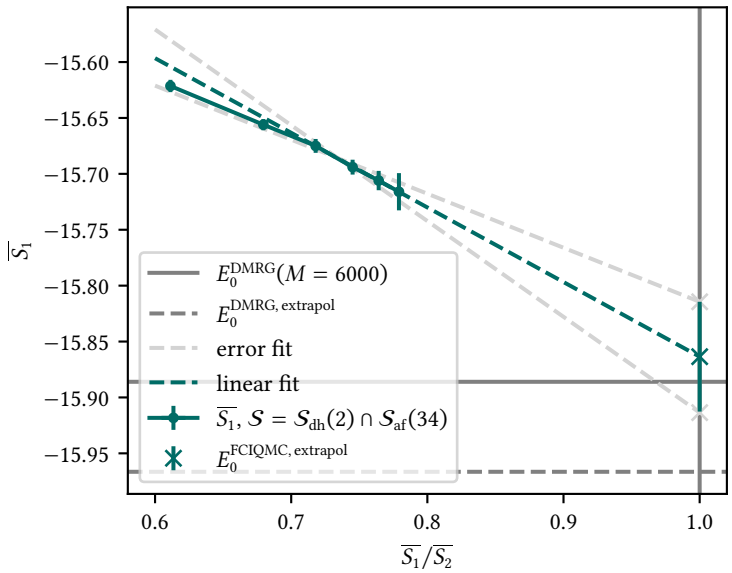
8.3.1 *18-Site Tilted Lattice*

I will first look at the 18-site tilted lattice at $U/t = 8$ again but now with one hole. This system has a stoquastised gap of $\Delta E^{\text{stoq}} = 5.287$, compared to $\Delta E^{\text{stoq}} = 3.697$ for the half-filled system. With importance sampling with $g = 0.15$, N_{min} is located at 3.36×10^7 walkers.

Figure 8.11 shows both $\overline{S_1}$ as a function of N_{tot} and the extrapolation of $\overline{S_1}$ as a function of $\overline{S_1}/\overline{S_2}$. The subspace used for the type-1a determinants is $S_{\text{dh}}(1)$ without a truncation based on the antiferromagnetic criterion. From three calculations with walker numbers $N_{\text{tot}} = 3 \times 10^6$, 4×10^6, and 5×10^6, a linear extrapolation to $E_0^{\text{extrapol}} = -11.728(2)$ is possible. When comparing to the converged ground-state energy $E_0 = -11.7902(8)$, this is a relative error



(a) Total walker numbers used.



(b) Extrapolation.

Figure 8.12. Extrapolation to the exact ground-state energy for the 32-site honeycomb system with one hole at $U/t = 8$ with $g = 0.15$ for a subspace choice $\mathcal{S}_{\text{dh}}(2) \cap \mathcal{S}_{\text{af}}(34)$. Linearly extrapolating to $\overline{S_1}/\overline{S_2}$ leads to an estimate of the ground-state energy that agrees with an $M = 6000$ DMRG reference within statistical errors. The errors were obtained using linear error fits (light grey) that were chosen to match with the statistical errors of the individual calculations.

of $\Delta E/E_0 = 0.5\,\%$ with $\Delta E = E_0^{\text{extrapol}} - E_0$. This is a significant reduction when comparing to a single-shot two-shift calculation with 5×10^6 walkers which yields $E_0\left(N_{\text{tot}} = 5 \times 10^6\right) = -11.357$ which corresponds to a relative error of $\Delta E/E_0 = 3.7\,\%$.

8.3.2 *32-Site Honeycomb Lattice*

The half-filled 32-site problem on a honeycomb structure, with its lattice structure depicted in figure 7.11, has been calculated in an exact manner using importance-sampled FCIQMC in section 7.3.2. For the system with one hole, this is no longer possible. Even when applying importance sampling using the Gutzwiller-like guiding wavefunction, N_{min} cannot be reduced sufficiently to match available computational resources. Due to the very weak sign problem due to the six-site smallest loop (see section 5.2.2) however, the two-shift method together with the extrapolation scheme can be applied to get a ground-state energy estimate.

Figure 8.12 again shows $\overline{S_1}$ as a function of the N_{tot} used (between 1×10^8 and 6×10^8 walkers) and the extrapolation of $\overline{S_1}$ as a function of $\overline{S_1}/\overline{S_2}$. Empirically, it is found that a subspace choice $\mathcal{S}_{\text{dh}}(2) \cap \mathcal{S}_{\text{af}}(34)$ is optimal for a walker number regime on the order of 10^8 walkers. Especially, it is superior to restricting the subspace to $\mathcal{S}_{\text{dh}}(1)$ and releasing the $\mathcal{S}_{\text{af}}$ constraint. Extrapolating between 3×10^8 and 6×10^8 walkers leads to an estimated ground-state energy $E_0^{\text{FCIQMC, extrapol}} = -15.864(49)$. The error was obtained using the error fits shown in light grey. A DMRG benchmarks yields a variational ground-state energy $E_0^{\text{DMRG}} = -15.8861$ for $M = 6000$ which lies within statistical errorbars of the FCIQMC extrapolated estimate. Linearly extrapolating the DMRG energies as a function of $1/M$ from $M = 5000$ to 6000 yields $E_0^{\text{DMRG, extrapol}} = -15.9665$. This is not a variational estimator however and may not be reliable as the linear regime of DMRG might not have been reached for these bond dimensions.

9 *Summary & Outlook*

In this thesis, the applicability of FCIQMC was enlarged beyond the scope of *ab initio* systems and the reciprocal-space Hubbard model. The special characteristics of the real-space Hubbard model in conjunction with the FCIQMC algorithm were studied. Due to this, it was possible to develop algorithms to reduce or even remove systematic biases that are unusually strong in these kinds of systems.

9.1 *Summary*

Firstly, the emergence of a sign problem in the aforementioned lattice models was studied. It was established that there are certain non-trivial sign-problem-free lattice systems in FCIQMC: Apart from the already widely known fact that the 2-d Heisenberg model is sign-problem-free, it was proven that there are certain 1-d Hubbard systems that also do not exhibit a sign problem. A rule for when this is the case that only depends on the number of ↑- and ↓-spins was given: For periodic (antiperiodic) boundary conditions, a 1-d Hubbard chain is sign-problem-free for an odd (even) number of ↑- and an odd (even) number of ↓-electrons. In 1-d Hubbard systems that do show a sign problem, it was found that the stoquastised gaps, which are important and easy-to-calculate estimators of the strength of the sign problem, decrease with system size, i.e. the sign problem is non-size-extensive. The influence of other system parameters such as filling and on-site interaction strength was studied. Moving to 2-d systems, in FCIQMC there is an inevitable size-extensive sign problem. Compared to *ab initio* systems and the Hubbard model in a reciprocal-space basis however, the sign problem is weak in comparison. Its strength is mainly determined by the number and the size of the innermost sign-problematic loops. Therefore, it was found that 2-d ladder systems and the honeycomb lattice structure both show especially weak sign problems.

Secondly, when trying to solve large sign-problem-free systems, it was discovered that the results are biased due to a systematic effect in the FCIQMC algorithm. This had previously been masked by other systematic biases mainly caused by the sign problem. It was established that the newly discovered bias is caused by a non-vanishing covariance between the population

control parameter, the shift, and the sampled wavefunction. Thus, it was named population control bias. It was found that the bias could be reduced substantially by introducing importance sampling to FCIQMC. Already a simple Gutzwiller-like guiding wavefunction that could be evaluated with almost no computational overhead sufficed to remove the population control bias by large amounts. The Gutzwiller-like guiding wavefunction was compared to the full Gutzwiller wavefunction that generates system-size-dependent overhead. By employing an a-posteriori reweighting procedure that removed the remaining correlation between the shift and the wavefunction, the bias could be removed practically entirely. With this, it was possible to calculate the ground-state energies of the 1-d Hubbard model at $U/t = 8$ and 4 with 102 lattice sites at half-filling and with four holes and of the half-filled 150-site Hubbard model at $U/t = 8$ in good agreement with DMRG benchmarks and analytical results from the Bethe ansatz. Furthermore, the fundamental gaps of the 1-d Hubbard chains were calculated. They were calculable close to the thermodynamic limit at 102 sites due to the non-size-extensive character of the sign problem. The many-particle gaps calculated using FCIQMC are in good agreement with other high-accuracy methods.

Thirdly, the effect of applying importance sampling to sign-problematic systems was studied. It was discovered that applying a Gutzwiller-like guiding wavefunction significantly reduces the minimum number of walkers to obtain an unbiased ground-state energy in weakly sign-problematic cases. This happens even though the stoquastised gap remains unchanged when sampling according to the respective similarity-transformed Hamiltonian which led to the definition of the FCIQMC-related (relative) strength of the sign problem. The reason for the reduction of the minimum walker number to resolve the sign problem was found in the fact that the effectiveness of the annihilation process is significantly improved. This is because the ℓ_1 norm of the wavefunction is shifted towards Slater determinants with small diagonal element which corresponds to a compactification of the wavefunction. With this, it was possible to calculate the fundamental many-particle gaps of $2 \times \ell$ Hubbard ladder systems in the intermediate interaction regime up to $\ell = 12$. Additionally, the ground-state energy of a 32-site honeycomb lattice at $U/t = 8$, a large yet very weakly sign-problematic system, could be calculated using importance-sampled FCIQMC.

Fourthly, a new approximate method was developed that is specifically tailored towards weakly sign-problematic Hubbard systems: the two-shift method. The two-shift method is complementary to the usual population-

based initiator method which turns out to be a crude approximation in systems like the real-space Hubbard model with highly spread-out wavefunctions. The two-shift method is based on the predefinition of initiator subspaces with very weak sign problems that already contain a large fraction of the total wavefunction weight. The predefinition of initiator subspaces also allows for the use of the previously developed importance sampling using a Gutzwiller-like guiding wavefunction. To define these, Hubbard wavefunction ansatzes like the doublon–holon and the antiferromagnetic ansatz were used. Wavefunction contributions from outside the exactly treated subspace are included in a perturbative manner by applying a separate shift to them, still allowing them to contribute but preventing the dominating stoquastised signal from growing. It was also shown that extrapolation to the unbiased ground-state energy is possible which allowed ground-state calculations of systems with one hole with up to 32 sites in a honeycomb lattice geometry.

9.2 *Future Outlook*

The fundamental understanding and algorithmic developments about weak sign problems in real-space lattice systems in this thesis opens up related research topics and, among others, will have to be persued in future work:

- With the ability to calculate unbiased ground-state energies more sign-problem-free systems, especially systems containing bosons instead of fermions, can be studied. It remains an open question whether simple guiding wavefunctions, like the ones applied in this thesis, will be sufficient also in other systems.

- The effect of more sophisticated wavefunction ansatzes compared to the Gutzwiller-like guiding wavefunction could be evaluated for sign-problematic lattice models.

- The work on lattice model systems with quantum-chemical methods quite naturally leads to the question whether the newly developed concepts can be applied to molecular *ab initio* systems as well. In a previous work, it was shown that in a variety of systems choosing localised orbitals leads to weaker sign problems compared to a set of delocalised orbitals [39]. It will be interesting to see whether concepts like importance sampling or the two-shift method also work in these kinds of systems and which wavefunction ansatzes will work well.

A *Appendix*

A.1 *Lattice Geometries*

In figure A.1, a synopsis of all lattice structures used throughout this thesis is shown. Lattice sites highlighted in orange are part of the respective finite bulk. Grey solid lines indicate nearest-neighbour connections. Dashed lines indicate possible periodic boundary conditions.

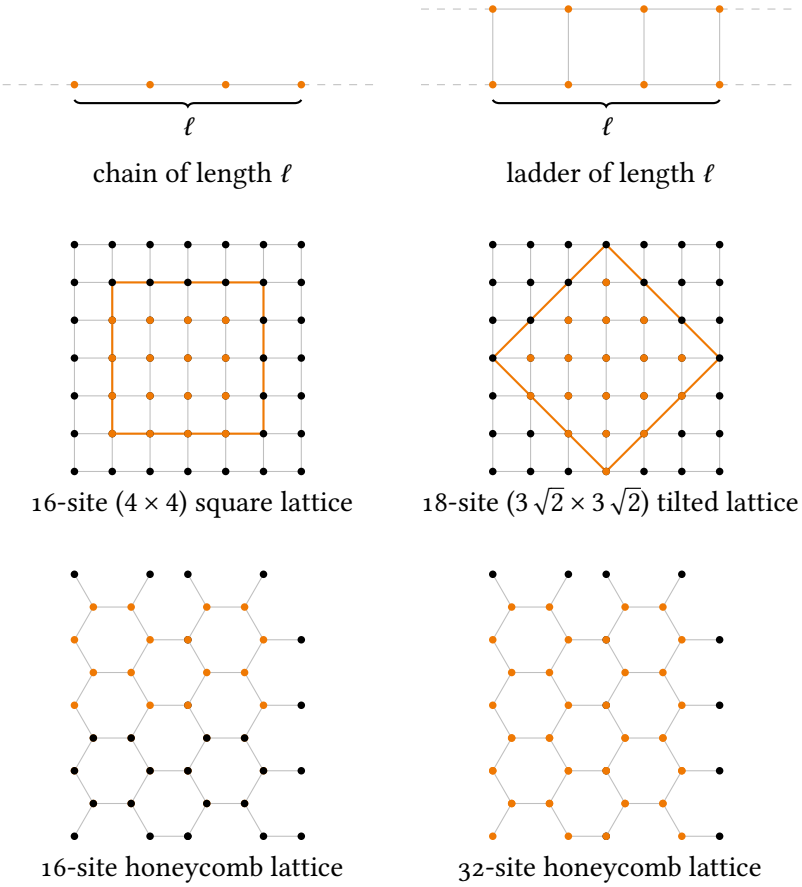


Figure A.1. Overview over the lattice geometries used throughout the thesis.

A.2 *Blocking Analysis*

Calculating the mean of a time series of data points $d(t_i)$ is simple. The sample mean is given by

$$\bar{d} = \frac{1}{n} \sum_{i}^{n} d(t_i) \tag{A.1}$$

where n is the number of data points. In this thesis, mostly energy estimators like the shift or projected energies are estimated using averaging. In practice, averaging estimators in FCIQMC runs is begun after an equilibration period when all simulation variables have settled to a constant value.

To determine the amount of dispersion of a data set, which can be used as a measure for the precision of the mean, the *standard deviation* can be used. It is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} = \sqrt{\frac{1}{N}\left(\overline{x^2} - \overline{x}^2\right)} \tag{A.2}$$

where N is the size of the entire population. When estimating the standard deviation from a sample with sample size n, one typically uses the *corrected sample standard deviation* which is given by

$$s_d = \frac{1}{n-1}\sum_{i=1}^{n}\left(d\left(t_i\right) - \overline{d}\right)^2 = \sqrt{\frac{1}{n-1}\left(\overline{d^2} - \overline{d}^2\right)}\,. \tag{A.3}$$

This is *Bessel's correction.*[14] The errorbars are then best estimated by

$$\Delta d = \frac{s_d}{\sqrt{n}} \tag{A.4}$$

[14] While s_d^2 is an unbiased estimator of the variance, s_d itself is still biased. Further corrections are dependent on the actual distribution of d.

It has to be noted that the mean value is independent of correlations within the time series [69, 207]. This is not true for the standard deviation of the mean. In FCIQMC calculations, one deals with time series with non-zero autocorrelation

$$R(s) = \frac{1}{n-2s}\sum_{i=1}^{n-s} d\left(t_i\right) d\left(t_{i+s}\right) \tag{A.5}$$

which is the covariance of the data with itself s instances later. For Markov processes, the autocorrelation falls off exponentially as a function of t, i.e.

$$R(t) \propto \exp\left(-t/t_{\mathrm{corr}}\right). \tag{A.6}$$

In this case, the best estimate of the standard deviation is given by

$$s_d = \sqrt{\frac{1 + \frac{2t_{\mathrm{corr}}}{\Delta t}}{n-1}\left(\overline{d^2} - \overline{d}^2\right)} \tag{A.7}$$

with the time step $\Delta t = t_{i+1} - t_i$. For $t_{\mathrm{corr}} \ll \Delta t$, the autocorrelated standard deviation estimator reverts back to the uncorrelated one.

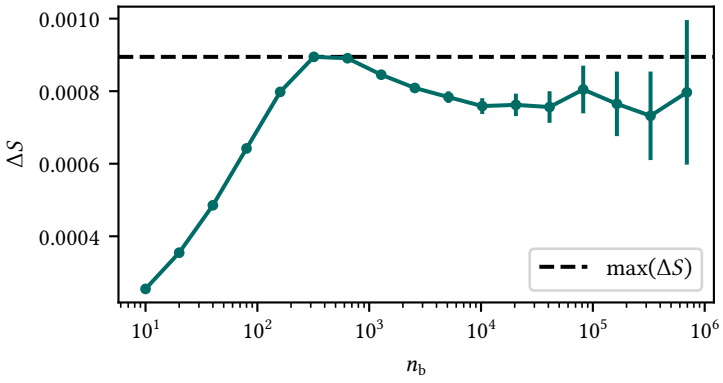


Figure A.2. Errorbars ΔS of the shift estimator as a function of block sizes n_b for the periodic 14-site Hubbard chain at $U/t = 8$. The dashed line indicates the value chosen for the errorbar of the mean $\overline{S}$.

In the usual case of $t_\mathrm{corr} \gg \Delta t$, evaluating equation (A.7) is impractical. Instead, one arranges the sample data into n_b blocks of increasing sizes n/n_b and for each block size calculates

$$s_d\left(n_\mathrm{b}\right) = \sqrt{\frac{1}{n_\mathrm{b} - 1} \sum_{j=1}^{n_\mathrm{b}} \left(d_j - \overline{d}\right)^2} \tag{A.8}$$

where d_j is the mean value in the j-th block. For small blocks, the standard deviation is underestimated due to the autocorrelation. The errorbar $s_d\left(n_\mathrm{b}\right)/\sqrt{n_\mathrm{b}}$ saturates if the block size is large enough such that the blocks become uncorrelated.

Therefore, in practice a calculation needs to be run until the saturation occurs. Subsequently, the largest $s_d\left(n_\mathrm{b}\right)/\sqrt{n_\mathrm{b}}$ value is chosen as the statistical errorbar for the respective calculation. An example of how the errorbar of the shift estimator in an FCIQMC calculation is chosen is shown in figure A.2.

Bibliography

[1] P. W. Anderson, “More Is Different”, *Science*, vol. 177, no. 4047, pp. 393–396, 1972.

[2] M. Le Bellac, *Quantum Physics*, Reprint Edition. Cambridge: Cambridge University Press, 2011, 606 pp.

[3] E. Schrödinger, “An Undulatory Theory of the Mechanics of Atoms and Molecules”, *Physical Review*, vol. 28, no. 6, pp. 1049–1070, 1926.

[4] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons: Theory and Computational Approaches*. Cambridge: Cambridge University Press, 2016.

[5] M. Born and J. R. Oppenheimer, “On the Quantum Theory of Molecules”, 1927.

[6] H. Haken and H. C. Wolf, *Molecular Physics and Elements of Quantum Chemistry: Introduction to Experiments and Theory*, Softcover reprint of hardcover 2nd ed. Berlin; Heidelberg: Springer, 2010, 620 pp.

[7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Revised ed. Mineola, New York: Dover Publications Inc., 1996, 480 pp.

[8] T. Helgaker, J. Olsen, and P. Jorgensen, *Molecular Electronic-Structure Theory*, Reprint Edition. Chichester, New York: Wiley-Blackwell, 2013, 940 pp.

[9] A. Avella and F. Mancini, *Strongly Correlated Systems: Theoretical Methods*. Berlin; Heidelberg: Springer, 2014, 496 pp.

[10] M. Amusia and V. Shaginyan, *Strongly Correlated Fermi Systems: A New State of Matter*, 1st ed. Basel: Springer, 2021, 404 pp.

[11] D. R. Hartree, “The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods”, *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 1, pp. 89–110, 1928.

[12] J. C. Slater, “Note on Hartree’s Method”, *Physical Review*, vol. 35, no. 2, pp. 210–211, 1930.

[13] W. Schattke and R. D. Muiño, *Quantum Monte-Carlo Programming: For Atoms, Molecules, Clusters, and Solids*, 1st ed. Wiley-VCH, 2013, 296 pp.

[14] J. Gubernatis, N. Kawashima, and P. Werner, *Quantum Monte Carlo Methods: Algorithms for Lattice Models*, Illustrated Edition. Cambridge: Cambridge University Press, 2016, 512 pp.

[15] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge: Cambridge University Press, 2017.

[16] 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*, vol. 131, no. 5, p. 054 106, 2009.

[17] 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*, vol. 132, no. 4, p. 041 103, 2010.

[18] K. Guther *et al.*, "NECI: N-Electron Configuration Interaction with an emphasis on state-of-the-art stochastic methods", *The Journal of Chemical Physics*, vol. 153, no. 3, p. 034 107, 2020.

[19] Zagoskin, *Quantum Theory of Many-Body Systems: Techniques and Applications*, 2nd ed. New York: Springer, 2014, 296 pp.

[20] W. Pauli, "The Connection Between Spin and Statistics", *Physical Review*, vol. 58, no. 8, pp. 716–722, 1940.

[21] I. Duck, *Pauli And The Spin-Statistics Theorem*. Singapore; River Edge, New Jersey: Wspc, 1998, 524 pp.

[22] W. Pauli, "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren", *Zeitschrift für Physik*, vol. 31, no. 1, pp. 765–783, 1925.

[23] G. E. Uhlenbeck and S. Goudsmit, "Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons", *Die Naturwissenschaften*, vol. 13, no. 47, pp. 953–954, 1925.

[24] P. A. M. Dirac and R. H. Fowler, "The quantum theory of the electron", *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 117, no. 778, pp. 610–624, 1928.

[25] H. Haken and H. C. Wolf, *Atomic and Quantum Physics: An Introduction to the Fundamentals of Experiment and Theory*, Softcover reprint of the original 2nd ed., trans. by W. D. Brewer. Berlin; Heidelberg: Springer, 1987, 474 pp.

[26] W. Heisenberg, "Mehrkörperproblem und Resonanz in der Quantenmechanik", *Zeitschrift für Physik*, vol. 38, no. 6, pp. 411–426, 1926.

[27] J. C. Slater, "The Theory of Complex Spectra", *Physical Review*, vol. 34, no. 10, pp. 1293–1322, 1929.

[28] P. A. M. Dirac and R. H. Fowler, "On the theory of quantum mechanics", *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 112, no. 762, pp. 661–677, 1926.

[29] O. Goldreich, *Computational Complexity: A Conceptual Perspective*, 1st edition. Cambridge; New York: Cambridge University Press, 2008, 632 pp.

[30] M. Troyer and U.-J. Wiese, "Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations", *Physical Review Letters*, vol. 94, no. 17, p. 170 201, 2005.

[31] M. Marvian, D. A. Lidar, and I. Hen, "On the computational complexity of curing non-stoquastic Hamiltonians", *Nature Communications*, vol. 10, no. 1, p. 1571, 2019.

[32] J. Klassen, M. Marvian, S. Piddock, M. Ioannou, I. Hen, and B. M. Terhal, "Hardness and Ease of Curing the Sign Problem for Two-Local Qubit Hamiltonians", *SIAM Journal on Computing*, vol. 49, no. 6, pp. 1332–1362, 2020.

[33] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 276, no. 1365, pp. 238–257, 1963.

[34] M. C. Gutzwiller, "Effect of Correlation on the Ferromagnetism of Transition Metals", *Physical Review Letters*, vol. 10, no. 5, pp. 159–162, 1963.

[35] J. Kanamori, "Electron Correlation and Ferromagnetism of Transition Metals", *Progress of Theoretical Physics*, vol. 30, no. 3, pp. 275–289, 1963.

[36] 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*, vol. 134, no. 2, p. 024 112, 2011.

[37] J. J. Shepherd, G. Booth, A. Grüneis, and A. Alavi, "Full configuration interaction perspective on the homogeneous electron gas", *Physical Review B*, vol. 85, no. 8, p. 081 103, 2012.

[38] J. S. Spencer, N. S. Blunt, and W. M. Foulkes, "The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method", *The Journal of Chemical Physics*, vol. 136, no. 5, p. 054 110, 2012.

[39] N. S. Blunt, "Fixed- and Partial-Node Approximations in Slater Determinant Space for Molecules", *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6092–6104, 2021.

[40] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, "A diffusion Monte Carlo algorithm with very small time-step errors", *The Journal of Chemical Physics*, vol. 99, no. 4, pp. 2865–2890, 1993.

[41] G. L. Warren and R. J. Hinde, "Population size bias in descendant-weighted diffusion quantum Monte Carlo simulations", *Physical Review E*, vol. 73, no. 5, p. 056 706, 2006.

[42] N. Nemec, "Diffusion Monte Carlo: Exponential scaling of computational cost for large systems", *Physical Review B*, vol. 81, no. 3, p. 035 119, 2010.

[43] M. Boninsegni and S. Moroni, "Population size bias in diffusion Monte Carlo", *Physical Review E*, vol. 86, no. 5, p. 056 712, 2012.

[44] M. P. Nightingale and H. W. J. Blöte, "Gap of the linear spin-1 Heisenberg antiferromagnet: A Monte Carlo calculation", *Physical Review B*, vol. 33, no. 1, pp. 659–661, 1986.

[45] N. Trivedi and D. M. Ceperley, "Ground-state correlations of quantum antiferromagnets: A Green-function Monte Carlo study", *Physical Review B*, vol. 41, no. 7, pp. 4552–4569, 1990.

[46] D. M. Ceperley and B. J. Alder, "Ground State of the Electron Gas by a Stochastic Method", *Physical Review Letters*, vol. 45, no. 7, pp. 566–569, 1980.

[47] D. Ceperley and B. Alder, "Quantum Monte Carlo", *Science*, vol. 231, no. 4738, pp. 555–560, 1986.

[48] P. J. Reynolds, J. Tobochnik, and H. Gould, "Diffusion Quantum Monte Carlo", *Computers in Physics*, vol. 4, no. 6, pp. 662–668, 1990.

[49] I. Kosztin, B. Faber, and K. Schulten, "Introduction to the diffusion Monte Carlo method", *American Journal of Physics*, vol. 64, no. 5, pp. 633–644, 1996.

[50] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, "Monte Carlo calculations of coupled boson-fermion systems. I", *Physical Review D*, vol. 24, no. 8, pp. 2278–2286, 1981.

[51] G. Sugiyama and S. E. Koonin, "Auxiliary field Monte-Carlo for quantum many-body ground states", *Annals of Physics*, vol. 168, no. 1, pp. 1–26, 1986.

[52] S. Zhang, "Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems", in *Emergent Phenomena in Correlated Matter*, ser. Modeling and Simulation, vol. 3, Jülich: Schriften des Forschungszentrums Jülich, 2013.

[53] M. Motta and S. Zhang, "Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method", *WIREs Computational Molecular Science*, vol. 8, no. 5, e1364, 2018.

[54] Simons Collaboration on the Many-Electron Problem *et al.*, "Solutions of the Two-Dimensional Hubbard Model: Benchmarks and Results from a Wide Range of Numerical Algorithms", *Physical Review X*, vol. 5, no. 4, p. 041 041, 2015.

[55] R. Levy and B. K. Clark, "Mitigating the Sign Problem through Basis Rotations", *Physical Review Letters*, vol. 126, no. 21, p. 216 401, 2021.

[56] N. S. Blunt, "A hybrid approach to extending selected configuration interaction and full configuration interaction quantum Monte Carlo", *The Journal of Chemical Physics*, vol. 151, no. 17, p. 174 103, 2019.

[57] K. Ghanem, A. Y. Lozovoi, and A. Alavi, "Unbiasing the initiator approximation in full configuration interaction quantum Monte Carlo", *The Journal of Chemical Physics*, vol. 151, no. 22, p. 224 108, 2019.

[58] K. Ghanem, K. Guther, and A. Alavi, "The adaptive shift method in full configuration interaction quantum Monte Carlo: Development and applications", *The Journal of Chemical Physics*, vol. 153, no. 22, p. 224 115, 2020.

[59] N. S. Blunt, "Communication: An efficient and accurate perturbative correction to initiator full configuration interaction quantum Monte Carlo", *The Journal of Chemical Physics*, vol. 148, no. 22, p. 221 101, 2018.

[60] E. U. Condon, "The Theory of Complex Spectra", *Physical Review*, vol. 36, no. 7, pp. 1121–1133, 1930.

[61] M. N. Rosenbluth, "Genesis of the Monte Carlo Algorithm for Statistical Mechanics", vol. 690, pp. 22–30, 2003.

[62] Rubinstein, *Simulation and the Monte Carlo Method*, 3rd ed. Hoboken, New Jersey: Wiley, 2016, 432 pp.

[63] T. Kloek and H. K. van Dijk, "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo", *Econometrica*, vol. 46, no. 1, pp. 1–19, 1978.

[64] 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*, vol. 21, no. 6, pp. 1087–1092, 1953.

[65] W. K. Hastings, "Monte Carlo Sampling Methods using Markov Chains and their Applications", *Biometrika*, vol. 57, pp. 97–109, 1970.

[66] A. J. Walker, "An Efficient Method for Generating Discrete Random Variables with General Distributions", *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 253–256, 1977.

[67] D. Ceperley, G. V. Chester, and M. H. Kalos, "Monte Carlo simulation of a many-fermion study", *Physical Review B*, vol. 16, no. 7, pp. 3081–3099, 1977.

[68] A. Ciric, *A Guide to Monte Carlo and Quantum Monte Carlo Methods: Quantum Monte Carlo: Variational and Diffusion; MC in General; Markov Chain; Statistics; Random Number Generators; Hidden Monte Carlo*, 1st ed. CreateSpace Independent Publishing Platform, 2016, 132 pp.

[69] P. R. C. Kent, *Techniques and Applications of Quantum Monte Carlo*. University of Cambridge, 1999.

[70] F. Becca, "Variational Wave Functions for Strongly Correlated Fermionic Systems", in *Many-Body Methods for Real Materials*, ser. Modeling and Simulation, vol. 9, Jülich: Schriften des Forschungszentrums Jülich, 2019.

[71] G. C. Wick, “Properties of Bethe-Salpeter Wave Functions”, *Physical Review*, vol. 96, no. 4, pp. 1124–1134, 1954.

[72] J. T. Krogel and D. M. Ceperley, “Population control bias with applications to parallel diffusion monte carlo”, *Advances in Quantum Monte Carlo*, ACS Symposium Series, pp. 13–26, 2012.

[73] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, “Fixed-node quantum Monte Carlo for molecules”, *The Journal of Chemical Physics*, vol. 77, no. 11, pp. 5593–5603, 1982.

[74] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, “Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule”, *The Journal of Chemical Physics*, vol. 144, no. 15, p. 151 103, 2016.

[75] M. Levy, “Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem”, *Proceedings of the National Academy of Sciences*, vol. 76, no. 12, pp. 6062–6065, 1979.

[76] D. M. Ceperley and B. J. Alder, “Quantum Monte Carlo for molecules: Green’s function and nodal release”, *The Journal of Chemical Physics*, vol. 81, no. 12, pp. 5833–5844, 1984.

[77] A. Lüchow and J. B. Anderson, “Accurate quantum Monte Carlo calculations for hydrogen fluoride and the fluorine atom”, *The Journal of Chemical Physics*, vol. 105, no. 11, pp. 4636–4640, 1996.

[78] N. M. Tubman, J. L. DuBois, R. Q. Hood, and B. J. Alder, “Prospects for release-node quantum Monte Carlo”, *The Journal of Chemical Physics*, vol. 135, no. 18, p. 184 109, 2011.

[79] R. L. Stratonovich, “On a Method of Calculating Quantum Distribution Functions”, *Soviet Physics Doklady*, vol. 2, p. 416, 1957.

[80] J. Hubbard, “Calculation of Partition Functions”, *Physical Review Letters*, vol. 3, no. 2, pp. 77–78, 1959.

[81] J. E. Hirsch, “Two-dimensional Hubbard model: Numerical simulation study”, *Physical Review B*, vol. 31, no. 7, pp. 4403–4419, 1985.

[82] M. Qin, H. Shi, and S. Zhang, “Benchmark study of the two-dimensional Hubbard model with auxiliary-field quantum Monte Carlo method”, *Physical Review B*, vol. 94, no. 8, p. 085 103, 2016.

[83] S. Zhang, J. Carlson, and J. E. Gubernatis, "Constrained Path Quantum Monte Carlo Method for Fermion Ground States", *Physical Review Letters*, vol. 74, no. 18, pp. 3652–3655, 1995.

[84] S. Zhang, J. Carlson, and J. E. Gubernatis, "Constrained path Monte Carlo method for fermion ground states", *Physical Review B*, vol. 55, no. 12, pp. 7464–7477, 1997.

[85] M. P. Nightingale and C. J. Umrigar, *Quantum Monte Carlo Methods in Physics and Chemistry* (NATO ASI Series). Dordrecht: Kluwer Academic, 1999, 467 pp.

[86] S. Zhang and H. Krakauer, "Quantum Monte Carlo Method using Phase-Free Random Walks with Slater Determinants", *Physical Review Letters*, vol. 90, no. 13, p. 136 401, 2003.

[87] J. Shee, S. Zhang, D. R. Reichman, and R. A. Friesner, "Chemical Transformations Approaching Chemical Accuracy via Correlated Sampling in Auxiliary-Field Quantum Monte Carlo", *Journal of Chemical Theory and Computation*, vol. 13, no. 6, pp. 2667–2680, 2017.

[88] J. Shee, E. J. Arthur, S. Zhang, D. R. Reichman, and R. A. Friesner, "Phaseless Auxiliary-Field Quantum Monte Carlo on Graphical Processing Units", *Journal of Chemical Theory and Computation*, vol. 14, no. 8, pp. 4109–4121, 2018.

[89] S. R. White, "Density matrix formulation for quantum renormalization groups", *Physical Review Letters*, vol. 69, no. 19, pp. 2863–2866, 1992.

[90] S. R. White, "Density-matrix algorithms for quantum renormalization groups", *Physical Review B*, vol. 48, no. 14, pp. 10 345–10 356, 1993.

[91] K. G. Wilson, "The renormalization group: Critical phenomena and the Kondo problem", *Reviews of Modern Physics*, vol. 47, no. 4, pp. 773–840, 1975.

[92] K. G. Wilson, "The renormalization group and critical phenomena", *Reviews of Modern Physics*, vol. 55, no. 3, pp. 583–600, 1983.

[93] G. K.-L. Chan and D. Zgid, "Chapter 7 The Density Matrix Renormalization Group in Quantum Chemistry", in *Annual Reports in Computational Chemistry*, R. A. Wheeler, Ed., vol. 5, Elsevier, 2009, pp. 149–162.

[94] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states", *Annals of Physics*, January 2011 Special Issue, vol. 326, no. 1, pp. 96–192, 2011.

[95] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, "Matrix product states and projected entangled pair states: Concepts, symmetries, theorems", *Reviews of Modern Physics*, vol. 93, no. 4, p. 045 003, 2021.

[96] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, "Matrix product operator representations", *New Journal of Physics*, vol. 12, no. 2, p. 025 012, 2010.

[97] G. D. Chiara, S. Montangero, P. Calabrese, and R. Fazio, "Entanglement entropy dynamics of Heisenberg chains", *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 03, P03001, 2006.

[98] J. Eisert, M. Cramer, and M. B. Plenio, "Colloquium: Area laws for the entanglement entropy", *Reviews of Modern Physics*, vol. 82, no. 1, pp. 277–306, 2010.

[99] G. K.-L. Chan and M. Head-Gordon, "Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group", *The Journal of Chemical Physics*, vol. 116, no. 11, pp. 4462–4476, 2002.

[100] U. Schollwöck, "The density-matrix renormalization group", *Reviews of Modern Physics*, vol. 77, no. 1, pp. 259–315, 2005.

[101] U. Schollwöck, "The density-matrix renormalization group: A short introduction", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1946, pp. 2643–2661, 2011.

[102] M. H. Kalos, "Monte Carlo Calculations of the Ground State of Three- and Four-Body Nuclei", *Physical Review*, vol. 128, no. 4, pp. 1791–1795, 1962.

[103] M. A. Lee and K. E. Schmidt, "Green's Function Monte Carlo", *Computers in Physics*, vol. 6, no. 2, pp. 192–197, 1992.

[104] K. J. Runge, "Finite-size study of the ground-state energy, susceptibility, and spin-wave velocity for the Heisenberg antiferromagnet", *Physical Review B*, vol. 45, no. 21, pp. 12 292–12 296, 1992.

[105] M. Calandra Buonaura and S. Sorella, "Numerical study of the two-dimensional Heisenberg model using a Green function Monte Carlo technique with a fixed number of walkers", *Physical Review B*, vol. 57, no. 18, pp. 11 446–11 456, 1998.

[106] J. S. Spencer *et al.*, "The HANDE-QMC Project: Open-Source Stochastic Quantum Chemistry from the Ground State Up", *Journal of Chemical Theory and Computation*, vol. 15, no. 3, pp. 1728–1742, 2019.

[107] S. Sharma. "Dice 0.1". (2022), [Online]. Available: `https://sanshar.github.io/Dice/` (visited on 08/30/2022).

[108] J. Brand. "Rimu 0.8.0". (2022), [Online]. Available: `https://juliahub.com/ui/Packages/Rimu/nXI2P/0.8.0` (visited on 08/30/2022).

[109] F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, "Semistochastic Projector Monte Carlo Method", *Physical Review Letters*, vol. 109, no. 23, p. 230 201, 2012.

[110] 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*, vol. 151, no. 9, p. 094 104, 2019.

[111] G. Li Manni, W. Dobrautz, and A. Alavi, "Compression of Spin-Adapted Multiconfigurational Wave Functions in Exchange-Coupled Polynuclear Spin Systems", *Journal of Chemical Theory and Computation*, vol. 16, no. 4, pp. 2202–2215, 2020.

[112] S. Yun, W. Dobrautz, H. Luo, and A. Alavi, "Benchmark study of Nagaoka ferromagnetism by spin-adapted full configuration interaction quantum Monte Carlo", *Physical Review B*, vol. 104, no. 23, p. 235 102, 2021.

[113] W. Dobrautz, O. Weser, N. A. Bogdanov, A. Alavi, and G. Li Manni, "Spin-Pure Stochastic-CASSCF via GUGA-FCIQMC Applied to Iron–Sulfur Clusters", *Journal of Chemical Theory and Computation*, vol. 17, no. 9, pp. 5684–5703, 2021.

[114] W. Dobrautz, V. M. Katukuri, N. A. Bogdanov, D. Kats, G. Li Manni, and A. Alavi, "Combined unitary and symmetric group approach applied to low-dimensional Heisenberg spin systems", *Physical Review B*, vol. 105, no. 19, p. 195 123, 2022.

[115] A. A. Holmes, H. J. Changlani, and C. J. Umrigar, "Efficient Heat-Bath Sampling in Fock Space", *Journal of Chemical Theory and Computation*, vol. 12, no. 4, pp. 1561–1571, 2016.

[116] 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*, vol. 15, no. 1, pp. 127–140, 2019.

[117] A. A. Kunitsa and S. Hirata, "Grid-based diffusion Monte Carlo for fermions without the fixed-node approximation", *Physical Review E*, vol. 101, no. 1, p. 013 311, 2020.

[118] M. Yang, E. Pahl, and J. Brand, "Improved walker population control for full configuration interaction quantum Monte Carlo", *The Journal of Chemical Physics*, vol. 153, no. 17, p. 174 103, 2020.

[119] 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*, vol. 146, no. 24, p. 244 105, 2017.

[120] G. H. Booth, S. D. Smart, and A. Alavi, "Linear-scaling and parallelisable algorithms for stochastic quantum chemistry", *Molecular Physics*, vol. 112, no. 14, pp. 1855–1869, 2014.

[121] A. S. Mikhayhu, *Embarrassingly Parallel*. Tempor, 2012.

[122] T. F. Chan and T. P. Mathew, "Domain decomposition algorithms", *Acta Numerica*, vol. 3, pp. 61–143, 1994.

[123] A. Toselli and O. Widlund, *Domain Decomposition Methods - Algorithms and Theory*, Softcover reprint of hardcover 1st ed. Berlin; Heidelberg: Springer, 2010, 468 pp.

[124] B. Smith, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Revised ed. Cambridge: Cambridge University Press, 2008, 240 pp.

[125] V. Dolean, P. Jolivet, and F. Nataf, *An Introduction to Domain Decomposition Methods: Algorithms, Theory, and Parallel Implementation*. Philadelphia: Society for Industrial and Applied Mathematics, 2016, 262 pp.

[126] MPI-Forum, *MPI: A Message-Passing Interface Standard, Version 3.0*. High-Performance Computing Center Stuttgart, 2012.

[127] J. J. Eriksen *et al.*, "The Ground State Electronic Energy of Benzene", *The Journal of Physical Chemistry Letters*, vol. 11, no. 20, pp. 8922–8929, 2020.

[128] O. Weser, K. Guther, K. Ghanem, and G. Li Manni, "Stochastic Generalized Active Space Self-Consistent Field: Theory and Application", *Journal of Chemical Theory and Computation*, vol. 18, no. 1, pp. 251–272, 2022.

[129] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, "Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling", *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 3674–3680, 2016.

[130] 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*, vol. 13, no. 4, pp. 1595–1604, 2017.

[131] J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, "Fast semistochastic heat-bath configuration interaction", *The Journal of Chemical Physics*, vol. 149, no. 21, p. 214 110, 2018.

[132] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands. II. The degenerate band case", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 277, no. 1369, pp. 237–259, 1964.

[133] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands III. An improved solution", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 281, no. 1386, pp. 401–419, 1964.

[134] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands - IV. The atomic representation", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 285, no. 1403, pp. 542–560, 1965.

[135] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands V. A perturbation expansion about the atomic limit", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 296, no. 1444, pp. 82–99, 1967.

[136] J. Hubbard and B. H. Flowers, "Electron correlations in narrow energy bands VI. The connexion with many-body perturbation theory", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 296, no. 1444, pp. 100–112, 1967.

[137] F. H. L. Essler, H. Frahm, F. Göhmann, A. Klümper, and V. E. Korepin, *The One-Dimensional Hubbard Model*. Cambridge: Cambridge University Press, 2005.

[138] E. Dagotto, "Correlated electrons in high-temperature superconductors", *Reviews of Modern Physics*, vol. 66, no. 3, pp. 763–840, 1994.

[139] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, "The Hubbard Model: A Computational Perspective", *Annual Review of Condensed Matter Physics*, vol. 13, no. 1, pp. 275–302, 2022.

[140] F. C. Zhang and T. M. Rice, "Effective Hamiltonian for the superconducting Cu oxides", *Physical Review B*, vol. 37, no. 7, pp. 3759–3761, 1988.

[141] D. J. Scalapino, "The 2D Hubbard Model and the High Tc Cuprate Problem", *Journal of Superconductivity and Novel Magnetism*, vol. 19, no. 3, pp. 195–200, 2006.

[142] C. J. Halboth and W. Metzner, "*d*-wave superconductivity and pomeranchuk instability in the two-dimensional hubbard model", *Physical Review Letters*, vol. 85, no. 24, pp. 5162–5165, 2000.

[143] C. Honerkamp, H. C. Fu, and D.-H. Lee, "Phonons and *d*-wave pairing in the two-dimensional hubbard model", *Physical Review B*, vol. 75, no. 1, p. 014 503, 2007.

[144] S. R. White and D. J. Scalapino, "Stripes on a 6-Leg Hubbard Ladder", *Physical Review Letters*, vol. 91, no. 13, p. 136 403, 2003.

[145] E. W. Huang, C. B. Mendl, H.-C. Jiang, B. Moritz, and T. P. Devereaux, "Stripe order from the perspective of the Hubbard model", *npj Quantum Materials*, vol. 3, no. 1, pp. 1–6, 1 2018.

[146] D. J. Scalapino, "Numerical Studies of the 2D Hubbard Model", in *Handbook of High-Temperature Superconductivity*, J. R. Schrieffer and J. S. Brooks, Eds., New York: Springer New York, 2007, pp. 495–526.

[147] 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", *Reviews of Modern Physics*, vol. 68, no. 1, pp. 13–125, 1996.

[148] G. Knizia and G. K.-L. Chan, "Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory", *Physical Review Letters*, vol. 109, no. 18, p. 186 404, 2012.

[149] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, "Quantum cluster theories", *Reviews of Modern Physics*, vol. 77, no. 3, pp. 1027–1080, 2005.

[150] A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, "Dual fermion approach to nonlocal correlations in the Hubbard model", *Physical Review B*, vol. 77, no. 3, p. 033 101, 2008.

[151] N. V. Prokof'ev and B. V. Svistunov, "Polaron Problem by Diagrammatic Quantum Monte Carlo", *Physical Review Letters*, vol. 81, no. 12, pp. 2514–2517, 1998.

[152] K. Van Houcke, E. Kozik, N. Prokof'ev, and B. Svistunov, "Diagrammatic Monte Carlo", *Physics Procedia*, Computer Simulations Studies in Condensed Matter Physics XXI, vol. 6, pp. 95–105, 2010.

[153] G. D. Purvis and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples", *The Journal of Chemical Physics*, vol. 76, no. 4, pp. 1910–1918, 1982.

[154] G. E. Scuseria, A. C. Scheiner, T. J. Lee, J. E. Rice, and H. F. Schaefer, "The closed-shell coupled cluster single and double excitation (CCSD) model for the description of electron correlation. A comparison with configuration interaction (CISD) results", *The Journal of Chemical Physics*, vol. 86, no. 5, pp. 2881–2890, 1987.

[155] R. J. Bartlett and M. Musiał, "Coupled-cluster theory in quantum chemistry", *Reviews of Modern Physics*, vol. 79, no. 1, pp. 291–352, 2007.

[156] R. Rodríguez-Guzmán, K. W. Schmid, C. A. Jiménez-Hoyos, and G. E. Scuseria, "Symmetry-projected variational approach for ground and excited states of the two-dimensional Hubbard model", *Physical Review B*, vol. 85, no. 24, p. 245 130, 2012.

[157] E. Stoudenmire and S. R. White, "Studying Two-Dimensional Systems with the Density Matrix Renormalization Group", *Annual Review of Condensed Matter Physics*, vol. 3, no. 1, pp. 111–128, 2012.

[158] C.-C. Chang and S. Zhang, "Spatially inhomogeneous phase in the two-dimensional repulsive Hubbard model", *Physical Review B*, vol. 78, no. 16, p. 165 101, 2008.

[159] C.-C. Chang and S. Zhang, "Spin and Charge Order in the Doped Hubbard Model: Long-Wavelength Collective Modes", *Physical Review Letters*, vol. 104, no. 11, p. 116 402, 2010.

[160] 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*, vol. 99, no. 7, p. 075 119, 2019.

[161] S. Yun, W. Dobrautz, H. Luo, V. Katukuri, N. Liebermann, and A. Alavi, "Ferromagnetic domains in the large-U Hubbard model with a few holes: A full configuration interaction quantum Monte Carlo study", *Physical Review B*, vol. 107, no. 6, p. 064 405, 2023.

[162] P. A. M. Dirac and R. H. Fowler, "Quantum mechanics of many-electron systems", *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 123, no. 792, pp. 714–733, 1929.

[163] J. H. Van Vleck, "The Dirac Vector Model in Complex Spectra", *Physical Review*, vol. 45, no. 6, pp. 405–419, 1934.

[164] W. Heisenberg, "Zur Theorie des Ferromagnetismus", in *Original Scientific Papers Wissenschaftliche Originalarbeiten*, ser. Werner Heisenberg Gesammelte Werke Collected Works, W. Blum, H. Rechenberg, and H.-P. Dürr, Eds., Berlin; Heidelberg: Springer, 1985, pp. 580–597.

[165] C. L. Cleveland and R. Medina A., "Obtaining a Heisenberg Hamiltonian from the Hubbard model", *American Journal of Physics*, vol. 44, no. 1, pp. 44–46, 1976.

[166] P. R. Hammar *et al.*, "Characterization of a quasi-one-dimensional spin-1/2 magnet which is gapless and paramagnetic for $g\mu_B H \lesssim J$ and $k_B T \ll J$", *Physical Review B*, vol. 59, no. 2, pp. 1008–1015, 1999.

[167] B. Lake *et al.*, "Multispinon Continua at Zero and Finite Temperature in a Near-Ideal Heisenberg Chain", *Physical Review Letters*, vol. 111, no. 13, p. 137 205, 2013.

[168] M. Mourigal, M. Enderle, A. Klöpperpieper, J.-S. Caux, A. Stunault, and H. M. Rønnow, "Fractional spinon excitations in the quantum Heisenberg antiferromagnetic chain", *Nature Physics*, vol. 9, no. 7, pp. 435–441, 7 2013.

[169] E. Manousakis, "The spin-½ heisenberg antiferromagnet on a square lattice and its application to the cuprous oxides", *Reviews of Modern Physics*, vol. 63, no. 1, pp. 1–62, 1991.

[170] M. Greven *et al.*, "Spin correlations in the 2d heisenberg antiferromagnet $Sr_2CuO_2Cl_2$: Neutron scattering, monte carlo simulation, and theory", *Physical Review Letters*, vol. 72, no. 7, pp. 1096–1099, 1994.

[171] F. M. Woodward, A. S. Albrecht, C. M. Wynn, C. P. Landee, and M. M. Turnbull, "Two-dimensional $S = \frac{1}{2}$ heisenberg antiferromagnets: Synthesis, structure, and magnetic properties", *Physical Review B*, vol. 65, no. 14, p. 144 412, 2002.

[172] R. Coldea, R. A. Cowley, T. G. Perring, D. F. McMorrow, and B. Roessli, "Critical behavior of the three-dimensional heisenberg antiferromagnet $RbMnF_3$", *Physical Review B*, vol. 57, no. 9, pp. 5281–5290, 1998.

[173] A. Salazar, M. Massot, A. Oleaga, A. Pawlak, and W. Schranz, "Critical behavior of the thermal properties of $KMnF_3$", *Physical Review B*, vol. 75, no. 22, p. 224 428, 2007.

[174] F. Bonechi, E. Celeghini, R. Giachetti, E. Sorace, and M. Tarlini, "Heisenberg XXZ model and quantum Galilei group", *Journal of Physics A: Mathematical and General*, vol. 25, no. 15, pp. L939–L943, 1992.

[175] F. H. L. Essler, V. E. Korepin, and K. Schoutens, "Fine structure of the Bethe ansatz for the spin-1/2 Heisenberg XXX model", *Journal of Physics A: Mathematical and General*, vol. 25, no. 15, pp. 4115–4126, 1992.

[176] M. Karabach, G. Müller, H. Gould, and J. Tobochnik, "Introduction to the Bethe Ansatz I", *Computers in Physics*, vol. 11, no. 1, pp. 36–43, 1997.

[177] M. Karbach, K. Hu, and G. Müller, "Introduction to the Bethe Ansatz II", *Computers in Physics*, vol. 12, no. 6, pp. 565–573, 1998.

[178] S. Belliard and N. Crampé, "Heisenberg XXX Model with General Boundaries: Eigenvectors from Algebraic Bethe Ansatz", *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, vol. 9, p. 072, 2013.

[179] N. Liebermann, K. Ghanem, and A. Alavi, "Importance-sampling FCIQMC: Solving weak sign-problem systems", *The Journal of Chemical Physics*, vol. 157, no. 12, p. 124 111, 2022.

[180] M. H. Kolodrubetz, J. S. Spencer, B. K. Clark, and W. M. Foulkes, "The effect of quantization on the full configuration interaction quantum Monte Carlo sign problem", *The Journal of Chemical Physics*, vol. 138, no. 2, p. 024 110, 2013.

[181] M. J. Allen, V. C. Tung, and R. B. Kaner, "Honeycomb Carbon: A Review of Graphene", *Chemical Reviews*, vol. 110, no. 1, pp. 132–145, 2010.

[182] A. D. Ghuge, A. R. Shirode, and V. J. Kadam, "Graphene: A Comprehensive Review", *Current Drug Targets*, vol. 18, no. 6, pp. 724–733, 2017.

[183] B. K. Clark, D. A. Abanin, and S. L. Sondhi, "Nature of the Spin Liquid State of the Hubbard Model on a Honeycomb Lattice", *Physical Review Letters*, vol. 107, no. 8, p. 087 204, 2011.

[184] Q. Chen, G. H. Booth, S. Sharma, G. Knizia, and G. K.-L. Chan, "Intermediate and spin-liquid phase of the half-filled honeycomb Hubbard model", *Physical Review B*, vol. 89, no. 16, p. 165 134, 2014.

[185] W. Wu and A.-M. S. Tremblay, "Phase diagram and Fermi liquid properties of the extended Hubbard model on the honeycomb lattice", *Physical Review B*, vol. 89, no. 20, p. 205 128, 2014.

[186] K. Ghanem, N. Liebermann, and A. Alavi, "Population control bias and importance sampling in full configuration interaction quantum Monte Carlo", *Physical Review B*, vol. 103, no. 15, p. 155 135, 2021.

[187] N. Cerf and O. C. Martin, "Finite population-size effects in projection Monte Carlo methods", *Physical Review E*, vol. 51, no. 4, pp. 3679–3693, 1995.

[188] 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*, vol. 151, no. 6, p. 061 101, 2019.

[189] K. Guther, A. J. Cohen, H. Luo, and A. Alavi, "Binding curve of the beryllium dimer using similarity-transformed FCIQMC: Spectroscopic accuracy with triple-zeta basis sets", *The Journal of Chemical Physics*, vol. 155, no. 1, p. 011 102, 2021.

[190] K. Liao, T. Schraivogel, H. Luo, D. Kats, and A. Alavi, "Towards efficient and accurate ab initio solutions to periodic systems via transcorrelation and coupled cluster theory", *Physical Review Research*, vol. 3, no. 3, p. 033 072, 2021.

[191] 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*, vol. 14, no. 3, pp. 1403–1411, 2018.

[192] S. F. Boys, N. C. Handy, and J. W. Linnett, "The determination of energies and wavefunctions with full electronic correlation", *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 310, no. 1500, pp. 43–61, 1969.

[193] S. F. Boys, N. C. Handy, and J. W. Linnett, "A condition to remove the indeterminacy in interelectronic correlation functions", *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 309, no. 1497, pp. 209–220, 1969.

[194] N. C. Handy, "Towards an understanding of the form of correlated wavefunctions for atoms", *The Journal of Chemical Physics*, vol. 58, no. 1, pp. 279–287, 1973.

[195] V. S. Ryaben'kii and S. V. Tsynkov, *A Theoretical Introduction to Numerical Analysis*, 1st ed. Boca Raton, Florida: Chapman & Hall/CRC, 2006, 537 pp.

[196] J. A. Storer, *An Introduction to Data Structures and Algorithms*. Berlin; Heidelberg: Springer Science & Business Media, 2001, 632 pp.

[197] P. S. Goldbaum, "Existence of Solutions to the Bethe Ansatz Equations for the 1D Hubbard Model: Finite Lattice and Thermodynamic Limit", *Communications in Mathematical Physics*, vol. 258, no. 2, pp. 317–337, 2005.

[198] J. M. P. Carmelo, T. Čadež, and P. D. Sacramento, "One-particle spectral functions of the one-dimensional Fermionic Hubbard model with one fermion per site at zero and finite magnetic fields", *Physical Review B*, vol. 103, no. 19, p. 195 129, 2021.

[199] T. Tohyama, Y. Inoue, K. Tsutsui, and S. Maekawa, "Exact diagonalization study of optical conductivity in the two-dimensional Hubbard model", *Physical Review B*, vol. 72, no. 4, p. 045 113, 2005.

[200] M. Fiedler, "Eigenvectors of acyclic matrices", *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 607–618, 1975.

[201] M. Juvan and B. Mohar, "Optimal linear labelings and eigenvalues of graphs", *Discrete Applied Mathematics*, vol. 36, no. 2, pp. 153–168, 1992.

[202] H. Yokoyama, Y. Tanaka, M. Ogata, and H. Tsuchiura, "Crossover of Superconducting Properties and Kinetic-Energy Gain in Two-Dimensional Hubbard Model", *Journal of the Physical Society of Japan*, vol. 73, no. 5, pp. 1119–1122, 2004.

[203] T. Yanagisawa, "Crossover from Weakly to Strongly Correlated Regions in the Two-dimensional Hubbard Model — Off-diagonal Wave Function Monte Carlo Studies of Hubbard Model II —", *Journal of the Physical Society of Japan*, vol. 85, no. 11, p. 114 707, 2016.

[204] H. Yokoyama, M. Ogata, and Y. Tanaka, "Mott transitions and *d*-wave superconductivity in half-filled-band hubbard model on square lattice with geometric frustration", *Journal of the Physical Society of Japan*, vol. 75, no. 11, pp. 114 706–114 706, 2006.

[205] P. Prelovšek, J. Kokalj, Z. Lenarčič, and R. H. McKenzie, "Holon-doublon binding as the mechanism for the Mott transition", *Physical Review B*, vol. 92, no. 23, p. 235 155, 2015.

[206] 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*, vol. 141, no. 24, p. 244 117, 2014.

[207] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*, Illustrated ed. Oxford, New York: Oxford University Press, USA, 1999, 496 pp.

Curriculum Vitae

Niklas Julian Liebermann
born 31st July 1993 in Tuttlingen (Germany)

Education

2018–2023 PhD Student, Theoretical Chemistry
Max Planck Institute for Solid State Research

Thesis: *The FCIQMC Sign Problem in the Real-Space Hubbard Model*
Department of Electronic Structure Theory
Advisor: Prof. Dr. Ali Alavi

2014–2017 Master of Science, Physics
University of Stuttgart

Thesis: *Exceptional Points in Billiard Systems by Applying the Semiclassical Theory of Periodic Orbits*
1st Institute for Theoretical Physics
Advisor: Prof. Dr. Jörg Main

2011–2014 Bachelor of Science, Physics
University of Stuttgart

Thesis: *Localization Accuracy in Nanoscale Magnetometry Experiments*
3rd Institute of Physics
Advisors: Prof. Dr. Jörg Wrachtrup, Dr. Ilja Gerhardt

2003–2011 Abitur

List of Publications

2023 Sujun Yun, Werner Dobrautz, Hongjun Luo, Vamshi Katukuri, Niklas Liebermann and Ali Alavi:
"Ferromagnetic domains in the large-U Hubbard model with a few holes"
Physical Review B 107, 064405

2023 Giovanni Li Manni, Daniel Kats and Niklas Liebermann:
"Resolution of Electronic States in Heisenberg Cluster Models within the Unitary Group Approach"
Journal of Chemical Theory and Computation 19 (4), 1218–1230

2022 Niklas Liebermann, Khaldoon Ghanem and Ali Alavi:
"Importance-sampling FCIQMC: solving weak sign-problem systems"
The Journal of Chemical Physics 157 (12), 124111

2022 Oskar Weser, Niklas Liebermann, Daniel Kats, Ali Alavi and Giovanni Li Manni:
"Spin Purification in Full-CI Quantum Monte Carlo via a First-Order Penalty Approach"
The Journal of Physical Chemistry A 126 (12), 2050–2060

2021 Khaldoon Ghanem, Niklas Liebermann and Ali Alavi:
"Population control bias and importance sampling in full configuration interaction quantum Monte Carlo"
Physical Review B 103 (15), 155135

2020 Kai Guther *et al.* (including Niklas Liebermann):
"NECI: N-Electron Configuration Interaction with an emphasis on state-of-the-art stochastic methods"
The Journal of Chemical Physics 153 (3), 034107

2017 Niklas Liebermann, Jörg Main and Günter Wunner:
"Exceptional points in the elliptical three-disk scatterer using semiclassical periodic orbit quantization"
Europhysics Letters 118 (3), 30006