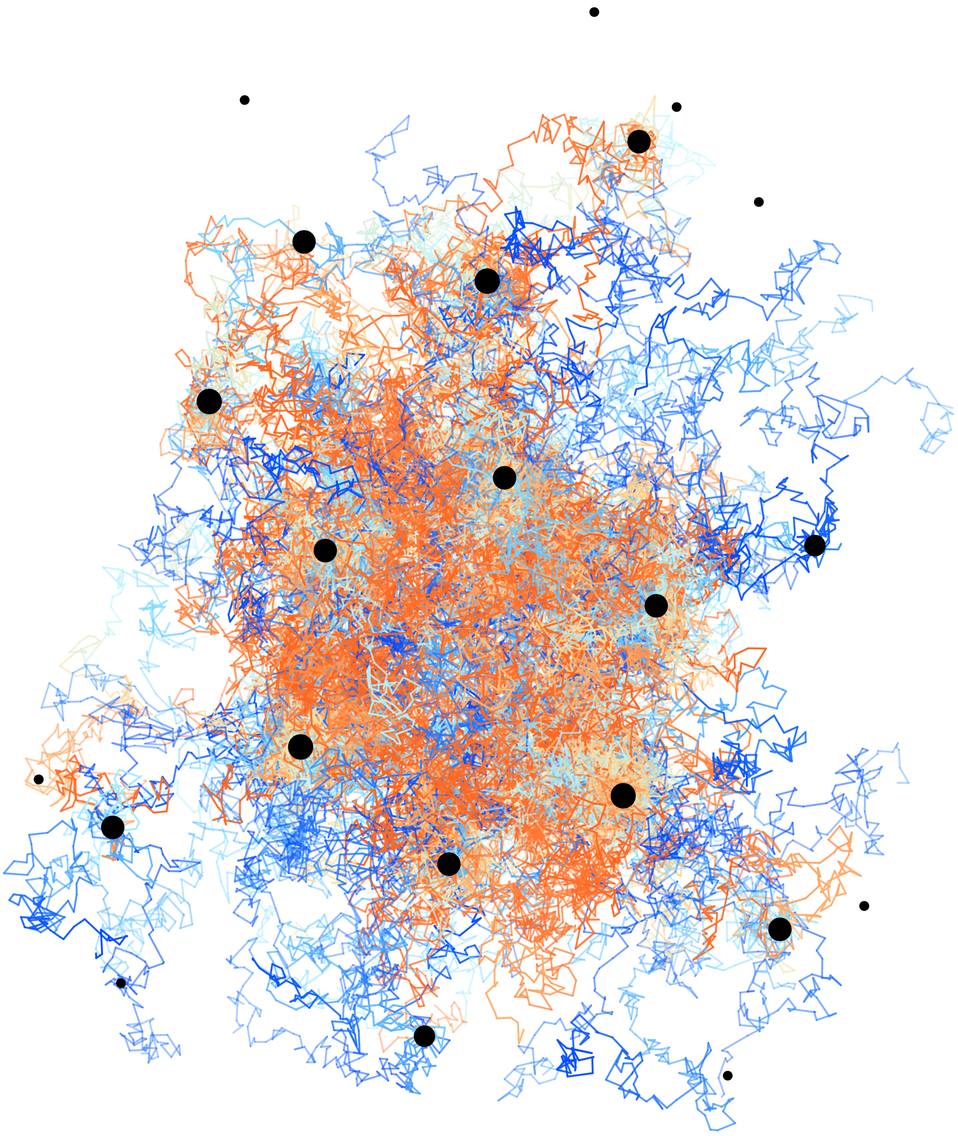
Dissertation

Real-space orbitals for Quantum Monte Carlo and quasi-exact densities for Kohn-Sham inversions

Alexander Kaiser



UNIVERSITÄT BAYREUTH

Real-space orbitals for Quantum Monte Carlo and quasi-exact densities for Kohn-Sham inversions

Von der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

von

Alexander Kaiser

aus Dillingen a.d. Donau

1. Gutachter: Prof. Dr. Stephan Kümmel
2. Gutachter: Prof. Dr. Harald Oberhofer

Tag der Einreichung: 03.03.2025
Tag des Kolloquiums: 18.07.2025

Abstract

Quantum mechanics is at the core of many modern devices and applications, spanning diverse fields such as the design of efficient integrated circuits, mimicking photosynthesis for efficient solar cells, or medical applications like the prediction of the effectiveness of vaccines. Computationally efficient and accurate methods to describe the quantum nature of matter are required to predict material properties and stimulate the search for new materials with the aid of theory and computation. A variety of different computational methods exist to solve quantum mechanical problems. Density functional theory (DFT) is the most prominent of these frameworks, offering accurate predictions at moderate computational cost. While DFT is computationally efficient, its accuracy is limited by approximate exchange-correlation functionals. Compared to DFT, quantum Monte Carlo (QMC) methods provide a more precise description of quantum effects based on classical random walks. This exact treatment of the correlation effects and its wave function based character make QMC an ideal candidate for computing a wide range of electronic structure properties with high precision. QMC calculations require a higher computational cost compared to DFT but nevertheless represent a viable approach to the electronic structure problem for larger systems relevant to real-world applications.

This thesis provides a self-contained review of these methods, especially variational and diffusion Monte Carlo fundamentals, and focuses on the advantages that can be gained in combining DFT and QMC methods. To correctly describe exchange and correlation effects within DFT, sophisticated approximations to the exchange-correlation functional are required, whose quality directly affects the quality of the computed results. The development of more powerful density functional approximations is traditionally accelerated by comparison with exact results from wave function based methods like QMC. The comparison of exact densities with densities from DFT enables the assessment of how accurately a density functional reproduces the exact density. Even more detailed insights into the properties of the true exchange-correlation functional are available if the exact densities are inverted. The inversion is an intricate process in which the central quantity of DFT, the true exchange-correlation potential, is recovered from the density. The true exchange-correlation potential determines the true Kohn-Sham system and enables detailed comparison with density functional approximations.

In the research that led to this thesis, a novel QMC code was developed that allows for a precise representation of the true wave function with high computational efficiency. While previous QMC methods focus on atom-centered basis sets or plain waves, a premier part of this work is the representation of the orbitals on a real-space grid. The grid is specially tailored to allow highly accurate all-electron calculations for small systems. With these orbitals, QMC energies are calculated, recovering 90 to 97% of the correlation, confirming the accuracy of the QMC wave function. The same grid is utilized to calculate the quasi-exact QMC densities with the binning method. This method constructs a density histogram with significantly reduced variance due to the specialized grid. Suggestions for an improved density estimation from first principles and beyond the construction of a histogram are made.

The accurate densities are inverted to recover the quasi-exact exchange-correlation potentials of DFT. The general limitations of the inversion of the Hohenberg-Kohn map are discussed in detail. New inversion techniques are presented that extend the scope of present methods. A special focus is put on densities that exhibit statistical fluctuations, and regularization schemes are devised to lower and even remove statistical fluctuations. The inversion of the densities gives insight into the exact structural properties of the exchange-correlation potential. Finally, QMC and inversion methods are combined in a novel approach. Therein, refined orbitals from the density inversion are leveraged to improve the wave function used in QMC calculations. The partially surprising results of this approach are discussed and put into perspective.

Kurzdarstellung

Die Quantenmechanik bildet das Fundament vieler moderner Geräte und Anwendungen, die sich über verschiedenste Bereiche erstrecken – von der Entwicklung effizienter Mikroelektronik über die Nachahmung der Photosynthese für effiziente Solarzellen bis hin zu medizinischen Anwendungen, wie der Vorhersage der Wirksamkeit von Impfstoffen. Um ausgehend von theoretischen Überlegungen und computergestützten Verfahren Materialeigenschaften vorherzusagen und neue Materialien zu finden, werden recheneffiziente und präzise Methoden benötigt, die die quantenmechanische Natur der Materie beschreiben. Zur Lösung quantenmechanischer Probleme steht eine Vielzahl verschiedener Rechenmethoden zur Verfügung. Die Dichtefunktionaltheorie (DFT) ist hierbei das bekannteste Verfahren und ermöglicht präzise Vorhersagen bei moderaten Rechenkosten. Obwohl DFT recheneffizient ist, wird ihre Genauigkeit durch approximierte Austauschkorrelationsfunktionale begrenzt. Im Vergleich zur DFT liefern Quantum-Monte-Carlo-Methoden (QMC) eine präzisere Beschreibung quantenmechanischer Effekte, basierend auf klassischen Zufallsbewegungen. Aufgrund der exakten Behandlung von Korrelationseffekten ist das wellenfunktionsbasierte QMC ideal für die hochpräzise Berechnung vieler Eigenschaften der elektronischen Struktur. QMC-Berechnungen erfordern zwar einen höheren Rechenaufwand als DFT, stellen aber dennoch einen praktikablen Ansatz zur Berechnung der elektronischen Struktur in größeren Systemen dar, die für reale Anwendungen relevant sind.

Diese Dissertation bietet eine in sich geschlossene Übersicht der verwendeten Methoden, insbesondere die Grundlagen der Variations- und Diffusions-Monte-Carlo-Methoden, und konzentriert sich auf die Vorteile, die sich aus der Kombination von DFT- und QMC-Methoden ergeben. Im Zentrum dieser Arbeit stehen die Vorteile, die durch eine Kombination von DFT- und QMC-Methoden erzielt werden können. Um Austausch- und Korrelationseffekte innerhalb der DFT korrekt zu beschreiben, sind anspruchsvolle Näherungen an das Austauschkorrelationsfunktional erforderlich, deren Qualität die Güte der berechneten Eigenschaften direkt beeinflusst. Die Entwicklung leistungsfähigerer Dichtefunktional-Näherungen wird traditionell durch den Vergleich mit exakten Ergebnissen aus wellenfunktionsbasierten Methoden wie QMC vorangetrieben. Der Vergleich exakter Dichten mit DFT-Dichten ermöglicht eine Bewertung, wie akkurat ein Dichtefunktional die exakte Dichte reproduziert. Noch detailliertere Einblicke in die wahren Eigenschaften der Dichtefunktionale werden gewonnen, wenn exakte Dichten invertiert werden, um eine zentrale Größe der DFT, das wahre Austausch-Korrelationspotential, zu rekonstruieren. Das wahre Austausch-Korrelationspotential bestimmt das exakte Kohn-Sham-System und ermöglicht eine detaillierte Bewertung von Dichtefunktional-Näherungen.

In der Arbeit, die dieser Dissertation zugrunde liegt, wurde ein neuer QMC-Code entwickelt, der eine genaue Darstellung der wahren Wellenfunktion mit hoher Recheneffizienz erlaubt. Während bisherige QMC-Methoden auf atomzentrierte Basisfunktionen oder ebene Wellen verwenden, ist ein zentraler Bestandteil dieser Arbeit die Darstellung der Orbitale auf einem Ortsraumgitter. Dieses Gitter ist speziell auf hochpräzise Berechnungen für kleine Systeme einschließlich der Kernelektronen ausgelegt. Mit diesen Orbitalen werden QMC-Energien berechnet, wobei 90 bis 97% der Korrelationseffekte wiedergewonnen werden, was die Genauigkeit der QMC-Wellenfunktion bestätigt. Dasselbe Gitter wird verwendet, um die quasi-exakte QMC-Dichten mittels Histogrammen zu berechnen. Diese Methode erstellt ein Dichtehistogramm mit erheblich reduzierter Varianz aufgrund des spezialisierten Gitters. Ein grundsätzlich verbessertes Konzept zur Schätzung der Dichte jenseits von Histogrammen wird vorgestellt.

Die genauen Dichten werden invertiert, um die quasi-exakten Austausch-Korrelationspotentiale der DFT zu rekonstruieren. Die generellen Grenzen der inversen Hohenberg-Kohn Abbildung werden ausführlich diskutiert. Neue Inversionstechniken werden vorgestellt, die den Rahmen der bisherigen Methoden erweitern. Ein besonderer Schwerpunkt liegt dabei auf der Inversion von Dichten mit statistischen Fluktuationen. Es werden Regularisierungsschemata entwickelt, um diese Fluktuationen zu reduzieren oder sogar vollständig zu entfernen. Die Inversion der Dichten gibt Aufschluss über die exakten strukturellen Eigenschaften des Austausch-Korrelationspotentials. Schließlich werden QMC- und Inversionsmethoden in einem neuartigen Ansatz kombiniert. Dabei werden verfeinerte Orbitale aus der Dichteinversion genutzt, um die in QMC-Berechnungen verwendete Wellenfunktion zu verbessern. Die teilweise überraschenden Ergebnisse dieses Ansatzes werden diskutiert und eingeordnet.

Contents

1. Introduction

It has been a long scientific journey from discovering quantum effects to today's understanding of the microcosm, with unprecedentedly accurate predictions like the standard model of particle physics. While the fundamental interactions of the standard model's particles are very precisely described in modern physics, understanding extended composite systems is challenging. Beyond the nano scale, at the length and time scales of extended molecular or periodic systems, the prediction of electronic properties faces major challenges. From the sizes of a few atoms to small biological units, classical theories are inaccurate and a full quantum mechanical description is computationally infeasible. Exchange and correlation effects, genuine properties of matter due to their quantum nature, play an important role in many aspects of such systems and are not yet sufficiently understood. To tackle these challenges, a diverse landscape of computational methods has emerged over the past decades, each with specific strengths, limitations, and further refinements and extensions.

Understanding the quantum world allowed for inventions at the core of almost every modern application. Many applications enabled by quantum physics originate from systems beyond the scale of single atoms, such as solid-state physics or large molecular systems up to the protein scale. Given the amount of scientific effort spent on understanding the effects of quantum mechanics and developing its computational methods, it is no surprise that this laid the foundations of today's technologies and gave rise to the present information age. A prominent example is enabling the miniaturization of transistors by exploitation of quantum effects in semiconductors to arrive at length scales of a few tens of nanometers. The improved understanding of quantum effects is also promising to mark the beginning of an age of energy. Examples on the energy-production side are a better understanding of light-harvesting physics in photosynthesis or solar cells [1–3] or the various quantum effects that play a role in nuclear fusion [4–11]. Additionally, energy consumption can significantly be reduced by a more efficient design of electronic devices like integrated circuits or by replacing electronic devices with optical technologies. The importance of understanding the electronic structure of such systems cannot be underestimated. To give clear evidence, from the ten most internally cited[1] articles of the Physical Review journals, nine cover electronic structure [12]. The four most internally cited articles [13–16] cover without exception the topic of density functional theory (DFT) or quantum Monte Carlo (QMC) [12], with a total of more than 123000 citations[2].

The two approaches under consideration in this thesis are DFT and QMC. Due to its favorable computational cost, the former is the workhorse for a vast community of quantum chemistry users. While DFT is a density-based method without a proper description of the wave function, QMC focuses on the wave function explicitly. QMC methods are less frequently used but are much more accurate in revealing true insights into the nature of correlation effects. While both approaches in principle allow the study of the exact properties of the wave function, the DFT approach inevitably relies on approximations, while the QMC approach is more straightforwardly based on first principles. Besides the differences between DFT and QMC regarding methodology, precision and computation, a common variable between both is the density. In this thesis, accurate densities are generated from the QMC approach. In a process called inversion, these exact densities are mapped to central quantities of DFT to assess the quality of the approximations

---

[1] Internal citations cover only citations from within the Physical Review journals.

[2] According to Google scholar and mendeley.com, Feb. 2025.

made in DFT. The calculation of accurate densities and the inversion process are challenging tasks on their own. However, combining information from DFT and QMC offers vast mutual benefits for both methods. With the quasi-exact quantities from inversion, the QMC calculations might be enhanced to higher accuracy.

This introductory chapter is followed by an overview of the electronic structure problem of this thesis in Chapt. 2 and the DFT method in Chapt. 3. The variational Monte Carlo (VMC) method and properties of the wave function are presented in Chapt. 4 and Chapt. 5, respectively. The diffusion Monte Carlo (DMC) method is explained in Chapt. 6. In Chapt. 7, methods to calculate quasi-exact densities are presented. Some results of the QMC methods are presented in Chapt. 9, which reports energies and densities. The densities are subsequently used in Chapt. 8 and Chapt. 10, which links the two approaches of DFT and QMC to not only gain insights into the correlation effects of an interacting system but also to assess the accuracy of the exchange-correlation approximations and accelerate their development. Finally, Chapt. 11 paves the way to further explore combinations of DFT and QMC, presents first results, and discusses their implications.

2. Electronic structure calculations

This thesis covers time-independent, non-relativistic molecular systems exclusively. The standard Hamiltonian for a molecular system with N electrons is given in the form

$$\hat{\mathrm{H}} = \sum_{k=1}^{N} \left(-\frac{1}{2} \nabla^2_{\mathbf{r}_k} + v_{\text{ext}}(\mathbf{r}_k) \right) + U(\mathbf{r}_1, \dots \mathbf{r}_N). \tag{2.1}$$

Throughout this thesis, Hartree atomic units, $e = \hbar = m_e = 4\pi\epsilon_0 = 1$, are used. The static external potential

$$v_{\text{ext}}(\mathbf{r}) = \sum_{\eta=1}^{L} \frac{-Z_\eta}{|\mathbf{r} - \boldsymbol{\xi}_\eta|} + v_e(\mathbf{r}) \tag{2.2}$$

includes L classical ions of charges Z_η at fixed coordinates $\boldsymbol{\xi}_\eta$ in the Born-Oppenheimer approximation [17]. A non-ionic external potential v_e, modeling, for example, the interaction of the electrons with an external field, might be present. Introducing the compact notation $\mathbf{R} = (\mathbf{r}_1, \dots \mathbf{r}_N) \in \mathbb{R}^{3N}$ for all electron coordinates, the full external potential can be written as

$$V(\mathbf{R}) = \sum_{k=1}^{N} v_{\text{ext}}(\mathbf{r}_k), \tag{2.3}$$

and the energy of the electron-electron interaction reads

$$U(\mathbf{R}) = \sum_{k=1}^{N} \sum_{l=k+1}^{N} \frac{1}{|\mathbf{r}_k - \mathbf{r}_l|}. \tag{2.4}$$

With the operator for the kinetic energy

$$\hat{\mathrm{T}} = -\frac{1}{2} \sum_{k=1}^{N} \nabla^2_{\mathbf{r}_k} = -\frac{1}{2} \nabla^2_{\mathbf{R}}, \tag{2.5}$$

the Hamiltonian can be written in compact notation,

$$\hat{\mathrm{H}} = \hat{\mathrm{T}} + V + U. \tag{2.6}$$

A note on the notation: For a simplified notation, the gradient or Laplacian regarding the full configuration $\mathbf{R}$ is used. The gradient with respect to $\mathbf{R}$, denoted as $\nabla_{\mathbf{R}}$, is the combined gradients of all coordinates $\nabla_{\mathbf{R}} = (\nabla_{\mathbf{r}_1}, \dots \nabla_{\mathbf{r}_N})$, while the Laplacian is the sum $\nabla^2_{\mathbf{R}} = \sum_k \nabla^2_{\mathbf{r}_k}$. No distinction of $\nabla_{\mathbf{R}}$ and ∇ will be made unless necessary.

The Schrödinger equation [18] for the Hamiltonian is the eigenvalue problem

$$\hat{\mathrm{H}} \Psi_i(\mathbf{R}) = E_i \Psi_i(\mathbf{R}) \tag{2.7}$$

with the ground state Ψ_0 being the state of lowest energy E_0. While this work is restricted to Hamiltonians independent of spin, the spin coordinates enter the wave function due to the spin-statistic theorem [19, 20]. With spin-space coordinates $\mathbf{x}_k = (s_k, \mathbf{r}_k)$ for every electron of the system and again the compact notation $\mathbf{X} = (\mathbf{x}_1, \ldots \mathbf{x}_N)$ accounting for all coordinates, the particles' statistic can be applied. For fermions, which are the only particle kind treated here, the symmetry character of Ψ is defined. Interchanging particle coordinates with some permutation π from the symmetric group $\mathbb{S}(N)$ of degree N, the wave function transforms as

$$\Psi(\pi\mathbf{X}) = \Psi(\mathbf{x}_{\pi(1)}, \ldots \mathbf{x}_{\pi(N)}) = \operatorname{sign}(\pi)\Psi(\mathbf{X}). \tag{2.8}$$

Due to the non-separable form of the electron-electron interaction terms in U, the Schrödinger equation, a $3N$-dimensional partial differential equation (PDE), has no analytical solution for most systems of interest. Many modern approaches to finding sufficiently accurate solutions of the Schrödinger equation start with some non-interacting Hamiltonian excluding any two- or higher-body terms. Omitting the many-body interactions separates the Schrödinger equation in a set of single-particle, i.e. three-dimensional, Schrödinger-like equations. The set of orbitals $\{\varphi_i\}$ that solve the Schrödinger equation for a single-particle Hamiltonian, can be constructed by significantly simpler means. Slater determinants Φ can be constructed from any selection of orbitals,

$$\Phi(\mathbf{X}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{x}_1) & \varphi_1(\mathbf{x}_2) & \ldots & \varphi_1(\mathbf{x}_N) \\ \varphi_2(\mathbf{x}_1) & \varphi_2(\mathbf{x}_2) & \ldots & \varphi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_N(\mathbf{x}_1) & \varphi_N(\mathbf{x}_2) & \ldots & \varphi_N(\mathbf{x}_N) \end{vmatrix}, \tag{2.9}$$

to arrive at non-interacting wave functions, i.e. wave functions neglecting correlation effects but satisfying the spin-symmetry of Eq. (2.8). The wave function can be expanded in the Slater determinant space to restore the lost correlation,

$$\Psi = \sum_i c_i \Phi_i. \tag{2.10}$$

This expansion is guaranteed to fully recover the correlation of the original Hamiltonian if it covers the complete Slater determinant space. Expansions in determinant space are computationally expensive and often show slow convergence in the number of Slater determinants. Therefore, many schemes to combine the single-particle orbitals more efficiently have been proposed to yield insight into the properties of the interacting wave function. The strong dependency on Slater determinants renders the orbitals a key ingredient in almost any quantum mechanical approach. A popular example of a single-particle approach is the Hartree-Fock method. It minimizes the energy expectation value of the fully interacting Hamiltonian evaluated with a single Slater determinant by variation of its single-particle orbitals. The Hartree-Fock equations determine the orbitals to construct the single Slater determinant with.

While the Hartree-Fock method is based on an energy minimization, in DFT the density plays the central role. A special auxiliary single-particle Hamiltonian is constructed, such that its orbitals give rise to a density that is the density of the exact ground state of the interacting many-body Hamiltonian.

3. Density functional theory

In DFT, the central quantity is the electronic one-body density. The accurate description of the density enables to access many electronic properties of the system. The density is accompanied by the in principle exact ground state energy of the electronic structure problem. This exactness is one of the key features of DFT and allows a detailed analysis of the electronic and energetic structure.

3.1. Fundamentals

At the core of DFT is the connection between the external potential and the density. The Hohenberg-Kohn theorem [14] guarantees that the ground state density n_0 of a fermionic system, corresponding to the Hamiltonian of Eq. (2.1) with external potential v_{ext} and ground state Ψ_0, holds all information about its external potential, i.e. there is a unique mapping connecting the density n_0 and the potential v_{ext}. Although the Hohenberg-Kohn theorem does not hold for all densities, it holds for densities that originate from a Hamiltonian with a local potential. Such densities are called v-representable. The set of densities can be generalized by the Levy-Lieb constrained search [21–23] to any density that stems from a ground state wave function, the so-called N-representable densities. Starting from an external potential v_{ext} with either interacting or non-interacting Hamiltonian and calculating its (non-degenerate) ground state wave function Ψ_0 together with the corresponding unique density n_0, the ground state wave function is rendered a functional of the density, $\Psi[n]$. This functional is defined on all N-representable densities. The Levy-Lieb constrained search does not only reveal the unique mapping between different potentials and the density. It further guarantees the existence of a universal density functional for the total energy of the ground state for any v_{ext},

$$E[n] = T[n] + U[n] + \int n v_{\text{ext}} = F[n] + \int n v_{\text{ext}}, \tag{3.1}$$

whose minimizer is the ground state density. To simplify the notation, the arguments of functions inside an integral and the integration variable are suppressed whenever reasonable in the notation here and throughout this thesis. $F[n]$ combines the universal functionals for the interacting kinetic energy $T[n] = \langle \Psi[n] | \hat{\mathrm{T}} | \Psi[n] \rangle$ and the electron-electron interaction $U[n] = \langle \Psi[n] | U | \Psi[n] \rangle$. If the true energy functional was known, a minimization process would yield the true ground state energy and the true density. Since the density allows the straightforward evaluation of wave function expectation values of any multiplicative one-body observable, it poses significant insight into the electronic structure of a system. Despite these encouraging prospects, no sufficiently accurate direct density functional approximation $E[n]$ has been found. One important obstacle is that it is very hard to find an accurate density functional for the interacting kinetic energy $T[n]$.

The Kohn-Sham (KS) scheme [13], which is the workhorse of DFT, can remedy this shortcoming as presented below. To this end, one can distinguish the space of v-representable densities by interacting and non-interacting v-representable densities. Interacting v-representable densities stem from the wave functions of interacting Hamiltonians, whereas the non-interacting v-representable densities come from non-interacting Hamiltonians. Suppose a ground state density is non-interacting v-representable. Then there exists a unique potential, denoted by v_{KS} for the reasons explained below, that reproduces the density in a non-interacting Hamiltonian $\hat{\mathrm{H}} = \hat{\mathrm{T}} + v_{\mathrm{KS}}$. Suppose, on the other hand, a ground state density is interacting v-representable. Then there exists a unique potential v_{ext} that leads to the density via an interacting Hamiltonian $\hat{\mathrm{H}} = \hat{\mathrm{T}} + V + U$. A ground state density that simultaneously is interacting v-representable and non-interacting v-representable uniquely links the two potentials v_{KS} and v_{ext}. This assumption is the working hypothesis of KS-DFT. The KS approach utilizes the non-interacting Hamiltonian with potential v_{KS} that corresponds to the density n_0 to implement a reasonable functional approximation to Eq. (3.1). To this end, the energy functional of the non-interacting system $E_{\mathrm{s}}[n]$ with some effective potential v_{KS} is

$$E_{\mathrm{s}}[n] = T_{\mathrm{s}}[n] + \int n v_{\mathrm{KS}}, \tag{3.2}$$

where the non-interacting kinetic energy T_{s} is an explicit functional of the orbitals,

$$T_{\mathrm{s}} = -\frac{1}{2} \sum_i \int \varphi_i^* \nabla^2 \varphi_i. \tag{3.3}$$

The minimizing non-interacting density must, by construction of the KS scheme, be n_0 and is given by the density of the Slater determinant of the non-interacting system, viz., the sum over the orbitals' absolute value squares,

$$n_0(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{r})|^2 . \tag{3.4}$$

In contrast to the interacting energy functional $E[n]$ of Eq. (3.1), the non-interacting functional $E_{\mathrm{s}}[n]$ is much more accessible due to its single-particle character. The minimization of E_{s} with respect to the orbitals and subject to orthonormalization constraints results in the KS eigenvalue problem

$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{KS}}(\mathbf{r})\right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}), \tag{3.5}$$

which is a standard computational problem on $\mathbb{R}^3$ and can be solved by various means. The KS scheme is based on the connection of the interacting functional $E[n]$ (with external potential v_{ext}) and the non-interacting functional E_{s} (with the potential v_{KS}). By construction, both functionals have the same minimizing density, n_0.

$$\left.\frac{\delta E_{\mathrm{s}}[n]}{\delta n}\right|_{n_0} = \left.\frac{\delta E[n]}{\delta n}\right|_{n_0} = 0 \tag{3.6}$$

Therefore, the effective single-particle potential is fully determined,

$$v_{\mathrm{KS}}(\mathbf{r}) = v_{\mathrm{ext}} + \frac{\delta T[n]}{\delta n(\mathbf{r})} - \frac{\delta T_{\mathrm{s}}[n]}{\delta n(\mathbf{r})} + \frac{\delta U[n]}{\delta n(\mathbf{r})}, \tag{3.7}$$

up to a constant. Naturally, this non-interacting potential is not a trivial functional of the density since it is burdened to maintain the unique map between a fully interacting Hamiltonian and a conceptually much simpler non-interacting Hamiltonian. Therefore, it must also maintain the uniqueness of the map of the corresponding wave functions. This complexity is reflected in the contributions from the functional derivative of the interacting kinetic energy $\delta T/\delta n$ and the electron-electron interaction $\delta U/\delta n$. Despite this difficulties the KS approach allows for calculating reasonable densities since many approximate xc functionals are able to capture a large fraction of physically relevant effects. One major contribution of $U[n]$ is its mean-field approximation, the Hartree energy

$$U_{\mathrm{H}}[n] = \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \tag{3.8}$$

an explicit density functional with the functional derivative

$$v_{\mathrm{H}}(\mathbf{r}) = \frac{\delta U_{\mathrm{H}}[n]}{\delta n(\mathbf{r})} = \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}'. \tag{3.9}$$

Separating the Hartree energy U_{H} from v_{KS} in Eq. (3.7) and subsequently collecting the remaining inaccessible contributions in the so-called exchange-correlation (xc) functional E_{xc} is the final step to cast the effective potential in the common form

$$v_{\mathrm{KS}}(\mathbf{r}) = v_{\mathrm{ext}} + v_{\mathrm{H}}(\mathbf{r}) + \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(\mathbf{r})}. \tag{3.10}$$

The exchange-correlation energy functional

$$E_{\mathrm{xc}}[n] = T[n] - T_{\mathrm{s}}[n] + U[n] - U_{\mathrm{H}}[n] \tag{3.11}$$

models all “remainder” exchange-correlation effects: Any electron-electron interaction beyond the mean-field Hartree interaction, $U - U_{\mathrm{H}}$, and the difference between the fully interacting kinetic and non-interacting kinetic energy, $T_{\mathrm{c}} = T - T_{\mathrm{s}}$. The energetic contributions of E_{xc} are comparably small to those of the external potential, the non-interacting kinetic energy, or the Hartree energy in many relevant cases. Therefore, already an approximate representation of E_{xc} might yield a decent energy or density. This significantly contributes to the success of DFT. Note that for small energy differences of chemical bonds, the contribution of E_{xc} is nevertheless often decisive.

The functional derivative of $E_{\mathrm{xc}}[n]$ can be written as a local potential, leading to the exchange-correlation potential, or xc potential,

$$v_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(\mathbf{r})}. \tag{3.12}$$

The KS equations take their final form

$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\mathbf{r}) + v_{\mathrm{H}}[n](\mathbf{r}) + v_{\mathrm{xc}}[n](\mathbf{r})\right)\varphi_i(\mathbf{r}) = \varepsilon_i\varphi_i(\mathbf{r}). \tag{3.13}$$

From the KS orbitals the KS Slater determinant Φ follows. The obvious challenge is an accurate description of the xc energy functional E_{xc}. Although a general true xc density functional is guaranteed to exist, it is unknown in practice. The development of accurate approximations to it remains a key task in the DFT research field.

The practical application of the KS scheme follows a reasoning process that is the reverse of its derivation. First, a density functional approximation to the xc functional is chosen. With the known form of some E_{xc}, the functional derivative can be carried out. Then, the xc potential is used in the KS equations with some initial guess for the density. The resulting orbitals give rise to a new density, and the process is iterated. The self-consistent solution to the KS equations is then the approximate ground state density of the approximate xc functional. The energy functional of DFT is given by the interacting energy functional $E[n]$ with $F = T + U$ replaced by Eq. (3.11),

$$E[n] = T_\mathrm{s} + U_\mathrm{H} + \int n v_\mathrm{ext} + E_\mathrm{xc} = \sum_i \varepsilon_i + E_\mathrm{xc} - U_\mathrm{H} - \int n v_\mathrm{xc}. \tag{3.14}$$

Substituting the approximate density into the interacting energy functional $E[n]$ yields the corresponding approximate ground state energy. Thus, the xc functional $E_\mathrm{xc}[n]$ takes the double role of generating a reasonable ground state density in the KS equations on the non-interacting side of the KS scheme while simultaneously approximating the energetic xc contributions beyond the mean-field in the total energy functional on the interacting side [24].

While the development of increasingly sophisticated xc functionals continues, the predictive power of DFT combined with its low computational cost makes it one, if not the, most widely used computational approach in quantum mechanics. Many extensions of the above outline of the DFT method exist, including time-dependent DFT [25], ensemble densities with fractional orbital occupation [26], the generalized Kohn-Sham scheme [27, 28], relativistic DFT for the Dirac equation [29, 30], etc. A relatively straightforward extension used in this thesis is the inclusion of spin [31, 32]. Even though the above formulation does not treat a spin variable, the spin-polarized KS scheme can readily be obtained by denoting spin-dependent quantities, like the density, orbitals, and potentials, with a spin index.

3.2. Exact exchange

The xc energy can be split into contributions from exchange and correlation, $E_\mathrm{xc} = E_\mathrm{x} + E_\mathrm{c}$. The correlation energy [33] is defined to capture all effects of the interacting wave function Ψ that are beyond the description of a single Slater determinant Φ.

$$E_\mathrm{c} = \langle\Psi|\hat{\mathrm{H}}|\Psi\rangle - \langle\Phi|\hat{\mathrm{H}}|\Phi\rangle \tag{3.15}$$

From this definition, $E_\mathrm{c} = T_\mathrm{c} + U - \langle\Phi|U|\Phi\rangle$ follows and upon comparison with Eq. 3.11 an corresponding definition for the exchange follows [33],

$$E_\mathrm{x} = \langle\Phi|U|\Phi\rangle - U_\mathrm{H}. \tag{3.16}$$

The determinant Φ is the KS determinant build from the KS orbitals. The exchange energy is no explicit functional of the density but can be expressed solely in terms of the orbitals.

$$E_\mathrm{x}[\{\varphi_{\sigma i}\}] = -\frac{1}{2}\sum_\sigma \sum_{i,j=1}^{N_\sigma} \int\int \frac{\varphi^*_{\sigma i}(\mathbf{r}')\varphi^*_{\sigma j}(\mathbf{r})\varphi_{\sigma i}(\mathbf{r})\varphi_{\sigma j}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \tag{3.17}$$

Since E_x is a functional of the orbitals instead of the density, evaluating its functional derivative $\delta E_x/\delta n$ is an intricate and demanding task. Nevertheless, the calculation of the functional derivative of the exact exchange can be carried out by the chain rule. Various approximations [34–38] and even exact schemes [39–43] exist to arrive at a self-consistent density accounting for the exact exchange. In this thesis, the exact exchange is approximated by the Krieger-Li-Iafrate (KLI) [34, 35] approximation and the corresponding calculations are denoted by the term xKLI. A beneficial property of the xc potential arising from the exact exchange functional (even in the KLI approximation) is its correct asymptotic behavior. The true xc potential has a Coulomb-like decay [44] of

$$v_{\mathrm{xc}\sigma}(\mathbf{r}) \to -\frac{1}{|\mathbf{r}|} + \mathcal{O}\left(\frac{1}{|\mathbf{r}|^3}\right), \quad |\mathbf{r}| \to \infty \tag{3.18}$$

up to a constant irrespective of spin-channel or -polarization. For systems with a nodal structure in the highest occupied molecular orbital (HOMO) that extends to infinity, a spin-dependent constant offset in the xc potential might arise along the nodal surface, see Sec. 9.3 or Ref. [45].

3.3. Local density approximation

A class of xc functionals that shows tremendous success for its simplicity and computational efficiency is the class of local density approximations (LDA). Due to this class, large systems with several thousands of electrons can be analyzed, accounting for their quantum mechanical nature. The LDAs cast the xc functional in the form

$$E_{\mathrm{xc}}[n] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}(n(\mathbf{r}))\,\mathrm{d}\mathbf{r}, \tag{3.19}$$

with ϵ_{xc} being a function, not functional, of the density. This representation allows for a fast evaluation of the xc potential and an accelerated self-consistent solution. From considerations of the homogeneous electron gas [46] an approximation to the exact exchange functional can be made to arrive at

$$E_{\mathrm{x}}^{\mathrm{LDA}}[n] = -\frac{3}{4}\int n(\mathbf{r})\left(\frac{3}{\pi}n(\mathbf{r})\right)^{\frac{1}{3}}\mathrm{d}\mathbf{r} \tag{3.20}$$

with the corresponding exchange potential

$$v_{\mathrm{x}}(\mathbf{r}) = -\left(\frac{3}{\pi}n(\mathbf{r})\right)^{\frac{1}{3}}. \tag{3.21}$$

While the assumption of a constant density or at least a slowly varying density seems crude, the LDA exchange has proven to yield reasonable results for many systems beyond slowly varying densities [46]. To accompany the exchange by a suitable functional for the correlation energy, the results of QMC calculation of the homogeneous electron gas by Ceperley and Alder [16] are used to parameterize the correlation as a function of the density. In this thesis, the Perdew-Wang parameterization [47] is used for the LDA correlation. Although the LDA is not a reasonable approximation for many systems [48, 49], and far better approximations exist, it represents a sweet spot between computational cost and predictive power.

3.4. Limitations of DFT

Despite its great success, DFT faces severe challenges and fundamental shortcomings. One example is the density-based nature of DFT. Most observables $\hat{\mathrm{A}}$ that are not explicit functionals of the density are difficult or even impossible to evaluate, although the unique map from the density to the wave function should in principle allow the evaluation of $\langle\Psi_0[n]|\hat{\mathrm{A}}|\Psi_0[n]\rangle$. The KS Slater determinant is a rather crude approximation to the ground state wave function, due to its auxiliary character that is only intended to provide the correct density. Another limitation is that the density is bound to the ground state wave function, making statements about excited states a complicated objective. Furthermore, there seems to be a clear trade-off between the accuracy of a density functional approximation and its computational cost. Suppose the limitations of a certain functional or class of functionals are known, for example, regarding the prediction of bond lengths, response to electrical fields, van-der-Waals interaction, etc. In that case, the DFT user can try to choose the proper functional for a given task. This approach, however, requires a posteriori information of the system under study and also the density functional approximation. For systems not adequately representable by a single Slater determinant, i.e. that have a strong static correlation character, many common DFT functionals might yield poor results [48, 49]. In the subsequent chapters, the VMC and DMC methods are outlined, which alleviate the shortcomings of DFT at the cost of sacrificing DFT's favorable computational efficiency.

4. Variational Monte Carlo

The VMC method is a wave function-based method for finding an explicit and accurate representation of the wave function. Known exact properties of the wave function guide the search for such representations. Sophisticated parametrizations of the wave function enable computational optimization of the parameters. The optimization is a minimization of the wave function's energy expectation value in the spirit of the variational principle. The computationally expensive evaluation of the energy expectation value is accomplished by Monte Carlo (MC) integration.

4.1. Variational principle

To start with, the energy expectation value for systems with Hamiltonian $\hat{\mathrm{H}}$ is a functional of the wave function,

$$E[\psi] = \frac{\langle\psi|\hat{\mathrm{H}}|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\int \psi^* \hat{\mathrm{H}} \psi}{\int |\psi|^2}. \tag{4.1}$$

In this thesis, wave functions with spin-arguments are denoted by a capital Ψ, whereas wave functions only depending on the spatial coordinates are denoted by a lowercase ψ. For reasons explained in Sec. 5.1 and Sec. A.1, the above energy functional is spin-free and therefore a spin-free wave function ψ is its minimizer. While minimization of $E[\psi]$ with normalization constraints leads to the Schrödinger equation (2.7), the variational principle instead searches for the ground state only. The search space for the wave function ψ is some subspace $H \subset \mathbb{H}(\mathbb{R}^{3N})$ of the corresponding Hilbert space. Note that symmetry restrictions due to spin-statistics already constrain the full Hilbert space $\mathbb{H}(\mathbb{R}^{3N})$. For fermions, the subspace H respects the anti-symmetry of the wave function. By definition, the minimum of the Hamiltonian's expectation value is the ground state energy

$$E_0 = \min_{\psi \in \mathbb{H}(\mathbb{R}^{3N})} E[\psi] \tag{4.2}$$

Moreover, the minimizing wave function is guaranteed to be the ground state ψ_0. In the case of a degenerate ground state, the energy $E[\psi]$ is minimized by some element in the subspace spanned by the degenerate ground states.

Suppose a sophisticated analytical ansatz for the wave function is available, including adjustable parameters $\mathbf{p} = (p_1 \ldots p_q) \in \mathbb{R}^q$. Then, the wave function $\psi_T[\mathbf{p}]$ probes the subspace H by variation of the parameters. The wave function ψ_T is therefore called a trial wave function. Note that a vast number of different forms of trial wave functions have been proposed, incorporating different physical effects and exploiting numerical strategies [50–57]. Therefore, a variation in the analytical form of the trail wave functions also probes the variational space. A parameterized trial wave function implies that the probability distribution $\psi_T^2[\mathbf{p}](\mathbf{R})$ and the Hamiltonian applied to the trial wave function $\hat{\mathrm{H}}\psi_T[\mathbf{p}](\mathbf{R})$ are known expressions of the parameters, and can be evaluated at any coordinate $\mathbf{R}$. Although the expressions are known, their evaluation is not necessarily computationally cheap. With a given form and parametrization the energy functional becomes an ordinary function of the parameter vector, $E[\psi_T[\mathbf{p}]] = E(\mathbf{p})$. Various standard q-dimensional optimization strategies are readily available to find not only the minimizing parameters $\mathbf{p}_0$ but also reasonable approximations to the ground state energy $E(\mathbf{p}_0)$ and the ground state wave function $\psi_T[\mathbf{p}_0]$. The VMC energy

$$E_{\mathrm{VMC}} = E(\mathbf{p}_0) = \min_{\mathbf{p}\in\mathbb{R}^q} E(\mathbf{p}) \geq E_0 \tag{4.3}$$

is an upper bound to the ground state energy by the variational principle, and more appropriate parameterizations $\psi_T[\mathbf{p}]$ allow for better estimates of E_0.

While the idea is simple, evaluating $E(\mathbf{p})$ is a major computational issue. The exponential scaling of the integral of the energy expectation value in Eq. (4.1) with the number of electrons renders the evaluation of a general trial wave function next to impossible with standard approaches like finite differences or basis sets for the full configuration space, even for small systems. MC methods are required to estimate $E(\mathbf{p})$ and are discussed in the next section.

Note that the true ground state is generally beyond any ansatz with a finite number of parameters. Nevertheless, very accurate trial wave functions exist that achieve chemical accuracy for real systems [52] and can recover the correlation energy to more than 99%. The percentage of correlation energy is a standard measure for the accuracy of a QMC energy. It relates the calculated energy E with the Hartree-Fock energy E_{HF} and an accurate estimate of the true energy in the non-relativistic limit E_{nrl}. The recovered correlation energy E_{c} for some calculation is defined as the difference of the calculated energy to the Hartree-Fock energy, $E_{\mathrm{c}} = E - E_{\mathrm{HF}}$. The maximum correlation energy to recover is $E_{\mathrm{nrl}} - E_{\mathrm{HF}}$. Thus the percentage of correlation energy recovered is given by

$$E_{\mathrm{c}}(\%) = \frac{E - E_{\mathrm{HF}}}{E_{\mathrm{nrl}} - E_{\mathrm{HF}}}. \tag{4.4}$$

However, in the absence of more accurate ground state energies as reference, the gap between the VMC and the ground state energy remains unknown. Note that the above definition of the correlation energy differs from the correlation energy in DFT. The correlation energy in DFT is defined by Eq. (3.15) and evaluated with the KS orbitals, and not with the Hartree-Fock orbitals, see Sec. 3.2.

The major advantage of the VMC method is the freedom in the analytic representation of the trial wave function. This freedom allows more accurate representations beyond linear combinations of Slater determinants. Its wave function can thus be more accurate compared to methods like coupled cluster, configuration interaction, multi configurational self-consistent field, etc. In Chapt. 5, properties of the exact wave function are studied and appropriate trail wave functions are constructed in Sec. 5.3 to incorporate exact properties in the ansatz. In this work, the variational ansatz only includes a single parameter since the focus of this thesis does not lie in VMC optimization. A straightforward least squares minimization optimizes the single parameter. Other properties of the trial wave function are fixed and not subject to an optimization. For more sophisticated trial wave functions with a large number of parameters, however, a variety of standard optimization strategies is available. Extensions of these strategies exist, that are explicitly tailored to MC applications [50, 58–60].

4.2. Monte Carlo integration

The composite configuration $\mathbf{R} = (\mathbf{r}_1, \ldots \mathbf{r}_N)$ collecting all electron coordinates is an element of the configuration space Ω^N, where Ω might be the whole $\mathbb{R}^3$ or a finite region. Irrespective of finiteness, the dimension of Ω^N grows exponentially with the number of electrons N. The major obstacle in utilizing the variational principle is the multidimensional integral in Eq. (4.1). Suppose the integral $E(\mathbf{p})$ for a single electron can be evaluated with m degrees of freedom, e.g., by discretization on a grid with m grid points. For a many-body wave function of N electrons, $E(\mathbf{p})$ requires integration over m^N degrees of freedom. This exponential wall limits the evaluation of $E(\mathbf{p})$ with standard measures to only a few electrons. For example, the relatively small nitrogen molecule with 14 electrons already has a configuration space of dimension 42. To represent the full wave function of the nitrogen molecule on a grid, even for the smallest possible grid of only two grid points per dimension, more than 4×10^{12} grid points are required to represent Ω^N. Therefore, standard methods to solve the Schrödinger equation as a PDE on its full configuration space are impossible for all but the tiniest systems.

The efficient evaluation of integrals in high-dimensional spaces is a key competence of MC methods. In the following, the variational principle is rewritten to fit in the MC framework. A detailed discussion of the MC approach is presented in Sec. C.2. The central step to make the energy functional $E[\psi_T]$ of Eq. (4.1) compatible with MC integration is rewriting

$$E[\psi_T] = \langle\psi_T|\psi_T\rangle^{-1} \int\limits_{\Omega^N} \psi_T^* \hat{\mathrm{H}} \psi_T = \int\limits_{\Omega^N} \frac{|\psi_T(\mathbf{R})|^2}{\langle\psi_T|\psi_T\rangle} \frac{\hat{\mathrm{H}}(\mathbf{R})\psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} \,\mathrm{d}\mathbf{R} = \int\limits_{\Omega^N} f E_L[\psi_T]. \tag{4.5}$$

The normalized wave function probability $f(\mathbf{R}) = \langle\psi_T|\psi_T\rangle^{-1}|\psi_T(\mathbf{R})|^2$ qualifies as a probability function on Ω^N. Note that with a real representation of the trial wave function $|\psi_T(\mathbf{R})|^2 = \psi_T(\mathbf{R})^2$. The second term in the integral is the so-called local energy,

$$E_L[\psi_T](\mathbf{R}) = \psi_T(\mathbf{R})^{-1}\hat{\mathrm{H}}(\mathbf{R})\psi_T(\mathbf{R}). \tag{4.6}$$

It results from applying the Hamiltonian to the wave function and dividing the result by the wave function's value. The local energy is no well-behaved quantity on the complete configuration space. It shows discontinuities and different kinds of singularities at special points or hypersurfaces of the configuration space, see Sec. 5.6. It is nevertheless evaluated numerically and various measures are taken to remedy the consequences of its pathological behavior, cf. Sec 5.1, Sec. 5.4 and Sec. 6.13.

The integral can now be evaluated with the MC method and reduces to an average of $E_L[\psi_T]$. Suppose an infinite series $\mathbf{R}_1, \mathbf{R}_2, \ldots$ of random configurations is available, that are all distributed according to f. In compact notation this reads $\mathbf{R}_i \sim f$ for all i. The integral can be evaluated by the law of large numbers [61],

$$E[\psi_T] = \int\limits_{\Omega^N} f E_L[\psi_T] = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} E_L[\psi_T](\mathbf{R}_i). \tag{4.7}$$

In practice, the series is finite, and the number of configurations M can be interpreted as the degrees of freedom in the MC integration, which is limited to a large number in real applications. The construction of such a series of configurations, of course, requires a method to draw random variates, i.e., a numerical realizations of a random variable, that are distributed according to f. The direct generation of random variates, however, is only possible for a limited number of elementary probability distributions. The trial wave function is, in almost all cases, far too complicated to sample it directly. The general purpose tool to generate such a series is the Metropolis-Hastings (MH) algorithm [62, 63]. It constructs a Markov chain, $\mathbf{R}_1, \mathbf{R}_2, \cdots \sim f$, with superior efficiency compared to, for example, rejection sampling. It is described in detail in C.4.

In short, the MH algorithm works as follows: First, it requires a target probability function f and additionally a proposal transition probability $G(\mathbf{R}' \leftarrow \mathbf{R})$ that assigns a probability for the transition from any configuration $\mathbf{R}$ to any other configuration $\mathbf{R}'$. This transition probability must allow the efficient generation of a random configuration from the conditional probability function $G(\cdot \leftarrow \mathbf{R})$. The dot in the first argument of the Green function is a placeholder, while the second argument is actually inserted into the Green function to render it a probability distribution of a single variable. The MH algorithm then iteratively updates a configuration $\mathbf{R}_i$, with a random initial configuration $\mathbf{R}_1$. First, from the transition probability $G(\cdot \leftarrow \mathbf{R}_i)$ a configuration $\mathbf{R}'$ is drawn. With the two configurations $\mathbf{R}_i$ and $\mathbf{R}'$ the so-called acceptance probability

$$p = A(\mathbf{R}' \leftarrow \mathbf{R}_i) = \min\left[1, \frac{f(\mathbf{R}')G(\mathbf{R}_i \leftarrow \mathbf{R}')}{f(\mathbf{R}_i)G(\mathbf{R}' \leftarrow \mathbf{R}_i)}\right] \tag{4.8}$$

is calculated, cf. Sec. C.4. Next, the new element of the series is set to the proposed configuration $\mathbf{R}_{i+1} = \mathbf{R}'$ with probability p or is left unchanged $\mathbf{R}_{i+1} = \mathbf{R}_i$ with probability $1-p$. This process is repeated to form a Markov chain of configurations. After a starting phase, in which the dependence of the series on the initial configuration $\mathbf{R}_1$ has vanished, the series is guaranteed [63] to be distributed according to f. Formally, the MH algorithm requires certain conditions on G regarding ergodicity, cf. Sec. C.2, which are satisfied for a broad range of transition probability distributions. After a reasonable number of configurations have been generated, the set of configurations allows for the accumulation of the local energy by the law of large numbers.

Instead of an integer index i to label the elements of the Markov chain, a notion of time τ can be introduced. Reinterpreting the MH algorithm as a stochastic equation of motion that advances a given configuration by a time step $\Delta\tau$, the Markov chain can be written as $\mathbf{R}(0), \mathbf{R}(\Delta\tau), \mathbf{R}(2\Delta\tau), \ldots$. The obvious interpretation would identify the configuration as a classical particle subject to stochastic forces, in short $\mathbf{R}(\tau)$. Although this shorthand notation suggests a continuous time, the time evolution of the Markov chain is still discrete. The sampling of integral (4.5), i.e. the generation of a Markov chain $\mathbf{R}(\tau)$, can be parallelized trivially. Instead of a single one, M configurations $\mathbf{R}_1(\tau), \ldots \mathbf{R}_M(\tau)$ can be instantiated and propagated simultaneously, making the MC integration a powerful tool on modern distributed machines and cluster computers. The normalized stochastic representation of ψ_T^2 for a time step τ is then given by

$$f(\mathbf{R}, \tau) = \frac{1}{M}\sum_{k=1}^{M} \delta(\mathbf{R} - \mathbf{R}_k(\tau)) \tag{4.9}$$

This form will reappear in the DMC method with a different and more physical interpretation of the time τ. Here $\mathbf{R}_k(0)$ are randomly chosen configurations. The initial ensemble of configurations has to reach equilibrium at some time $\tau_{\text{eq}} = T_{\text{eq}}\Delta\tau$. At equilibrium, the averages over all observables become stationary in time τ, which is used to detect the equilibrium in actual computations. The non-equilibrium distribution is usually discarded, and in the following, the time variable τ denotes the time passed since equilibrium. After the equilibration phase, which formally ensures the expectation value $\operatorname{Exp}[f(\cdot, \tau)] = \operatorname{Exp}[f(\cdot, \tau + \Delta\tau)]$ for all $\tau > 0$, statistical certainty can be build up by sampling the equilibrium probability distribution for increasing number of propagation steps T. The distribution at time $\tau = T\Delta\tau$ including every configuration reads

$$F(\mathbf{R}) = \frac{1}{T}\sum_{\tau=1}^{T} f(\mathbf{R}, \tau) = \frac{1}{MT}\sum_{\tau=1}^{T}\sum_{k=1}^{M} \delta(\mathbf{R} - \mathbf{R}_k(\tau)) \tag{4.10}$$

and the energy of Eq. (4.7) can be estimated by

$$\begin{aligned} E[\psi_T] &= \int \frac{1}{MT}\sum_{\tau=1}^{T}\sum_{k=1}^{M} \delta(\mathbf{R} - \mathbf{R}_k(\tau)) E_L[\psi_T](\mathbf{R}) \\ &= \frac{1}{MT}\sum_{\tau=1}^{T}\sum_{k=1}^{M} E_L[\psi_T](\mathbf{R}_k(\tau)) \end{aligned} \tag{4.11}$$

This equation aligns with the law of large numbers in Eq. (4.7). Note the equivalence of the law of large numbers and the representation of the probability distribution by an ensemble of Dirac distributions. In probability theory, representations using the Dirac measure are not uncommon since the Dirac measure represents a single realization or measurement in continuous space, bridging between discrete and continuous measures. For every configuration used in the law of large numbers, a Dirac contribution is added to the corresponding representation of the probability distribution F. With the target distribution being the trial wave function probability ψ_T^2, the MH algorithm only requires a proposal transition probability. Candidates are discussed in the following section.

4.3. Green functions

The proposal transition probability kernel of the MH algorithm can be cast as a Green function. It propagates a probability distribution $f(\mathbf{R}, \tau)$ by a time step $\Delta\tau$ to yield

$$f(\mathbf{R}', \tau + \Delta\tau) = \int G(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau) f(\mathbf{R}, \tau) \, \mathrm{d}\mathbf{R} \tag{4.12}$$

and reduces to a Dirac distribution for a vanishing time step, $G(\mathbf{R}' \leftarrow \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$. Instead of the notation $G(\mathbf{R}', \mathbf{R}, \Delta\tau)$ the arguments of the Green function are separated by an arrow, $G(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau)$, to indicate the direction of the transition from configuration $\mathbf{R}$ to $\mathbf{R}'$. While the MH algorithm admits transition kernels that are not compatible with the notion of a time step and do not reduce to a Dirac distribution, most reasonable transition kernels do, and the terms proposal transition kernel and Green function will be used synonymously.

Any proposal kernel for some non-trivial f is situated between two limiting cases. The most simple Green function is uniform sampling, $G(\mathbf{R}' \leftarrow \mathbf{R}) = \text{const.}$, but it is inefficient in almost all cases and essentially resembles uniform rejection sampling. The acceptance rate, i.e., the average of the acceptance probability A from Eq. (4.8) over the ensemble of configurations, will be considerably smaller than one since most of the proposed moves are rejected, and the configuration space is explored inefficiently.

The optimal case is the direct sampling from f, with $G(\mathbf{R}' \leftarrow \mathbf{R}) = f(\mathbf{R}')$. Of course, this case contradicts the assumption that f is a non-trivial probability distribution, which cannot be drawn from directly. Nevertheless, in this case, the acceptance rate is unity. New configurations are independent of the previous ones, the configurations space is probed with high efficiency and the serial correlation vanishes, cf. Sec. C.5. In general, the goal is to approximate the case of optimal sampling $G(\mathbf{R}' \leftarrow \mathbf{R}) \approx f(\mathbf{R}')$ resulting in little correlation and fast exploration of the configuration space. At the same time, the Green function must allow the efficient drawing of configurations. There are many choices for the Green function, balancing computational expense and accurate description of the target distribution. If, for example, f is a function that is computationally expensive to evaluate, a more accurate Green function might be appropriate. If f is computationally cheap, a rather simple Green function can be utilized to the contrary. Many sophisticated transition probabilities are parameterized in terms of elementary probability distributions or their combinations with cleverly chosen parameters.

One simple but commonly used Green function is a uniform proposal in a confined region enclosing $\mathbf{R}$, for example, the Green function that is uniformly distributed in an n-dimensional sphere around the center $\mathbf{R}$. The sphere's radius can be used to vary the acceptance rate. For a proposal configuration in a small volume around $\mathbf{R}$, the acceptance rate is higher compared to a proposal in a larger volume. Furthermore, the maximum distance $|\mathbf{R}' - \mathbf{R}|$ for a single propagation is limited by the extent of the sphere, limiting exploration of the state space. A $3N$-dimensional normal distribution,

$$G(\mathbf{R}' \leftarrow \mathbf{R}, \sigma) = \left(2\pi\sigma^2\right)^{-3N/2} \exp\left(-\frac{1}{2\sigma^2} \left|\mathbf{R}' - \mathbf{R}\right|^2\right), \tag{4.13}$$

is another prominent candidate. Centered around the coordinate $\mathbf{R}$ and equipped with an adjustable scalar variance σ^2, the normal distribution can be drawn from efficiently. It is also possible to use the truly multidimensional normal distribution with an $3N$-dimensional covariance matrix instead of the scalar σ^2, providing more adjustable parameters that can be adapted to f. Both normal and uniform distributions have a key limitation. They propose updates with strictly diffusive character, only differing in their underlying probability distribution. However, they incorporate no information about the target function f.

To include information from f in the Green function, one can, for example, take a closer look at the stationary Fokker-Planck equation [64, 65] of $f(\mathbf{R}, \tau)$ with some diffusion constant D and a spatially varying drift vector $\mathbf{u}$, reading

$$0 = D\nabla^2 f(\mathbf{R}) - \nabla(\mathbf{u}(\mathbf{R}) \cdot f(\mathbf{R})) = \nabla \cdot \left[D\nabla f(\mathbf{R}) - \mathbf{u}(\mathbf{R}) \cdot f(\mathbf{R})\right]. \tag{4.14}$$

The divergence of the square brackets vanishes in equilibrium. Assuming in addition that also the square brackets vanish leads to an expression for $\mathbf{u}$,

$$\mathbf{u}(\mathbf{R}) = -D^{-1}\frac{\nabla f(\mathbf{R})}{f(\mathbf{R})}. \tag{4.15}$$

With a known function $\mathbf{u}$, an approximate solution to the time-dependent version of the above equilibrium equation is given by

$$G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) = (4\pi D\tau)^{-3N/2} \exp\left(-\frac{1}{4D\tau}\left|\mathbf{R}' - \mathbf{R} - \mathbf{u}(\mathbf{R})\tau\right|^2\right), \tag{4.16}$$

cf. Sec. D.4. This Green function generates very reasonable proposal configurations with a high acceptance rate. The proposal of a new configuration $\mathbf{R}'$ is not only governed by a purely diffusive random displacement but also exhibits the drift term $\mathbf{u}(\mathbf{R})\tau$, which points along the direction of the steepest gradient of f. Although the Green function based on the Fokker-Planck equation is a reasonable approximation, drawing a configuration from it is straightforward,

$$\mathbf{R}' = \mathbf{R} + \mathbf{u}(\mathbf{R})\tau + \xi \sim G(\cdot \leftarrow \mathbf{R}, \tau), \tag{4.17}$$

with a normally distributed $3N$-dimensional random variable ξ with vanishing expectation value and variance $2D\tau$. The coordinate proposed by the drift alone, $\mathbf{R} + \mathbf{u}(\mathbf{R})\tau$, often has a higher probability than $f(\mathbf{R})$, which increases the acceptance rate. The additional diffusion around the drifted position allows for the non-deterministic exploration of configuration space, for example, by escaping local maxima or advancing into low-probability regions. With the Green function based on the Fokker-Planck equation, larger time steps are possible, which de-correlates the Markov chain. The parameters D and τ of the proposal Green function are readily interpreted from the Fokker-Planck equation to be the diffusion constant and the diffused time, respectively. The time step, which equally affects drift and diffusion, should be limited to prevent $\mathbf{u}(\mathbf{R})\tau$ from overshooting the region of high probability. For a large diffusion constant and small time step, the Green function character is almost pure diffusion, resembling the normally distributed proposal. In the limit of vanishing D, the transition probability is dominated by the drift and becomes purely deterministic, disqualifying as a proposal due to violation of ergodicity, see Sec. C.4.

Since in almost all applications $\mathbf{R} = (\mathbf{r}_1, \dots \mathbf{r}_N)$ is a configuration involving multiple electron coordinates, one can devise a Green function that moves each electron coordinate separately. Starting with the first electron, an updated position to $\mathbf{r}_1$ is proposed, the acceptance probability is calculated, and the electron's new coordinates position is set accordingly. This process is repeated until all electron coordinates in $\mathbf{R}$ are updated. In a QMC application, such a scheme enables tailoring the proposal Green's function to the update of a single electron. Instead of simultaneously drawing new random coordinates for each electron independently, the step-by-step update only changes one electron coordinate at a time, which might allow a higher acceptance rate. For example, it is possible to stochastically avoid updates that propose a configuration in which two electrons are close to each other and are likely to be rejected since the wave function is zero for electron coincidences. This is a reasonable approach for systems of many electrons [66]. Furthermore, the Sherman-Morrison formula [67] can be invoked to arrive at an efficient update of the Slater determinants involved. For systems of only some electrons, the benefit of the electron-by-electron update is marginal and therefore this scheme was not used in this thesis.

In a QMC code that features the DMC method, an equally convenient and efficient proposal Green function for the VMC is a DMC Green function. The DMC Green function originates from the imaginary-time Schrödinger equation and is usually quite accurate. Therefore, it can also be applied within the VMC framework.

4.4. Observables

As seen in Sec. 4.2 the law of large numbers is utilized to calculate the energy of some parameterized wave function ψ_T. Other observables can be calculated as well using the same ensembles of configurations. For any local or semi-local observable $\hat{\mathrm{O}}$, its analogon to the local energy is

$$O_L[\psi_T](\mathbf{R}) = \psi_T(\mathbf{R})^{-1}\hat{\mathrm{O}}\psi_T(\mathbf{R}), \tag{4.18}$$

where any operator action in $\hat{\mathrm{O}}$ is carried out on the trial wave function. Thus, O_L is a multiplicative function without operator character. The expectation value of $\hat{\mathrm{O}}$ is then evaluated as in Eq. (4.11) by

$$\frac{\langle\psi_T|\hat{\mathrm{O}}\psi_T\rangle}{\langle\psi_T|\psi_T\rangle} = \frac{1}{MT}\sum_{\tau=1}^{T}\sum_{k=1}^{M} O_L[\psi_T](\mathbf{R}_k(\tau)). \tag{4.19}$$

Note that here and in the following, the notation for the time argument is slightly abused: Integer time steps are not discriminated from the actual time. For example, while the summation index $\tau = 1, \dots T$ is integer, it is used in the summation as the actual time in the expression $\mathbf{R}_k(\tau)$ instead of the formally correct $\mathbf{R}_k(\tau\Delta\tau)$. While this estimation of the observable's expectation value is fully justified, the same quantity can be estimated with less uncertainty by the following consideration. In the MH algorithm an acceptance probability p is calculated to decide whether the proposed configuration $\mathbf{R}'$ or the last configuration $\mathbf{R}(\tau)$ is used as the next element $\mathbf{R}(\tau+\Delta\tau)$ of the Markov chain and the Markov chain reads

$$\begin{cases} \mathbf{R}(0), \mathbf{R}(\Delta\tau), \dots, \mathbf{R}(\tau), \mathbf{R}' & \text{with probability } p \\ \mathbf{R}(0), \mathbf{R}(\Delta\tau), \dots, \mathbf{R}(\tau), \mathbf{R}(\tau) & \text{with probability } 1-p. \end{cases} \tag{4.20}$$

Thus, the Markov chain at time $\tau + \Delta\tau$ is itself a random variable. By collapsing the Markov chain to an actual variate, which is, of course, necessary to propagate the configurations further, information is eliminated. However, both Markov chains can be accounted for when estimating the observable's expectation value. The corresponding representation of $f(\cdot, \tau + \Delta\tau)$ is given by

$$f(\mathbf{R}, \tau + \Delta\tau) = \frac{1}{M}\sum_{k=1}^{M}\left[p_k(\tau)\delta(\mathbf{R} - \mathbf{R}_k'(\tau)) + (1 - p_k(\tau))\delta(\mathbf{R} - \mathbf{R}_k(\tau))\right], \tag{4.21}$$

where $p_k(\tau)$ is the acceptance probability for the proposed configuration $\mathbf{R}_k'(\tau)$ at time τ. This representation for f has a lower variance since the configurations' contributions are weighted according to their acceptance probabilities and, therefore, pose more statistical certainty compared to the unweighted configurations. The estimate for the observable reads

$$\frac{\langle\psi_T|\hat{\mathrm{O}}\psi_T\rangle}{\langle\psi_T|\psi_T\rangle} = \frac{1}{MT}\sum_{\tau=1}^{T}\sum_{k=1}^{M}\left[p_k(\tau)O_L(\mathbf{R}'_k(\tau)) + (1-p_k(\tau))O_L(\mathbf{R}_k(\tau))\right]. \tag{4.22}$$

This expression is of particular interest since all the required ingredients are already calculated in the MH algorithm and the extra certainty is generated at no additional computational cost.

A note on the notation is appropriate here. While the estimators from Eq. (4.19) as well as the above Eq. (4.22) have the same expectation value, they are not the same random variables, which is already apparent from their differing variance. Therefore, they deserve a different label and more precise notation to highlight their differences. Consistently maintaining this notational precision throughout this thesis would overload the expressions and hinder clarity and readability. In the following, the notation is kept short and no discriminating symbols for the estimators are introduced, but sufficient explanation is given in the relevant cases.

Another central quantity alongside the total energy is the quantum variance Σ^2 of the state ψ_T, given by

$$\Sigma^2[\psi_T] = \frac{\left\langle\psi_T\middle|\left(\hat{\mathrm{H}} - E[\psi_T]\right)^2\middle|\psi_T\right\rangle}{\langle\psi_T|\psi_T\rangle} \tag{4.23}$$

If ψ_T approaches an eigenstate of the Hamiltonian, $\Sigma^2[\psi_T]$ approaches zero. This measure gives additional insight into the accuracy of the trial wave function. In general, the variance of the trial wave function will only approach zero but not be zero identically since any parametrization for non-trivial systems cannot fully recover the true ground state. Since the variance is bound from below by zero and has a defined limit for the ground state, it is a reasonable proxy for quality of the trial wave function. In VMC optimization, the variance can be helpful to find the optimal parameters [50, 58–60, 68]. The variance can be calculated from the identity

$$\left\langle\psi_T\middle|\left(\hat{\mathrm{H}} - E_{\mathrm{ref}}\right)^2\middle|\psi_T\right\rangle = \left\langle\left(\hat{\mathrm{H}} - E_{\mathrm{ref}}\right)\psi_T\middle|\left(\hat{\mathrm{H}} - E_{\mathrm{ref}}\right)\psi_T\right\rangle, \tag{4.24}$$

and its local representation reads

$$\Sigma_L^2[\psi_T](\mathbf{R}) = \left(E_L[\psi_T](\mathbf{R}) - E_{\mathrm{ref}}\right)^2, \tag{4.25}$$

with some reference energy E_{ref} that is the best estimate for the VMC ground state energy available. The estimation of the one-body electron density is another example of an estimator and is analyzed in detail in Sec. 7.

So far, only the expectation values of the observables have been addressed. Stemming from the summation of random variables, their estimators are random variables as well. They are thus subject to statistical fluctuation and only reliable within some error range. The assessment of the statistical error of the estimated expectation value is equally important to the interpretation of the results of an MC calculation. The statistical error $\mathrm{s}[X]$ of some random variable X is the square root of the variance of X, see Sec. C.1. While the estimation of the expectation value is quite simple, this is not the case for the variance, cf. Sec. C.3. The variance can only be evaluated trivially via the law of large numbers for independent and identically distributed (i.i.d.) random variables. Even though the Markov chains $\mathbf{R}_k(\tau)$ are independent for different configurations k, they exhibit a natural, non-vanishing correlation in time τ. Sections C.5 and C.6 discuss the calculation of the variance and its limitations.

4.5. Convergence and analysis

From the Markov chain of configurations $(\mathbf{R}_k(\tau))$ a series of local observations $(O_L(\mathbf{R}_k(\tau))$ follows for every observable. With this series a histogram of the distribution of the random variable O_L can be calculated. The histogram grants insight into many aspects of the underlying distribution. In Fig. 4.1 a histogram for the local energy of diatomic lithium Li_2 is shown. The plot additionally shows the minimal electron-ion-distance,

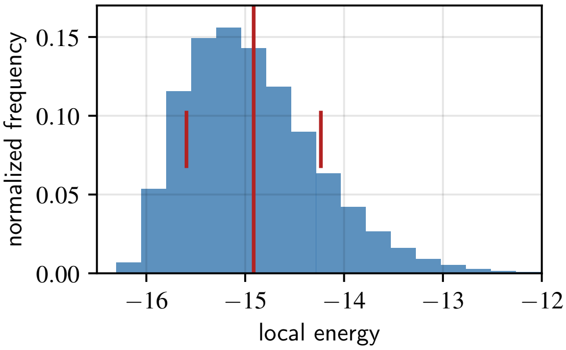


Figure 4.1.: The histograms of the local energy and the minimal distance of Eq. (4.26) are shown with different bin sizes. The VMC calculation belongs to the diatomic Li_2 and is based on orbitals from LDA calculation and a simple Padé Jastrow factor for the electron-electron interaction. The middle red line indicates the expectation value of the distribution, while the shorter lines indicate the standard error centered around the expectation value. The differing resolution or bin size of the left and right panel is due to the long tail and outliers in the local energy.

$$d_{\mathrm{en}}(\mathbf{R}) = \min_{i,\eta} \left|\mathbf{r}_i - \boldsymbol{\xi}_\eta\right| , \tag{4.26}$$

where the minimization is over all electrons i of the configuration $\mathbf{R}$ and all ions η in the system. This distance is an example of a regular observable. It is continuous and bound from below and above. The evaluation of the local energy, on the other hand, is a highly non-trivial task. The wave function's Laplacian includes derivatives of functions with cusps, and evaluating the Hamiltonian's potentials $U + V$ is complicated by the Coulomb singularity near particle coincidences. If, for example, an electron approaches a nucleus, the singularity in the kinetic energy has to cancel with the Coulomb singularity in V. While it is readily shown that the singularities cancel analytically with a correct electron nuclear cusp, cf. Sec. B.1, an accurate numerical cancellation is much more difficult to achieve. These pathologically difficult characteristics of the local energy are derived in Chapt. B and visualized in Sec. 5.6. They are independent of any MC calculation since they are analytical expressions of the trial wave function.

Fig. 4.2 represents a typical time evolution of the estimated observables. After an initial phase of equilibration, the observables converge to their equilibrium value. The error of the estimate decreases during the accumulation of more configurations. Fig. 4.2 visualizes an important fact of the MC method. Although the VMC energy is an upper bound to the true ground state energy, the naive interpretation of the estimated VMC energy might be non-variational: Unless the VMC energy is converged and its statistical error is respected, it might be more negative than the true ground state energy.

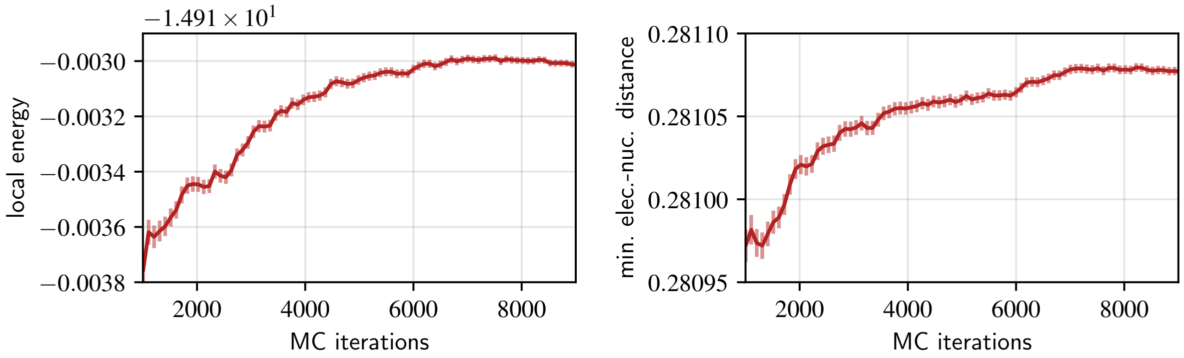


Figure 4.2.: Time evolution of the local energy and the minimal electron-nucleus-distance for the Li_2 system based on a single-reference Slater-Jastrow trail wave function with LDA orbitals. The errors and fluctuations are quite small, since a total of $M = 3.2 \times 10^5$ configurations are propagated in parallel. The initial 1000 MC iterations are not shown, since the estimators show larger fluctuations in this early phase of equilibration.

5. Wave function

Although the exact wave function remains unknown for almost all systems of physical interest, some of its properties are known rigorously. An example already mentioned in Sec. 4.5 is the nuclear cusp of the wave function. In this chapter, the behavior of the exact and a single-reference Slater-Jastrow trial wave function, their gradients, and their Laplacians near critical points in the configuration space are derived.

To begin with, however, the standard representation of the wave function is modified to arrive at its usual representation in DMC. The great redundancy of the wave function due to the anti-symmetry can be reduced if the Hamiltonian is spin-independent. The wave function is then represented in an equivalent but numerically more feasible way.

The conclusions of the analysis lead to construction principles for high-quality trail wave functions.

5.1. Wave function representation

The total anti-symmetry of the wave function is crucial for describing indistinguishable particles but also introduces redundancy in the representation of a quantum state. In the following, only wave functions of definite spin are treated, as well as observables that respect the indistinguishability of particles and are therefore symmetric upon coordinate exchange. Additionally, the Hamiltonian is required to be spin-independent. If the Hamiltonian had spin-contributions, the eigenstates, especially the ground state, could have entangled spin states, i.e. correlated spin states of different spin quantum numbers. This case is explicitly excluded in this thesis.

Consider a general Slater determinant of spin-space orbitals $\varphi_i(\mathbf{x})$ in its explicitly anti-symmetrized form

$$\Phi(\mathbf{X}) = \sqrt{N!}\hat{\mathrm{A}} \prod_{i=1}^{N} \varphi_i(\mathbf{x}_i), \tag{5.1}$$

with the anti-symmetrizer $\hat{\mathrm{A}}$ of N coordinates, that sums over signed permutations of the symmetric group $\mathbb{S}(N)$,

$$\hat{\mathrm{A}} = \frac{1}{N!} \sum_{\pi \in \mathbb{S}(N)} \mathrm{sign}(\pi)\pi. \tag{5.2}$$

A Slater determinant of defined spin $S = \frac{1}{2}(N_\alpha - N_\beta)$ representing N fermions, with N_α electrons in the α-channel and $N_\beta = N - N_\alpha$ in the β-channel, reads

$$\Phi(\mathbf{X}) = \sqrt{N!}\hat{\mathrm{A}} \prod_{\sigma=\alpha,\beta} \prod_{i=1}^{N_\sigma} \sigma(s_i)\varphi_{\sigma i}(\mathbf{R}_{\sigma i}), \tag{5.3}$$

where the spin-functions $\{\alpha, \beta\}$ form an orthonormal spin basis. The symbols α and β serve as both, spin functions and spin indices, with the intended meaning inferred from the context. The spin-spatial orbitals $\varphi_i(\mathbf{x})$ are reordered with respect to their corresponding spin-channel to $\varphi_{\sigma i}(\mathbf{r})$. To simplify the notation, the configuration $\mathbf{R}$ is split in two configurations: The configuration of the spin-α-channel, $\mathbf{R}_\alpha = (\mathbf{r}_1, \ldots, \mathbf{r}_{N_\alpha})$, and the configuration of the spin-β-channel, $\mathbf{R}_\beta = (\mathbf{r}_{N_\alpha+1}, \ldots, \mathbf{r}_N)$. The separation of the configuration into spin channels is of mathematical and artificial character. Due to the anti-symmetrizer, any mapping of an electron to a definite spin is undone, and the determinant remains inherently anti-symmetric and cannot distinguish electrons or spins. However, the representation of Eq. (5.3) is more complicated than necessary. It is shown in the appendix A.1 that the following partially anti-symmetrized representation $\tilde{\Phi}$ holds the same physical results as the fully symmetrized Slater determinant Φ but is written as a product of spatial and spin coordinates.

$$\tilde{\Phi}(\mathbf{X}) = \left(\prod_{i=1}^{N_\alpha} \alpha(s_i)\right) \left(\prod_{i=N_\alpha+1}^{N} \beta(s_i)\right) \phi(\mathbf{R}) \tag{5.4}$$

The partially anti-symmetric spatial Slater determinant ϕ is a product of a normalized α- and β-determinant

$$\phi(\mathbf{R}) = \phi_\alpha(\mathbf{R}_\alpha)\phi_\beta(\mathbf{R}_\beta). \tag{5.5}$$

This representation fixes the first $1, \dots N_\alpha$ coordinates, or equivalently the coordinates of $\mathbf{R}_\alpha$, to the spin state of α, while all coordinates in $\mathbf{R}_\beta$ are affiliated with the β channel. Therefore, the representation of Eq. (5.4) is not totally anti-symmetric with respect to particle exchange. Nevertheless, this form can be used equivalently for almost all calculations, namely overlap integrals and expectation values, cf. A.1. The numerical advantage of this form is its computational cost. While a determinant is usually calculated in $\mathcal{O}(N^3)$ complexity, with the above representation, only $\mathcal{O}(2(N/2)^3)$ are required in a spin-unpolarized system. Since in QMC applications, the most time is spent calculating the determinants and their derivatives, a gain in computational efficiency of 300% is achieved.

This simplified representation is not restricted to a single Slater determinant. By linear combination of Slater determinants, any correlated wave function with definite spin S can be cast in the same partially anti-symmetric form since all building determinants of a correlated wave function of definite spin must also be of definite spin. The representation analogous to Eq. (5.4) for the correlated wave function is

$$\tilde{\Psi}(\mathbf{X}) = \left(\prod_{i=1}^{N_\alpha} \alpha(s_i)\right) \left(\prod_{i=N_\alpha+1}^{N} \beta(s_i)\right) \psi(\mathbf{R}). \tag{5.6}$$

While ψ is not a product, as is ϕ from Eq. (5.5), it has the same symmetries. For any two permutations $\pi_\alpha \in \mathbb{S}(\{1, \dots N_\alpha\})$ and $\pi_\beta \in \mathbb{S}(\{N_\alpha + 1, \dots N\})$, i.e. permutations of the set of the α- or β-channel coordinates, the wave function is anti-symmetric.

$$\pi_\alpha \psi(\mathbf{R}) = \operatorname{sign}(\pi_\alpha)\psi(\mathbf{R}) \tag{5.7}$$

$$\pi_\beta \psi(\mathbf{R}) = \operatorname{sign}(\pi_\beta)\psi(\mathbf{R}) \tag{5.8}$$

For any permutation $\pi \in \mathbb{S}(\{1 \dots N\})$ that does not reduce to two spin-channel-wise permutations, the wave function has no definite permutation symmetry. To illustrate this symmetry, note that the wave function evaluated with two identical electron coordinates $\mathbf{r}_i = \mathbf{r}_j$ does not vanish if i and j belong to different spin channels. This is in stark contrast to the usual form of the wave function. The full symmetry is readily restored by anti-symmetrizing and subsequent normalization

$$\Psi = \sqrt{\frac{N!}{N_\alpha! N_\beta!}} \hat{\mathrm{A}} \tilde{\Psi} \tag{5.9}$$

where the normalization accounts for symmetrizing the parts that are already partially symmetrized by the spin permutations.

In the following, the difference of $\tilde{\Phi}$ and Φ (or $\tilde{\Psi}$ and Ψ), that is, full or partial anti-symmetry wave functions, is omitted. Since every wave function in this work has definite spin and particle number, it is also sufficient to treat $\psi(\mathbf{R})$ instead of $\tilde{\Psi}(\mathbf{X})$. As an example of the consequences of the wave function's form, the Schrödinger Eq. (2.7) reduces to

$$\hat{\mathrm{H}}\psi(\mathbf{r}_1, \dots \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots \mathbf{r}_N). \tag{5.10}$$

The explicit spin dependence of the wave function is completely omitted, while the total spin S dictates the symmetries of ψ.

5.2. Cusp conditions

Whenever two (or more) particles from the Hamiltonian of Eq. (2.1) meet, their Coulomb potential becomes singular. In order to satisfy the Schrödinger Eq. (5.10), the kinetic energy has to cancel the singularity in the potential, which introduces distinct characteristics in the wave function ψ at particle-coalescence points [69]. The cancellation of the singularity is a strong enough requirement to impose conditions on the wave function that do not depend on the location of the other non-incident particles. These are analyzed in the following to construct trial wave functions that obey the properties of the exact wave function and thus improve the quality of the trial wave function as well as regularize its drift and local energy required in the QMC methods. A detailed derivation is presented in Sec. B. In this chapter, the only coincidences discussed are electron-electron and electron-nucleus coincidences. At higher orders, e.g., electron-electron-nucleus coincidences, the wave function also has a definite but more complicated behavior. In principle, every order of particle coalescence could be included in the trial wave function. However, due to electrons repelling each other, a higher-order coalescence with several electrons involved is very unlikely.

For the electron-nucleus coalescence with a nucleus of charge Z, mass M and a separation vector $\boldsymbol{\varepsilon}$ of nucleus and electron, the cusp condition reads

$$\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[1 - Z\mu|\boldsymbol{\varepsilon}| + \mathbf{w}\cdot\boldsymbol{\varepsilon}\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \tag{5.11}$$

where the effects of the reduced mass $\mu = (1+M^{-1})^{-1} \approx 1$ are omitted due to the large difference of electron to nuclear masses. The vector $\mathbf{w}$ is unknown in general. The singlet electron-electron coalescence with a reduced mass of $\mu = 1/2$ of two electrons of opposite spin reads

$$\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[1 + \frac{1}{2}|\boldsymbol{\varepsilon}| + \mathbf{w}\cdot\boldsymbol{\varepsilon}\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \tag{5.12}$$

where $\boldsymbol{\varepsilon}$ is the separation of the incident electrons. The triplet electron incident, however, naturally differs from the above form. Due to the anti-symmetry of the determinant, the wave function must vanish. The corresponding cusp condition reads

$$\psi(\boldsymbol{\varepsilon}) = \mathbf{w}\cdot\boldsymbol{\varepsilon}\left[1 + \frac{1}{4}|\boldsymbol{\varepsilon}|\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2). \tag{5.13}$$

These conditions are satisfied by the exact wave function, i.e. any solution to Eq. (5.10), but also offer constructive insight, from which improved trial wave functions can be constructed. As a consequence of the cusp condition, the drift and local energy of trial wave functions exhibit characteristic behavior, cf. Sec. B for a detailed derivation or Tab. 5.1 for an overview of the irregularities.

5.3. Trial wave functions

High-quality trial wave functions do not only improve the variational energy from VMC, in some cases recovering almost all correlation effects [50–52] but also allow for more accurate results from DMC due to more efficient numerics, lowered time-step and extrapolation-estimator errors, etc. Therefore, it is desirable to incorporate the above exact cusp conditions into the trial wave function. As will become apparent, exact cusp conditions are relatively straightforward to implement in QMC. Any method based on the linear combination of Slater determinants that genuinely do not show the electron-electron cusps would have to include an inefficiently large amount of determinants to resemble the cusp. Although a vast number of sophisticated trial wave functions exist, this work focuses on simple trial wave functions. The most straightforward trial wave function would be a single Slater determinant ϕ. Since a single determinant can be calculated much more efficiently without MC methods, this choice would be of little use. The advantage of using MC integration is the unrestricted form of the trial wave function. It allows the single Slater determinant to be combined with another function in an arbitrary way. A very successful modification to apply to the Slater determinant is the Jastrow factor. The Jastrow factor J alters the magnitude of the determinant ϕ by a simple multiplication.

$$\psi_T(\mathbf{R}) = \exp\left(J(\mathbf{R})\right)\phi(\mathbf{R}) \tag{5.14}$$

The Jastrow factor is non-negative on the entire configuration space due to the exponentiation, does not alter the nodal structure of the determinant, and must be symmetric upon spin-particle exchange since the determinant is anti-symmetric already. This form is very flexible and can account for multiple electron correlation effects. The above form of the wave function is called a single-reference Slater-Jastrow wave function. It is commonly used in QMC applications and is the only trial wave function form used in this thesis.

In contrast, multi-determinant Slater-Jastrow wave functions replace the single determinant with an expansion of determinants and allow for a more accurate description of the correlated wave function, especially in cases where a single determinant is not sufficient to describe the state, i.e. in systems dominated by static correlation. Another approach is a coordinate transformation T, called back-flow [54, 55], that alters the coordinates that enter the trial wave function as arguments, introducing additional optimizable parameters. Such multi-determinant Slater-Jastrow wave functions with back-flow take the powerful form

$$\psi_T(\mathbf{R}) = \exp\left(J(\mathbf{R})\right)\sum_k c_k\phi_k(T(\mathbf{R})) \tag{5.15}$$

and are able to describe many systems with high accuracy in VMC and DMC. Note that the Jastrow factor in a multi-determinant wave function significantly reduces the number of determinants required to describe the state. The linear combination of determinants, as well as the back-flow transformation, are able to modify the topology of the wave function's nodal structure, which is an important capability in the DMC method, see Sec. 6.9.

5.4. Jastrow factors

The Jastrow factor defined above aims to incorporate properties of the exact wave function into the trial wave function. There are many suggestions for elaborate forms [56, 70–72]. Jastrow factors can lower the variational energy in VMC and simultaneously allow parametrization of the trial wave function by parameterized Jastrow factors. A systematic way of categorizing different forms is to split the Jastrow factor by the order of particle coalescence

$$J(\mathbf{R}) = J_{en}(\mathbf{R}) + J_{ee}(\mathbf{R}) + J_{een}(\mathbf{R}) + J_{eee}(\mathbf{R}) + \ldots, \tag{5.16}$$

where most expansions do not go beyond electron-electron-nucleus terms. The electron-nucleus Jastrow factor takes the general form

$$J_{en}(\mathbf{R}) = \sum_{\eta=1}^{L}\sum_{i=1}^{N}\beta_\eta u_{en}(|\boldsymbol{\xi}_\eta - \mathbf{r}_i|), \tag{5.17}$$

which sums over all electron-nucleus pairs and thus is symmetric in the electrons. In this form, u_{en} is some function of the electron-nucleus distance and is restricted to have unit derivative at particle coalescence, $u'_{en}(0) = 1$. Upon comparison with the cusp condition (5.11), the parameter β_η is the Jastrow factor's contribution to the wave functions nuclear cusp. However, the trial wave function's total cusp also holds a contribution from its Slater determinant, see Sec. B.1. Slater determinants built from smooth orbitals, e.g., gaussian basis set orbitals, do not hold a nuclear cusp. In such a case, the Jastrow factor enforces the correct cusp alone. The DFT orbitals in the grid representation used in this work, however, already exhibit the correct nuclear cusp. The electron-nucleus Jastrow factor then requires a vanishing cusp and therefore vanishes.

The electron-electron Jastrow factor J_{ee} models two electrons approaching each other. J_{ee} splits in singlet and triplet terms including u^S_{ee} for a total spin of $S = 0$ for the incident electrons and u^T_{ee} for $S = \pm 1$,

$$\begin{aligned} J^S_{ee}(\mathbf{R}) &= \beta^S_{ee} \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} u^S_{ee}(|\mathbf{R}_{\alpha i} - \mathbf{R}_{\beta j}|) \\ J^T_{ee}(\mathbf{R}) &= \beta^T_{ee} \sum_{\sigma=\alpha,\beta} \sum_{i=1}^{N_\sigma} \sum_{j=i+1}^{N_\sigma} u^T_{ee}(|\mathbf{R}_{\sigma i} - \mathbf{R}_{\sigma j}|), \end{aligned} \tag{5.18}$$

where again with $u^{S,T\,\prime}_{ee}(0) = 1$ and by comparison with the cusp conditions of Eq. (5.13) and Eq. (5.12) $\beta^S_{ee} = 1/2$ and $\beta^T_{ee} = 1/4$.

The functions u_{en} and $u^{S,T}_{ee}$ can be modeled with a variety of representations, e.g., polynomials, Chebychev polynomials, B-Splines or Padé-forms, might include cut-off distances, and can be rewritten to use modified variables to respect certain coordinate transformations [56]. Representations that are linear in the parameters, like Chebychev polynomials and B-Splines, simplify the parameter optimization in VMC. However, the simplest Jastrow factor preserving the exact cusp conditions is used here. The Padé form

$$u[b](r) = \frac{r}{1 + br}, \tag{5.19}$$

with numerator and denominator degree of one, is the minimal Padé form that qualifies as a Jastrow factor. It has unit derivative at particle coalescence and a single parameter b. Despite or even because of its simplicity, it is used frequently in the literature [50, 51, 56, 73] and exhibits excellent results for its simplicity [72].

From the variational principle and the cusp condition, one expects a lower energy if the cusp condition is satisfied exactly and the variable parameters, b^S and b^T, are independent. If the spatial coordinates $\mathbf{R}_\alpha$ are discriminated from $\mathbf{R}_\beta$ in the Jastrow factor, the wave function is not an eigenfunction of the spin operator $\hat{\mathrm{S}}^2$, as discussed in detail in Ref. [74]. However, the introduced spin contamination is quite small. In contrast, if singlet and triplet cases are treated identically, i.e. with identical value for the cusp parameters, $\beta^S_{ee} = \beta^T_{ee}$, and the variable parameters, $b^S = b^T$, the Jastrow factor becomes symmetric. Then, the spin contamination vanishes at the cost of violating the cusp condition. The resulting variational energy is higher compared to the spin-contaminated case. According to Ref. [74], this dilemma can be resolved by restricting the representation of the Jastrow factor. In this work, the issue is remedied with an intermediate solution. The cusp condition is strictly enforced for singlet and triplet cusp separately, but identical parameters are used $b^T = b^S$. Therefore, the complete Jastrow factor used in the following is the electron-electron Jastrow factor in the Padé form, enforcing the correct singlet and triplet cusp conditions

$$J(\mathbf{R}) = J_{ee}[b](\mathbf{R}) = J^S_{ee}[b](\mathbf{R}) + J^T_{ee}[b](\mathbf{R}) \tag{5.20}$$

with $J^{S,T}_{ee}$ from Eq. (5.18) and $u^S_{ee} = u^T_{ee}$ from Eq. (5.19).

5.5. Drift and local energy

In addition to the value of the trial wave function, its drift and local energy have to be evaluated. In VMC, the drift is required when using a drift-diffusion Green function. With the single-reference Slater-Jastrow wave function given in Eq. (5.14), the gradient reads

$$\nabla\psi_T = e^J\left[\phi\nabla J + \nabla\phi\right], \tag{5.21}$$

which implies a drift of the form

$$\mathbf{u} = 2D\left[\nabla J + \phi^{-1}\nabla\phi\right] \tag{5.22}$$

Due to the exponential form of the Jastrow factor, its contributions to the gradient are given by the gradient of the Jastrow factor alone. For the determinantial part, the gradient must be divided by the determinant's value. This already hints at issues at the determinant's nodes, see Sec. B.2 for details. The Laplacian reads

$$\nabla^2\psi_T = e^J\left[2\nabla\phi\cdot\nabla J + \phi\nabla^2 J + \nabla^2\phi + \phi|\nabla J|^2\right]. \tag{5.23}$$

It mixes the contributions from the Jastrow factor and the determinant. The resulting kinetic energy is given by

$$K_L = -D\psi_T^{-1}\nabla^2\psi_T = -D\left[\nabla^2 J + |\nabla J|^2 + 2\phi^{-1}\nabla\phi\cdot\nabla J + \phi^{-1}\nabla^2\phi\right] \tag{5.24}$$

For a numerically efficient evaluation of the determinant's derivatives, see Sec. A.2. The evaluation of the Slater determinant, its gradient and its Laplacian, of course, requires the evaluation of the building orbitals. The orbitals used in this thesis are represented on a non-Euclidean but uniform grid. See Sec. E for a detailed description of the method. In the conversion of the orbitals' grid representation to a continuous representation compatible with the QMC approaches, numerical artifacts arise in the vicinity of the nuclei. Sec. E.4 explains how the artifacts arise and Sec. F analyzes their magnitude for different grids.

The derivatives of the electron-electron Jastrow factor (in compact notation) can be calculated analytically:

$$\begin{aligned}\nabla J &= \hat{\mathbf{e}}_k\nabla_k\sum_{i,j}\beta_{ij}u(r_{ij}) = \hat{\mathbf{e}}_k\sum_{i,j}\beta_{ij}u'(r_{ij})\nabla_k r_{ij}\\ &= \hat{\mathbf{e}}_k\sum_{i,j}\beta_{ij}u'(r_{ij})\left[\hat{\mathbf{r}}_{ij}\delta_{ki} - \hat{\mathbf{r}}_{ij}\delta_{kj}\right] = \hat{\mathbf{e}}_k\sum_{i,j}[\beta_{ik}+\beta_{ki}]u'(r_{ki})\hat{\mathbf{r}}_{ki}\end{aligned} \tag{5.25}$$

with $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $\hat{\mathbf{r}}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/r_{ij}$. The values for β_{ij} must be chosen to match the triplet or singlet case. The Laplacian is given by

$$\begin{aligned}\nabla^2 J &= \nabla_k\sum_{i,j}[\beta_{ik}+\beta_{ki}]u'(r_{ki})\hat{\mathbf{r}}_{ki} = \sum_{i,j}[\beta_{ik}+\beta_{ki}]\left[u''(r_{ki})\hat{\mathbf{r}}_{ki}\cdot\hat{\mathbf{r}}_{ki} + u'(r_{ki})\nabla_k\hat{\mathbf{r}}_{ki}\right]\\ &= \sum_{i,j}[\beta_{ik}+\beta_{ki}]\left[u''(r_{ki}) + \frac{2u'(r_{ki})}{r_{ki}}\right]\end{aligned} \tag{5.26}$$

5.6. Trial wave function characteristics

In the previous sections, some exact properties are incorporated into the trial wave function to more accurately approximate the exact wave function. Improvements to the trial wave function

Table 5.1.: Classification of the irregularities of the true wave function and a trial wave function satisfying the cusp conditions. The vector $\boldsymbol{\varepsilon}$ represents the distance to the critical point, the symbol $\mathbf{w}$ and $\mathbf{k}$ represents some unknown vectors, Z is the nuclear charge, a some unknown scalar, and E_0 the exact ground state energy.

	electron-nucleus	electron-electron singlet	electron-electron triplet	wave function node
ψ_T	$a(1 - Z\lvert\boldsymbol{\varepsilon}\rvert) + \boldsymbol{\varepsilon} \cdot \mathbf{w}$	$a(1 + \lvert\boldsymbol{\varepsilon}\rvert/2)$	$\boldsymbol{\varepsilon} \cdot \mathbf{w}$	$\boldsymbol{\varepsilon} \cdot \mathbf{w}$
ψ_0	$a(1 - Z\lvert\boldsymbol{\varepsilon}\rvert) + \boldsymbol{\varepsilon} \cdot \mathbf{w}$	$a(1 + \lvert\boldsymbol{\varepsilon}\rvert/2)$	$\boldsymbol{\varepsilon} \cdot \mathbf{w}$	$\boldsymbol{\varepsilon} \cdot \mathbf{w}$
$\mathbf{u}[\psi_T]/2D$	$-Z\hat{\boldsymbol{\varepsilon}}$	$\frac{1}{2}\hat{\boldsymbol{\varepsilon}}$	$\frac{1}{4}\hat{\boldsymbol{\varepsilon}} + \frac{\mathbf{w}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$	$\frac{\mathbf{w}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$
$\mathbf{u}[\psi_0]/2D$	$-Z\hat{\boldsymbol{\varepsilon}}$	$\frac{1}{2}\hat{\boldsymbol{\varepsilon}}$	$\frac{1}{4}\hat{\boldsymbol{\varepsilon}} + \frac{\mathbf{w}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$	$\frac{\mathbf{w}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$
$E_L[\psi_T]$	$\mathbf{w} \cdot \hat{\boldsymbol{\varepsilon}}$	$\mathbf{w} \cdot \hat{\boldsymbol{\varepsilon}}$	$\frac{\mathbf{k} \cdot \boldsymbol{\varepsilon}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$	$\frac{\mathbf{k} \cdot \boldsymbol{\varepsilon}}{\mathbf{w} \cdot \boldsymbol{\varepsilon}} + \frac{a}{\mathbf{w} \cdot \boldsymbol{\varepsilon}}$
$E_L[\psi_0]$	E_0	E_0	E_0	E_0

affect the QMC methods in various ways. For the reasons discussed above, the cusp conditions are the most prominent example of how an improved trial wave function affects the QMC method.

In Chapt. B, the behavior of the exact ground state wave function and a trial wave function that satisfies the cusp conditions is examined in detail. Tab. 5.1 provides an overview of the kind of irregularity the wave function value, drift, or local energy shows in the cases of nuclear and electron-electron cusps and nodes. Fixing the cusp conditions with proper Jastrow factors renders the local energy more regular but does not entirely remove irregular behavior. The local energy exhibits discontinuities in any case. For triplet electron incidences (and also at nodes), the local energy still diverges, however, not due to the Coulomb singularity, which is remedied by the cusp condition, but due to the node (induced by anti-symmetry in the triplet case). The drift vector of a trial wave function with correct cusps points away from electron-electron incidences as well as wave function nodes, whereas it points toward the nucleus for an electron-nucleus coalescence. This behavior is much appreciated since thus the walker is stochastically repelled from regions of low probability and, at the same time, notorious behavior of the local energy.

The behavior of the trial wave function of Eq. (5.14), its drift, and local energy can be visualized near critical points. While the full configuration space cannot be plotted conveniently, slices or hypersurfaces through the configuration space allow insights into the wave function's behavior. To this end, a configuration $\mathbf{R}$ is constructed, such that its first electron moves in the plain given by $\mathbf{r}_1(\lambda, \mu) = \boldsymbol{\eta} + \lambda\delta\mathbf{u} + \mu\delta\mathbf{v}$, where the vectors $\delta\mathbf{u}$ and $\delta\mathbf{v}$ are random vectors with $\delta\mathbf{v} \cdot \delta\mathbf{u} = 0$ and $|\delta\mathbf{v}| = |\delta\mathbf{u}|$. The coordinate $\boldsymbol{\eta}$ represents the incident point. By sampling the parameters $(\lambda, \mu) \in [-1, 1]^2$ the vicinity of the incident point can be depicted for any observable. The particle coincidence is located at $(\lambda, \mu) = (0, 0)$. The other electrons $2, \ldots N$ have randomly chosen but fixed coordinates, such that every electron pair is well separated. The trail wave function used for figures 5.1, 5.2, 5.3, and 5.4 is of the form of Eq. (5.14), was built from DFT orbitals from the LDA, and represents the nitrogen molecule.

In Fig. 5.1 the nuclear cusp is plotted. The local energy exhibits a discontinuous jump and numerical artifacts confined to a small region around the nucleus, cf. Sec. E.4. The singlet electron incident is plotted in Fig. 5.2. While the trail wave function value is non-zero, it shows a cusp, which leads to a discontinuity in the local energy. The triplet electron incident is plotted in Fig. 5.3. Here, the wave function shows a node in the form of a straight line, i.e. is in accordance with $\boldsymbol{\varepsilon} \cdot \mathbf{w}$. Since the local energy is divided by the value of the wave function, a singularity arises in the local energy along the node. Additionally, a point discontinuity is present at the incident point. Finally, the behavior of the local energy along a node is plotted in Fig. 5.4. Along the nodal line are two kinds of singularities. The kind that is present along the entire nodal line is $a/\mathbf{w}\cdot\boldsymbol{\varepsilon}$, where $\mathbf{w}$ indicates the direction of the nodal line. The other kind of singularity, $\mathbf{k}\cdot\boldsymbol{\varepsilon}/\mathbf{w}\cdot\boldsymbol{\varepsilon}$, only occures at special points of the nodal line.

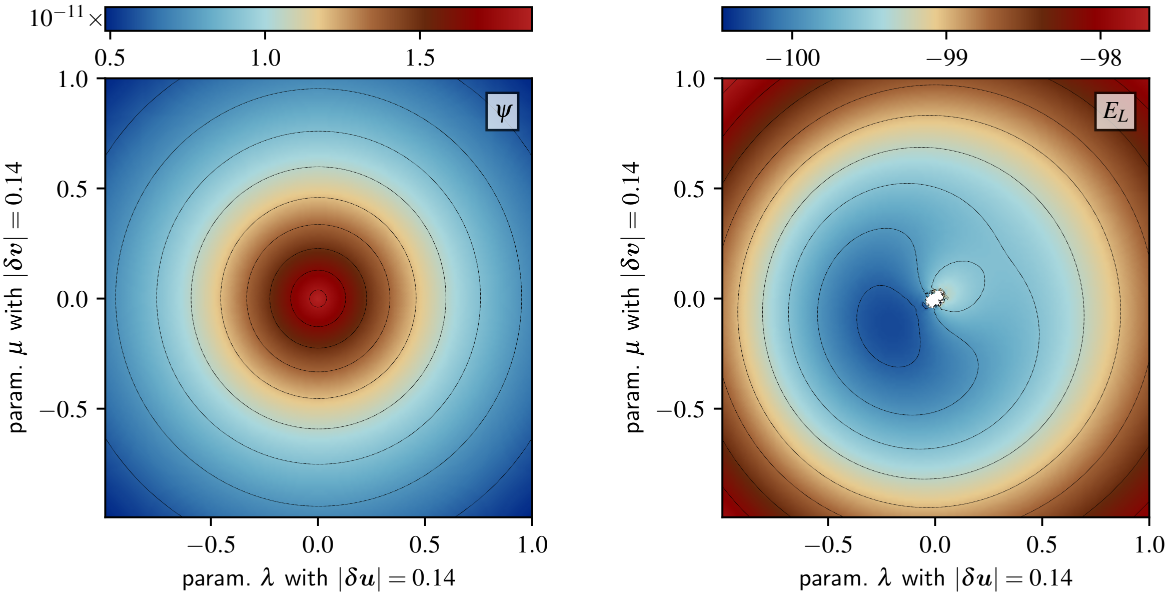


Figure 5.1.: The value of the trial wave function (left panel) and its local energy (right panel) on a two-dimensional plain through the nucleus are shown. The wave function value has the form of a cusp. The dipole-type discontinuity $\mathbf{w}\cdot\hat{\varepsilon}$ of the local energy is visible from the iso-lines and the coloring. In addition to the discontinuity, numerical artifacts at the incident point are visible. The white area indicates numerical values of the local energy that are diverging and exceed the range of the color scale. These artifacts are discussed in Sec. E.4. The incident point $\boldsymbol{\eta}$ is the nuclear coordinate.

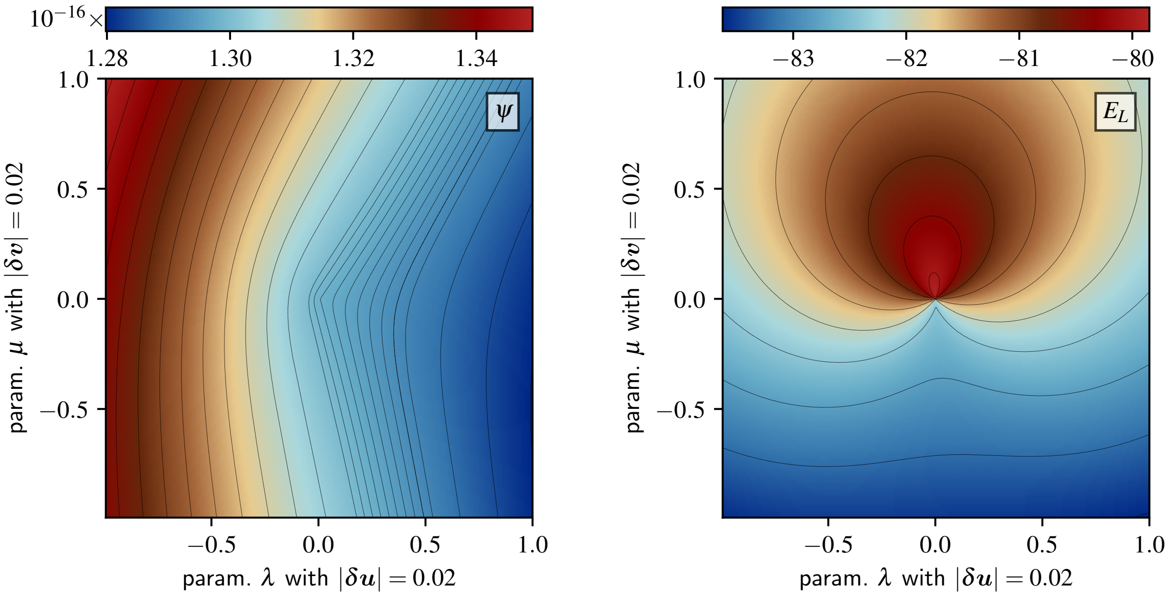


Figure 5.2.: The value of the trial wave function (left panel) and its local energy (right panel) on a two-dimensional plane shows the singlet electron-electron incident. The value of the wave function has a slightly visible cusp, highlighted by the iso-lines. Although no node is present in the singlet incident, the local energy develops a dipole-type discontinuity, $\mathbf{w} \cdot \hat{\boldsymbol{\varepsilon}}$. The incident point is given by the last electron coordinate $\boldsymbol{\eta} = \mathbf{r}_2$. Additional iso-lines are added to highlight the cusp.

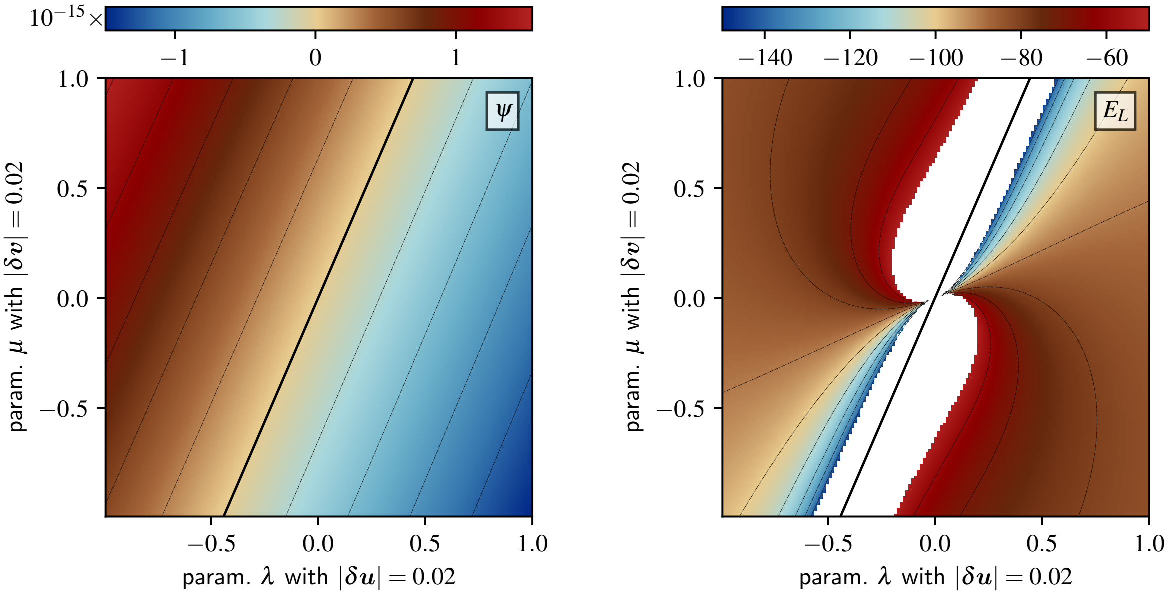


Figure 5.3.: The value of the trial wave function (left panel) and its local energy (right panel) on a two-dimensional plane shows the triplet electron-electron incident. The large black line indicates the wave function's nodes. The cusp in the wave function is not visible. In the local energy, however, the severe singularity of type $\mathbf{k} \cdot \boldsymbol{\varepsilon}/\mathbf{w} \cdot \boldsymbol{\varepsilon}$ is a distinct characteristic. The straight iso-line is the direction of the vector $\mathbf{k}$, while the wave function's node is aligned with the vector $\mathbf{w}$. The white area indicates numerical values of the local energy that are diverging and exceed the range of the color scale. The incident point is given by the last electron coordinate $\boldsymbol{\eta} = \mathbf{r}_N$.

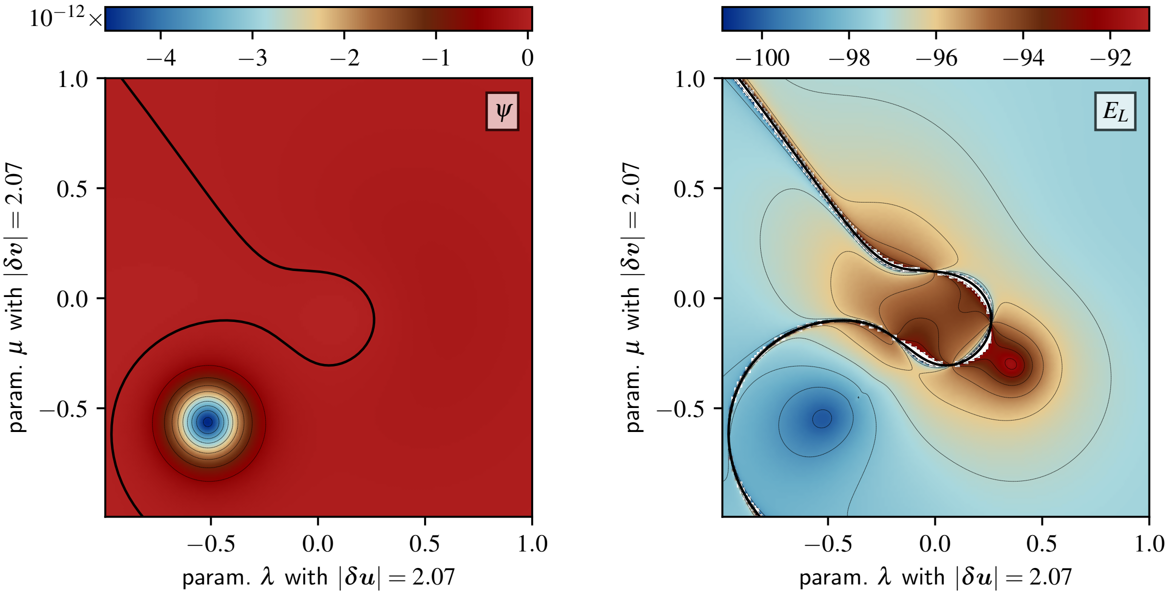


Figure 5.4.: The value of the trial wave function (left panel) and its local energy (right panel) on a two-dimensional plain showing the trial wave function's nodal structure sliced by the plain. The large black line indicates the wave function's nodes. Two types of singularities are present near the nodes. Along the node a singularity of the form $a/\mathbf{w} \cdot \boldsymbol{\varepsilon}$ is always present. The vector $\mathbf{w}$ represents the tangent along the node at every point. The second kind of singularity, $\mathbf{k} \cdot \boldsymbol{\varepsilon}/\mathbf{w} \cdot \boldsymbol{\varepsilon}$, is also present in the triplet electron incidences and does only appear at distinct locations along the node. It is characterized by the sign change of the local energy. At these distinct locations, the local energy does not diverge along a single path across the node. The direction of this path is given by the vector $\mathbf{k}$ and the path itself is given by the iso-lines crossing the node. There are in total four such points in the figure, all located close to the central nodal arc.
To visualize the nodes, the incident point $\boldsymbol{\eta}$ was chosen randomly. Then, in almost all cases, nodes appear naturally. The well-separatedness of the electrons ensures that the node is not due to anti-symmetry.

6. Diffusion Monte Carlo

Optimizing a reasonable parameterized wave function within VMC generates a good approximation of the true ground state. In general, however, no trial wave function with a finite set of parameters can reproduce the true ground state, leaving a gap between observables calculated in VMC and the true ground state results. Although more elaborate trial wave functions can significantly reduce this gap and reach errors smaller than chemical accuracy, the quality of the VMC ground state is always bound by the quality of the parametrization. Since many aspects of correlated wave functions might not yet be fully understood, it is tedious to adapt the parametrization and generate ever more sophisticated trial wave functions.

A method that reaches beyond the limits of VMC and nevertheless benefits from sophisticated trial wave functions and VMC methods is diffusion Monte Carlo (DMC). This method iterates an initial guess for the wave function to an approximate ground state that improves upon the trial wave function of VMC. While the DMC approach has certain limits, it is exact in principle and provides results that are among the most accurate at a moderate computational cost. In DMC, the MC approach is used to solve the Schrödinger equation by a diffusive process or more precisely random walks. In opposition to VMC, the DMC method is not a mere minimization or fit of parameters, as it is based on the Schrödinger equation directly instead of the variational principle. Therefore, DMC extends over VMC by many means.

This chapter establishes a connection between the Schrödinger equation and a classical diffusive process that can efficiently be carried out numerically.

6.1. Imaginary-time transformation

There is a long known and instructive link [75–77] from the trivially time-dependent Schrödinger equation with $\partial_t \hat{\mathrm{H}} = 0$,

$$i\partial_t \psi(\mathbf{R}, t) = \hat{\mathrm{H}}\psi(\mathbf{R}, t), \tag{6.1}$$

to a propagation scheme for the stationary Schrödinger equation: the imaginary-time propagation. It maps the eigenvalue problem of Eq. (2.7) to a PDE of time and space that yields the ground state. To begin with, the physical time t in the Eq. (6.1) is substituted with an imaginary time $\tau = it$ and the energy scale is shifted by an offset energy E_N.

$$-\partial_\tau \psi(\mathbf{R}, \tau) = \left(\hat{\mathrm{H}} - E_N\right) \psi(\mathbf{R}, \tau) \tag{6.2}$$

This yields a real-valued PDE with the formal solution

$$\psi(\mathbf{R}, \tau) = \hat{\mathrm{G}}(\tau)\psi(\mathbf{R}, 0), \tag{6.3}$$

where $\psi(\mathbf{R}, 0)$ is the initial state, and the imaginary-time propagator is given by

$$\hat{\mathrm{G}}(\tau) = \exp\left(-\tau\left(\hat{\mathrm{H}} - E_N\right)\right). \tag{6.4}$$

The action of $\hat{\mathrm{G}}(\tau)$ on the initial state can be seen by the expansion in the basis of the Hamiltonian's eigenfunctions, $\hat{\mathrm{H}}\psi_k = E_k\psi_k$. Assume the ground state has a non-vanishing contribution in the expansion, $\langle\psi(\cdot,0)|\psi_0\rangle \neq 0$. Then the action of $\hat{\mathrm{G}}(\tau)$ can be written as a time evolution of the coefficients

$$\begin{aligned}\psi(\mathbf{R},\tau) &= \hat{\mathrm{G}}(\tau)\psi(\mathbf{R},0) = \hat{\mathrm{G}}(\tau)\sum_k c_k(0)\psi_k \\ &= \sum_k e^{-\tau(E_k-E_N)}c_k(0)\psi_k = \sum_k c_k(\tau)\psi_k,\end{aligned} \tag{6.5}$$

where the last equality defines the time dependence of the coefficients. The propagator amplifies coefficients of low energy, $E_k < E_N$ and magnifies the ground state coefficient most. As a result, the ground state dominates the expansion for large τ. In the limit of infinite imaginary time, only the ground state survives the action of the propagator, and in the case of degeneracy, a mixture of ground states. Note that the propagator $\hat{\mathrm{G}}(\tau)$ is not norm-conserving, and the propagated wave functions are not normalized. For the steady-state solution, this implies

$$\begin{aligned}\lim_{\tau\to\infty}|\psi(\mathbf{R},\tau)|^2 &= \lim_{\tau\to\infty}|c_0(\tau)|^2 = \lim_{\tau\to\infty} e^{-\tau(E_0-E_N)}c_0(0) \\ &= \begin{cases} 0, & E_0 > E_N \\ |c_0(0)|^2, & E_0 = E_N \\ \infty, & E_0 < E_N \end{cases}\end{aligned} \tag{6.6}$$

From the norm of $\psi(\mathbf{R},\tau)$, the role of the offset energy E_N becomes apparent. It serves as a normalization constant for the ground state and therefore is referred to as the normalization energy. It is utilized to control the growth or decay of the propagated wave function's norm in Sec. 6.6. Solving Eq. (6.3) for increasing τ, will propagate the initial wave function $\psi(\cdot,0)$ to its steady state ψ_0. To solve Eq. (6.3), however, the exponential of the Hamiltonian has to be approximated. In DMC, this approximation is the expansion of the exponential for short times [53, 78]. The Baker-Campbell-Hausdorff [79] formula can be applied to the Hamiltonian to arrive at the first-order approximation

$$\hat{\mathrm{G}}(\tau) = e^{-(U+V-E_N)}e^{-\tau\hat{\mathrm{T}}} + \mathcal{O}(\tau^2) \tag{6.7}$$

or the more accurate second-order approximation

$$\hat{\mathrm{G}}(\tau) = e^{-\frac{\tau}{2}(U+V-E_N)}e^{-\tau\hat{\mathrm{T}}}e^{-\frac{\tau}{2}(U+V-E_N)} + \mathcal{O}(\tau^3), \tag{6.8}$$

while the latter formula is used in DMC in practice. Since the expansion is more accurate for small arguments, the time step τ should be chosen small enough. To nevertheless reach the long-time limit, the Trotter-Suzuki formula [80, 81] is applied to the approximate exponential with a small time step $\Delta\tau$,

$$\hat{\mathrm{G}}(n\Delta\tau) = \hat{\mathrm{G}}(\Delta\tau)^n. \tag{6.9}$$

Note that many alternative operators to $\hat{\mathrm{G}}(\tau)$ also reveal the ground state. Such operators are required to magnify the ground state coefficient more than any coefficient of states of higher energy. Therefore, the approximations made here are not as severe and crucial as they might appear as long as the propagator exhibits this property.

In the steady state, the propagator formally will not alter $\psi(\tau)$ from one step to another, but it will become apparent in Sec. 6.4, that the numerical representations of $\psi(\tau)$ and $\psi(\tau+\Delta\tau)$ differ. The solution of the propagator Eq. (6.3) in the long-time limit, i.e. the ground state, is, of course, also the solution of the imaginary-time PDE of Eq. (6.2). The molecular Hamiltonian of Eq. (2.1) expressed in terms of the kinetic energy, which includes the Laplacian, and local potentials renders the explicit imaginary-time PDE to be

$$-\partial_\tau\psi(\mathbf{R},\tau) = -D\nabla^2_{\mathbf{R}}\psi(\mathbf{R},\tau) + (V(\mathbf{R}) + U(\mathbf{R}) - E_N)\,\psi(\mathbf{R},\tau), \tag{6.10}$$

where the constant $D = {}^{1}\!/_{2}$ is introduced to represent the prefactor of the kinetic energy and the electron mass, in SI units $D = {}^{\hbar^2}\!/_{2m_e}$. This PDE manifestly is a diffusion equation. The potential contributions, $V + U$, act as a spatially varying sink-source term that creates or annihilates total probability. This form of the stationary Schrödinger equation is the core of DMC since the connection to the diffusion equation offers the opportunity to recast the Schrödinger equation as a diffusion process or random walk. Although solving Eq. (6.10) for large τ would yield the ground state, the PDE is still an equation of $3N$ coordinates, infeasible to solve with standard PDE methods. In the next sections, Eq. (6.10) is therefore reformulated to fit in the framework of MC methods, which enable the actual calculation of the ground state or at least reasonable approximations.

6.2. The Green function

On the way to a framework compatible with MC, the PDE (6.10) has to be reformulated as an integral equation. To this end, the Green function of the Eq. (6.10) is readily given by a kernel representation for the propagator $\hat{\mathrm{G}}(\tau)$,

$$\psi(\mathbf{R}', \tau) = \int G(\mathbf{R}' \leftarrow \mathbf{R}, \tau)\psi(\mathbf{R}, 0)\,\mathrm{d}\mathbf{R}, \tag{6.11}$$

or equivalently

$$-\partial_\tau G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) = \hat{\mathrm{H}}(\mathbf{R}')G(\mathbf{R}' \leftarrow \mathbf{R}, \tau), \tag{6.12}$$

subject to the initial condition $G(\mathbf{R}' \leftarrow \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$. The exact Green function incorporates the exact time evolution of an arbitrary initial wave function $\psi(\cdot, 0)$ to the ground state, comprising an extensive amount of information in a rather direct form. However, no true Green function is known for non-trivial systems, nor is the integration of the Green function in the $3N$ dimensional configuration space by any means simple. The true Green function is a theoretical tool generally unknown for most applications. For small times τ, however, there is a simple approximation to the true Green function reading

$$\begin{aligned} G(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau) = (4\pi D\Delta\tau)^{-3N/2} \exp\left(-\frac{|\mathbf{R}' - \mathbf{R}|^2}{4D\Delta\tau}\right) \\ \times \exp\left(-\frac{\Delta\tau}{2}\left(V(\mathbf{R}') + U(\mathbf{R}') + V(\mathbf{R}) + U(\mathbf{R}) - 2E_N\right)\right), \end{aligned} \tag{6.13}$$

see Sec. D.4 or Ref. [78, 82]. This form resembles the second-order expansion of $\hat{\mathrm{G}}$ of Eq. (6.8). While this form might not reveal its usefulness yes, it gives rise to approximations and reinterpretations of the PDE, which are covered in Sec. 6.4.

The first goal is now achieved, namely an integral representation of the wave function in the long-time limit, and even a limiting case of the true Green function is available. However, a severe caveat of the above Green function lies in the Coulomb singularities of the Hamiltonian whenever two particles meet, leading to a singular contribution and, ultimately, reducing the convenience of this approach substantially. In the next section this problem is accounted for explicitly.

6.3. Importance sampling

One of the most commonly used techniques in the repertoire of MC theory is importance sampling. It allows including prior knowledge about the solution in the problem. In the case of the diffusion

Eq. (6.10), it allows remedying the singularities of the Hamiltonian to yield a Green function without a singular exponent, which improves the quality of the results significantly. To this end, Eq. (6.10) is multiplied with a time-independent wave function ψ_T. The role of ψ_T will become apparent later. For now, ψ_T is some ansatz or candidate solution to the Schrödinger equation. After some algebraic manipulation, a PDE for the quantity for $f = \psi_T\psi$ follows. The contributions from the multiplicative potential and the time-derivative are readily recast to act on f, while the Laplacian transforms as

$$\begin{aligned}\psi_T\nabla^2\psi &= \nabla^2(\psi_T\psi) - \psi\nabla^2\psi_T - 2\nabla\psi_T\cdot\nabla\psi = \\ &= \nabla^2(\psi_T\psi) + \psi\nabla^2\psi_T - 2\nabla\cdot(\psi\nabla\psi_T) = \\ &= \nabla^2(\psi_T\psi) + (\psi_T\psi)\frac{\nabla^2\psi_T}{\psi_T} - 2\nabla\cdot\left((\psi_T\psi)\psi_T^{-1}\nabla\psi_T\right) \\ &= \nabla^2 f + f\psi_T^{-1}\nabla^2\psi_T - 2\nabla\cdot\left(f\psi_T^{-1}\nabla\psi_T\right)\end{aligned} \tag{6.14}$$

The appearing gradient $\nabla\psi_T$ is included in the definition of the drift vector, already known from the Green functions in VMC in Sec. 4.3 in Eq. (6.15). The drift vector reads

$$\mathbf{u}(\mathbf{R}) = 2D\psi_T^{-1}(\mathbf{R})\nabla\psi_T(\mathbf{R}). \tag{6.15}$$

Inserting the importance-sampled Laplacian into the importance-sampled version of Eq. (6.10) yields

$$\begin{aligned}-\partial_\tau f &= -D\nabla^2 f + f\frac{\nabla^2\psi_T}{\psi_T} + \nabla\cdot\mathbf{u}f + (V + U - E_N)\, f = \\ &= -D\nabla^2 f + f{\psi_T}^{-1}\left(-D\nabla^2 + V + U\right)\psi_T + \nabla\cdot\mathbf{u}f - E_N f\end{aligned} \tag{6.16}$$

and again a known quantity, ${\psi_T}^{-1}\left(-D\nabla^2 + V + U\right)\psi_T$ can be identified in the equation. It is the local energy $E_L = {\psi_T}^{-1}\hat{\mathrm{H}}\psi_T$ from Eq. (4.6). Finally, the PDE takes the form of a the sink-source drift-diffusion equation

$$-\partial_\tau f = -D\nabla^2 f + \nabla\cdot(\mathbf{u}f) - Sf \tag{6.17}$$

with a sink-source term

$$S(\mathbf{R}) = E_N - E_L(\mathbf{R}). \tag{6.18}$$

Note that the derivatives acting on f are the time derivative and the Laplacian, and also the second term on the right-hand side of Eq. (6.17), including the gradient, that acts on the product of f and $\mathbf{u}$. The diffusion Eq. (6.17) is almost of the same form as Eq. (6.10). The major differences are the source term and the additional drift. In contrast to the non-importance sampled equation (6.10) with source term $V + U$, the importance sampled source term S above does not necessarily diverge at particle coincidences and can be a much more regular quantity. Chapt. 5 is dedicated to constructing wave functions ψ_T, that render $\mathbf{u}$ and S more regular functions. The above sink-source drift-diffusion equation describes an non-normalized distribution or density profile subject to diffusion with diffusion constant D and drifting with a locally varying drift vector $\mathbf{u}$ while gaining or losing total mass or normalization due to a local sink-source term S.

The stationary Schrödinger equation is now cast in a form, that can be solved using MC methods. Solving this PDE is equivalent to solving the Schrödinger equation within some limitations further discussed in Sec. 6.9. As for the non-importance sampled diffusion equation, there exists an integral kernel $\tilde{G}$,

$$-\partial_\tau\tilde{G}(\mathbf{R}'\leftarrow\mathbf{R},\tau) = \left[-D\nabla'^2 + \nabla'\cdot\mathbf{u}(\mathbf{R}') - S(\mathbf{R}')\right]\tilde{G}(\mathbf{R}'\leftarrow\mathbf{R},\tau) \tag{6.19}$$

that has the same initial condition as G. The Green function $\tilde{G}$ propagates the importance sampled probability f

$$f(\mathbf{R}', \tau + \Delta\tau) = \int \tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau) f(\mathbf{R}, \tau)\, \mathrm{d}\mathbf{R}. \tag{6.20}$$

The kernel $\tilde{G}$ differs from the kernel G but can be retrieved by insertion of unity.

$$\begin{aligned} f(\mathbf{R}', \tau) &= \psi_T(\mathbf{R}')\psi(\mathbf{R}', \tau) \\ &= \psi_T(\mathbf{R}') \int G(\mathbf{R}' \leftarrow \mathbf{R}, \tau)\psi(\mathbf{R}, 0)\, \mathrm{d}\mathbf{R} \\ &= \int \psi_T(\mathbf{R}') G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \psi_T(\mathbf{R})^{-1} f(\mathbf{R}, 0)\, \mathrm{d}\mathbf{R}, \end{aligned} \tag{6.21}$$

from which $\tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}, \tau) = \psi_T(\mathbf{R}') G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \psi_T(\mathbf{R})^{-1}$ follows. Since the propagator $\hat{\mathrm{G}}(\tau)$ is not norm-conserving, the Green function can be split into a norm-conversing part $\tilde{G}_D$, that describes drift and diffusion, and a branching part $\tilde{G}_B$ that adjusts the norm of f,

$$\tilde{G} = \tilde{G}_B \tilde{G}_D. \tag{6.22}$$

The reason for the naming of the branching Green function will become apparent in the following.

The short-time approximation to $\tilde{G}_D$ is derived in the appendix in Sec. D.1 for the case of constant drift and source term and is also available from Refs. [53, 83]. In contrast to the non-importance sampled Green function G, the importance sampled Green function $\tilde{G}$ is given by the drift-diffusion part

$$\tilde{G}_D(\mathbf{R}', \mathbf{R}, \Delta\tau) = (4\pi D \Delta\tau)^{-3N/2} \exp\left(- \frac{\left|\mathbf{R}' - \mathbf{R} - \Delta\tau \mathbf{u}(\mathbf{R})\right|^2}{4 D \Delta\tau} \right) \tag{6.23}$$

and its sink-source contribution

$$\tilde{G}_B(\mathbf{R}', \mathbf{R}, \Delta\tau) = \exp\left(\frac{\Delta\tau}{2} \left(S(\mathbf{R}') + S(\mathbf{R}) \right) \right). \tag{6.24}$$

The Green function $\tilde{G}$ does explicitly include a drift term, providing a deterministic directedness in the transition probability, that modulates the Green function's probability distribution to have a higher probability in the regions where f has a higher probability. This is a major improvement of the Green function on its own. Another great improvement of the importance sampling is the new form of the branching Green function $\tilde{G}_B$. Its source term $S(\mathbf{R}) = E_N - E_L(\mathbf{R})$ can be much more regular compared to the source term of Eq. (6.13) with bare singularities. If exact conditions for particle coincidences are incorporated into the trial wave function, as discussed in Sec. 5.2 and in detail in the appendix in Sec. B, the local energy and the sink-source term show significantly less severe irregularities.

It is no coincidence the importance sampling function shares the symbol ψ_T with the parameterized trial wave function of VMC. While in general, both functions, i.e. the trial wave function of Chapt. 4 and the importance sampling function of this chapter, need not be identical, in the majority of cases, the VMC trial wave function is used as the importance sampling function, since it usually is the best approximation to the wave function available in DMC. In the limiting case of a perfect trial wave function $\psi_T = \psi_0$ with perfect normalization energy $E_N = E_0$, the source term is given by $S = E_N - E_L = 0$ and $\tilde{G}_B = 1$, and Eq. (6.17) describes the probability of the exact ground state.

In the following sections, it will become clear what specific properties the trial wave function has to obey in order to yield the true ground state in the long-time limit.

6.4. The diffusive process

Chapt. D gives a detailed overview of the reinterpretation of the diffusion equation and the resulting numerical scheme to solve it, including drift and sink-source terms. Here, the key elements are emphasized within their context. With an integral representation of the ground state wave function and even a well-behaved Green function from the last section, the majority of the formal derivation of the DMC scheme is complete. Next is the conceptual reinterpretation of the integral equation. The notable difference between the ground state eigenvalue problem and the diffusion equation is that the diffusion equation admits a probabilistic interpretation and, with importance sampling, even a practical MC algorithm. To this end, assume the numerical representation of the solution f is of the form

$$f(\mathbf{R},\tau) = \sum_{k=1}^{M} \lambda_k(\tau)\delta(\mathbf{R} - \mathbf{R}_k(\tau)) \tag{6.25}$$

with positive numbers $\lambda_k(\tau)$, called weights, and norm $\Lambda(\tau) = \sum_k \lambda_k(\tau)$. Note that f is not normalized since $\hat{\mathrm{G}}(\tau)$ is not norm-conserving, and therefore $\Lambda(\tau)$ has an intrinsic time dependence. The true solution f is certainly not composed of (a finite number of) Dirac distributions. Nevertheless, this representation can be understood as an application of the law of large numbers: Instead of the whole analytical function f, defined on the entire domain Ω^N, the distribution is replaced by a large but representative set of realizations, $(\lambda_k(\tau), \mathbf{R}_k(\tau))$, where $k = 1, \ldots M$ enumerates the parallelly propagated weights and configurations and τ acts as a serial index for the Markov chain. The representation of Eq. (6.25) allows to store the distribution of f only by its weights and configurations. To propagate this representation to the ground state, any analytically known Green function can propagate $f(\cdot,\tau)$ by Eq. (6.20) resulting in an analytical expression for $f(\cdot,\tau+\Delta\tau)$ due to the Dirac distribution.

$$f(\mathbf{R}',\tau+\Delta\tau) = \sum_{k=1}^{M} \lambda_k(\tau)\tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}_k(\tau), \Delta\tau) \tag{6.26}$$

It is not possible in general to further propagate the above expression for $f(\cdot,\tau+\Delta\tau)$ by inserting it into Eq. (6.20) again, since the integral over two Green functions,

$$f(\mathbf{R}'',\tau+2\Delta\tau) = \sum_{k=1}^{M} \lambda_k(\tau) \int \tilde{G}(\mathbf{R}'' \leftarrow \mathbf{R}', \Delta\tau)\tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}(\tau), \Delta\tau)\,\mathrm{d}\mathbf{R}', \tag{6.27}$$

cannot be calculated in almost all cases. Therefore, the complexity of the above expression (6.26) has to be reduced. The central idea that makes DMC practical is the collapse of the distribution $f(\cdot,\tau+\Delta\tau)$ by drawing random variates from it. That is, for every configuration k in $f(\cdot,\tau+\Delta\tau)$, a new configuration $\mathbf{R}_k(\tau+\Delta\tau)$ is drawn from the distribution $\tilde{G}(\cdot \leftarrow \mathbf{R}_k(\tau), \Delta\tau)$ and subsequently the distribution $\tilde{G}(\cdot \leftarrow \mathbf{R}_k(\tau), \Delta\tau)$ is replaced by a Dirac distribution $\delta(\cdot - \mathbf{R}_k(\tau+\Delta\tau))$ centered at the new configuration $\mathbf{R}_k(\tau+\Delta\tau)$. The replacement of the full distribution $\tilde{G}(\cdot \leftarrow \mathbf{R}_k(\tau), \Delta\tau)$ by a single Dirac peak is a harsh approximation, but justified due to the conserved expectation value.

$$\mathrm{Exp}\left[\mathbf{R}_k(\tau+\Delta\tau)\right] = \int \mathbf{R}'\tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}_k(\tau), \Delta\tau)\,\mathrm{d}\mathbf{R}' \tag{6.28}$$

Without conservation of the norm, $\tilde{G}$ is not a probability distribution and a configuration cannot be drawn from it directly. In practice, the new configuration is drawn from the norm-conserving drift-diffusion kernel $\tilde{G}_D$ only.

$$\mathbf{R}_k(\tau+\Delta\tau) \sim \tilde{G}_D(\cdot \leftarrow \mathbf{R}_k(\tau), \Delta\tau) \tag{6.29}$$

With the new configuration, the branching Green function $\tilde{G}_B$ is evaluated to update the weight

$$\lambda_k(\tau + \Delta\tau) = \lambda_k(\tau)\tilde{G}_B(\mathbf{R}_k(\tau + \Delta\tau) \leftarrow \mathbf{R}_k(\tau)). \tag{6.30}$$

Thus, the probability distribution of the next time step becomes

$$f(\mathbf{R}', \tau + \Delta\tau) = \sum_{k=1}^{M} \lambda_k(\tau + \Delta\tau)\delta(\mathbf{R}' - \mathbf{R}_k(\tau + \Delta\tau)). \tag{6.31}$$

The propagated distribution is of the same form as Eq. (6.25), and the propagation can be repeated iteratively. For every time step, a new distribution of the form of Eq. (6.25) is generated with differing weights and configurations. This allows for propagation f for an arbitrary amount of time. The numerical representation of f is essentially solving the PDE (6.17) of f in the long-time limit, revealing the ground state in a stochastic sense. Although this approximation is one of the most central steps in DMC, it also poses great consequences. While the expectation value is conserved, the information about the variance and any higher moment is eliminated and the resulting method requires a large amount of configurations to reduce statistical fluctuations.

Note that updating the configuration and weight as outlined above is an approximation. The approximation's character is explained in Sec. D.5 in detail. Even though approximate the corresponding error contributes to the time-step error, vanishing in the limit $\Delta\tau \to 0$.

The pair of weight and configuration, $(\lambda_k(\tau), \mathbf{R}_k(\tau))$, is subject to an iterative update process with combined deterministic and stochastic character. It thus performs a random walk in the configuration space Ω^N with some notion of size due to the weight $\lambda_k(\tau)$. It is therefore often called a walker in the literature. At this point, the intended computational approach becomes transparent. Instead of solving the imaginary-time PDE, for example, on a grid, the problem is rewritten such that an ensemble of walkers explores the configuration space and approximates the wave function by a set of Dirac distributions. This maps a high-dimensional PDE problem to the problem of solving classical trajectories, a conceptual and computational improvement that cannot be underestimated.

6.5. Approximate Green functions and the acceptance probability

At this point, it is worthwhile to recall that $\tilde{G}$ from the above sections is the true Green function of the PDE (6.17) and therefore not only unknown in general but also a highly complicated function. In applications, the Green function has to be approximated, with the short-time approximation being the starting point. As the true Green function follows directly from the Hamiltonian, an approximated Green function might correspond to a different Hamiltonian. Consequently, an approximation to the Green function results in a different stationary state f and thus a different ground state. Even worse, if the Green function does not satisfy detailed balance, the existence of a stationary state is not guaranteed. This poses a severe mismatch in the expectation values for any observable. Given that any Green function reduces to a Dirac distribution in the limit of vanishing time, the error in f reduces with smaller time steps $\Delta\tau$. At the same time, small steps increase serial correlation and limit exploration of the configuration space. This error can significantly be reduced by an acceptance-rejection procedure [53, 83], previously introduced for the MH algorithm in the VMC method.

In this regard, consider some approximate Green function and the limit of a perfect trial wave function $\psi_T \to \psi_0$, and equivalently $E_L = E_0$ or $\tilde{G}_B = 1$, in the whole configuration space. In the absence of a branching term, the MH acceptance probability

$$A(\mathbf{R}' \leftarrow \mathbf{R}) = \min\left(1, \frac{\psi_T(\mathbf{R}')^2}{\psi_T(\mathbf{R})^2}\frac{\tilde{G}_D(\mathbf{R} \leftarrow \mathbf{R}')}{\tilde{G}_D(\mathbf{R}' \leftarrow \mathbf{R})}\right) \tag{6.32}$$

preserves the correct stationary distribution $\psi_T^2 = \psi_0^2$ for any value of τ used in the approximate Green function. Therefore, an accept-reject step is necessary in the DMC algorithm, even in the case of a perfect trial wave function. Note, however, that the above argument is only valid in the limit of perfect trial wave function. For any deviation from the true ground state, the branching term deviates from unity. The acceptance probability cannot include the branching Green function, since it is mathematically only intended for norm-conserving transition probabilities. Even if the scope of the acceptance probability is extended to cover unnormalized distributions, $f = \psi_T \psi_0$ cannot be evaluated, and thus, the corresponding acceptance probability cannot be calculated. The acceptance probability of Eq. (6.32) then usually is the best approximation to the true acceptance probability. Some authors argue [84] that with the accept-reject step, the equilibrium state of the drift-diffusion equation is conserved, while the error is attributed to the reweightening due to the branching term.

The key observation is that for an approximate Green function, there remains a time-step error in the equilibrium distribution despite the acceptance step. Thus, any improvement to the quality of the Green function further reduces this time-step error, highlighting the importance of an accurate Green function.

When rejecting some of the proposed walker updates, the formal time step $\Delta\tau$ does not correspond to the effective time step $\Delta\tau_{\text{eff}}$ for that the walker actually diffuses. The latter can be calculated by observation of the accepted and rejected walker displacements. A simple approximation is $\Delta\tau_{\text{eff}} = \text{Exp}\,[A]\,\Delta\tau$, with the average acceptance probability, or acceptance rate, $\text{Exp}\,[A]$, but more elaborate approximations exist. For example, the diffused time can be calculated from the length of the diffusive displacement by

$$\Delta\tau_{\text{eff}} = \frac{\text{Exp}\left[p|\Delta\mathbf{R}_{\text{diff}}|^2\right]}{\text{Exp}\left[|\Delta\mathbf{R}_{\text{diff}}|^2\right]}\Delta\tau, \tag{6.33}$$

where $\Delta\mathbf{R}_{\text{diff}}$ is the length of the proposed displacement from the diffusion part of the Green function alone (excluding the displacement due to the drift), and p is the acceptance probability of the proposed move. The effective time step should be used in the calculation of the branching term in order to account for the actual diffusion time only. In many applications, however, the effective time step is very close to the formal time step.

Furthermore, the acceptance probability introduces a notion of probability for the new configuration $\mathbf{R}'$, namely probability p, and the old configuration $\mathbf{R}$, probability $1-p$. This can be exploited to arrive at a more accurate approximation to the updated weight compared to Eq. (6.30) with the branching Green function of Eq. (6.24). In considering both outcomes with their respective probabilities p and $1-p$, respectively, the new weight reads

$$\lambda(\tau+\Delta\tau) = \lambda(\tau)\left[p\tilde{G}_B(\mathbf{R}_k(\tau+\Delta\tau),\mathbf{R}_k(\tau)) + (1-p)\tilde{G}_B(\mathbf{R}_k(\tau),\mathbf{R}_k(\tau))\right] \tag{6.34}$$

By first-order expansion of the branching terms in the above equation and rearrangement into an exponential, the weight reduces to

$$\lambda(\tau+\Delta\tau) = \lambda(\tau)\exp\left[\Delta\tau_{\text{eff}}\Big(\frac{p}{2}\left(S(\mathbf{R}_k(\tau+\Delta\tau)) - S(\mathbf{R}_k(\tau))\right) + S(\mathbf{R}_k(\tau))\Big)\right] \tag{6.35}$$

where the effective time step was used. This update of the weight incorporates the probabilities for both configurations, increasing the statistical certainty and reducing the fluctuation of the weight. The utilization of an acceptance probability in DMC has further consequences resembling those in VMC, for example, an improved estimator for the observables, cf. Sec. 6.11.

6.6. Population control

During the time evolution of the diffusion process, the weights experience growth in regions of low local energy (below the normalization energy) and decay for the opposite case. The normalization energy E_N can be adjusted to control the evolution of the global weight $\Lambda(\tau)$ and to force $\Lambda(\tau)$ to resemble some target weight Λ_0. With the mean-field argument that the weight of the propagated ensemble $\Lambda(\tau + \Delta\tau)$ is roughly given by the averaged branching Green function applied to the current weight,

$$\langle \Lambda(\tau + \Delta\tau) \rangle = \langle \tilde{G}_B \Lambda(\tau) \rangle \approx \exp\left(\Delta\tau \bar{S}\right) \Lambda(\tau), \tag{6.36}$$

and the condition $\langle \Lambda(\tau + \Delta\tau) \rangle = \Lambda_0$. The adjustment of the normalization energy for the time step $\tau + \Delta\tau$ can be estimated with $\bar{S} = E_N - \bar{E_L}$ as

$$E_N(\tau + \Delta\tau) = E_{\text{ref}}(\tau) - \frac{\Delta\tau}{\Delta\tau_{\text{eff}}} \ln\left(\frac{\Lambda(\tau)}{\Lambda_0}\right). \tag{6.37}$$

Usually, the target weight is the number of walkers that should be propagated in parallel, $\Lambda_0 = M$, where M is determined considering the reduction of correlation, hardware limitations, and computational efficiency. The energy $E_{\text{ref}}(\tau)$ is inserted for the averaged local energy $\bar{E_L}$ and is the best available guess for the ground state energy in at time τ. The normalization is adapted in each time step to control the walker ensemble's total weight.

However, there is a flaw in the branching diffusion process presented so far. To see it, note that the branching term affects each walker's weight in a multiplicative way. After T iterations the weight of a walker takes the form

$$\lambda(T\Delta\tau) = \prod_{\tau=1}^{T} \tilde{G}_B(\tau)\lambda(0). \tag{6.38}$$

For the sake of clarity, the dependence of $\tilde{G}_B$ on the walker trajectory is omitted in the notation. For a short analysis, assume the initial weight $\lambda(0) = 1$ and that all branching terms $\tilde{G}_B(\tau)$ for $\tau = 1, \dots T$ are normally i.i.d. random variables, each with expectation value 1 and variance σ^2. This models a diffusive process with branching terms close to one, including fluctuations that scale with σ^2. Then, from an elementary calculation, the expectation value of $\lambda(T\Delta\tau)$ is unity. Its variance to the contrary is increasing exponentially with T as $(1 + \sigma^2)^T - 1$. This example illustrates the unpredictable nature of products of random variables, which is not limited to the instructive example of normally distributed random variables.

Therefore, in an ensemble of several thousand walkers, a few will eventually build up excessive weights during the random walk and dominate the distribution. The distribution of walkers is then practically reduced to the few walkers with large weights, and the overall distribution loses a large amount of statistical certainty. Serial correlation only increases the exponential scaling of the variance.

A simple yet efficient reinterpretation of the distribution f alleviates this issue. Assume for simplicity a distribution of a single walker,

$$f(\mathbf{R}, \tau) = \lambda(\tau)\delta(\mathbf{R} - \mathbf{R}(\tau)). \tag{6.39}$$

If the walker's weight exceed some reasonable threshold $\lambda(\tau) > \lambda_{\max} > 1$, the single walker can be split,

$$f(\mathbf{R}, \tau) = \alpha\lambda(\tau)\delta(\mathbf{R} - \mathbf{R}(\tau)) + (1 - \alpha)\lambda(\tau)\delta(\mathbf{R} - \mathbf{R}(\tau)) \tag{6.40}$$

with some value for $\alpha \in [0, 1]$, but ideally $\alpha = 1/2$. This is a most trivial rewriting. If the diffusion process, however, is continued for both walkers independently, this manipulation ensures bound weights. Another approach must be pursued to treat walkers of little weight, which practically do not contribute to the distribution. While it is possible to remove walkers with weight $\lambda(\tau) < \lambda_{\min} < 1$ with probability $\lambda(\tau)$, this introduces additional fluctuations in the global normalization constant. A method that does not change the total weight is to join two lightweight walkers and conditionally replace one with the other with probabilities corresponding to their respective weights.

$$(\lambda_1, \mathbf{R}_1), (\lambda_2, \mathbf{R}_2) \mapsto \begin{cases} (\lambda_1 + \lambda_2, \mathbf{R}_1) & \text{with probability } p = \lambda_1/(\lambda_1 + \lambda_2) \\ (\lambda_1 + \lambda_2, \mathbf{R}_2) & \text{with probability } p = \lambda_2/(\lambda_1 + \lambda_2) \end{cases} \tag{6.41}$$

The merging of walkers is a “winner-takes-it-all” game in which one walker is removed from the ensemble. The total weight is conserved, and a soft lower bound is imposed on the weights. The size of the walker ensemble M at a given time τ is no longer constant but becomes a variable of time $M(\tau)$. With an updated normalization energy and a birth-death process for the walkers, there is a slight interaction between them. When walkers with large weights lower the normalization energy, walkers with less weight are more prone to be removed. This evolutionary process promotes walkers in beneficial regions with positive branching terms $S(\mathbf{R}) = E_N - E_L(\mathbf{R}) > 0$ and hence in regions of low local energy. In the long-time limit of the diffusion process, the walkers do not migrate to the beneficial regions, but walkers outside beneficial regions become extinct. This is an important distinction between the spatial directedness of the walkers due to the drift, which stems from the gradient and value of the trial wave function, and the branching term, which originates from the local energy.

Without limiting the weights with a birth-death process, the VMC walker ensemble could readily be retrieved from the DMC ensemble by neglecting the walkers’ weights. Thus, one could simultaneously calculate DMC and VMC results with a single ensemble. With the birth-death process, however, this is no longer possible since the weights have an actual impact on the walker distribution by removing or adding walkers.

6.7. Simple propagation scheme

With the above sections in mind, a simple propagation scheme can be laid out. First, an ensemble of some $M(0)$ walkers with randomly chosen configuration and unit weight must be prepared. The following procedure is repeated for each time step. For each walker k in the ensemble, the walker must be propagated by first proposing a new configuration $\mathbf{R}' \sim \tilde{G}_D(\cdot \leftarrow \mathbf{R}_k(\tau), \Delta\tau)$. This is done by the proposal

$$\mathbf{R}' = \mathbf{R}_k(\tau) + \Delta\tau \mathbf{u}(\mathbf{R}_k(\tau)) + \boldsymbol{\chi} \tag{6.42}$$

where $\boldsymbol{\chi}$ is a normally distributed $3N$-dimensional vector $\boldsymbol{\chi} \sim \mathcal{N}^{3N}(0, 2D\Delta\tau)$. The configuration $\mathbf{R}'$ would be distributed according to $\tilde{G}_D$, not the MH acceptance-corrected distribution $A \times \tilde{G}_D$, therefore the acceptance probability of Eq. (6.32) is evaluated and the configuration is accepted with probability A. In the case of rejection, the configuration is not updated.

$$\mathbf{R}_k(\tau + \Delta\tau) = \begin{cases} \mathbf{R}' & \text{with probability } A \\ \mathbf{R}_k(\tau) & \text{with probability } 1 - A \end{cases} \tag{6.43}$$

Next, the weight is adjusted by Eq. (6.35). Note that the weight is adjusted irrespective of whether the move was accepted. After every walker was iterated, the new weight $\Lambda(\tau + \Delta\tau)$ is calculated, and with it, a new normalization energy from Eq. (6.37). The joining or duplication of walkers is carried out, resulting in a new number of walkers $M(\tau + \Delta\tau)$. If sufficiently many iterations have passed, the dependence on the initial configuration has vanished and the equilibrium distribution is reached. From there on, the walker ensemble resembles the function $f = \psi_T\psi_0$, and observables can be calculated by the law of large numbers, cf. Sec. 6.11. This represents the essence of the DMC algorithm, although various improvements to this simple scheme are possible as discussed in Sec. 6.13.

Although this simple Green function can give reasonable results, it reliably fails in many physical systems for the following reasons: The approximation of constant drift and branching term can always be made more accurate if the time step is lowered, but its accuracy breaks down near a singularity in both the drift and the branching term independent of the time step. Such singularities are common near particle coincidences or wave function nodes as discussed in Chapt. 5.

Since the local energy and the drift term play an active role in the time evolution of the walkers' stochastic trajectory, in most cases, their evaluation requires special care. With singularities in both quantities near nodes or nuclei, they reach excessive values in certain critical points of configuration space. Thus, their values have to be truncated to reasonable limits. Two different truncation schemes are proposed in Ref. [85] and in Ref. [53].

As might have become apparent from the above sections and the various approximations, there is some room left for improvements to the Green function. While the simple Green function of Eq. (6.23) and Eq. (6.24) yields reasonable results, its time-step error, can be reduced significantly, with the improvements discussed in Sec. 6.13.

The next sections reveal a conceptual weakness in the DMC method outlined so far and adapt the DMC approach to remedy the problem.

6.8. Fermionic sign problem

Until now, it was silently assumed that the DMC probability distribution

$$f(\mathbf{R}) = \psi_T(\mathbf{R})\psi_0(\mathbf{R}) \tag{6.44}$$

indeed qualifies as a probability distribution, i.e. is normalizable and non-negative everywhere. However, this is only true for the ground state of bosonic systems and generally violated for fermionic systems. Fermionic many-body wave functions must have nodes due to their anti-symmetry. These nodes are easily found whenever the two coordinates for the same spin coincide. If, for some reason, ψ_T and ψ_0 have identical nodes, f is non-negative. A short analysis of the nodes of many-body wave functions is required to assess this possibility.

The configuration space of N electrons each moving in d dimensions is $\mathbb{R}^{dN}$ (or a subset of it). The manifold of the nodes induced by anti-symmetry is $dN - d$ since the condition that two electrons meet is given by a vector equality in d dimensions. The manifold of the general wave function nodes, given by the single equation $\psi = 0$, has dimension $dN - 1$. The full and the symmetry-induced nodal manifolds are of the same dimension only for the case $d = 1$, but in $d = 3$ dimensions, the symmetry-induced nodes are a sub-manifold of the full nodal hypersurface. The nodal hypersurfaces of ground state wave functions indeed make use of the full dimensionality of $3N - 1$ and are not restricted to the symmetry nodes by some strange incident. Therefore, there is no hope that ψ_T and ψ_0 share nodes by accident [86, 87].

If the sign problem is ignored and the diffusion process presented so far is utilized irrespective of this limitation, the DMC distribution converges to the bosonic ground state, $f = \psi_T\psi_0^+$, since it has lower energy than the fermionic ground state ψ_0^-. The action of the operator

$\hat{\mathrm{G}}(\tau) = \exp\left(-\tau\hat{\mathrm{H}}\right)$ on a mixture of bosonic and fermionic ground state reads

$$\begin{aligned}\psi(\tau) &= e^{-\tau\hat{\mathrm{H}}}\left(c_+(0)\psi_0^+ + c_-(0)\psi_0^-\right) = \\ &= c_+ e^{-\tau E_0^+}\psi_0^+ + c_- e^{-\tau E_0^-}\psi_0^- = \\ &= c_+(\tau)\psi_0^+ + c_-(\tau)\psi_0^-\end{aligned} \tag{6.45}$$

The fermionic ground state has a higher energy $E_0^- > E_0^+$ and therefore the coefficient of the bosonic ground state outcompetes the fermionic coefficient in magnitude in the long-time limit.

$$\frac{|c_+(\tau)|}{|c_-(\tau)|} = \frac{|c_+(0)|}{|c_-(0)|} e^{\tau\left(E_0^- - E_0^+\right)} \to \infty, \tau \to \infty. \tag{6.46}$$

Furthermore, due to statistical fluctuations in the diffusion process, one cannot generate a representation of the wave function with vanishing bosonic coefficients. There have been some attempts to bypass the fermion sign problem, each with its benefits and limitations [88–91].

6.9. Fixed-node approximation

The most intuitive approach to acknowledge the fermion nodes is the fixed-node approximation (FNA). It essentially enforces the non-negativity $f \geq 0$ for an arbitrary but fixed trial wave function ψ_T. Demanding a non-negative $f = \psi_T \psi$ and at the same time keeping the trial wave function fixed, imposes additional constraints on the candidate ground state ψ. This constraint must ensure non-negative f for any trial wave function given. Therefore, the nodes of ψ must be identical to the nodes of ψ_T.

$$\forall \mathbf{R} \in \Omega^N : \quad \psi(\mathbf{R}, \tau) = 0 \Leftrightarrow \psi_T(\mathbf{R}) = 0 \tag{6.47}$$

In general, the propagated wave function $\psi(\cdot, \tau)$ cannot converge to the true ground state ψ_0 with this constraint. It instead converges to the state of lowest energy that satisfies the additional constraints and hence shares the nodes of ψ_T. The state the diffusion process converges to in the FNA is called the fixed-node ground state, ψ_0^{fn}, with a corresponding fixed-node energy E_0^{fn}.

The nodal surface of a wave function can be defined as

$$S[\psi] = \left\{\mathbf{R} \in \Omega^N \middle| \psi(\mathbf{R}) = 0\right\}. \tag{6.48}$$

The FNA energy is variational, i.e. the additional nodal constraints increase the ground state energy by reducing the search space in the Hilbert space, cf. Eq. (4.2).

$$E_0^{\mathrm{fn}}[S] = \min_{\psi \in H, S[\psi] = S} \frac{\langle\psi|\hat{\mathrm{H}}\psi\rangle}{\langle\psi|\psi\rangle} \geq E_0 \tag{6.49}$$

For any given nodes S this introduces the fixed-node error of the energy, $E_0 - E_0^{\mathrm{fn}}[S]$, which is of quadratic order in the error of the fixed-node ground state $\delta\psi = \psi_0 - \psi_0^{\mathrm{fn}}$,

$$\begin{aligned}E_0^{\mathrm{fn}} &= \langle\psi^{\mathrm{fn}}|\hat{\mathrm{H}}^{\mathrm{fn}}\psi^{\mathrm{fn}}\rangle = \langle\psi_0 - \delta\psi|\hat{\mathrm{H}}(\psi_0 - \delta\psi)\rangle = \\ &= \langle\psi_0|\hat{\mathrm{H}}\psi_0\rangle - 2\langle\delta\psi|\hat{\mathrm{H}}\psi_0\rangle + \langle\delta\psi|\hat{\mathrm{H}}\delta\psi\rangle = \\ &= E_0 - 2E_0\langle\delta\psi|\psi_0\rangle + \mathcal{O}^2(\delta\psi) = \\ &= E_0 + \mathcal{O}^2(\delta\psi),\end{aligned} \tag{6.50}$$

since $\langle\delta\psi|\psi_0\rangle = 0$. The probability distribution f does also show a fixed-node error, $\psi_T\delta\psi = \psi_T(\psi_0 - \psi_0^{\mathrm{fn}})$, however, linear in $\delta\psi$.

If the trial nodes $S_T = S[\psi_T]$ approach the true nodes $S_0 = S[\psi_0]$, the nodal constraint is satisfied already and the true ground state energy is recovered, $E_0^{\mathrm{fn}}[S_0] = E_0$. Note that the trial wave function is not required to approach the true ground state. The trial nodes approaching the true nodes are sufficient. Therefore, the ground state energy and the ground state are functionals of the nodal surface alone.

Unfortunately the analysis of topological characteristics and general structure of exact nodes remains a challenging field of research [86, 87, 92–94]. This is not surprising, since the full information of the ground state must be present in the nodal structure. The complexity of the mapping from the nodes to the ground state resembles the complexity of the Hohenberg-Kohn map, discussed in Sec. 3.1. Thus, only a few exact properties of ground state nodes are known.

So far, the diffusion process does not respect the nodes imposed by the trial wave function, and a modification to it is required. Assume a walker that is subject to the previously derived diffusion process with some finite time step is crossing a node of ψ_T during its random walk. Since ψ shares the trial wave function's nodes, f has quadratic nodes at S_T. A node-crossing random walk may be unlikely due to the little probability near the nodes and a drift vector repelling the walker from the node, but it is still possible. With an increasing time step, the probability of node-crossing increases. By reducing the time step further and further, the probability of the walker crossing the trial nodes vanishes, since the acceptance probability is proportional to ψ_T^2. In the limit of a vanishing time step, the FNA holds automatically. However, any attempt to cross the trial nodes must be rejected for a finite time step. This is accomplished by modification of the acceptance probability, which is adjusted to vanish if nodes are crossed. For any proposed move from $\mathbf{R}$ to $\mathbf{R}'$ replace the MH acceptance probability by

$$A^{\mathrm{fn}}(\mathbf{R}' \leftarrow \mathbf{R}) = A(\mathbf{R}' \leftarrow \mathbf{R}) \times \begin{cases} 1, & \operatorname{sign}(\psi_T(\mathbf{R}')) = \operatorname{sign}(\psi_T(\mathbf{R})) \\ 0, & \operatorname{sign}(\psi_T(\mathbf{R}')) \neq \operatorname{sign}(\psi_T(\mathbf{R})) \end{cases} \tag{6.51}$$

This confines the random walk within a nodal pocket. A nodal pocket ω around some configuration $\mathbf{R}_0$ is the region of space that is reachable by any continuous path that starts at $\mathbf{R}_0$ and does not cross a node.

$$\omega[\psi, \mathbf{R}_0] = \left\{ \mathbf{R} \in \Omega^N \middle| \mathbf{R} \in \gamma(I) \text{ with } \gamma : I \to \Omega^N \backslash S[\psi] : \gamma(0) = \mathbf{R}_0 \right\} \tag{6.52}$$

The configuration $\mathbf{R}_0$ is restricted such that it does not lie on a node, i.e., $\psi_T(\mathbf{R}_0) \neq 0$. The configuration space is thus segmented into a number of mutually non-overlapping nodal pockets. The trial wave function has constant sign inside a single nodal pocket. Using the manipulated diffusion process has side effects on the resulting equilibrium distribution. These are discussed next.

The Schrödinger equation solved by the fixed-node wave function differs from the original Schrödinger equation in general and has a different Hamiltonian. The fixed-node Hamiltonian [95] can be written down explicitly for any imposed nodes S.

$$\hat{\mathrm{H}}^{\mathrm{fn}}[S] = \hat{\mathrm{H}} + V^{\mathrm{fn}} \begin{cases} 0, & \mathbf{R} \in S \\ 1, & \mathbf{R} \notin S \end{cases} \tag{6.53}$$

In addition to the molecular Hamiltonian $\hat{\mathrm{H}}$ of the physical problem, it incorporates a non-passable potential barrier, $V^{\mathrm{fn}} \to \infty$, at the nodes. Any wave function minimizing the action of $\hat{\mathrm{H}}^{\mathrm{fn}}[S]$ must vanish at the potential barriers. Solving the Schrödinger equation for the fixed-node Hamiltonian in the whole configuration space is equivalent to solving it for each nodal pocket separately since the nodes impose zero boundary conditions for each nodal pocket, and there is no interaction between the solutions of different pockets. With an infinite potential barrier, the wave function is also allowed to have discontinuous first-order derivatives at the node. Let the set of all nodal pockets be $\{\omega_\alpha\}$ with $\Omega^N \backslash S = \cup_\alpha \omega_\alpha$. In each pocket α, an independent wave function

$$\psi^{\text{fn}}_\alpha : \Omega^N \to \mathbb{R}_+ \quad \text{with} \quad \psi^{\text{fn}}_\alpha|_{\Omega^N \backslash \omega_\alpha} = 0 \tag{6.54}$$

solves the fixed-node Hamiltonian as an eigenstate with some energy E^{fn}_α. Since each pocket wave function is nodeless inside its pocket, one can utilize the standard diffusion process to converge it to the bosonic ground state within the pocket.

$$\hat{\mathrm{H}}^{\text{fn}} \psi^{\text{fn}}_\alpha = \hat{\mathrm{H}} \psi^{\text{fn}}_\alpha = E^{\text{fn}}_\alpha \psi^{\text{fn}}_\alpha \tag{6.55}$$

A wave function on the whole domain is given by some linear combination of the pocket wave functions.

$$\psi^{\text{fn}} = \sum_\alpha c_\alpha \psi^{\text{fn}}_\alpha \tag{6.56}$$

A wave function's time evolution under the action of the propagation operator $\hat{\mathrm{G}}^{\text{fn}}(\tau)$ corresponding to the fixed-node Hamiltonian is given by

$$\begin{aligned}
\psi^{\text{fn}}(\tau) = \hat{\mathrm{G}}^{\text{fn}}(\tau)\psi^{\text{fn}}(0) &= \sum_\alpha c_\alpha(0) \exp\left(-\tau\left(\hat{\mathrm{H}} - E_N\right)\right)\psi^{\text{fn}}_\alpha \\
&= \sum_\alpha c_\alpha(0) \exp\left(-\tau\left(E^{\text{fn}}_\alpha - E_N\right)\right)\psi^{\text{fn}}_\alpha \\
&= \sum_\alpha c_\alpha(\tau)\psi^{\text{fn}}_\alpha
\end{aligned} \tag{6.57}$$

Absorbing the exponential into the time dependence of the coefficient, it becomes apparent that only the coefficients of minimal energy contribute to the wave function in the long-time limit. The normalization enters by adjustment of the normalization energy E_N, which has to be set equal to the lowest E^{fn}_α in order to maintain normalization in the long-time limit. Thus, the branching part of the diffusion process depopulates nodal pockets of higher energy. While $\psi^{\text{fn}}(\tau)$ is not necessarily an eigenfunction of the fixed-node Hamiltonian for any time τ,

$$\hat{\mathrm{H}}^{\text{fn}} \psi^{\text{fn}}(\tau) = \sum_\alpha c_\alpha(\tau) E^{\text{fn}}_\alpha \psi^{\text{fn}}_\alpha \to (\min_\alpha E^{\text{fn}}_\alpha)\psi^{\text{fn}}_{\text{eq}} \tag{6.58}$$

it is in the limit $\tau \to \infty$, since only coefficients with minimal pocket energy survive. Denote the index of a nodal pocket with minimal energy by $\tilde{\alpha}$. The fixed-node energy is then given by $E^{\text{fn}}_0 = \min_\alpha E^{\text{fn}}_\alpha = E^{\text{fn}}_{\tilde{\alpha}}$. One might stumble upon the fact that now the wave function has a rather strange form. It vanishes everywhere except for the nodal pocket $\omega_{\tilde{\alpha}}$. In this form, the wave function does not obey anti-symmetry and thus disqualifies as a solution. Anti-symmetrization from Eq. (5.2) can restore the anti-symmetry.

$$\psi^{\text{fn}}(\tau) = \frac{1}{N!} \sum_{\pi \in \mathbb{S}(N)} \operatorname{sign}(\pi) \pi \psi^{\text{fn}}_{\tilde{\alpha}}(\tau) \tag{6.59}$$

The question arises whether, after the anti-symmetrization, some regions of configuration space still have vanishing wave function amplitude, i.e. whether the anti-symmetrization is sufficient to generate a qualified wave function from a single pocket. To this end, fortunately, one important topological property of ground state nodes comes to aid: the tiling property [92]. It states that there is only one unique nodal pocket, irrespective of the system under study. For any nodal pocket ω and any configuration $\mathbf{R} \notin \omega$ that does not lie on a node, a permutation π can be found such that $\pi\mathbf{R} \in \omega$. Due to the tiling property, there are still an unknown number of nodal pockets, but all are permuted copies of a single unique pocket. The two-nodal-domains conjecture [93] even support significant evidence that there are at most two nodal domains, a positive and a negative one. With the tiling property at hand, the solution of the fixed-node Hamiltonian on a single nodal pocket is sufficient. The remaining configuration space is "filled" or tiled with the wave function $\psi^{\text{fn}}_{\tilde{\alpha}}$ by permutations. The minimization of the energy over all pockets α becomes trivial. In fact, all nodal pockets are equivalent. This implies that the diffusive process rejecting node-crossings indeed yields an equilibrium state that respects the nodes, represents the fixed-node ground state, and results in an DMC energy variational in the nodes. The DMC method based on the FNA is called fixed-node diffusion Monte Carlo (FNDMC).

Note that even without the rejection of node crossings in the standard diffusion process, node crossings at small time steps are not probable. First, the acceptance probability lowers node crossing probabilities, as mentioned above. Second, the drift vector near nodes points away from the local nodal plane, cf. Eq. (B.31), and its magnitude scales inversely with the distance to the node.

The diffusive process with node-crossing rejection is still not an algorithm free of errors for finite time steps. Assume two configurations, $\mathbf{R}_1$ and $\mathbf{R}_2$, to be in the same nodal pocket as depicted in Fig. 6.1. The straight line connecting both configurations passes through another nodal pocket, crossing two nodal surfaces. With a sufficiently large time step, a random walk might lead from $\mathbf{R}_1$ to $\mathbf{R}_2$ by crossing and recrossing the nodal surface in a single time step. The diffusive process has no indication that a node has been crossed and accepts the proposed move. With a smaller time step, double node-crossings become less probable, and the random walk eventually cannot step across the other nodal pocket within one move. Thus, the double-crossing violates the FNA. Assume that the nodeless path of least length between $\mathbf{R}_1$ and $\mathbf{R}_2$ is the one indicated in Fig. 6.1. A random walk along this path might be assigned a much smaller probability since more distance has to be traveled. The shortcut of crossing and recrossing therefore corrupts the probability distribution, introducing another contribution to the time-step error, that can be extrapolated, see Sec. 6.15.

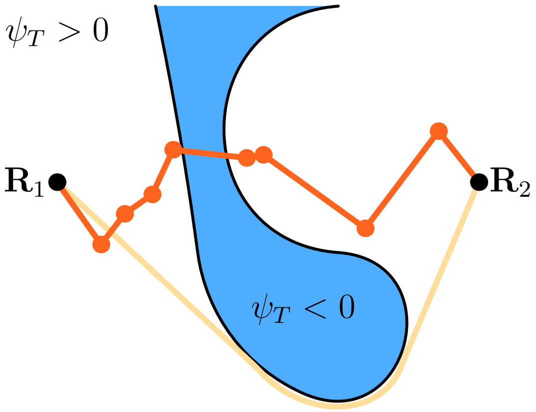


Figure 6.1.: Illustration of the cross-recross error. While moves across nodes are forbidden in FNDMC, the DMC algorithm cannot efficiently notice double crossings. While a nodeless random walk connecting $\mathbf{R}_1$ and $\mathbf{R}_2$ might have a rather different probability assigned to it, the shortcut across the nodes introduces an error in the probability distribution.

In the following, there will be no notational reminder for the FNA. Every DMC wave function in this thesis is a fixed-node wave function and the terms FNDMC and DMC will be used interchangeably.

6.10. Trial wave function nodes

The nodal structure of a trial wave function determines the fixed-node energy in a DMC calculation. Therefore, the nodes play one of the most important roles in FNDMC. In a single-determinant wave function without back-flow, the orbitals alone determine the nodal structure. There is, however, a severe shortcoming of single-determinant nodes. The representation of a single-reference wave function is based on the product of two determinants, one for each spin channel, $\phi = \phi_\alpha \phi_\beta$. Therefore, the nodes of a single-reference wave function are given by the condition $\phi_\alpha(\mathbf{R}_\alpha) = 0 \vee \phi_\beta(\mathbf{R}_\beta) = 0$. The corresponding nodal structure

$$S[\phi] = S[\phi_\alpha] \cup S[\phi_\beta] = \left\{\mathbf{R} \in \Omega^N \middle| \phi_\alpha(\mathbf{R}_\alpha) = 0\right\} \cup \left\{\mathbf{R} \in \Omega^N \middle| \phi_\beta(\mathbf{R}_\beta) = 0\right\} \tag{6.60}$$

cannot represent the nodal structure of a general wave function due to its topology. A simple instructive example is the two-dimensional case, where some function $f(x, y)$ could exhibit nodes in the form of, for example, a circle. A product function $g(x)h(y)$ necessarily has nodes that are given by intersecting horizontal and vertical lines in the x-y-plane and cannot resemble a circle-like shape.

Studies on the topology of interacting and non-interacting wave functions have found several cases where the non-interacting nodes also violate other topological characteristics of the interacting wave function [86, 87, 93, 94]. With this limitation in mind, the expectations about the FNDMC energy of a single-reference trial wave function must be reconsidered. The corresponding FNDMC energy should not be expected to converge to the true ground state energy. The interesting question is, instead, how close the FNDMC energy of the simple trial wave function is to the true ground state energy compared to more elaborate methods, including Slater determinant expansions. An important note on expansions in Slater determinant space is given in Ref. [86]. If an “inappropriate” determinant is included in the expansion the nodal structure might get spoiled and the FNDMC energy increases. Thus, simply adding more determinants does not guarantee to systematically improve the energy. However, a multi-reference trial wave function can of course be highly beneficial to more accurately resemble the true wave function’s nodes.

For reasons explained in Chapt. 11 no determinantial superposition and no back-flow is used in this thesis, and the trial wave function is restricted to the form of Eq. (5.14). With a single determinant, its building single-particle orbitals play a vital role. They alone define the nodal structure of the trial wave function. In this thesis, orbitals from DFT are used, which are represented on a spatial grid that exploits the cylindrical symmetry of diatomic molecules, cf. Chapt. E for details. The density functionals used to self-consistently calculate the orbitals are the LDA in the Perdew-Wang parametrization [47] and the exact exchange functional in the KLI approximation [34, 35].

The comparison of wave functions with different nodes and therefore DMC energies is given in Sec. 9.1. The wave functions include the single-determinant Slater-Jastrow wave function discussed in Sec. 5.3 generated with LDA or xKLI orbitals and multi-determinant Slater-Jastrow wave functions.

6.11. Observables

Now that the diffusive process of DMC is adapted to deal with the fermion sign problem and an ensemble of walkers that resemble the (fixed-node) ground state can be generated efficiently, the evaluation of observables from the ensemble is discussed.

In VMC, the evaluation of observables beyond the energy is straightforward. ψ_T represents the wave function and $f = \psi_T^2$ is accessable by the walker ensemble. In DMC, on the other hand, the wave function ψ_0 is not accessible analytically. The realized ensemble is $f = \psi_T \psi_0$. Whenever a walker ensemble is available for a certain distribution, expectation values can be estimated by the law of large numbers, cf. Eq. (4.22) for the VMC case. In DMC, the walker ensemble additionally holds weights, which have to be respected in the observable estimation. Although the weights are genuinely random variables due to their coupling to the random configuration, they are treated like deterministic weights in the statistical analysis of the observables. This is solely due to the fact that the inclusion of the random nature of the weights would prohibit the analysis of most of the observables' statistical properties.

While in VMC the notion of time was an artificial relabeling with no physical meaning, in DMC, time becomes a more central variable for several reasons. Recall the representation of the walker ensemble at time τ is

$$f(\mathbf{R},\tau) = \sum_{k=1}^{M(\tau)} \lambda_k(\tau)\delta(\mathbf{R}-\mathbf{R}_k(\tau)) \tag{6.61}$$

with a normalization constant

$$\Lambda(\tau) = \sum_{k=1}^{M(\tau)} \lambda_k(\tau). \tag{6.62}$$

Time-averaged from the equilibration time, which is assumed to be $\tau_{\text{eq}} = 0$, to the time $T\Delta\tau$, the complete walker ensemble reads

$$F(\mathbf{R},T) = \frac{1}{T}\sum_{\tau'=1}^{T}\sum_{k=1}^{M(\tau')} \lambda_k(\tau')\delta(\mathbf{R}-\mathbf{R}_k(\tau')) \tag{6.63}$$

with normalization

$$\Lambda(T) = \frac{1}{T}\sum_{\tau'=1}^{T}\sum_{k=1}^{M(\tau')} \lambda_k(\tau') = \frac{1}{T}\sum_{\tau'=1}^{T}\Lambda(\tau'). \tag{6.64}$$

Instead of a time-average over non-normalized distributions $f(\cdot,\tau)$ one could argue to average over normalized distributions $\Lambda(\tau)^{-1}f(\cdot,\tau)$, therefore enforcing normalization of $f(\cdot,\tau)$ in every time step. The counter-argument is, that once the equilibrium state is reached, the operator $\hat{\mathrm{G}}$ is norm-conserving for the right choice of E_N and thus the differing normalization constants at different times are only the statistical fluctuations of the over all ensemble. A refinement of this approach is required since the normalization energy is adjusted every time step to keep a constant population of walkers, cf. Sec. 6.6. This refinement [53] will be omitted here for simplicity and discussed in Sec. 6.13.

Next, a DMC estimator can be derived by generalizing Eq. (4.22) to a weighted ensemble by inserting the ensemble F in the definition of the expectation value.

$$\frac{\langle\psi_0|\hat{\mathrm{O}}\psi_T\rangle}{\langle\psi_0|\psi_T\rangle} = \frac{\sum\limits_{\tau'=1}^{T}\sum\limits_{k=1}^{M(\tau')} \lambda_k(\tau')O_L(\mathbf{R}_k(\tau'))}{\sum\limits_{\tau'=1}^{T}\sum\limits_{k=1}^{M(\tau')} \lambda_k(\tau')} \tag{6.65}$$

Depending on the character of the observable $\hat{\mathrm{O}}$, it might be necessary to exclude some of the observations $O_L(\mathbf{R}_k(\tau'))$ from the above sum. The local energy and the drift vector are two such observables that become very large in magnitude near either nodes or points of particle-coincidences, cf. Sec. B. Truncation schemes have been proposed [53, 85] to limit the observed values allowed for accumulation within some reasonable range.

With the same arguments made in the derivation of Eq. (4.22), the acceptance probabilities $p_k(\tau)$ can be incorporated in the estimator to reduce its variance, which reads

$$\frac{\sum_{\tau'=1}^{T}\sum_{k=1}^{M(\tau')}\lambda_k(\tau')\left[p_k(\tau')O_L(\mathbf{R}_k(\tau'))+(1-p_k(\tau'))O_L(\mathbf{R}_k(\tau'-1))\right]}{\sum_{\tau'=1}^{T}\sum_{k=1}^{M(\tau')}\lambda_k(\tau')}. \tag{6.66}$$

6.12. The extrapolation estimator

Now that an estimator for the distribution $f=\psi_T\psi_0$ is found, estimates over this distribution can be calculated. However, the target quantity to obtain, the ground state energy, is not related to $\psi_T\psi_0$ but to ψ_0^2. The ground state energy can nevertheless be calculated, even with $f=\psi_T\psi_0$. To this end, the trial wave function is expanded in eigenstates of the Hamiltonian, $\psi_T=\sum_i c_i\psi_i$, and the overlap integral

$$\frac{\langle\psi_0|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi_0|\psi_T\rangle}=\frac{\langle\psi_0|\sum_i c_iE_i\psi_i\rangle}{\langle\psi_0|\sum_i c_i\psi_i\rangle}=\frac{\langle\psi_0|\hat{\mathrm{H}}\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} \tag{6.67}$$

yields the ground state energy due to orthogonality, $\langle\psi_i|\psi_j\rangle=\delta_{ij}$. In this work, the DMC energy is calculated from the local energy with the estimator of Eq. (6.66). The growth estimator [53, 96], an energy estimate based on the evolution of the total weight common in DMC, is not used. An unconverged wave function $\psi=\psi_0+\delta\psi$ with $\langle\psi|\delta\psi\rangle=0$, however, is not an eigenstate of the Hamiltonian and the above expectation value is contaminated.

$$\begin{aligned}\int fE_L=\frac{\langle\psi|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi|\psi_T\rangle}&=\frac{\langle\psi_0|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi|\psi_T\rangle}+\frac{\langle\delta\psi|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi|\psi_T\rangle}=E_0\frac{\langle\psi_0|\psi_T\rangle}{\langle\psi|\psi_T\rangle}+\frac{\langle\delta\psi|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi|\psi_T\rangle}\\&=E_0\frac{\langle\psi_0|\psi_T\rangle}{\langle\psi_0|\psi_T\rangle+\langle\delta\psi|\psi_T\rangle}+\frac{\langle\delta\psi|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi|\psi_T\rangle}\\&=E_0\left(1-\frac{\langle\delta\psi|\psi_T\rangle}{\langle\psi_0|\psi_T\rangle}\right)+\frac{\langle\delta\psi|\hat{\mathrm{H}}\psi_T\rangle}{\langle\psi_0|\psi_T\rangle}+\mathcal{O}(\delta\psi^2)\\&=E_0+\frac{\langle\delta\psi|\hat{\mathrm{H}}-E_0|\psi_T\rangle}{\langle\psi_0|\psi_T\rangle}+\mathcal{O}(\delta\psi^2)\end{aligned} \tag{6.68}$$

Any error in the probability distribution f implies an error in the accumulated observables. The error is of first order in the deviation of the unconverged state to the ground state and simultaneously first order in the deviation of the trial wave function from the ground state.

In a QMC calculation, the two distributions available are ψ_T^2 and $\psi_T\psi_0$ with their respective estimators for the observable $\hat{\mathrm{A}}$

$$A_m=\frac{\langle\psi_0|\hat{\mathrm{A}}\psi_T\rangle}{\langle\psi_0|\psi_T\rangle} \tag{6.69}$$

$$A_T=\frac{\langle\psi_T|\hat{\mathrm{A}}\psi_T\rangle}{\langle\psi_T|\psi_T\rangle} \tag{6.70}$$

where A_m is called the mixed estimator [97]. However, the desired estimator is

$$A_0=\frac{\langle\psi_0|\hat{\mathrm{A}}\psi_0\rangle}{\langle\psi_0|\psi_0\rangle}. \tag{6.71}$$

The distribution ψ_0^2 is not realized in the VMC or DMC ensembles. Only in the limit $\psi_T \to \psi_0$ all estimators are equal. To analyze the order of errors in each estimator, define $\psi_0 = \psi_T + \delta\psi_T$ with $\langle\psi|\delta\psi_T\rangle = 0$. Thus, the trial wave function is exact in the limit $\delta\psi \to 0$. The accumulation procedure of Eq. (6.67) cannot be applied to any observable $\hat{\mathrm{A}}$. It requires the ground state ψ_0 to be an eigenstate of $\hat{\mathrm{A}}$. For observables that commute with the Hamiltonian, $[\hat{\mathrm{H}}, \hat{\mathrm{A}}] = 0$, the eigenstates of $\hat{\mathrm{A}}$ can be chosen to be the eigenstates of $\hat{\mathrm{H}}$. Then the same arguments as above apply, and $\hat{\mathrm{A}}$ can be estimated by the mixed estimator, $A_{\mathrm{m}} = A_0$.

For observables that do not commute with the Hamiltonian, like the density, kinetic energy, potentials, dipole moments, the above is not applicable. In this case, the mixed estimator is fundamentally biased. However, from

$$\begin{aligned}\langle\psi_0|\psi_0\rangle A_0 &= \langle\psi_0|\hat{\mathrm{A}}\psi_0\rangle \\ &= \langle\psi_0|\hat{\mathrm{A}}\psi_T\rangle + \langle\psi_0|\hat{\mathrm{A}}\delta\psi_T\rangle \\ &= A_{\mathrm{m}} + \langle\psi_T|\hat{\mathrm{A}}\delta\psi_T\rangle + \langle\delta\psi_T|\hat{\mathrm{A}}\delta\psi_T\rangle \\ &= A_{\mathrm{m}} + \langle\psi_T|\hat{\mathrm{A}}\psi_0\rangle - \langle\psi_T|\hat{\mathrm{A}}\psi_T\rangle + \langle\delta\psi_T|\hat{\mathrm{A}}\delta\psi_T\rangle \\ &= 2A_{\mathrm{m}} - \langle\psi_T|\hat{\mathrm{A}}\psi_T\rangle + \langle\delta\psi_T|\hat{\mathrm{A}}\delta\psi_T\rangle \\ &= 2A_{\mathrm{m}} - A_T + \langle\delta\psi_T|\hat{\mathrm{A}}\delta\psi_T\rangle\end{aligned} \tag{6.72}$$

together with $\langle\psi_0|\psi_0\rangle = \langle\psi_0|\psi_T\rangle + \langle\psi_0|\delta\psi_T\rangle$ a trivial but longer calculation shows

$$A_0 = 2A_m - A_T + \mathcal{O}(\delta\psi_T^2), \tag{6.73}$$

where the first order error cancels. This gives rise to an estimator for the ground state observable that is quadratic in the error of the trial wave function, the so-called extrapolation estimator [97]

$$A_e = 2A_m - A_T. \tag{6.74}$$

It can be verified, that A_e samples from the distribution $(2\psi_0 - \psi_T)\psi_T$ with an error of quadratic order in $\delta\psi_T$.

$$A_e = \frac{\langle(2\psi_0 - \psi_T)|\hat{\mathrm{A}}\psi_T\rangle}{\langle(2\psi_0 - \psi_T)|\psi_T\rangle} + \mathcal{O}(\delta\psi_T^2) \tag{6.75}$$

The corresponding calculation is again elementary. The extrapolation estimator will become particularly important for the evaluation of the density in Sec. 7.

6.13. The Umrigar-Nightingale-Runge Green function

So far, a reasonable approximation to the true Green function has been presented in Sec. 6.7. It stems from the short-time approximation of a Schrödinger equation. However, there are some flaws in the Green function, especially its behavior at special points or hypersurfaces of the configuration space, like electron-nuclear coincidences or wave function nodes. As mentioned in Sec. 6.2, the ideal Green function $G(\mathbf{R}' \leftarrow \mathbf{R})$ is independent of its rightmost variable $\mathbf{R}$ and represents the stationary solution as a function of its leftmost variable $\mathbf{R}'$. The simple normally distributed Green function is hence a good approximation for free electrons, since free electrons move in Gaussian wave packages after localization. To also capture the electron behavior in the proximity of a nucleus, Umrigar, Nightingale and Runge (UNR) [53] extended the simple Green function by an exponential, hydrogen-like distribution to cover the limiting case of an electron confined by an ion. For the Green function's analytical form the reader is referred to the instructive Ref. [53]. Its physical interpretation is insightful and mentioned in the following.

Starting from the characteristic behavior of the wave function near critical points, cf. Sec. B, the improved Green function takes into account the correct asymptotic behavior of the wave function, its drift vector, and its local energy, going beyond the approximation of constant drift and branching term. From the correct asymptotic behavior, a truncation method for local energy and drift is devised to reject outliers more accurately. The drift-diffusion Green function, proposing single electron displacements, is adjusted to conditionally sample from two elementary probability distributions. In the vicinity of a nucleus, the Green function draws from an exponential distribution that resembles the hydrogen-like 1s orbital, with high probability. Sec. C.9 given a detailed explanation on how to draw samples from the exponential distribution in three-dimensional space. Further away from a nucleus, the probability to draw from the hydrogen-like distribution decreases while sampling from the normal distribution becomes more probable. Additional measures are taken to increase the de-correlation of the proposed moves.

The branching Green function is also subject to improvements by incorporating the acceptance probability weighting from Eq. (4.22) into the evaluation of the branching term of Eq. (6.24) and applies a truncation scheme for the local energy. While the requirement to account for rejected moves with an effective time step was explained in Sec. 6.5, the actual calculation of the effective time step is refined in Ref. [53].

Another important improvement is the conception of a correct treatment of the weights, together with the adjustment of the normalization energy. If the propagation operator $\hat{\mathrm{G}}$ is realized with a manually adjusted normalization energy for every time step, $E_N(\tau)$, the ensemble weight for this single time step $\lambda_k(\tau)$ are biased towards the normalization energy. The history of the adjustments has to be remembered and included in the weighting of observables to account for and remedy the bias of the manual adjustment.

6.14. Persistent walkers

> The following section is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

Over the course of the huge number of MC steps in a DMC run, when walkers explore the configuration space, some of them can appear to become immobilized in configuration space for many propagation steps. Proposed moves generated from the Green function are rejected due to the walkers local environment, making the walker a persistent one. While this does not influence the DMC run if the walker's weight is reducing (i.e., if it is located in a region where the local energy is above the normalization energy), it becomes a serious bias of the statistical accumulation if its weight increases, and therefore the walker duplicates itself. Umrigar et al. [53] chose to multiplicatively "boost" the acceptance probability by a constant factor to higher values after a fixed number of rejections to cope with this problem. We found that the boosting process in some cases requires too many steps to effectively prevent a persistent walker from harming the statistical accumulation. Therefore, we here adjust the notion of when a walker can be regarded as being persistent and propose appropriate actions.

Assume that in some calculation the average acceptance probability takes the value of $p = \mathrm{Exp}\,[A]$. A walker governed by this acceptance rate stays at its current location for one iteration with probability $P(1) = (1-p)$. The probability to stay at the same location for n consecutive iterations is $P(n) = (1-p)^n$. The probability decreases exponentially with the number of idle iterations n, resembling the geometric distribution. This equips every idle walker with a probability. In a calculation there might be walkers with only a few idle steps n and therefore (small but) reasonable probabilities $P(n)$. On the other hand, if walkers are idle for many iterations, this hints to some issue in their diffusive process. Their associated probability will become very small. If we chose a small threshold probability P_t, we can categorize the reasonable idle times by $P(n) > P_t$, and the un-physical idle times by $P(n) < P_t$. Solving $P(n) = P_t$ for the number of idle steps, $n_{\text{idle}}(p, P_t) = \frac{\ln(P_t)}{\ln(1-p)}$ yields an estimate of how many idle iterations can be regarded as physically reasonable, given the acceptance rate $\text{Exp}\,[A]$ and some predefined threshold probability P_t. Setting the threshold probability P_t to a small value, we can compute the corresponding maximum number of idle steps n_{idle} that a walker should be allowed to be immobilized for. Every walker immobile for more than the determined n_{idle} steps can be regarded as persistent. Of course, the acceptance rate $\text{Exp}\,[A]$ may be a poor indicator for the local acceptance rate in some places of configuration space. However, by choosing a small enough threshold P_t one can compensate for a locally lower acceptance rate.

The major advantage of this approach is incorporating the acceptance rate $\text{Exp}\,[A]$ into the definition of persistence. It relieves the user to set an actual number for n_{idle}. At the same time, the new criterion is more broadly applicable since varying acceptance rates readily adjust the notion of persistence. The new parameter is the threshold P_t. It is an absolute probability, valid independent of the system under consideration.

Once a walker is identified as persistent, we implement the boosting [53] of the acceptance probability. Additionally, we set the walker's branching term to one, essentially conserving its weight during the boost process. Thus no duplication of a persistent walker is possible. As this manipulates the diffusive process, an alternative would be to only ban the walker from contributing to the statistical accumulation of the observables, while waiting for an accepted move within the unchanged diffusion process.

6.15. Summary of errors

The DMC approach comes with various types of well understood errors, which the user has to respect to generate reliable results. While, intuitively, errors seem to be a threat to the accuracy of a method, all numerical calculations are bound to exhibit errors and understanding the errors' behavior is central to assess the reliability of the methods. The error of a DFT calculation, in contrast, usually cannot be assessed rigorously from within DFT.

The first and most prominent error of any MC application is the statistical error of the stochastic estimates. At the same, this is a relatively pleasant type of error since it can be estimated from the MC method. Methods to calculate the uncertainty of an MC estimate are presented in Chapt. C.

The time-step error of DMC originates from a finite time step and formally vanishes in the limit $\Delta\tau \to 0$. This limit can never be realized since the DMC approach requires a finite time step to explore the configuration space. There are many contributions to the time-step error from various approximations, e.g., the approximate Green function, the approximate drawing procedure discussed in Sec. D.5, the MH acceptance step, the cross-recross error, to only name a few. The time-step error can nevertheless be readily assessed with an extrapolation. Therefore, the DMC calculation is repeated several times with different time steps approaching zero. An extrapolation to zero then yields the estimates free of the time-step error. This extrapolation naturally is computationally quite expensive because multiple independent calculations are required for a single extrapolation. In this thesis, some results are evaluated using a time-step extrapolation. However, the time-step error can become small enough to be omitted entirely, especially if the improved Green function from Sec. 6.13 is used. By comparison of the time-step extrapolation, i.e. multiple values for some observable and its extrapolated value, the time step can be chosen, such that the time-step error of a single calculation is in an acceptable range, e.g., within the statistical error. In a conservative estimate, the time-step error is linear in the time step, but some indication of Ref. [53] point to a quadratic scaling.

The FNDMC energy is a functional of the nodal structure of the trial wave function, cf. Sec. 6.9. With the single-reference Slater-Jastrow trial wave function from Sec. 5.3 the true nodes of the exact ground state cannot be resembled, cf. Sec. 6.10. In contrast, topological inconsistencies of the trial wave function nodes are inevitable. Nevertheless, the FNDMC energy can reasonably estimate the true ground state energy. The fixed-node error is assessed in Sec. 9.1 by comparison to FNDMC energies in the literature.

The estimation of observables that do not commute with the Hamiltonian is biased by the mixed-estimator. Although the extrapolation estimator remedies a large amount of this bias, it might still be present in many observables, e.g., all observables, that directly stem from the density or equivalently and local multiplicative observable. This is particularly probable for large deviations of ψ_T from ψ_0. The results from Sec. 9.2 allow some insights into the mixed estimator bias and its extrapolation scheme.

7. Estimating the charge density

With VMC and DMC methods at the disposal, any observable can be estimated. The wave function based QMC approach promises to capture almost all correlation effects and the estimated observables represent quasi-exact ground state properties. One such observable is the density. It has its own connection to the wave function via the Hohenberg-Kohn theorem and is the central quantity of DFT, cf. Sec. 3.1. However, to generate an estimate for the spatially resolved density $n(\mathbf{r})$ a estimation strategy is given below.

7.1. Binning technique

> The following section is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

To capture the electron density with spatial resolution, we use the binning technique. This well-known approach generates a histogram of the underlying density by performing a Bernoulli experiment for each bin at each walker realization. For VMC, where the many-body probability function ψ_T^2 is known analytically, Ref. Assaraf et al. [98] introduced an improved estimator for the density, based on derivatives of the trial wave function. In DMC, to the contrary, the probability function $\psi_T \psi_0$ is only known in terms of its stochastic realizations, and thus such approaches are not possible. Another density estimation common for Markov chains in general is kernel density estimation, which can be understood as a regularization or smoothing of the Dirac distribution. It maps the walker's location in the configuration space (together with other available information about the walker state, such as derivatives) to some function that estimates the single walker‘s density. Often kernel density estimation produces visually more appealing representations of the density compared to a simple histogram. However, a smoothing of the density has no justification in QMC, e.g., using kernel density estimation near the nucleus, the underlying kernel would hardly reproduce the electron-nucleus-cusp condition. Therefore, we resort to the binning method in conjunction with our prolate spheroidal grid, see Sec. E.3.

To start with, the (symmetrized) one-body density operator reads

$$\hat{\mathrm{n}}(\mathbf{r}, \mathbf{R}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{R}_i) \tag{7.1}$$

with the ith electron coordinate denoted by $\mathbf{R}_i$. The operator is applied to the probability function f from Eq. (6.25). To represent the resulting continuous (mixed estimator) density

$$n(\mathbf{r}) = \int_{\Omega^N} f(\mathbf{R}) \hat{\mathrm{n}}(\mathbf{r}, \mathbf{R}) \, \mathrm{d}\mathbf{R} \tag{7.2}$$

we introduce a discretization or binning of the domain of a single electron Ω into subdomains or bins $\{\omega_j\}$, such that, $\omega_j \in \Omega$ pairwise disjoint and $\cup_j \omega_j = \Omega$. The density integrated over any such bin ω is

$$\begin{aligned} N(\omega) &= \int\limits_{\omega} n(\mathbf{r})\, \mathrm{d}\mathbf{r} = \int\limits_{\Omega^N} f(\mathbf{R}) \sum_i \int\limits_{\omega} \delta(\mathbf{r}-\mathbf{R}_i)\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{R} \\ &= \int\limits_{\Omega^N} f(\mathbf{R}) \sum_i \mathbb{1}_\omega(\mathbf{R}_i)\, \mathrm{d}\mathbf{R} = \mathrm{Exp}\left[\sum_i \mathbb{1}_\omega(\mathbf{R}_i)\right], \end{aligned} \tag{7.3}$$

where $\mathbb{1}_\omega(\mathbf{r})$ is unity when $\mathbf{r} \in \omega$, and zero otherwise. For sufficiently small bins, where the density has small variations over the bin's extent, the density at some point $\mathbf{r} \in \omega_j$ is approximated by the mean density $n(\mathbf{r}) = |\omega_j|^{-1} N(\omega_j)$. Thus the binning estimator for the density of bin j is

$$\check{n}_j(\mathbf{R}) = |\omega_j|^{-1} \sum_i \mathbb{1}_{\omega_j}(\mathbf{R}_i) \tag{7.4}$$

Evaluating the expectation of the density value with respect to the probability density f yields the estimator of the density at bin j,

$$\begin{aligned} n_j &= \int f(\mathbf{R}) \check{n}_j(\mathbf{R})\, \mathrm{d}\mathbf{R} = \\ &= |\omega_j|^{-1} \sum_k \lambda_k \int \delta\left(\mathbf{R}-\mathbf{R}_k\right) \sum_i \mathbb{1}_{\omega_j}(\mathbf{R}_i)\, \mathrm{d}\mathbf{R} = \\ &= |\omega_j|^{-1} \sum_{ki} \lambda_k \mathbb{1}_{\omega_j}(\mathbf{R}_{ki}). \end{aligned} \tag{7.5}$$

The coordinate $\mathbf{R}_{ki}$ is the ith electron coordinate from walker k. To accumulate the density on the entire domain, this results in the following scheme. For every electron i of every walker k, we map the electron's coordinate $\mathbf{R}_{ki}$ to its corresponding bin ω_j and increase the bin's density by the walker's weight over the bin's volume, $|\omega_j|^{-1}\lambda_k$ and repeat this for every time step.
We choose our bins ω_j to align with the prolate spheroidal (PS) grid cells, sketched in Fig. E.2. This has several advantages, as discussed in the next section.

7.2. Density fluctuations

> The following section is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

The density from any QMC approach comes with a statistical error. With the binning method, the density not only depends on the realization of the walkers' random walk and its inherent fluctuations, but also on the volume of the bins. From Eq. (7.5), electrons visiting small bins cause a greater change in the density than electrons in larger bins, which reflects in the variance of the density. The variance of $\check{n}_j$ with respect to a realization of f reads

$$\begin{aligned} \mathrm{Var}\left[\check{n}_j\right] &= \int f(\mathbf{R}) \check{n}_j^2(\mathbf{R})\, \mathrm{d}\mathbf{R} - n_j^2 = \\ &= |\omega_j|^{-2} \sum_k \lambda_k \sum_{i,i'} \mathbb{1}_{\omega_j}(\mathbf{R}_{ki}) \mathbb{1}_{\omega_j}(\mathbf{R}_{ki'}) - n_j^2 \end{aligned} \tag{7.6}$$

By splitting the double sum into diagonal and off-diagonal parts, we arrive at

$$\mathrm{Var}\left[\check{n}_j\right] = n_j \left(|\omega_j|^{-1} - n_j\right) + \rho_j \tag{7.7}$$

where

$$\rho_j = |\omega_j|^{-2} \sum_k \lambda_k \sum_{i \neq i'} \mathbb{1}_{\omega_j}(\mathbf{R}_{ki}) \mathbb{1}_{\omega_j}(\mathbf{R}_{ki'}). \tag{7.8}$$

The first term of Eq. (7.7) resembles the typical Bernoulli variance of $p(1-p)$, modified to account for the bin size. The additional term ρ_j represents the variance caused by statistical and electronic correlations. Its only contributions arise from two electrons of the same walker being in the same bin, and thus links the variance of the one-body density to the two-body density. The $|\omega_j|^{-2}$ scaling of the variance is apparent from Eqs. (7.5) and (7.7). Thus smaller bins result in higher fluctuations.

Near a focal point in the PS grid the bin size becomes small enough to induce larger fluctuations on a small scale, however, the vicinity of the nuclei is a region of high density and great statistical significance. To the contrary, with few walkers visiting the regions far from the nuclei, the increased bin volume counteracts the lack of statistical significance. Using the PS grid cells as bins, the number of walkers visiting an individual bin does not needed to span the same range of orders as the density, greatly lowering the fluctuations compared to bins of constant volume. This property of the variance of $\check{n}_j$ reflects in the statistical error of its expectation value n_j.

In Sec. G, we propose a conceptually simple a posteriori refinement procedure based on extrapolation of the density to remedy the fluctuations in the notorious domains of the QMC density. We will refer to those regularized or refined DMC (VMC) densities as rDMC (rVMC) densities. If not explicitly mentioned, we always report regularized densities.

7.3. Density propagation technique

The binning method offers a simple method to calculate the one-body density with spatial resolution. The random nature of the density estimation of the above sections is, however, governed by large fluctuations. For every walker of N electrons accumulated to the density, N weighted, independent Dirac distribution are added to the density field. Using kernel density estimation, every inserted electron contributes with a smooth, or at least controlled, whole function to the density. This renders densities from kernel estimation much more regular and well-behaved. However, calculating the density with an arbitrary kernel is an additional approximation, since the kernel is not justified from within QMC.

Upon second thought, however, there is a reformulation of the density estimator very similar to kernel estimation, but based on the fundamentals of the DMC approach. Recall, that the equilibrium state is not altered by a propagation step.

$$f(\mathbf{R}', \tau + \Delta\tau) = \int G(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau) f(\mathbf{R}, \tau) \, \mathrm{d}\mathbf{R} \tag{7.9}$$

For the sake of notation an ensemble of a single walker is considered. With the Dirac representation for the distribution at time τ,

$$f(\mathbf{R}, \tau) = \lambda(\tau) \delta(\mathbf{R} - \mathbf{R}(\tau)), \tag{7.10}$$

the propagated distribution at time $\tau + \Delta\tau$ reads

$$f(\mathbf{R}', \tau + \Delta\tau) = \lambda(\tau) G(\mathbf{R}' \leftarrow \mathbf{R}(\tau), \Delta\tau). \tag{7.11}$$

Then, the mixed-estimator density from Eq. (7.2) is given by

$$\begin{aligned} n(\mathbf{r}) &= \int f(\mathbf{R}', \tau + \Delta\tau) \hat{\mathrm{n}}(\mathbf{r}, \mathbf{R}') \, \mathrm{d}\mathbf{R}' \\ &= \lambda(\tau) \sum_i \int G(\mathbf{R}' \leftarrow \mathbf{R}(\tau), \Delta\tau) \delta(\mathbf{r} - \mathbf{R}'_i) \, \mathrm{d}\mathbf{R}'. \end{aligned} \tag{7.12}$$

Hence, the density is given by a smooth function, or at least a function of the same regularity as the Green function G. The Dirac character of the density operator $\hat{n}(\mathbf{r})$, however, does only collapse a single integral of the N integrals involved in the above equations. For Green functions, that move electron by electron, and thus can be cast as

$$G(\mathbf{R}' \leftarrow \mathbf{R}, \Delta\tau) = g_i(\mathbf{R}'_i \leftarrow \mathbf{R}) Q(\mathbf{R}_{1:N\setminus i} \leftarrow \mathbf{R}), \tag{7.13}$$

or equivalently, separate into a marginal transition kernel for some electron coordinate $\mathbf{R}_i$ and some remainder marginal transition kernel for the remaining coordinates $\mathbf{R}_{1:N\setminus i}$. Then, the above density can be evaluated.

$$\begin{aligned} n(\mathbf{r}) &= \lambda(\tau) \sum_i \int g(\mathbf{R}'_i \leftarrow \mathbf{R}(\tau)) Q(\mathbf{R}'_{1:N\setminus i} \leftarrow \mathbf{R}(\tau)) \delta(\mathbf{r} - \mathbf{R}'_i) \, \mathrm{d}\mathbf{R}' \\ &= \lambda(\tau) \sum_i g_i(\mathbf{r} \leftarrow \mathbf{R}(\tau)) \end{aligned} \tag{7.14}$$

With the marginal distribution Q integrated over the variables $\mathbf{R}_{1:N\setminus i}$ and its normalization as a transition kernel, the contributions of Q vanish from the integral. The index i on the marginal transition probability g_i only denotes, which of the electrons in the walker to move. The density above originates from a single walker, yet is given by a sum of functions, that are defined on a larger domain. The contributions $g_i(\mathbf{r} \leftarrow \mathbf{R}(\tau))$ are usually not single or multiple Dirac distributions, but regular functions. If a walker is accumulated to the density, every electron i is considered separately. For every position $\mathbf{r}$, at which the density should be calculated, the Green function is evaluated for the transition, in which electron i moves from $\mathbf{R}_i(\tau)$ to $\mathbf{r}$, while all other electrons are fixed at the positions $\mathbf{R}_{1:N\setminus i}(\tau)$. For an ensemble the density is straightforwardly extended to

$$n(\mathbf{r}) = \sum_{ki} \lambda_k(\tau) g_i(\mathbf{r} \leftarrow \mathbf{R}_k(\tau)). \tag{7.15}$$

The above estimator is called the propagating estimator in the following.

This method allows to estimate the density not only inside bins, but at distinct locations. However, the density can be discretized as in Sec. 7.1 to compare the results with the binning technique. To this end, define the binned density of bin ω_j with $\mathbf{r}_j \in \omega_j$ as

$$n_j = |\omega_j|^{-1} \sum_{ki} \lambda_k(\tau) g_i(\mathbf{r}_j \leftarrow \mathbf{R}_k(\tau)). \tag{7.16}$$

Comparing the above and Eq. (7.5) the true nature of the propagating estimator is apparent. Instead of the binning estimator $\mathbb{1}_{\omega_j}(\mathbf{R}_{ki})$, which only locates and counts the electrons in their respective bins, the propagating estimator propagates the Dirac distribution further in time to arrive at a continuous representation of every electron's contribution, $g_i(\mathbf{r}_j \leftarrow \mathbf{R}_k(\tau))$, to the density. The propagating estimator is guaranteed to have a smaller variance, compared to the Bernulli-type binning. In principle any Green function can be used to estimate the density, that is separable in the above sense of Eq. (7.13). This also allows for different Green functions for the density accumulation and ensemble propagation.

However, approximate Green functions have to be corrected with the MH acceptance probability. This acceptance step, as well as the branching contribution to the Green function in DMC complicate the application of the propagating method. Not only has the acceptance probability of any walker to be calculated for every bin j. Additionally the source term including the local energy is required, if the Green function should capture the branching. This requirements manifest in a large computational effort. The branching, and even the acceptance probability, could, of course, be omitted. However, the behavior of the resulting errors have yet to be examined. Note that the transition kernel including the MH acceptance step of Eq. (C.22) is itself singular for rejections. While the transitions to every other location can be calculated by $g_i(\mathbf{r}_j \leftarrow \mathbf{R}_k(\tau))$, the rejection probability for $\mathbf{r}_j = \mathbf{R}_{ki}(\tau)$ has to be calculated by Eq. (C.23). Therefore, a discontinuous contribution to the density enter through the backdoor of the rejection probability. With this consideration in mind, the propagating estimator is not entirely smooth. However, it is significantly less noisy compared to the binning estimation. With the binning estimator the contributions of a walker k to the density originate solely from the electron coordinates of $\mathbf{R}_k(\tau)$, while the propagated distribution of the walker contributes to the density in a broad region around the electron coordinates. Only the rejection part of the propagated distribution results in a discontinuous contribution at the non-propagated electron coordinates of $\mathbf{R}_k(\tau)$, while it has a significantly reduced magnitude. A promising advantage of the propagating estimator is that exact properties of the density are respected. Suppose the Green function accurately describes the electron-nucleus cusp. Since the propagating estimator inherits the characteristics of the Green function, it also exhibits an accurate cusp. Compared to the binning method, where an accurate cusp requires the accumulation of usually several billions of walkers, with the propagating estimator an accurate cusp is given with a few walkers already. The propagating estimator thus also directly benefits from the development of more accurate Green functions.

The proposal part of the Green function is usually a function that allows fast evaluation. Therefore, it can efficiently pivot the regions of space or the set bins for low probability. Then, truncation schemes can be deviced to only evaluate the acceptance probability or the source term for locations with probabilities above a certain threshold. This reduces the computational effort for regions that are unlikely to more than marginally contribute to the density.

To the best of the authors knowledge, a propagating estimator has not yet been mentioned in the literature. Although first attempts to implement the above estimator were made in the preparation of this thesis, and a proof of the concept was successful, some practical aspects regarding possible biases and efficiency of the propagating estimator are not yet fully understood. Further research is required to understand the resulting density estimates and to benchmark the computational efficiency of the new estimator. The concept of the propagating estimator is not limited to QMC applications but applies to general density estimation in MC applications.

8. Inversion theory

In DFT, the central approximation that almost solely determines the quality of the results besides numerical accuracy is the exchange-correlation energy functional $E_{\mathrm{xc}}[n]$ from Eq.(3.11). The true xc functional would allow for accurately calculating the energy and density of an interacting system with the computational effort of calculating a non-interacting system. To guide the development of xc functionals, comparison with quasi-exact methods is a traditional concept. Comparison of energy values, i.e., DFT energies with true ground state energies from highly accurate quantum chemistry methods, is the most apparent benchmark of an xc functional and offers some indication of the deficits of the functional. However, there is little information to compare, and the qualities that characterize a good xc functional remain shadowed. Unfortunately, the true xc energy as a density functional is almost impossible to calculate numerically since it incorporates the response to any density variation. While the xc functional determines the total energy, its functional derivative, the xc potential v_{xc} from Eq. (3.12), has a one-to-one correspondence to the density. The xc potential therefore holds a significant fraction of the information encapsulated in the xc energy functional but offers a richer comparison. The true xc potential can be calculated from a given exact density in a process called inversion. Since the xc potential is spatially resolved, the qualitative characteristics and even quantitative values of an approximated xc potential can be evaluated against the true inverted xc potential. For this reason, inversion has a long history and marked prominent changes in DFT development [99–115].

With quasi-exact densities from the QMC method, density inversion is the natural next step to better understand and construct density functional approximations.

> The following chapter is an excerpt from the author's publication [116] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

8.1. The Kohn-Sham map

The Hohenberg-Kohn theorem (HKT) [14] establishes a one-to-one correspondence between a density n and a potential v up to a constant. Therefore any KS potential directly leads to a unique density and, in principle, any density must also lead to a unique (up to a constant) KS potential. For the sake of clarity we omit spin indices here and in the following whenever it is reasonable, even though some expressions might genuinely be spin-resolved. Since the Hartree and the external potential (given the particle number) have their own one-to-one mapping with the density in finite systems, the potential-density-map does not only connect the KS potential with the density, but also the xc potential itself. Thus, we use the symbols for forward and inverse maps for the KS as well as the xc potentials. The (forward) map from potential to density is numerically rather simple, requiring the solution of the KS equations for all orbitals contributing to the density,

$$F[v_{\mathrm{xc}}] = n. \tag{8.1}$$

The inverse map

$$I[n] = v_{\mathrm{xc}}, \tag{8.2}$$

although guaranteed to exist and be unique by the HKT, is a much more intricate and difficult to realize mapping. It belongs into the class of inverse problems [117] and is ill-posed [118], where well-posedness requires the existence of a solution, its uniqueness, and the continuous dependence on the initial conditions. A general approach to inverse problems is based on regularization strategies [117], i.e. some well-chosen transformation of the input data to arrive at a well-posed problem or at least render the inverse problem less ill-posed. With too little regularization, the ill-posedness is still present, while too much regularization puts an unnecessary bias onto the inverted potential. Between these limiting cases, there usually is an optimal range for the regularization and often much effort has to be spend to find it. Determining the corresponding potential for a given density (or given spin-densities) is a challenging task. On the other hand, the quality of any candidate potential can readily be evaluated by its density's difference to the target density, since the forward map is realized with high accuracy at low computational cost. HKT guarantees the one-to-one correspondence mathematically, based on the variational principle by an energy difference. In the proof of the HKT one exploits that analytically one can identify an infinitesimal amount of energy. In the numerical picture the guaranteed one-to-one correspondence becomes volatile to the point where is breaks down, because all real numerics works with a limited precision. A simple, well-known example for the non-uniqueness of the potential in a finite-precision numerical realization is the following. Take any molecular potential-density pair and add any positive function of small spatial extent to the original potential in the far asymptotic region. The additional repulsive potential does not alter the density in a numerically noticeable way, because in the far asymptotic region, the density already has decayed to a number that is negligible in a calculation with finite numerical precision. Thus, the density corresponding to the manipulated potential will be identical to the original density within numerical accuracy, disproving the uniqueness of the potential to density in a numerical representation. Analytically, however, the manipulated potential is linked to a new density that differs from the original density, even if only ever so slightly. Revisiting the proof of the HKT with the intention to allow for differing (spin-)densities and KS-Hamiltonians, yields insight to the degree of uniqueness the HKT guarantees. With this strong form of the HKT [119–121], our aim is to probe the environment around the potential and its corresponding density and make the consequences of differing potentials with similar densities more transparent. For any two KS potentials that share the same external potential, $v_{\mathrm{KS}\sigma}$ and $v'_{\mathrm{KS}\sigma}$, and their corresponding densities, n_σ and n'_σ, the argument of the HKT is based on the quantity

$$J = \sum_\sigma \int \left(v_{\mathrm{KS}\sigma} - v'_{\mathrm{KS}\sigma}\right)\left(n_\sigma - n'_\sigma\right) \leq 0. \tag{8.3}$$

Here and in the following, integrals without explicit notation of the variables that are being integrated over denote volume integrations over the functions' full domain. If the two potentials are not equal, $v_{\mathrm{KS}\sigma} \neq v'_{\mathrm{KS}\sigma}$, then J has to be strictly negative, $J < 0$, due to the variational principle [14, 119]. Identical densities would contradict a negative J, hence $n_\sigma \neq n'_\sigma$, uniqueness is guaranteed, and the HKT follows. When the density (or potential) of one density-potential pair approaches the density (or potential) of the other pair, the absolute value of J decreases until it vanishes in the limiting case of identical densities (or potentials up to a constant). J is at least quadratic in the density-potential variation, since a linear variation in density (or potential) induces a linear variation in the potential (or density). Eq. (8.3) allows to quantify the limitations on the numerical realization of the uniqueness of the potential-density map. While the proof of the HKT starts from a given potential, J treats density and potential equally and allows to assess the consequences of small differences in the density. With the numerical resolution restricted to a small positive number ϵ, representing the floating point precision, one has $J < -\epsilon$. In this line of thought, any two numerical densities with $J \in [-\epsilon, 0]$ cannot be distinguished from each other. While ϵ might be very small, the exponentially decaying asymptotic tail of the densities difference in Eq. (8.3) allows for significant variations in their corresponding potentials.

Splitting the KS potential in is constituents, allowing the external potential to cancel and carrying out spin summation in the Hartree part, the fraction of J corresponding to the xc potential is extracted,

$$\begin{aligned} J = & \int \left(v_{\mathrm{H}} - v'_{\mathrm{H}}\right)\left(n - n'\right) \\ & + \sum_{\sigma} \int \left(v_{\mathrm{xc}\sigma} - v'_{\mathrm{xc}\sigma}\right)\left(n_{\sigma} - n'_{\sigma}\right) \\ = & J_{\mathrm{H}} + \sum_{\sigma} J_{\mathrm{xc}\sigma}. \end{aligned} \tag{8.4}$$

The Hartree contribution is a positive quantity $J_{\mathrm{H}} > 0$, which can be seen by the relation $-4\pi n = \nabla^2 v_{\mathrm{H}}$ and the first Green identity. J_{H} plays a role in many inversion methods and has repeatedly been identified [101, 122–124]. The above arguments also transfer to J_{xc} and show that xc potentials for densities that agree within numerical resolution can differ significantly, e.g., in regions of low density.

A weaker form of ambiguity of potential-to-density map stems from the Coulomb singularity at the nuclei. In the presence of an $Z/|\mathbf{r} - \boldsymbol{\xi}|$ divergence for a nucleus at $\boldsymbol{\xi}$ with charge Z, the value of the xc potential at the nuclei becomes irrelevant. Although those points build a null set, the Coulomb divergence dominates in close vicinity of the nuclei and renders the values of the xc potential less important in those regions. Thus, the forward map is rather insensitive to variations of the potential at the nuclei. For example, the xc potential associated with Generalized Gradient Approximations (GGAs) typically diverges at the nuclei, yet GGAs can successfully predict energies and other observables. However, the same consideration for the inverse map yields a very different picture: A small variation of the density at the nuclei typically results in a very large variation in the potential. In order for the density to correspond to a molecular Hamiltonian, it has to satisfy the cusp condition,

$$\frac{\mathrm{d}}{\mathrm{d}|\boldsymbol{\epsilon}|} \ln \hat{n}(|\boldsymbol{\epsilon}|) = -2Z \tag{8.5}$$

with a separation vector $\boldsymbol{\epsilon} = \mathbf{r} - \boldsymbol{\xi}$ and $|\boldsymbol{\epsilon}|$ the distance between the nucleus and some coordinate $\mathbf{r}$, where $\hat{n}$ is the density spherically-averaged with respect to the nucleus' position. This fixes the density to an exponentially decaying form, with exponent proportional to Z. If the density does obey the cusp condition, but with a slightly different exponent $\tilde{Z} \neq Z$, the xc potential has to account for this deviation, as the external potential has a fixed singularity $Z/|\boldsymbol{\epsilon}|$. As a consequence, the xc potential will develop a singularity of the type $(\tilde{Z} - Z)/|\boldsymbol{\epsilon}|$. Thus, even for tiny deviations $\tilde{Z} - Z \neq 0$, the spatial form of the inverted xc potential will be massively affected: Small variations of the density yield diverging differences in the potential. The inversion is thus ill-posed: While the forward map is straight forward, the inverse map is sensitive to small changes in the density and many candidate potentials produce densities very close to the reference density.

Numerically realizing an inversion therefore requires one to handle the ambiguity in the potentials as good as possible. With ambiguity there is no predetermination of some characteristics of the xc potential. This seemingly unfavorable aspect of inversion can be reinterpreted as an advantage because it ultimately allows the user to select those characteristics. This is part of the regularization that is necessary to handle inverse problems. Since the regularization influences the inversion outcome, inversion schemes have to be chosen carefully. Fig. 8.1 illustrates how different inversion methods that start from the same density can yields different inverted potentials. For example, the asymptotic behavior of the xc potential is known to be $-1/|\mathbf{r}|$ from analytic considerations [44, 125], therefore this exact property can be incorporated into the inverted potential manually. This is often realized with the Fermi-Amaldi potential [126, 127]

$$v_{\mathrm{xc}}^{\mathrm{FA}} = -N^{-1} v_{\mathrm{H}} \tag{8.6}$$

or other prototype xc potentials with correct asymptotic decay.

The ambiguity in the domain of the potentials significantly simplifies the development of E_{xc} functionals. While tailored variations in the xc energy functional result in desired variations in the energy (and less considered variations in the xc potential), the density might remain almost unchanged. This offers a path to lower the functional-driven error [128] of a density functional approximation, without having to worry too much whether one might increase the density-driven error.

With a substantial part of the mathematical rigor of the HKT hidden in numerically inaccessible features of the density, numerical realization of the inverse map may be beyond reach in many cases. With these limitations in mind, let us examine the use case of inversions, namely the generation of xc potentials from high quality densities. Most of the time, such densities originate from a method other than DFT, i.e. coupled cluster, Møeller-Plesset perturbation theory, configuration interaction or QMC methods to only name a few. Those densities are never the true density, but might be a very reasonable approximation to it for two reasons. In any methods to calculate accurate densities there remains some conceptual convergence error. These can be the truncation of an expansion in Slater determinants or the fixed-node approximation in DMC, for example. Another error is given by the numerical representation, most often due but not limited to a finite basis set. Any inverted xc potential can only refer to the (high-accuracy but approximate) numerical density, which will be called the reference density from here on. While the errors in the reference densities might be negligible for almost all applications and especially for visual comparison, inverse problems are pathologically susceptible to numerical errors of the input data [117], i.e. the reference density. We will discuss this in detail for the case of the statistical fluctuations in DMC densities. In order to reproduce the slight inaccuracies of the reference density, the inverted potential tends to build up significant spurious features, in some cases to an extent where the inverted potential is useless.

In addition to the inverse problem being ill-posed, the v-representability problem can also prevent one from being able to solve the inversion problem. Even when the exact interacting density is non-interacting v-representable, its numerical representation might not be. In such a case, all of the criteria for well-posedness, i.e., existence, uniqueness, and stability of the solution, might be violated. The same applies if the exact density itself is not non-interacting v-representable. In the following, however, we always assume that the densities that we feed into the inversion are non-interacting v-representable.

8.2. Inversion techniques

The problem of inverting the KS map has been addressed many times and with various methods. Wang and Parr [99] were amongst the first to compose an iterative scheme, that updates an initial potential with feedback from the difference of the current and the reference density. Zhao et al. [101] devised an extrapolation scheme to yield an inverted potential in the limit that the Hartree potential of the reference and the inverted density be identical, i.e. $J_H = 0$. The work of Ref. [129, 130] retrieves the inverted potential from the comparison of a suitable DFT energy density with the corresponding wave function energy density. Many more approaches to the inverse problem exist [99–114].

In the following we present an iterative scheme to improve some initially guessed xc potential by repeated application of an update rule until convergence. While our reference density is

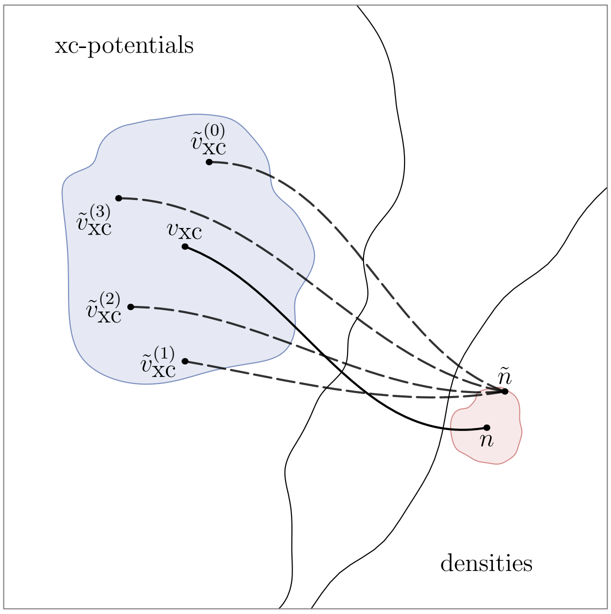


Figure 8.1.: Ambiguity: A schematic sketch of the mapping from xc potential space to density space. The true map from the potential v_{xc} to the density n is indicated by the solid line. Any inversion (dashed lines) starts from the numerical approximation or representation $\tilde{n}$ of the true density n. Using different inversion methods, the approximate density $\tilde{n}$ maps back to different inverted potentials. While the differences in the density domain might appear small, the differences between the inverted potentials can be large. In other words, a “small basin” in density space can map to a “large basin” in potential space. Conversely, distinct potentials can result in visually very similar densities. The question arises how to measure or quantify the distinctness of densities in contrast to the distinctness of the corresponding potentials.

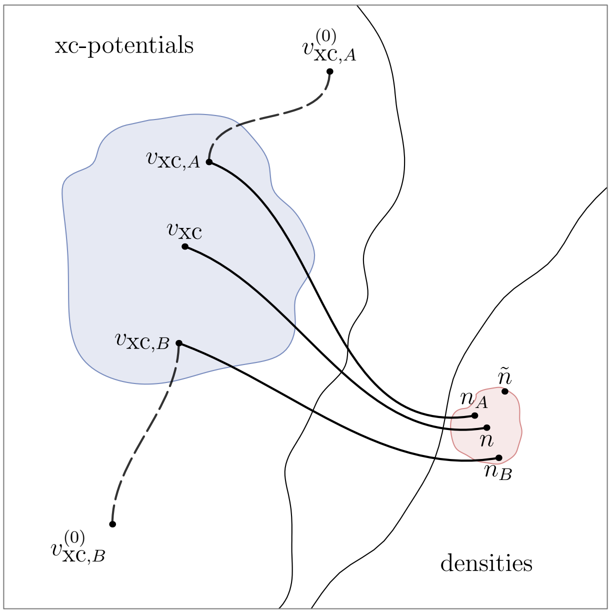


Figure 8.2.: Starting point dependence: In the framework of iterative potential updates, the initial xc potential poses a significant bias to the converged inverted potential. For example, features in the asymptotic region of the target xc potential corresponding to the reference density cannot be uniquely reconstructed.

denoted by n, the KS-density of the current iteration is denoted by $n^{(i)}$ or n' if the precise number of the iteration step does not matter. Any update rule can be written as

$$v_{\mathrm{xc}\sigma}^{(i+1)} = v_{\mathrm{xc}\sigma}^{(i)} + \epsilon \Delta v_{\mathrm{xc}\sigma}[n^{(i)}, n] \tag{8.7}$$

where ϵ is some feedback strength to manipulate the speed of convergence and $\Delta v_{\mathrm{xc}}[n', n]$ aims to improve the potential with information from the current and reference densities, n and n', respectively.

$$v_{\mathrm{xc}}^{(0)} \mapsto n_{\sigma}^{(0)} \mapsto \Delta v_{\mathrm{xc}\sigma}[n^{(0)}, n] \mapsto v_{\mathrm{xc}\sigma}^{(1)} \mapsto \dots \tag{8.8}$$

Any iterative update scheme has an intrinsic starting point dependence via the initial xc potential $v_{\mathrm{xc}}^{(0)}$. Two different starting potentials can converge to different self-consistent inverted potentials with an almost identical density, see Fig. 8.2. Since the inverted potential is almost invariant under changes of potential in low-density regions, as discussed in Sec. 8.1, the asymptotic behavior of the initial density is practically left unaltered throughout the iterations. Solving the inverse problem with a method based on update rules has an important technical advantage. Most update rules fit into existing DFT codes with little effort, as the update of the xc potential in the inverse problem is nothing more than the update of the xc potential in a self-consistent forward problem. This is applicable to update rules that fall into the class of LDAs, GGAs, and even meta-GGAs, where we denote an update rule as “falling into the LDA, GGA, or meta-GGA class” when the update procedure requires only knowledge of the density, or density and gradient, or in addition also the kinetic energy density, respectively.

The major quantity of the update rule is the update functional $\Delta v_{\mathrm{xc}}[n', n]$. It should provide the reasonable corrections to the potential in order to change the current density n' in such a way that it gets closer to the reference density n. The optimal update functional is the well-known inverse response function χ^{-1}, which is generally not known as an explicit functional of the density

$$\delta v(\mathbf{r}) = \int \chi^{-1}(\mathbf{r}, \mathbf{r}') \delta n(\mathbf{r}') \, \mathrm{d}\mathbf{r}'. \tag{8.9}$$

A successful update functional needs to satisfy several challenging requirement. First, the density spans a wide range of values, often over several orders of magnitude, while the xc potential has a very limited range of values (confined in the interval $[-10, 1]$ for the potentials found in this work). Second, the update functional has to convert the input densities to physically reasonable potential updates. For efficient updates this requires the update functional to have some incorporated knowledge of the non-locality of the KS equations, i.e. how a variation of the potential alters the density. The more of this knowledge is implemented into an update rule, the better can it predict reasonable potentials.

A common idea [120, 131] is the a priori determination of the direction of the update $\Delta v_{\mathrm{xc}}[n', n]$. Imagine at some point $\mathbf{r}$ the current density $n'(\mathbf{r})$ is less than reference density $n(\mathbf{r})$. For the densities to align, the update to the current potential should be negative, since a more negative potential attracts more density, and thus increases $n'(\mathbf{r})$. If Δv_{xc} has this property globally, i.e. $\Delta v_{\mathrm{xc}}[n', n](\mathbf{r}) < 0$ when $n'(\mathbf{r}) < n(\mathbf{r})$ and vice versa for all $\mathbf{r}$, we call Δv_{xc} sign-consistent. Sign-consistency is no necessary requirement for an update rule to be successful, since the KS equations themselves and the normalization of the density are non-local operations. Thus, even if the update rule is sign-inconsistent, the KS scheme could (by the semi-local Laplacian, the convoluting non-local Hartree potential and normalization) decrease the density in the next iteration at some point despite the potential was decreased at the same point. Usually (in our experience) the sign-consistency helps to accelerate convergence and update rules, while a large violation of the sign-consistency tends to give less accurate results.

Many choices for update rules can be found in the literature [101, 112, 123, 124, 129–132]. In the course of the research that led to this paper, we have tested various update rules, but in the following we focus on a few ones that we discuss below, because they appeared as particularly advantageous to us.

To start with, the Hartree potential offers a special access to the inversion procedure. Up to a constant, the Hartree potential has a unique relation to the density, since the Poisson equation grants uniqueness for given boundary conditions. If the Hartree potentials of two densities align, the densities must be equal. Since in addition, the Hartree potential readily converts a density to a physical potential, the update

$$\Delta v_{\mathrm{xc}\sigma}^{\mathrm{HAR}} = v_{\mathrm{H}}[n'] - v_{\mathrm{H}}[n] \tag{8.10}$$

is well-known in inversion. The update is connected to the minimization of J_H from Eq. (8.4), where J_H plays the central role of a minimizing functional in the inversion method devised by Zhao et al. [101]. The Hartree update to the potential is arbitrarily smooth [133] due to the integration of the density, i.e. it does not inherit any cusps or numerical artifacts of the reference density. Its non-locality usually accelerates convergence, but simultaneously violates sign-consistency. While the Hartree update is locally sign-consistent in a large portion of the density's domain, at points of rapid density changes, including its nuclear cusps, the smoothing character of the Hartree potential prevents it to resemble those sharp details. The Hartree potential at a nuclear cusp is smooth and of quadratic order with the distance to the nucleus. From a multipole expansion of the Hartree potential, the update has the form of a dipole potential in the asymptotic domain, $\Delta v_{\mathrm{xc}\sigma}^{\mathrm{HAR}} \to \mathbf{p} \cdot \hat{\mathbf{r}}/|\mathbf{r}|^2$ with some vector $\mathbf{p}$. Therefore slight changes in the asymptotic regions are possible with the Hartree update without "overwriting" an already existing $-1/|\mathbf{r}|$ behavior from the initial xc potential. The Hartree potential has no spin dependence as it models electrostatic repulsion from electrons independent of the spin. However, as an update in the inversion for a spin-polarized system, utilization of spin-dependent Hartree potentials, $v_{\mathrm{H}\sigma} = v_{\mathrm{H}}[n_\sigma]$, is advantageous, since the density feedback is resolved in each spin channel. Furthermore, the one-to-one correspondence due to the Poisson equation only holds per spin channel in the spin-polarized case.

Another conceptually very transparent update rule [131] derives from a purely local approximation of the inverse response function.

$$\Delta v_{\mathrm{xc}\sigma}^{\mathrm{DEN}} = n'_\sigma - n_\sigma \tag{8.11}$$

The update to the potential at some point in space stems only from the density difference at this point. This update is sign-consistent and needs to gain any non-local character from repeated updates, that include non-locality from the KS equations into the inverted potential. The way how this update rule handles the asymptotic region in unfortunate: Since $\Delta v_{\mathrm{xc}\sigma}^{\mathrm{DEN}}$ decays exponentially, as does the density, it renders the update unable to alter the potential even in the near-asymptotic regions.

Additionally, the update does not map the wide range of density values to moderate xc potential values, which leads to severe short-comings of the resulting update cycle. Not only is the asymptotic domain of the potential almost unaffected by a DEN update. Additionally at the nuclei the density differences are usually large and do not qualify as updates to the potential. In an all electron calculation, the density is largest at the nuclei and usually also changes its values at the nuclei from iteration to iteration in a self-consistent calculation upon distant variations of the density due to normalization. The density differences in Eq. (8.11) at the nuclei inherit the large magnitude and map it to the potential update. This leads to large spikes in the potential at the nuclei, easily overshooting the optimal update. In the subsequent iteration the overshooting is corrected for with a maybe even larger overshooting in the opposite direction. This leads to oscillations in the iterative scheme and often introduces characteristic wiggles near the nuclei not bound to the close vicinity of the nuclei. To prevent such behavior, the feedback strength ϵ has to be lowered, even if the update would need strong feedback elsewhere, e.g., the asymptotic domain. Modifications to scale the update by multiplication with powers of the distance, $|\mathbf{r}|(n'_\sigma - n_\sigma)$, or in general some fixed local weights, $w(\mathbf{r})(n'_\sigma - n_\sigma)$, to counteract the imbalance from nuclear to asymptotic density values are presented in Ref. [131]. While these in principle can help to improve the update, we have not found sufficient relieve from the above mentioned disadvantages of this method.

To address the above issue, we introduce a weightening of the density difference with the inverse Hartree potential $w = v_{\mathrm{H}}^{-\beta}[n]$ and some $-1 < \beta < \infty$. Due to is smoothness and its strong tendency to follow the magnitude of the density, the Hartree potential is an optimal candidate for a weight function, without the necessity for the user to define the weight function $w(\mathbf{r})$. Note that the Hartree potential incorporates non-locality into the weight. Large density differences, that stem from large absolute density values alone, are moderated by the weightening. However, a large feedback due to a large relative density difference is not reduced by $v_{\mathrm{H}}^{-\beta}$. One could argue that a weightening that accomplishes the same and is yet even simpler is the inverse density $w(\mathbf{r}) = n^{-1}$. But such a weight breaks down in the asymptotic region of molecular densities, where the density decays exponentially, since the resulting potential update $(n'_\sigma - n_\sigma)/n_\sigma$ will eventually diverge exponentially. The potential update resulting from the “density over Hartree” approach is

$$\Delta v_{\mathrm{xc}\sigma}^{\mathrm{DoH}} = \frac{n'_\sigma - n_\sigma}{v_{\mathrm{H}}^{\beta}[n]}. \tag{8.12}$$

The scaling parameter β allows for adaptation of the method, affecting the damping at the nuclei as well as the asymptotic behavior, mimicking the same parameter from Ref. [131]. For the purpose of this work, we only present results for $\beta = 1$.

Another approach to yield reasonable potential updates is to utilize a density functional with potential v_{xc} for the update potential, $\Delta v_{\mathrm{xc}} = v_{\mathrm{xc}}[n'] - v_{\mathrm{xc}}[n]$. Since the xc potential of any reasonable density functional maps densities to potentials, one can utilize them in the inversion process. Unfortunately, many density functionals yield potentials that have spurious features (e.g., GGAs close to the nucleus). Furthermore, it has been argued that some recent developments in DFT have led to better ground-state energies, but did not necessarily lead to improved densities [134]. Therefore, and due to numerical considerations, we restrain to local density approximations here.

$$\Delta v_{\mathrm{xc}\sigma}^{\mathrm{LDA}} = -\left(\mathrm{v}_{\mathrm{x}\sigma}^{\mathrm{LDA}}[n'_\sigma] - \mathrm{v}_{\mathrm{x}\sigma}^{\mathrm{LDA}}[n_\sigma]\right) \tag{8.13}$$

For the sake of transparency we employ the exchange part of the LDA $\mathrm{v}_{\mathrm{x}}^{\mathrm{LDA}} = -A_x n^{\frac{1}{3}}$ in our update rule omitting a correlation potential. The one-third exponent decreases large densities and enhances lower densities, mapping the density scale to potential scales. While the LDA update is strictly local and sign-consistent, oscillations at the nuclei are still possible, even though much smaller compared to the DEN update.

Our final update rule to be presented combines the inverse Hartree weight with the LDA update with some adjustments. Apparently the density to the power of one third as in the exchange-only LDA is a good mapping of density values to potential values. In self-consistent LDA calculations the xc potential shows a range of xc potential values that is in good agreement with the values of the true potential. When dividing the LDA potential (differences) by the Hartree potential, this property is lost. Therefore we introduce exponents for the potentials as a parameter and tailor the expression to yield the LDA potential in the case of vanishing exponent. The resulting update reads

$$\frac{\left|\mathrm{v}_{\mathrm{x}\sigma}^{\mathrm{LDA}}[n'_\sigma]\right|^{(1+\beta)} - \left|\mathrm{v}_{\mathrm{x}\sigma}^{\mathrm{LDA}}[n_\sigma]\right|^{(1+\beta)}}{v_{\mathrm{H}}^{\beta}[n]} \tag{8.14}$$

The prefactors involved in $v_{x\sigma}^{\mathrm{LDA}}$, i.e., A_x and the factors related to spin scaling for polarized densities, only introduce a global prefactor for the potential update, that unnecessarily scales with β and forces the user to adapt the update strengths ϵ when varying β. We have found the update rule to be more consistent across different systems without the scaled prefactors, i.e. requiring less adjustment of the user. Since the Hartree potential is usually larger than the xc potential, the above update has the tendency to show globally lower amplitudes for larger β. We therefore also divide the Hartree potential in the denominator by the particle number N. The resulting "LDA over Hartree" update is given by

$$\Delta v_{\mathrm{xc}\sigma}^{\mathrm{LoH}} = \frac{{n'_\sigma}^{(1+\beta)/3} - {n_\sigma}^{(1+\beta)/3}}{N^{-\beta} v_{\mathrm{H}}^{\beta}[n]} \tag{8.15}$$

We denote the method by LoH(β), to state the used value for β. Then LoH(0) is equivalent to the LDA-based update rule. Though the LoH(1) update does not directly resemble the DoH update, it corrects the numerators exponents of the density to $2/3$ when using a denominator with a the non-exponentiated Hartree potential. The LoH rule is sign-consistent. Its update damps the density feedback at the nuclei and improves the decay of the update in the asymptotic region by weighting and adjustment of the density values with a power law. The parameter β can be chosen in a relatively large range without affecting the results too much. At the same time the dependence on β originates from the interplay of numerator and denominator and therefore, the effects that changing β has on the update rule's asymptotic behaviour are somewhat subtle and not always straightforward to foresee. We illustrate the update rule in Fig. 8.3, where the updates Δv_{xc} in the course of an inversion of the reference xKLI density (with initial LDA potential and the update rule HAR) at iteration 200 are plotted on the left panels, including several values of β. The panels to the right, show the same for a reference density from QMC subject to statistical fluctuations. The updates Δv_{xc} of course differ in absolute magnitude, therefore we normalize them by division by their largest absolute value. The HAR update rule is not sign-consistent, but follows the sign of the density difference roughly. The sign-consistent rules share nodes with the difference of current and reference density. The update rules HAR, LDA and LoH are able to update the inverted potential in the asymptotic regions compared to the exponentially decaying DEN and DoH. The rule LoH(0), or LDA equivalently, have reduced exponential decay (by a factor of $1/3$ and get magnified in the asymptotic region due to the inverse Hartree weight, which renders it capable of adjusting the asymptotic regions. The LoH rules of varying parameter show that small values for β enforce a preferable scaling in the asymptotic region. Larger values of β have fast exponential decay (with values of $\beta > 3$ even faster than for DEN or DoH) and focus on corrections to the regions around the nuclei. On the other hand, low β map the orders of magnitude of the density to an increasingly low range of potential values, squeezing the orders of magnitude in the density. A reasonable range of β is therefore $[0, 3]$, while values in the interval $]-1, 0]$ still yield a sign consistent update, but tend to become less efficient as they do not take into account the density any more as $\beta \to -1$.

It is possible and in general also a good idea to combine the above update rules instead of using a single one. We use different update rules in parallel, but sequential application is also possible. Since each rule can holds its own feedback strength ϵ, cf. Eq. (8.7), we are free to assign feedback parameters to the ingredient rules separately. This defines a large space of parameters to explore and we explain which choices we took in the following.

Since the Hartree update reaches out to the asymptotic regions, has non-locality included, and has superior smoothness, it is our base ingredient. Experience has shown that a feedback feedback strength for the HAR update of 0.5 yields good results. On top of the HAR rule, we try any other presented update rule. For example, we denote with "HAR+DEN" an inversion in which the HAR rule was used with the DEN rule in parallel. Due to the large values of

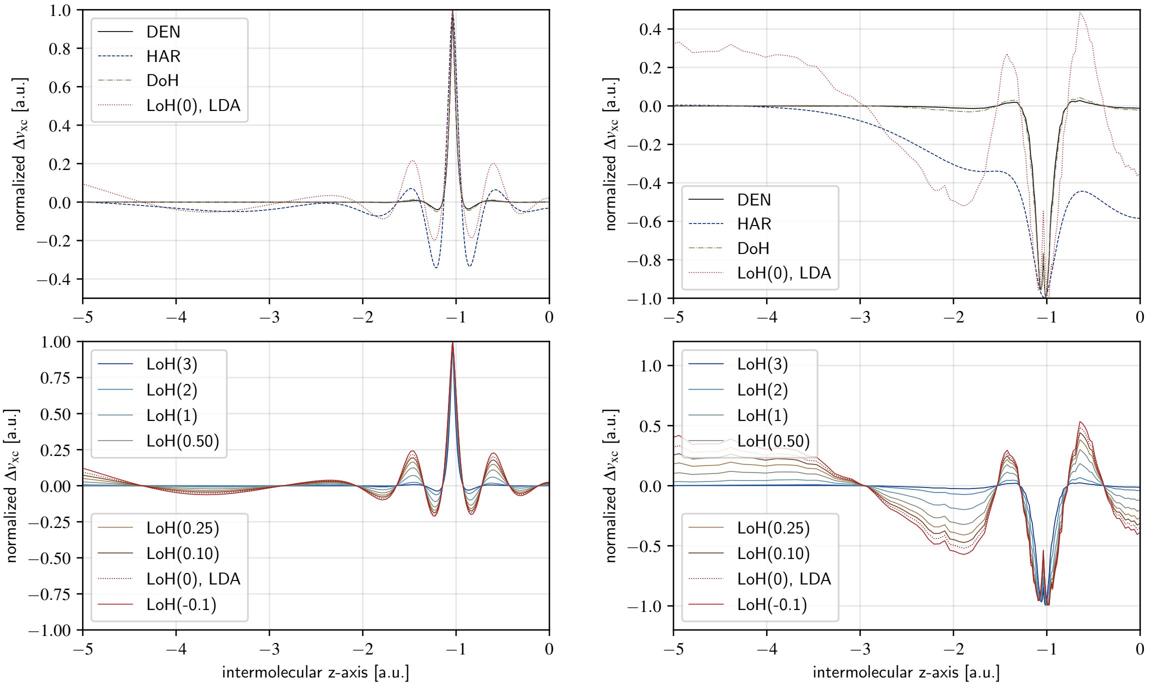


Figure 8.3.: The different update rules of Sec. 8.2 are compared, visualizing the properties associated with the respective update rules. The upper left panel shows the update rules DEN, HAR, DoH and LoH(0), while the lower left panel shows the update rule of LoH(β) for different values of β. The density difference is apparent from the DEN method. The current density n' to evaluate the update rules is taken from the 200th update cycle for the inversion based on the smooth HAR method for N_2. The initial xc potential is the LDA potential from a standard self-consistent DFT calculation. The reference density is the self-consistent density of a DFT calculation of the exact exchange functional within the KLI approximation. The right panels shows the same update rules as the left panels. However, the reference density is the QMC-A density for the nitrogen molecule and the initial potential is the Fermi-Amaldi potential. Therefore, the right panels visualize the effects of statistical noise in the reference density. The update potentials Δv_{xc} have been normalized to their maximum value on the plotted interval for easier visual comparability.

density differences of the other rules we choose a feedback strength of 0.1 for every rule but HAR. Note that by the combination of update rules, their effects on the potential could cancel each other, slowing or even preventing convergence, but if feedback strengths and rules are chosen carefully the opposite is more likely, i.e. combining the advantages of multiple rules. E.g. the label HAR+LoH(0.5) corresponds to an update rule of $\Delta v_{\mathrm{xc}\sigma} = 0.5 \times \Delta v^{\mathrm{HAR}}_{\mathrm{xc}\sigma} + 0.1 \times \Delta v^{\mathrm{LoH}}_{\mathrm{xc}\sigma}$ with $\beta = 0.5$. While we have defined the update rule LDA in Eq. (8.13), we do not report any results for this rule, since it is equivalent to LoH(0) up to a constant factor.

This concludes our presentation of the the update rules and their specific properties. In Sec. I we address further technical details of the iterative scheme. The iterative scheme for the inversion consists of two major choices: the initial xc potential and the update rule. Both affect the converged inverted potential and thus might be seen as biases or constituents of the regularization strategy. A third choice are the implementation details of the inversion procedure. While they are more likely to decide about the stability of the inversion, i.e. if and how fast convergence is achieved, they have less impact on the actual converged potential. Note that our selection of update rules is tailored to the inversion of QMC densities, as explained in Sec. 10.3, and special care was taken to ensure compatibility with statistical fluctuations in the reference density. In general, a smoother reference density, e.g., one that admits derivatives, allows for more sophisticated update rules, for example those of Refs. [112, 129, 130, 132]. While with less regular reference densities one might have to restrain to pure HAR updates.

8.3. Measuring inversion quality

It is a known fact from previous work on KS inversion, that significantly different xc potentials map to very similar densities [127, 135]. Consequently, similarity of densities in any metric disqualifies as a measure to quantify closeness for xc potentials. Unfortunately, in the inversion process the given reference density is the only reference point available for comparison and any quality criterion regarding the inverted xc potential must ultimately rely on the comparison of the densities. To this end we utilize two standard norms. The integrated absolute density difference

$$D_1[n', n] = \sum_\sigma \int \left| n'_\sigma - n_\sigma \right| \in [0, 2N] \tag{8.16}$$

and the infinity norm

$$D_\infty[n', n] = \max_{\sigma, \mathbf{r}} \left| n'_\sigma(\mathbf{r}) - n_\sigma(\mathbf{r}) \right| . \tag{8.17}$$

While it most probably is not possible to bypass the issue of similar densities by the choice of a superior metric, we additionally used the Wasserstein or Kantorovich–Rubinstein metric [136, 137] $W_2[n, n']$, that quantifies the Euclidean minimal transport of probability to align the two densities, to measure the closeness beyond points-wise arguments with a measure explicitly tailored to compare probability densities. We calculate the Wasserstein distance with the Python optimal transport library `POT`. Although the xc potentials show a much larger deviation from another, they cannot be compared with the metrics D_1 and D_∞, due to the ambiguous constant offset in the potential. Even if we calculate the above metrics with an offset minimizing the metric, due to the above discussed ambiguity in the asymptotic domains and at the nuclei, spatial comparison is not a reasonable tool to compare potentials. Fortunately, the potential rather directly gives rise to electronic properties such as the eigenvalues, energies, and all observables of the KS Slater determinant. Comparing those indirect observables of the potential filters out the auxiliary properties of the potential, that are not relevant for the actual physical interpretation. The inspection of the KS energy spectrum, is an elegant way to assess how different features in the xc potential affect the orbitals and their energies. For two potentials v and v', compare their respective eigenvalues, $(\varepsilon_{\sigma i})$ and $(\varepsilon'_{\sigma i})$, by the metric

$$\Delta\varepsilon[v, v'] = \frac{1}{N} \sum_{\sigma} \sum_{i=1}^{N_\sigma} \left| (\varepsilon_{\sigma i} - \varepsilon_{\sigma i-1}) - \left(\varepsilon'_{\sigma i} - \varepsilon'_{\sigma i-1}\right) \right| . \tag{8.18}$$

This metric compares the distance from one eigenvalue to the next for the two spectra and is invariant under any potential offset. To make the metric comparable for systems of different numbers of electrons, we divide by the particle number.

To further exploit the eigenvalue spectrum, one could consider the HOMO eigenvalue as a quantity, that is in principle the exact ionization potential, $I = -\varepsilon_{\text{HOMO}}$. With the eigenvalues being sensitive to a constant offset in the potential and therefore being connected to the asymptotic behavior of the xc potential and its high ambiguity, we cannot directly expect the HOMO eigenvalue to be close to the ionization potential. For example, an inversion initialized by the xc potential of the LDA, will never exhibit the true $-1/|\mathbf{r}|$ behavior of the xc potential for $|\mathbf{r}| \to \infty$. Its inverted potential will show some mixture of decay types, stemming from the initial potential in the far asymptotic domain, while in the near asymptotic domain the update rule could change the potential's shape. Thus the potential offset is hard to quantify and therefore its HOMO yields an incorrect ionization potential.

The next metric we discuss can only be applied if the true xc potential corresponding to the reference density is known. For two potential-density pairs (v_{xc}, n) and (v'_{xc}, n') direct evaluation of J_{xc} from Eq. (8.3) is possible.

$$J_{\text{xc}}[v_{\text{xc}}, v'_{\text{xc}}] = \sum_{\sigma} \int \left(v_{\text{xc}\sigma} - v'_{\text{xc}\sigma}\right) \left(n_\sigma - n'_\sigma\right) . \tag{8.19}$$

Evaluating J_{xc} with any current approximating potential-density pair and the true pair indicates how close the pairs approach each other. The metric scales quadratic with the distance of the pairs and reaches zero for identical pairs within the limitations discussed in Sec. 8.1.

We cannot directly apply the metric J_{xc} to a general inversion process, since the true xc potential is of course not known initially. However, we follow [132] and abuse the above scheme of the discriminator J_{xc}, inserting the xc potentials of the last two iterations together with the current density and the reference density to get a spatially resolved convergence indicator

$$\zeta_{\text{xc}\sigma}(\mathbf{r}) = \left(v^{(k+1)}_{\text{xc}\sigma}(\mathbf{r}) - v^{(k)}_{\text{xc}\sigma}(\mathbf{r})\right) \left(n_\sigma(\mathbf{r}) - n^{(k)}_\sigma(\mathbf{r})\right) \tag{8.20}$$

with integrated metrics

$$Z_{\text{xc}} = \sum_{\sigma} \int \zeta_{\text{xc}\sigma} \qquad \text{and} \qquad Z^{\max}_{\text{xc}} = \max_{\sigma,\ \mathbf{r}} \zeta_{\text{xc}\sigma}(\mathbf{r}) \tag{8.21}$$

The inserted new potential and the reference density do not correspond to each other, i.e. the reference density does not stem from the potential $v^{(k+1)}_{\text{xc}\sigma}$. Therefore there are no justifiable bounds for Z_{xc} as $J_{\text{xc}} < 0$. Although Z_{xc} is a convenient number to monitor the inversion iteration, it has some caveats. First a large constant potential offset from one iteration to the next might spoil the interpretation of Z_{xc}, if no measures are introduced to prevent the potential from jumping up and down during the iterative updates. Further, Z_{xc} can become zero even though the density difference does not vanish. This is the case if the inversion cycle reaches a stationary state with $v^{(k+1)}_{\text{xc}\sigma} = v^{(k)}_{\text{xc}\sigma}$ due to limitations of the update rules.

9. QMC results

> The following chapter is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

We have examined the energies and densities for the molecules Li_2, N_2 and the carbon atom. The equilibrium bond lengths are those of Huber and Herzberg [138] rounded to the fourth digit resulting in distances of $R = 5.051$ for Li_2 and $R = 2.0743$ for N_2. The atoms of the molecules are placed at $z = \pm R/2$ in the prolate spheroidal (PS) grid, while the carbon atom is located at $z = 1$ on the z-axis. Following our grid analysis in Sec. F, we choose 181×181 grid cells. Throughout this paper, we use Hartree atomic units unless specified otherwise.

9.1. QMC energies

To quantify the influence of the time-step error, we performed DMC calculations with the time steps $1, 2, 3, 5, 7.5, 10, 15$ and 20×10^{-3} a.u. and extrapolated to zero with a quadratic polynomial. Tab. 9.1 lists the resulting energies. The coefficient of determination of the fit was 0.96, 0.99 and 0.99 for Li_2, N_2 and C, respectively. For Li_2 the time-step error of the energy at a time step of 0.001 a.u. is less than 10^{-5} a.u., while 1.1×10^{-3} a.u. for N_2 and 4×10^{-4} a.u. for C. Our default time step for the density calculations of 0.001 a.u. is therefore within chemical accuracy ($1\text{kcal/mol} \approx 0.0016$ a.u.) of the extrapolated energy, indicating a negligible time-step error of the densities.

Optimization of the Padé Jastrow factor from Eq. (5.20) determines b to be 1.56, 3.60 and 3.20 for Li_2, N_2 and C, respectively. Table 9.1 shows reference energies and the results from this work. It confirms the well known fact that the total energies from LDA are not impressive, and the xKLI energies are close to the Hartree-Fock values. VMC, as expected with our simple trial wave function, improves the DFT results only to about 20% to 50% correlation energy, while Refs. [50, 56] recover up to 80% with a single determinant and high-quality Jastrow factors (not shown in the table). The DMC energies from the real-space grid are in very good agreement with the references based on basis sets, increasing the recovered correlation in all cases but Li_2@xKLI (where @xKLI here and in the following denotes the origin of the orbitals). The improvement ranges from about 0.5 to 0.8% for carbon, 0.1 to 0.6% for Li_2, and 2.3 to 5.7% for N_2.

While LDA- and xKLI-based trial wave functions recover significantly different amounts of correlation in VMC, the DMC energies are remarkably independent of the DFT functional, indicating a strong alignment of their nodal structure. We also found the KS orbitals of the LDA and xKLI to have remarkable spatial similarity (not shown). Although slight deviations in the orbitals could lead to larger differences in their determinant's nodes and thus the DMC energy, this does not seem to be the case in the systems we studied.

The review on single-determinant DMC results for Li_2 of Bressanini et al. [86], where basis sets up to $6s4p2d1f$ were investigated, exemplifies the challenges associated with finding a basis set that reliably represents the nodal structure. Inclusion of some orbital character even increases the DMC energy by decreasing the node quality, while careful optimization of the basis set results in a very accurate DMC energy of $-14.9923(2)$ [86]. Considerable effort must be invested to explore the large parameter space of the basis sets even in the case of a molecule as small as Li_2.

In the real-space grid approach, convergence of the orbitals and thus the nodal surface is very transparent and systematic, and a situation where including "unsuitable" basis functions lowers the nodal quality does not occur. We find a monotonous convergence of the DMC energy as a function of our grid spacing, i.e., there is only one parameter that is optimized. In practice, we do this in seven steps, cf. Chapt. F. It is further interesting to note that while on the one hand the one-determinant VMC energy of -109.4520(5) reported in Ref. [50] for N_2 is noticeably lower than our VMC energy of -109.10316(5) due to the better Jastrow factor that was used in Ref. [50], on the other hand our one-determinant DMC energy using the real-space grid, -109.5183(3), is lower than the literature value for the one-determinant DMC energy [50] of -109.5039(1).

The optimal nodes representable with a single determinant are generally unknown, except for small systems of high symmetry [86]. Taking the wave function of Ref. [86] as a proxy for this limit, the nodal manifold from our real-space grid orbitals is within a third of chemical accuracy to this proxy.

Regarding our treatment of persistent walkers, cf. Sec. 6.14, we note the following: In our DMC calculations we found a ratio of persistent to total walkers of about 1×10^{-9}, 2×10^{-7} and 8×10^{-8} for Li_2, N_2 and C, respectively, with a threshold probability of $P_t = 10^{-7}$. The threshold number of idle steps to identify persistent walkers are $n_{\text{idle}} = 3, 6$ and 5.

Table 9.1.: Energies and percentage of recovered correlation energy E_c from different methods. The referenced calculations in this table use basis set orbitals and employ elaborate Jastrow factors up to three-body terms. Our calculations are spin-polarized with orbitals on a PS grid of dimension 181 × 181, a simple Pade-Jastrow and a single determinant (SD). We use a time step of 0.001 a.u. and extrapolated to zero for the DMC run based on LDA orbitals. The bottom two rows show results for multi-determinantial wave functions (MD). References for the Hartree-Fock energy and the non-relativistic limit were taken from Refs. [139, 140], and Refs. [56, 57], respectively. The percentage values for the correlation energies given here may differ slightly from the ones given in the cited literature due to small differences in the reference values for the Hartree-Fock energies or the non-relativistic limits. For the sake of transparency and easier understanding of the table all results of this work are marked by an asterisk.

	$C(^3\Pi_0)$		$Li_2\left(^1\Sigma_g^+\right)$		$N_2\left(^1\Sigma_g^+\right)$	
	E	$E_c(\%)$	E	$E_c(\%)$	E	$E_c(\%)$
DFT@LDA*	−37.468 540	−144.39	−14.724 422	−118.82	−108.695 827	−54.33
DFT@xKLI*	−37.691 862	0.59	−14.870 580	−0.79	−108.985 178	−1.57
VMC[J+SD]@LDA*	−37.722 50(2)	20.47(2)	−14.913 02(1)	33.48(1)	−109.089 05(8)	17.36(1)
VMC[J+SD]@xKLI*	−37.735 69(2)	29.04(1)	−14.930 64(3)	47.70(3)	−109.103 16(5)	19.94(1)
DMC[J+SD][1]	−37.829 66(4)	90.04(3)	−14.991 67(2)	96.99(2)	−109.5039(1)	93.00(2)
DMC[J+SD]	−37.8295(2) [3]	89.9(1)	−14.9911(1) [2]	96.53(8)	−109.487(1) [2]	89.9(2)
DMC[J+SD]@LDA*	−37.830 37(1)	90.503(8)	−14.991 70(1)	97.02(1)	−109.516 75(3)	95.343(5)
DMC[J+SD]@xKLI*	−37.830 44(1)	90.548(7)	−14.991 43(4)	96.79(3)	−109.516 86(3)	95.361(5)
DMC[J+SD]@LDA, ext.*	−37.830 71(4)	90.72(3)	−14.991 83(9)	97.11(7)	−109.5183(3)	95.62(5)
DMC[J+MD][1]	−37.836 20(1)	94.287(5)	−14.994 56(1)	99.322(8)	−109.5206(1)	96.04(2)
DMC[J+MD]	−37.840 80(6) [3]	97.27(4)	−14.9938(1) [2]	98.71(8)	−109.505(1) [2]	93.2(2)

[1] Toulouse and Umrigar, (2008), Ref. [50]

[2] Filippi and Umrigar, (1996), Ref. [56]

[3] Buendía et al., (2009), Ref. [57]

9.2. QMC densities

The DMC density for Li_2 based on a calculation that started from LDA orbitals is shown in Fig. 9.1. We plot the density on a logarithmic scale to highlight its high resolution and global smoothness. The density accurately shows its distinct features: The nuclear cusps of purely exponential form are present at the Li ions. The asymptotic decay is continuous up to radii of about 12 to 15 a.u. and density values of order 10^{-7}. Beyond this domain the noise slowly starts to increase in the low-density regime - an inevitable feature of MC at low probabilities. Despite the density values ranging from 10^{-10} at boundaries of the simulation box to 10^{1} at the nuclei, the density is smooth throughout the physically relevant domain. Plotting the density with a linear instead of a logarithmic scale or with iso-surfaces could not render any of the mentioned fluctuations. Here, the beneficial effect of the spatially varying size of the PS grid cells comes into play, effectively sampling larger bins in regions of low density and small bin at the nuclei. The σ-type bond is visible between the atoms upon closer inspection, see Fig. 9.2 for a more detailed view of this region.

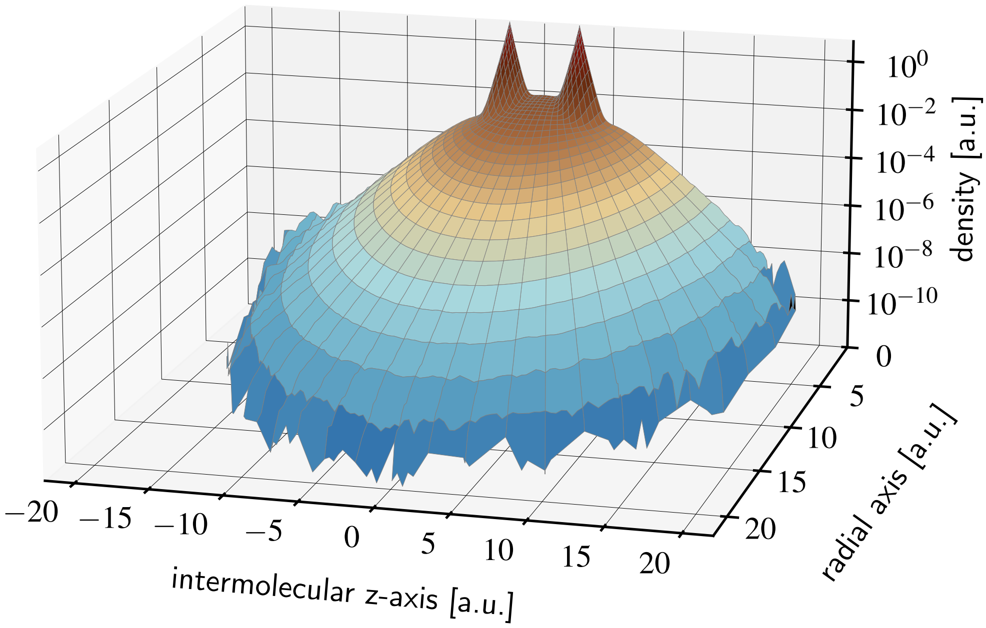


Figure 9.1.: Bare DMC density of Li_2 based on LDA orbitals. The density is smooth up to density values of the order 10^{-7}, rendering the density without significant noise to a radial extent of about 15 a.u. From there on stronger fluctuations build up until the density values are dominated by noise at distances of about 20 a.u. The color reflects the magnitude of the density. Note that the curved (gray) lines indicate straight lines in the PS coordinate system, but do not correspond to the PS grid, which is more fine-grained.

Let us shortly estimate the computational demands to recover a smooth density from QMC. With a total of N_r walker realizations in a QMC calculation, any bin ω is stochastistcally visited by an average of $N_r n|\omega|$ walkers, where n is the average density over the extent of the bin. For instance, bins at which the density of Fig. 9.1 has decayed to 10^{-8} have a volumes of about 0.1 to $1a_0^3$. For those bins to be visited 100 times during the QMC run, one requires a total number of walker realizations of about 10^{10} to 10^{11}. In our calculations of the density we aim for $N_r > 10^{11}$. To make efficient use of our computational infrastructure, 6.4×10^5 walkers are propagated in parallel. The number of propagation steps in our calculations is around 6×10^5. In practice, we reach reasonable computation times by parallelizing the calculations over several hundred cores. Although the serial correlation is greatly reduced by the parallel sampling, we used data blocking with a block length of 25 to reduce the remaining serial correlation in the sample averaged over the parallel walkers.

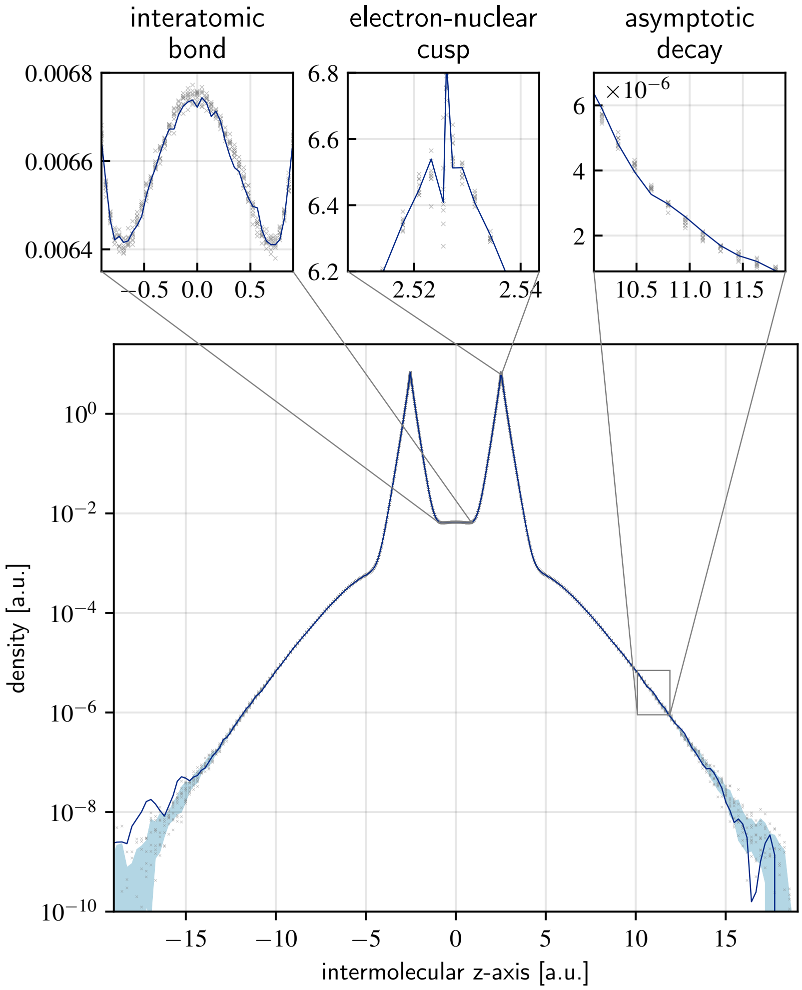


Figure 9.2.: The estimated error of the DMC@LDA density for Li_2. From independent density calculations, we estimated the standard deviation of a single density realization. The dark blue line shows a realization of the density, while the light blue halo region is its standard deviation. Note that the confidence band is the error of a single realization, not of the mean of all calculations. For clarity of visualization we plotted the confidence band around the mean of all density calculations. The gray crosses indicate the sampled densities. While the global plot scales the density logarithmically, the insets do not.

To estimate the statistical error of the DMC density we performed 10 independent calculations of the density. Standard statistics yield the error halo for the DMC density, shown in Fig. 9.2 along the molecular axis. The statistical error of the density in the physically relevant domains is not recognizable by eye. We again chose to show the logarithmic density to visualize the asymptotic realm, where the fluctuations of the density are bound to increase. We heavily zoomed in on the most intricate features of the density: the cusp, the inter-atomic partition and an asymptotic region, where the density is at the edge of actual physical relevance.

As expected, the error increases in the region far from any ion, since only a small fraction of electrons ever visits this region. However, even in the logarithmic representation the error at low density turns out to be quite small, with a relative value of about 6% at a density of 2×10^{-6}. In the entire moderate to high-density region, the relative error becomes negligible, effectively vanishing within the line width.

Between the ions the density takes values of about 6.6×10^{-3} and shows a variation stemming from the σ-bond on the order of 4×10^{-4} (difference from maximum to minimum), making it a challenging region to describe accurately. The density's absolute (statistical) error constitutes about 2×10^{-5}. Therefore, the relative error of the density is in the range of 3×10^{-3}. The density's σ-bond variation is resolved with an error of about 5%.

While increased stochastic fluctuations of the density close to the grid foci, or equivalently the nuclei, are inevitable due to smaller bins, the wiggles in the "cusp" inset are not of statistical origin only. As discussed in Chapt. F, some inaccuracies arise from converting the discrete grid-representation of the orbitals' derivatives to continuous space.

The refinement procedure, presented in Chapt. G, can be used to remedy any of the remaining artifacts discussed above. For example, a systematic error occurs in the immediate vicinity of the nuclei due to the larger variance that results from small bins, and the evaluation of the numerical derivatives, cf. Chapt. G. However, this error is embedded in the much larger domain of exponential decay due to the cusp condition, therefore the proper cusp can readily be restored with the extrapolation scheme that we discuss in detail in the Supplemental Material. This analysis shows that overall the acquired density is suitable not only for visual inspection, but also for subsequent analysis with high numerical requirements, e.g., inversion of the density to its KS potential.

The orbitals that build the trial wave function can have a major influence on the DMC density. With a single-determinant the orbitals alone define the nodal structure and therefore the quality of the DMC ground state. The electronic structure of the carbon atom with four electrons in the up- and two in the down-channel in combination with the orbital nodal line of the highest occupied (molecular) orbital (HOMO), further discussed in Sec. 9.3, make this system challenging to handle. The DFT densities of LDA and xKLI differ significantly from each other, and thus the carbon atom is a well-suited, natural test case for investigating how DMC results depend on the starting orbitals. The DFT and DMC densities (spin-up) are presented in Fig. 9.3 for the carbon atom. The DMC energies from LDA and xKLI are very similar, compare Tab. 9.1. Therefore, one would expect that their DMC densities should also be similar. We find this to be true in the high-density regime. However, the DMC density shows a non-negligible dependence on the functional at lower density, most prominently seen in the insets for the asymptotic decay in Fig. 9.3 at $z \approx 5$ a.u. For xKLI the DMC and DFT densities almost match, which underlines the quality of the DFT-density from xKLI. Since the xKLI approximation incorporates the correct $-1/|\mathbf{r}|$ asymptotic decay of the exact xc potential, it is not surprising that it performs better than LDA, for which the potential decays exponentially. The DMC@LDA density does not fully converge to the DMC@xKLI density in the asymptotic region, but one sees that the features stemming from the LDA are partially corrected for.

The densities show a different peak value at the nucleus, leading to the conclusion that the different methods locally predict different densities. This is a side-effect of the density normalization, since any non-norm-conserving change in the density leads to a global multiplicative scaling in the density (a constant additive offset in the logarithmic plot), and therefore different values at its cusp.

One might be surprised by the noticeable orbital-dependence of the DMC density, given that the DMC energies hardly showed such a dependence. However, the finding becomes understandable at least partially as a consequence of the different character of the observables. While both are evaluated with the mixed estimators, only the density suffers from the mixed estimator bias.

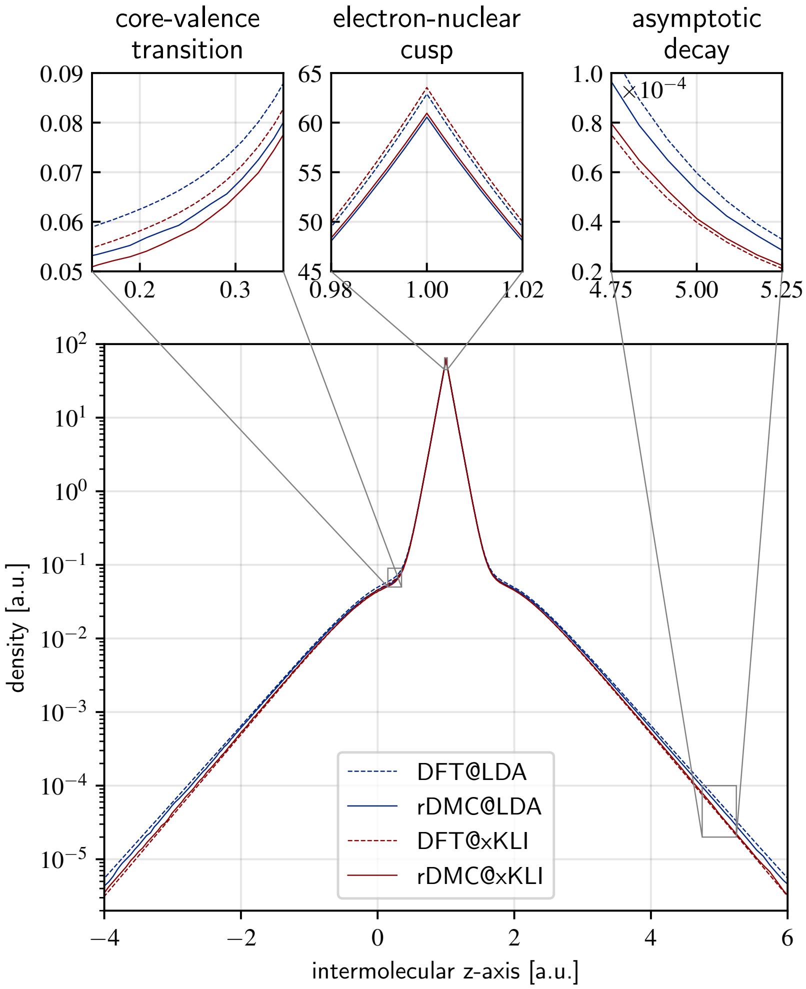


Figure 9.3.: Investigation of the orbital-dependence of DMC densities. We compare the DFT and refined DMC (mixed estimator) spin-up densities for LDA and xKLI orbitals for the carbon atom. Solid lines show the DMC densities, dashed lines their DFT counterpart. Values from LDA and xKLI are colored blue and red, respectively.

Therefore, the DMC densities in Fig. 9.3 have a stronger dependence on the trial wave function, as the DMC process samples $\psi_T\psi_0$, not ψ_0^2. Another reason for why the density shows a more pronounced dependence on the starting orbitals than the energy is the "probability regime" in which these observables are evaluated. While the energy is only marginally affected by regions of low wave function probability, the density is essentially capturing these regions with spatial resolution. Thus, the energy depends only weakly on the quality of the nodal manifold in low probability regions, if at all. The density, however, can discriminate nodal quality differences in those low-probability regions. Therefore, one can find different densities that yield very similar energies.

We have found qualitatively similar results for Li_2 and N_2, but on the quantitative level, the orbital dependence is much weaker for these molecules than for the carbon atom. This can at least partially be explained by the observation that the LDA and xKLI orbitals are very similar for Li_2 and N_2, and thus essentially represent the same starting point for DMC.

To assess the impact of the mixed estimator bias and its correction from Eq. (6.74), we next investigate the quality of the mixed estimator extrapolation scheme from Sec. 6.12. Therefore,

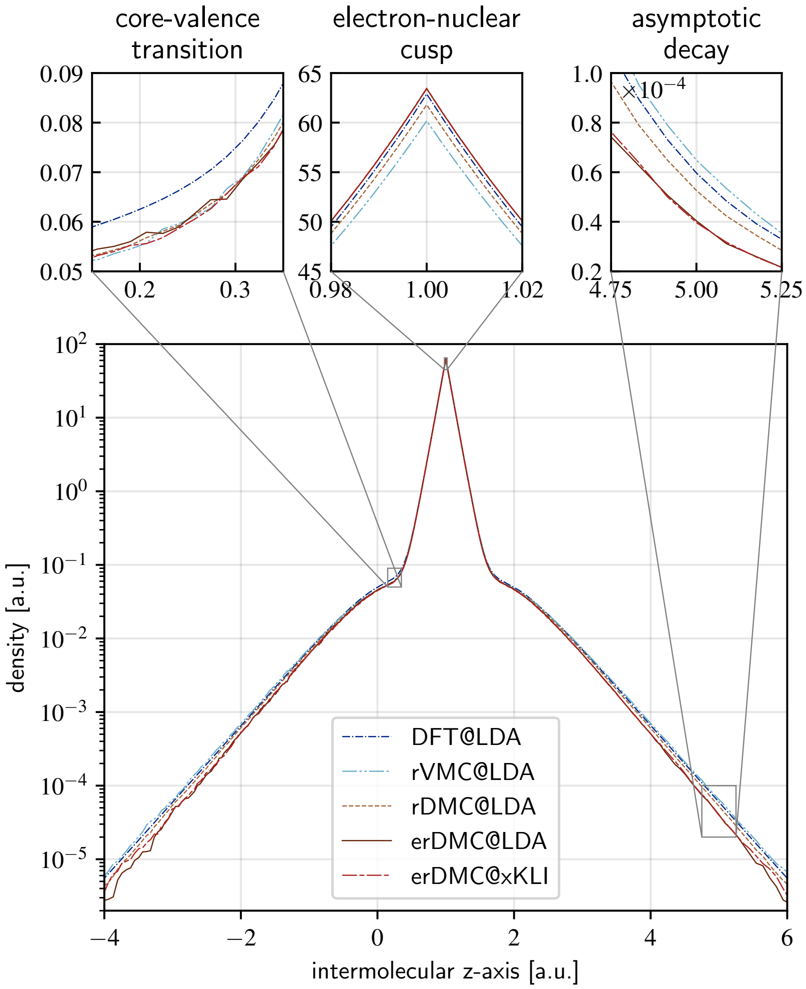


Figure 9.4.: Demonstration of the quality of the mixed estimator extrapolation: Plots of the DFT, rVMC, rDMC and erDMC densities for the carbon atom based on LDA and the erDMC density from xKLI orbitals. Although DFT, VMC and DMC lead to somewhat different densities in certain regions of space, the extrapolations based on xKLI and LDA align perfectly. This is not due to LDA and xKLI being the same starting point, see Fig. 9.3. All densities are spin-up densities.

we inspect the DFT, VMC, DMC and extrapolated density of the carbon atom based on LDA orbitals in Fig. 9.4. While the detailed behavior of the densities is not important to estimate the extrapolation quality, the differences of the densities are, and especially the comparison of DMC and VMC results. While small in the core-valence region, the density difference is noticeable at the cusp and in the asymptotics. If the deviation of the density due to $\psi_0 - \psi_T$ is beyond linear scaling, the extrapolation could not restore the unbiased density. From the densities generated with the LDA alone, one could not benchmark the extrapolation estimator, since a second-order deviation cannot be remedied and we have no indication of leaving the linear regime. Luckily, by comparison to the extrapolated densities generated from xKLI, we can assert such behaviour. The orbitals of xKLI differ significantly from the orbitals of the LDA, as discussed in Sec. 9.3. These differences are reflected in the QMC densities. If the extrapolation was incorrect, LDA and xKLI would yield different extrapolated densities. In Fig. 9.4 we plot all densitities, including the extrapolated ones. The extrapolated xKLI density is in perfect agreement with the one based on

LDA. This does not only indicate that the deviations remain within the linear regime, but also verifies the extrapolation. The bias of the mixed estimator therefore can efficiently be eliminated. We note in passing that also for Li_2 and N_2 the extrapolated densities from LDA and xKLI are in good agreement.

9.3. Orbital nodal features in the carbon atom density

The density that one obtains for the carbon atom with the LDA or EXX functionals is not spherically symmetric. The $^3\Pi_0$ ground state is reflected in a single-particle occupation scheme that follows Hund's rule, $1s^2 2s^2 2p^2_{\uparrow\uparrow}$. While the spin-down channel has no explicit p-orbital character, in the spin-up channel the two (degenerate) HOMO orbitals are of p-type with $l = |m| = 1$. Schematically, one can think about their angular dependence in the form

$$\varphi_{\mathrm{H}}(r, \vartheta, \phi) \sim r \sin\vartheta e^{-Zr/2} e^{\pm i\phi}, \tag{9.1}$$

though one should keep in mind that the orbitals realized in an actual DFT calculation are not of the exact hydrogen-orbital form of Eq. (9.1). However, this form illustrates the aspect that is decisive for the following discussion: This carbon-atom density realizes the special situation of a nodal surface in the orbital density of the highest occupied orbital(s) [141], with a nodal line along the z-axis, where $\vartheta = 0$. Situations with a nodal surface of the highest occupied orbital have found special interest in DFT [142–145], e.g., because of their possible consequences for the asymptotic structure of the xc potential.

The nodal surface induces a non-spherically symmetric decay of the carbon atom density of the schematic form

$$n = n_{\mathrm{R}} + n_{\mathrm{H}} = n_{\mathrm{R}} + 2|\varphi_{\mathrm{H}}|^2, \tag{9.2}$$

where n_{R} is the residual density without contributions from the HOMO orbital. While the HOMO orbital dominates the exponential decay anywhere except $\vartheta = 0$, the HOMO-1 orbital governs the decay along $\vartheta = 0$, or $z = 0$ equivalently. Since DFT guarantees an exact density for the exact xc potential, the density has solid physical meaning - beyond any doubts about the auxiliary character of KS quantities such as eigenvalues and orbitals. In the following we analyze densities that result from DMC calculations, i.e., from beyond the single particle representation, with respect to the question of whether they show the characteristic features that are associated with the nodal surface. To this end we calculate two densities, DMC@LDA and DMC@xKLI, and analyze their asymptotic decay. We implicitly define the spatially resolved asymptotic exponential decay rate γ for a density n by

$$n(\mathbf{r}) = n_0(\hat{\mathbf{r}}) \exp^{-\gamma(\hat{\mathbf{r}})r} \tag{9.3}$$

at $|\mathbf{r}| \to \infty$, where $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$. As any bound density decays exponentially [125], γ can at most depend on the angles, $\gamma(\hat{\mathbf{r}}) = \gamma(\vartheta, \phi)$, for large $|\mathbf{r}|$. Due to symmetry restrictions from the PS coordinates and the doubly occupied p-type orbital from Eq. (9.1), γ in our case can only depend on ϑ.

We now aim to quantify the value of $\gamma(\vartheta)$ for different angles ϑ. For every ϑ we consider a straight line starting from the carbon atom and forming an angle ϑ with the z-axis. For example, $\vartheta = 0$ represents a line from $(0, z_{\mathrm{carbon}})$ to $(0, z_{\mathrm{max}})$ in cylindrical coordinates, while $\vartheta = 90°$ yields the line from $(0, z_{\mathrm{carbon}})$ to $(\rho_{\mathrm{max}}, z_{\mathrm{carbon}})$. Along this straight line we construct the density by interpolating from its values on the PS grid. Such radially extending slices of the density are shown in Fig. 9.5. Although the densities at $\vartheta = 90°$ and 180° show a distinct gap in their magnitude, it is their slope that holds the information about the decay rates. For every slice we fit the density along the line to the exponential form of Eq. (9.3), omitting the core region

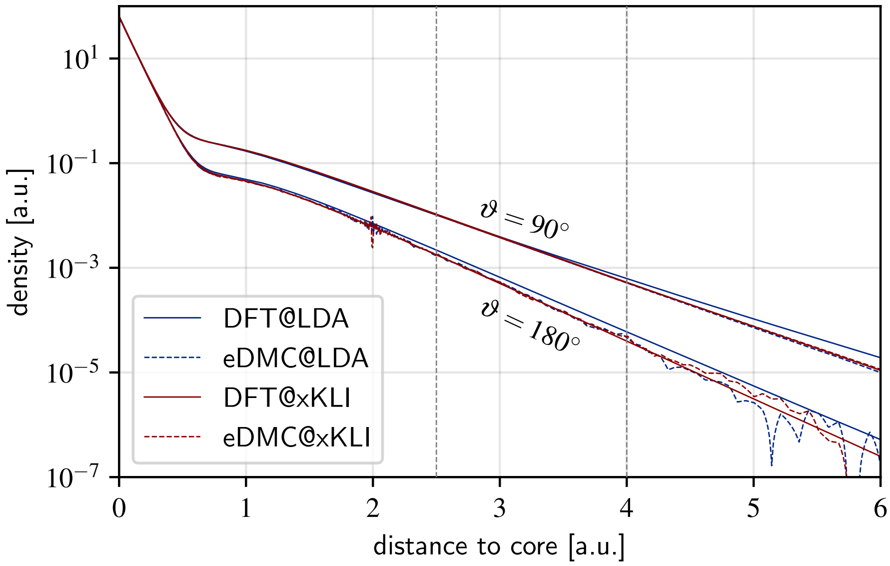


Figure 9.5.: The DFT and eDMC density along the z-axis ($\vartheta = 180°$) and the angle $\vartheta = 90°$ are shown for the functionals LDA and xKLI. We fit the graphs in the indicated interval from 2.5 to 4 a.u. Visually it is difficult to distinguish the decay rates γ (slope) from the amplitudes n_0 (offset), see Fig. 9.6 for the angle-dependence of γ. We plot the decay along $\vartheta = 180°$ instead of $\vartheta = 0°$, since the PS grid is more fine-grained on the negative part of the z-axis (from $-z_{\max}$ to z_{carbon}) than on the positive side (from z_{carbon} to $z_{\max}$), see Fig. E.2.

($r < 2.5$ a.u.) as well as the asymptotic region dominated by noise ($r > 4$ a.u.). This yields the decay rate as a function of the angle $\gamma(\vartheta)$.

The coefficient of determination in all fits exceeded 0.999, confirming that we consistently fitted only the pure asymptotic decay with our choice for the radial interval. Furthermore, we do not use the density from our refinement procedure to calculate the decay rate.

The results are shown in Fig. 9.6 for eDMC and DFT densities based on LDA and xKLI. The decay rates from eDMC align with the rate from DFT@xKLI. They approach their extrema at 0° and 90° with a quadratic behavior, which reflects the node of second order of the orbital-densities, since the HOMO orbitals have a node of first order and the HOMO density results from squared orbitals. The LDA underestimates the decay for all angles. This is a consequence of the decay rate being coupled to the HOMO eigenvalue [125] by $\gamma = 2\sqrt{-2\epsilon_{\text{H}}}$ and the well-known fact that the ϵ_{H} of LDA does not accurately reflect the true density decay. However, the decisive observation is that also the DMC densities show a substantial change in the decay rate when ϑ varies from 90° to 0° (or 90° to 180°). Thus, the DMC densities show the signature of the nodal line in a very similar way as the DFT densities. The results further show that this is not caused by the potential shortcoming of the DMC method to not be able to change the density in the asymptotic region: One can clearly see that the DMC calculation changes the density in the asymptotic region, as the decay rate of the DFT@LDA density is changed significantly by the subsequent DMC calculation.

The finding that the density obtained from the DMC method shows the characteristic features of the nodal line is of relevance in the context of the question whether the (exact) KS potential can have a uniform asymptotic decay, or must be expected to have unusual, potentially diverging features in some regions of space: These different options have been discussed in detail based on an expansion in Dyson orbitals in Ref. [143]. Following Gori-Giorgi et al., the observation that the ground-state density shows a different decay along the nodal surface leads to the conclusion that

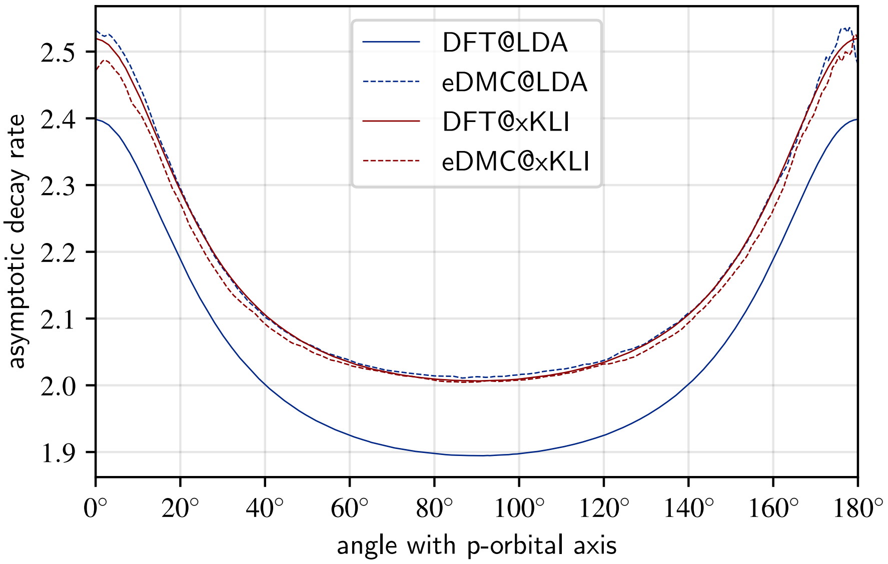


Figure 9.6.: The asymptotic decay of the densities is dominated by the lower rate of the HOMO $2p$ orbital over a large range of angles, but approaches the larger rate of the HOMO-1 decay of $2s$-type when approaching the z-axis or the angles 0 and 180°, equivalently. This behavior is observed irrespectively of the functional or method. The values of γ near an angle of $\vartheta = 0, 180°$ show some noise. This is due to the interpolation on the grid, which becomes a slightly inferior extrapolation when approaching the z-axis, since the grid points do not lie on the z-axis directly.

the corresponding (exact) KS potential can be expected to decay regularly, without unpleasant surprises on or in the vicinity of the nodal plane.

10. Inversion results

> The following chapter is an excerpt from the author's publication [116] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

10.1. The forward KS-map

Computing the forward KS map is a standard PDE eigenvalue problem, which can be done fast and at low computational cost. We use an elaborate and efficient numerical all-electron approach for the forward problem, with a prolate spheroidal coordinate system tailored to exploit the symmetry of diatomic molecules and atoms [146]. This two-dimensional grid concentrates grid points near the nuclei for accurate representation of the Coulomb potential. Our goal is to build inversions on this grid.

However, before studying the inverse problem it is natural to explore the properties of the forward problem. We examine the behavior of the forward map by varying the xc potential. Since by the HKT an auxiliary KS potential holds the full information of an interacting wave function, it should be assumed, that changes in the KS potential lead to complicated and nonlinear changes in the density. A natural way to perturb the KS potential is a mixing of different xc functionals, since artificial potential variations such as noise, bumps, kinks, ramps, etc. could introduce unnatural potentials. In Sec. J we examine the linear mixing of xc potentials. Our linear interpolation of two xc potentials $v_{\mathrm{xc}}^{(0)}$ and $v_{\mathrm{xc}}^{(1)}$, and their respective densities $n^{(0)}$ and $n^{(1)}$, is given by $v_{\mathrm{xc}}^{(\lambda)} = (1-\lambda)v_{\mathrm{xc}}^{(0)} + \lambda v_{\mathrm{xc}}^{(1)}$. We chose a large range for the values of λ not restricted to interpolation, for which $\lambda \in [0,1]$, but extrapolating the potentials. The linear perturbation allows to predict the linear response of eigenvalues and the total energy within DFT in advance, since the KS Slater determinant and DFT are based on a variational principle, and thus can be estimated with the Hellmann-Feynman theorem [147]. Due to the Hellmann-Feynman theorem only the explicit change in the potential enters the linear response, while the induced density variation does not contribute to the energy or eigenvalues. This allows the density to respond in a much more complicated way, because the energetic linear response is decoupled from the density's linear response. If we, for the moment, naively consider the density to only respond in the linear regime, the two known densities for $\lambda = 0$ and $\lambda = 1$ fix the linear response and allow a linear prediction of the density, $n_{\mathrm{lin}}^{(\lambda)} = (1-\lambda)n^{(0)} + \lambda n^{(1)}$. First, the range of the linear regime, i.e. the range for λ in which the density response linearly with negligible contributions from higher orders, is unknown and second higher order responses could become non-negligible even with small values of λ. Given the complexity of the true KS map and its requirement to compress the full information of an interacting wave function to a single density, one could assume that the response should be rather non-trivial. The results of this analysis are in detail shown in Sec. J. Surprisingly the density response is linear over an unexpectedly large interval for λ. Even if the xc potentials look nothing like a reasonable physical potential any more, the density remains in the linear response regime. A spatially resolved analysis of the density response shows how the density responds in a linear and *point-wise* way. The naive prediction $n_{\mathrm{lin}}^{(\lambda)}$ does in fact resemble the actual density to high accuracy. Thus a reasonable estimate of an

Table 10.1.: Metrics of a "simple" inversion of the carbon atom: The self-consistent xKLI density is chosen as the reference density and various inversions are performed starting from the xc potential of a self-consistent LDA calculation. We report inversions based on the update rules HAR, HAR+DEN, HAR+LoH(0) and HAR+LoH(1).

method	D_1 $\times 10^{-3}$	D_∞ $\times 10^{-2}$	W_2 $\times 10^{-5}$	J_{xc} $\times 10^{-5}$	optimal iteration
HAR	6.28	13.4	4.52	−20.9	254
HAR+DEN	4.97	3.45	4.32	−11.7	184
HAR+LoH(0.0)	3.09	9.32	1.98	−6.71	373
HAR+LoH(1.0)	3.13	5.48	3.01	−5.43	399

inter- and extrapolated density might be available without actual evaluation of the forward map. This has several implications, e.g., for the development of density-reproducing density functional approximations, various numerical applications, density inversion, etc. We suggest an inversion procedure based the local linearity of the KS map in Sec. K.

10.2. Simple inverse problems

In this section, we carry out inversions for simple cases. Our goal is to present insights independent of numerical artifacts of the reference density, e.g., the limitations of inversion in general and the capabilities of the update rules. We therefore invert densities, that itself stem from a DFT calculation. We use the same numerical approach, described in Sec. 10.1, for the generation of the reference densities and their inversion. First we do a standard DFT calculation for some xc functional, which yields energies, its density and its xc potential. Next, we do an inversion of this density as the reference density. While the xc potential to the reference density is available to us, we do not use this reference xc potential in the inversion process, as this would render the inversion trivial. Nevertheless, with a known reference xc potential we are in the the unique position to compare any inverted potential to the true xc potential, as well as its eigenvalues, eigenstates, etc. and calculate J_{xc}.

A challenging system even for simple inversions is the carbon atom. Due to spin maximizing, its occupation is spin-polarized with four spin-up electrons in the configuration $(1s2s)2p^2$ and two spin-down electrons in $(1s2s)$. The xKLI density shows a non-spherical asymptotic decay due to a nodal line of the HOMO orbital, see Sec. 9.3, and as a result its xc potential builds up a ridge along some axis [141–145]. The LDA xc potential does not show such complicated features, and therefore the xc potentials of LDA and xKLI are very distinct. With an reference xc potential (xKLI) exhibiting correct asymptotic decay, an asymptotic ridge, inter-shell bumps and quadratic behavior at the nuclei on one side and an initial xc potential (LDA) with exponential, spherically symmetric decay and cusps at the nuclei on the other, this could to the contrary be seen as an extraordinary hard inverse problem. In such a situation we do not expect to recover the reference xc potential by inversion, especially not for the case of the xKLI potential, since an asymptotic nodal ridge lies beyond the update rules capabilities and also beyond the general limitations of the inversion mentioned in Sec. 8.1.

With these prerequisites we inverted the xKLI density (reference density) with an initial xc potential from an LDA calculation. From the update rules introduced above we show the results for HAR, HAR+DEN, HAR+LoH(0) and HAR+LoH(1) in Fig. 10.1 alongside with the inversion metrics in Tab. 10.1. In these calculations, the HAR update rule uses spin-resolved Hartree potentials. From the initial LDA potential no update rule converges to the xKLI potential, as

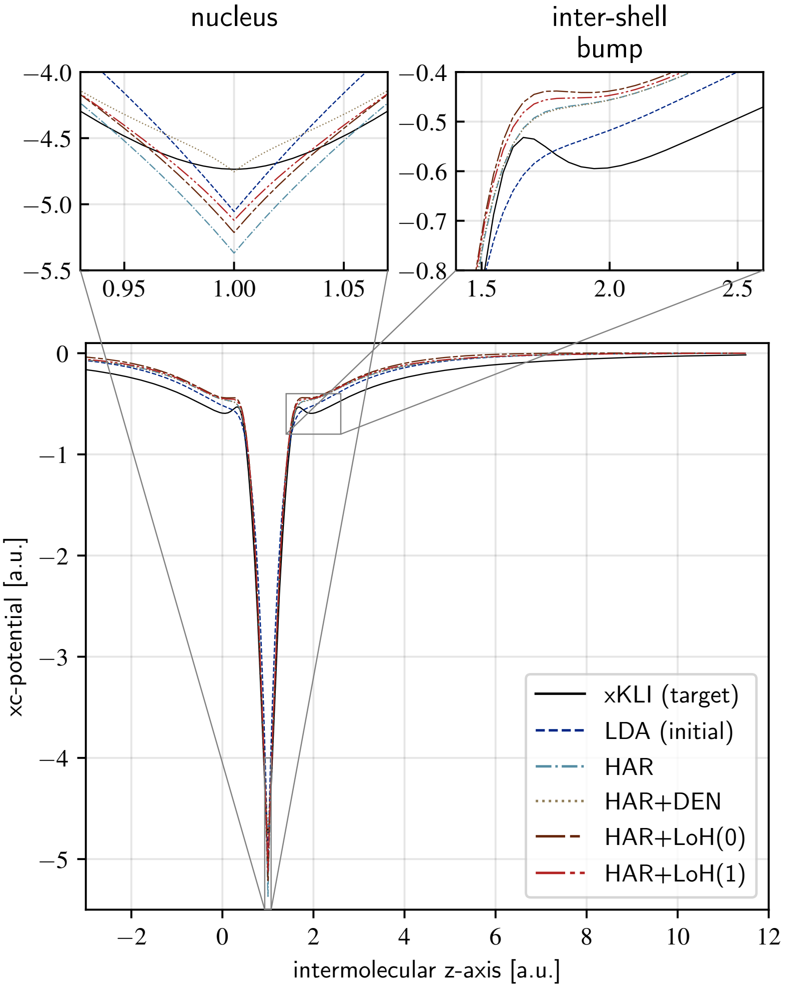


Figure 10.1.: “Simple” inversions for the carbon atom. The reference density originates from a standard DFT calculation for the exact exchange in KLI approximation (with its xc potential given by the solid black line), while the initial potential (dashed blue line) is the self-consistent LDA potential. Different update rules are examined with their xc potentials indicated by the lines of different colors and dash styles. The inverted potentials resemble only some of the features of the target xc potential. The upper insets zoom in on the nucleus and the location of an inter-shell bump.

expected, but to an intermediate potential, that holds some features of the LDA and some of the xKLI potential. For example, no inverted potential is quadratic at the nucleus, which ultimately stems from the construction of the update rules and the initial potential. However, the inter-shell bump, where the density takes values of about 7×10^{-2}, is partially generated by all update rules, but has a local maximum only for HAR+LoH(0) and a saddle point for HAR+LoH(1). The nodal ridge in the reference xKLI potential is not recovered by any update rule (not shown), as expected. Despite of this great structural difference in the inverted xc potentials, the density differences in Tab. 10.1 are rather small. However, since density differences tell little about the quality of the potentials, we also compare the occupied spectrum of the KS systems in Fig. 10.2. From Fig. 10.2 it becomes apparent that the structural deviations of the xc potentials do not greatly influence the eigenvalues or eigenstates. While the LDA shows distinctly different eigenvalues compared to xKLI, the inverted potentials recover great similarity to xKLI eigenvalue

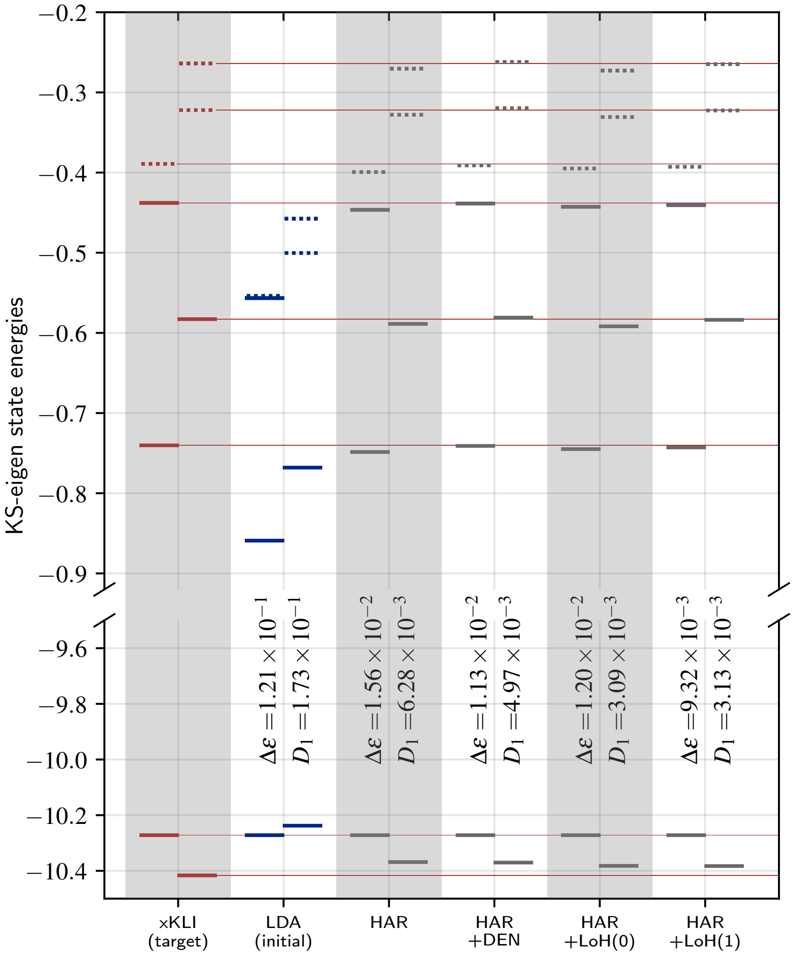


Figure 10.2.: Comparison of the eigenvalue spectrum of a “simple” inversion of the carbon atom with a reference density from xKLI and an initial potential from LDA. Left of the respective ticks on the horizontal axis are the spin-up eigenvalues, right of it the spin-down eigenvalues. Dashed lines correspond to unoccupied states. The eigenvalues of the referential xKLI calculation are highlighted with horizontal lines for comparison to the inverted eigenvalues. Note that the HOMO eigenvalues in the up-channel are doubly degenerate. Energies are given in Hartree units.

spectrum in the examined energy range. While the HAR and HAR+DEN rules show inferior density metrics, they lead to reasonable spectral properties, with slightly worse eigenvalues for HAR. The updates from HAR+LoH(1) as well as HAR+LoH(0) further improve the inversion for the density metrics as well as the eigenvalue metric.

The above methodology might be called an *inverse crime* [127, 148], i.e. evaluating the forward map with the same numerical methods that the inverse map relies on. In our case the above study is solely dedicated to the numerical understanding of the inverse map. In the following we apply the inversion procedure to densities from a different numerical approach. From comparing the inversions for the carbon atom to the inversions that we did for the dimers Li_2 and N_2, we learned that the HAR method discriminating spin channels is more accurate for truly spin-polarized systems, i.e., the carbon atom, than the HAR method that uses the spin-averaged potential. However, when we calculate the inversions for the spin-unpolarized systems Li_2 and N_2 with two independent spin channels, a spin-resolved Hartree potential update decreases the accuracy slightly. Therefore, we used the spin-resolved HAR method only with truly spin-polarized systems.

10.3. QMC density inversion

The reference densities of this section stem from QMC methods, more precisely from fixed-node diffusion Monte Carlo with a single-determinant Slater-Jastrow trial wave function. The densities are computed with the binning method and are extrapolated to reduce the mixed estimator bias. We use the densities published in Ref. [45]. As those densities are given on a large grid, infeasible for efficient inversion, we describe an interpolation scheme to reduce their dimensionality in Sec. H. The key challenge of inverting the QMC densities is the presence of statistical noise. This noise is point-wise in the density and, while unnoticeable to the bare eye for most regions of space, the asymptotic domain as well as the vicinity of the nuclei are affected by noise. As conjured above for the example of an incorrect nuclear cusp, inverting a density with even the slightest numerical artifacts might have severe impact on the inverted potential. Therefore we use the regularization scheme of Ref. [45] to remedy the critical statistical fluctuations of the density at the nuclei by invoking the cusp condition [69], cf. Sec. H for details.

We have inverted seven different reference densities from QMC calculations. The QMC method requires starting orbitals, for which we chose DFT orbitals from the LDA functional in the Perdew-Wang parametrization for the densities labeled “QMC-A” and exact exchange orbitals with a local potential in the KLI approximation for the density-label “QMC-B” for every system. The systems under study are the nitrogen molecule (with a bond length of 2.0743 from Ref. [138]), Li_2 (with a bond length of 5.051 from Ref. [138]) and the carbon atom. Additionally we calculated another density based on LDA orbitals for Li_2 labeled “QMC-C”. This density is entirely analog to the “QMC-A” density, but with different statistical fluctuations, i.e. a different seed for the random number generator. We carried out inversions with initial xc potentials from self-consistent LDA and xKLI calculation as well as from the Fermi-Amaldi potential. We have conducted inversions for every reference density, every initial potential and every update rule of Sec. 8.2 with values $\beta = -0.1, 0.1, 0.25, 0.5, 1, 2$ and 3 for the LoH update. However, only a subset of the results are reported due to the large number of inversions.

To begin with, we inspect the metrics of the inversions for the nitrogen molecule in Tab. 10.2 comparing different update rules and initial xc potentials.

One of the most “pleasant” inversion for nitrogen is generated by HAR+LoH(1)[xKLI], i.e. an inversion using the update rule HAR+LoH(1) and an initial xc potential from the xKLI approximation, and illustrated in Fig. 10.3. The inversion features inter-shell bumps and correct asymptotic decay inherited from the xKLI potential. It is smooth and shows no artifact even in the vicinity of the nuclei. The inverted density matches the reference density, up to $D_1/N < 0.05\%$. Thus, more than 99.95% of the recovered electron density per electron is distributed correctly. The inversion converged to $Z_{\mathrm{xc}} < 2 \times 10^{-7}$. This qualifies this specific inversion as a success. Nevertheless other inversions show similar metrics albeit with different features in the potential. An example of those is HAR+LoH(0)[xKLI], also shown in Fig. 10.4, which shows numerical artifacts in the asymptotic domain, and HAR+DEN[FA] from Fig. 10.5 with more deviating but still comparable metrics and much larger fluctuations at the nuclei.

Examination of Tab. 10.2 highlights some of the features of the update methods. The overall best results are those starting from the xKLI potential. The Fermi-Amaldi potential is still a good starting potential. The LDA shows inferior metrics, most properly due to the incorrect asymptotic behavior. The deviation among update rules is more subtile, except for HAR and HAR+DEN. The rule HAR does not fully recover the density, since it breaks sign-consistency on a relatively large scale, imposing a strong regularization on the inversion. Its metrics reflect

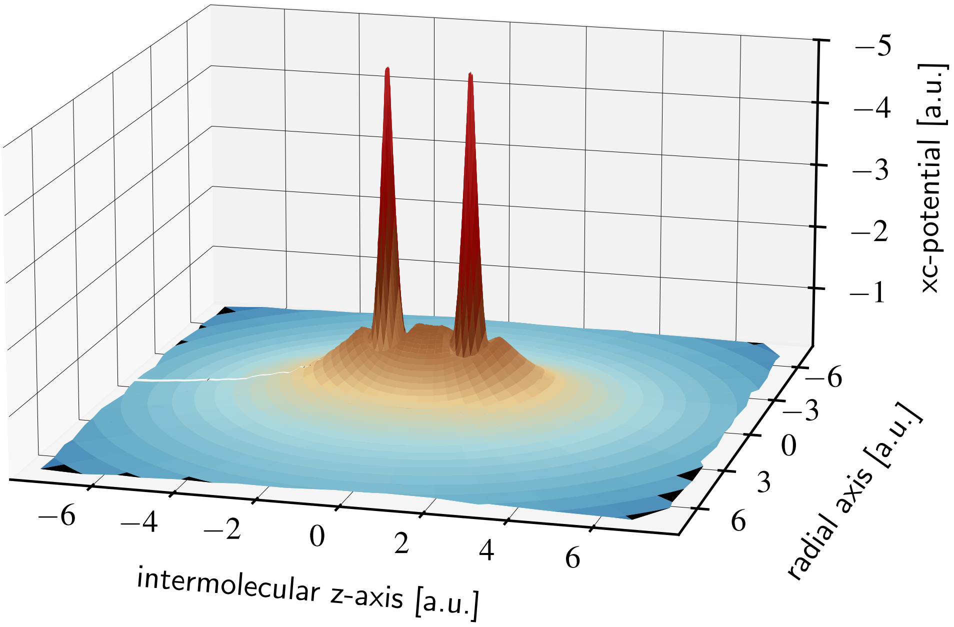


Figure 10.3.: A three-dimensional representation of the inverted xc potential of the nitrogen molecule in a plane through the molecular axis. The reference density QMC-B was inverted with the update rule HAR+LoH(1) with an initial xKLI xc potential. Despite the stochastic fluctuations in the reference density, the inverted potential has experienced enough regularization to exhibit a smooth behavior. The inter-shell bumps are clearly visible around the nuclei. Note, the inverted axis for the potential. The coloring was chosen to resolve the key features of the potential.

its limitations, while the resulting inverted potentials are smooth, see Fig. 10.4 and 10.5. With direct density feedback, the rule HAR+DEN excels to minimize the metric D_∞ compared to even the best other update rules. The D_∞ metric is most affected by the density deviations at the nuclei, where the density takes its maximum values, thus HAR+DEN best resembles the density at the nuclei. This may come at the cost of introducing artifacts in the inverted potential and missing to reproduce other properties, such as the eigenvalues. For the case of HAR+LoH(0.5)[xKLI], which is marked with a dagger in Tab. 10.2, the inversion cycles almost diverged and subsequently recovered slightly, see Sec. I.1. From the metric D_∞ alone, one would hardly recognize the failure of this inversion, which highlights the importance of different metrics to indicate the success of an inversion. The metrics D_1 and W_2 clearly mark the inversion as an outlier with respect to the others.

The parameter β in the rule LoH offers a way to continuously choose between smoothness of the resulting xc potentials at slightly unresolved density differences on the one hand and an accurate restoration of the reference density sacrificing regularity in the xc potential and introduction of artifacts. For the inversions started with an xKLI potential, except the outlier, the update rule LoH(β) shows monotonously decreasing D_∞ with increasing β, bridging from the rule HAR to HAR+DEN. The different LoH rules improve upon HAR or HAR+DEN in the Wasserstein metric W_2. The metric D_1 has a minimum for HAR+LoH(1) with maximum values for HAR and HAR+DEN, indicating a “sweet spot” for the parameter $\beta = 1$.

A spatial comparison for some of the inversions is given in Fig. 10.4 with the xKLI initial potential and Fig. 10.5 with the Fermi-Amaldi initial potential. Irrespective of the update method significant changes are made to the initial potential, in particular for the Fermi-Amaldi potential,

Table 10.2.: Metrics of inversion for the N_2 molecule. Inversions are performed for the density QMC-B. The entry marked by an asterisk (*), HAR+LoH(1), is an example of a very successful inversion and shown in Fig. 10.3. The rule HAR+LoH(0.5)[xKLI], marked by a dagger (†), is an example of an unstable inversion attempt. The inversion procedure did not converge with the same "soft" parameters of the other inversions, cf. Sec. I. While for further analysis the specific inversion attempt based on the rule HAR+LoH(0.5)[xKLI] should not be trusted, it is not impossible to carry out successful inversions for the combination HAR+LoH(0.5)[xKLI].

method		$v_{\text{xc}}^{(0)}$	D_1 $\times 10^{-3}$	D_∞ $\times 10^{-1}$	W_2 $\times 10^{-5}$	Z_{xc} $\times 10^{-7}$
HAR		FA	14.8	5.94	1.65	−6.2
HAR+LoH(0.0)		FA	11.1	6.38	0.81	−10.2
HAR+LoH(1.0)		FA	11.1	5.90	1.01	−7.2
HAR+DEN		FA	9.7	2.97	0.83	−5.3
HAR		LDA	16.5	8.32	1.47	−19.6
HAR+LoH(0.0)		LDA	13.8	8.10	1.15	−22.3
HAR+LoH(1.0)		LDA	12.7	7.22	1.21	−14.1
HAR+DEN		LDA	11.2	3.59	1.06	−9.2
HAR		xKLI	9.7	6.33	1.83	−1.3
HAR+LoH(-0.1)		xKLI	7.9	6.20	1.27	−2.3
HAR+LoH(0.0)		xKLI	7.6	6.11	1.24	−2.1
HAR+LoH(0.25)		xKLI	7.6	6.02	1.41	−2.0
HAR+LoH(0.5)	†	xKLI	10.1	6.32	2.15	−7.6
HAR+LoH(1.0)	*	xKLI	6.6	5.27	1.29	−1.5
HAR+LoH(2.0)		xKLI	7.9	4.72	1.46	−2.5
HAR+LoH(3.0)		xKLI	7.9	4.13	1.55	−2.5
HAR+DEN		xKLI	10.0	3.73	1.91	−6.1

and both initial potentials converge toward the same target potential. Inter-shell bumps build up in the inversions started from Fermi-Amaldi potential. This suggests that the inter-shell bumps are crucial to the potential in order to generate the reference density. In the case of HAR[xKLI] the inter-shell bumps in the inter-atomic bond region are accompanied by a repulsive bump in the xc potential at $z = 0$, which lowers the density in the σ-type bond region between the nuclei. This cannot be deduced from any of the other inversion due to the lack of such a repulsive bump or the magnitude of the surrounding artifacts. Focusing on the nuclei, HAR+DEN shows a diverging potential with strong wiggles in vicinity of a nucleus. The potentials of the other update rules approach the nuclei more regular. The regularity of the HAR update is outstanding with a smooth potential almost everywhere. The other updates show numerical artifact already in the inter-atomic bond region.

Note that from a single inversion at hand one would easily be led to the conclusion, that the numerical wiggles in either the inter-shell bumps or the inter-atomic bonds are pure noisy, with no physical meaning, and potentially very volatile to any changes in the numerical setup. With a closer look at the artifacts and the comparison of the two figures Fig. 10.4 and Fig. 10.5 it becomes clear that the wiggles are not chaotic numerical artifacts, but are highly correlated and share many characteristics. They serve the purpose to reproduce the statistical fluctuations of the reference density. The same holds for inversions started from LDA (not shown). It is ironic, that the numerical artifact associated with noise is one of the most consistent features across all inversions. This again highlights the very construction principle of the inversion - to reproduce a

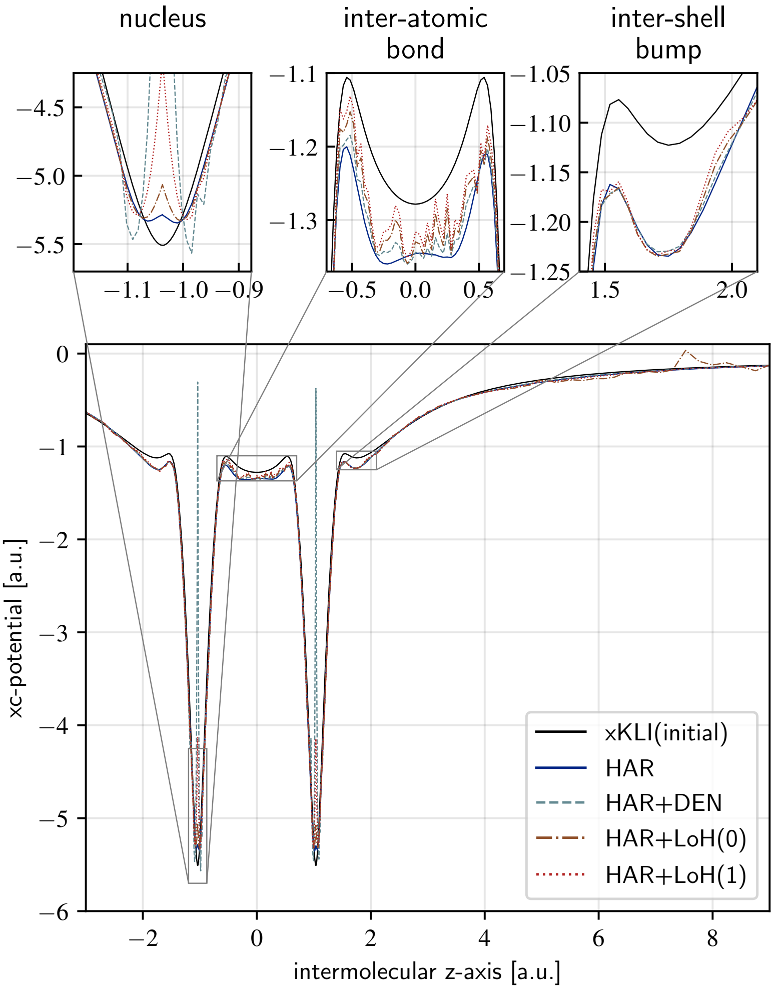


Figure 10.4.: Different inverted xc potentials for the reference density QMC-B of the N_2 molecule. Starting from the xc potential of an xKLI calculation, the inverted potentials stem from the update rules HAR, HAR+DEN, HAR+LoH(0) and HAR+LoH(1).

reference density - and aligns with the observation, that the numerical artifacts of the reference density tend to be overemphasized in the inverted potential.

The major conclusion to draw from this series of inversions for the nitrogen molecule, is the necessity of inter-shell bumps in the potential. Further, there is a potential bump at $z = 0$, that lowers the density in the σ-type bond region for the rule HAR[xKLI]. This bump is not present in the inverted xc potential of HAR[FA], which indicates a bias by the initial xc potential. Further conclusions, that go beyond analyzing the specific system, are for example the superiority of the xKLI and Fermi-Amaldi potentials compared to LDA as an initial xc potential. Within the set of update rules that we studied, the update rule LoH(β) with $\beta \in [0, 1]$ appear as reasonable intermediate candidates, capturing important aspects of both the smoothness of HAR as well as the density reconstruction of HAR+DEN. Any inverted potentials that are outliers with respect to the ensemble of the other inverted potentials shown should be treated with care. If some inverted potential shows qualitatively differing artifacts, and the origin thereof is not explained within the construction of the update rule or the initial potential, caution is advised.

In the presence of statistical fluctuations, it is difficult to decide which characteristics of the inverted potentials actually stem from the reference density's noise, and to distinguish such

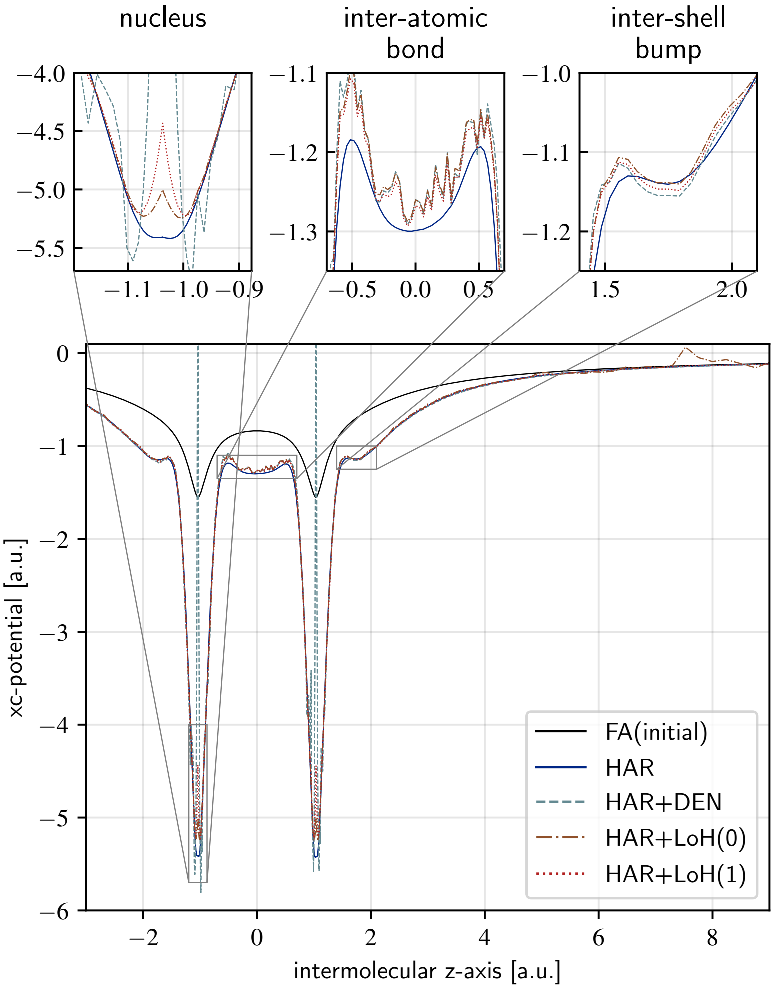


Figure 10.5.: Different inverted xc potentials for the reference density QMC-B of the N_2 molecule. Starting from the Fermi-Amaldi xc potential (generated from the reference density), the inverted potentials stem from the update rules HAR, HAR+DEN, HAR+LoH(0) and HAR+LoH(1).

artifacts from true characteristics that are independent of the noise. Therefore, we perform a series of identical inversions for two densities of the Li_2 system, QMC-A and QMC-C, that differ by their statistical fluctuations only. The QMC densities represent the same physical system and are methodically identical, their differences stem solely from a different set of random numbers. That being said, the densities can nevertheless slightly differ in quality, where we refer to the quality relevant for inversion, which can be unexpected as seen in the above examples. These difference could manifest in subtle features of the density, such as an unnatural bump in the density in the far asymptotic region. The obvious question to address, namely whether the statistical differences generate different inverted potentials, can partially be answered beforehand. The pure Hartree update HAR should yield a smooth potential, since its updates average over the fluctuations. The updates incorporating local density feedback, are bound to show larger dependencies on the noise. In Fig. 10.6 we compare the inverted potentials for the rules HAR, HAR+DEN and HAR+LoH(1) and find essentially the above prediction: The HAR updates are visually indistinguishable despite the noise, expect for the pathological vicinity of the nuclei, where they differ only slightly. The rule HAR+DEN does not improve over HAR much, since in the Li_2 system the absolute value of the density is close to the range of xc potential values and the direct density feedback from DEN therefore comparably weak. The rule HAR+LoH(1) exhibits a stronger dependence on the noise point-wise, but conserves the qualitative potential. While the capability of the update rule to locally adjust the potential from direct feedback of the density is an advantage for densities free of numerical artifacts, it can become a hindrance in the presence of noise, artificial wiggles, or spikes. To conclude with, the parts of the inverted potential, that stem from the densities' fluctuations, are clearly visible as noisy or artificial details and cannot be mistaken for distinct characteristics. These findings are supported by Tab. 10.3. The metrics for both densities are in close agreement to each other. The Wasserstein metric shows essentially no difference between the reference densities QMC-A and QMC-C. This indicates that the numerical artifacts of the reference densities, despite having impact on the inverted potentials, mostly spend this impact on less relevant, auxiliary characteristics of the inverted xc potential, that have little physical meaning. One can additionally assess the consequences of different fluctuations in the reference densities manifesting in the eigenvalues of the inverted xc potentials. For a non-trivial inversion (with unknown reference xc potential) it is not possible to measure the quality of the eigenvalues with respect to the true reference spectrum, as shown in Fig. 10.2. However, the eigenvalue spectra can be compared among different inversions. Fig. 10.7 shows the eigenvalues for the inverted potentials of Fig. 10.6 and Tab. 10.3. Clearly the eigenvalues are in very good agreement. Therefore the distinct numerical artifact visible in Fig. 10.6 must almost purely be of auxiliary type and do not influence the resulting physical interpretation of the KS system. This is reassuring to see, since apparently the xc potential equipped with artificial deviations below a certain magnitude does not change its physical implications. Another perspective on the small dependence of the physically meaningful quantities on the artifacts present in the xc potential is the one of averaging over the fluctuations. Consider the reference density, n, to be the true density, n_0, with an additional scalable artifact-density, thus $n = n_0 + \lambda n_a$ with scaling parameter λ. Then, within the linear approximation to the forward KS map of Sec. 10.1 the corresponding true xc potential is $v_{\mathrm{xc}} = v_{\mathrm{xc}0} + \lambda v_{\mathrm{xc}a}$. From the Hellmann-Feynman theorem the eigenvalues obey $\varepsilon_i(\lambda) = \varepsilon_i(0) + \lambda \langle \varphi_i | v_{\mathrm{xc}a} | \varphi_i \rangle$. The artificial contributions of the potential are averaged over the orbital's density, which allows fluctuations to cancel out. Even without the linear approximation to the forward map, the true density response, cf. Eq. (J.2), does also average fluctuations of the potential in the overlap integrals $\langle \varphi_i | v_{\mathrm{xc}a} | \varphi_j \rangle$ of the response function.

We cannot generalize these findings to other types of numerical artifacts, that are non-statistical in nature, such as limited basis set, incomplete configuration interaction or other biases. From the above it seems that noise in the reference density has mild consequences for the KS observables. Other types of artifacts might also influence the inverted potentials in a specific and systematic way. Such biases could, to the contrary, be smooth and thus be more subtle to detect, and may require a detailed analysis on their own. As a final comment on the oftentimes cumbersome inversion process, we show the evolution of the inversion over the course of the repeated application of the update rule in Sec. I.

Table 10.3.: Metrics of the inversion of two QMC densities (QMC-A and QMC-C) for the Li_2 molecule, that differ in their statistical fluctuations only, for the update rules HAR, HAR+DEN, HAR+DoH, HAR+LoH(0) and HAR+LoH(1). All inversions shown below were started from the Fermi-Amaldi potential.

ref.-den.	method	D_1 $\times 10^{-3}$	D_∞ $\times 10^{-2}$	W_2 $\times 10^{-4}$	Z_{xc} $\times 10^{-9}$
QMC-A	HAR	7.59	2.28	1.23	−4.07
QMC-A	HAR+DEN	7.12	1.73	1.21	−4.47
QMC-A	HAR+DoH	7.52	2.15	1.23	−5.32
QMC-A	HAR+LoH(0.0)	3.64	2.08	0.92	−3.14
QMC-A	HAR+LoH(1.0)	4.77	1.96	1.07	−3.49
QMC-C	HAR	8.63	2.21	1.24	−3.80
QMC-C	HAR+DEN	7.71	1.77	1.21	−3.67
QMC-C	HAR+DoH	6.58	2.03	1.17	−3.24
QMC-C	HAR+LoH(0.0)	2.90	1.95	0.89	−7.04
QMC-C	HAR+LoH(1.0)	4.26	1.88	1.06	−4.32

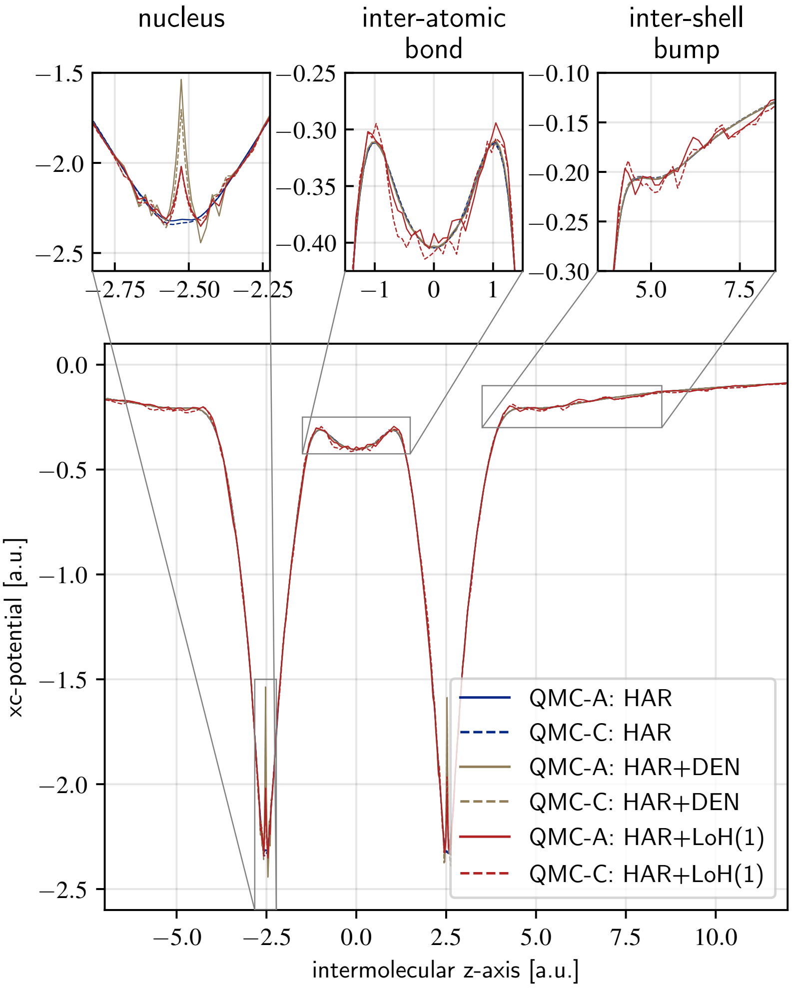


Figure 10.6.: Inverted xc potentials of the two densities QMC-A and QMC-C for the Li_2 molecule, that differ in their statistical fluctuations only, for the update rules HAR, HAR+DEN, and HAR+LoH(1) starting from the Fermi-Amaldi potential. The HAR densities are not distinguishable by the eye in the main plot and HAR+DEN only marginally differs from HAR (mostly in the high-density regions). While the HAR+LoH(1) updates are much more “noisy”, their corresponding densities are also closer to the reference density, see the metrics in Tab. 10.3 or Fig. 10.7 for the exact values.

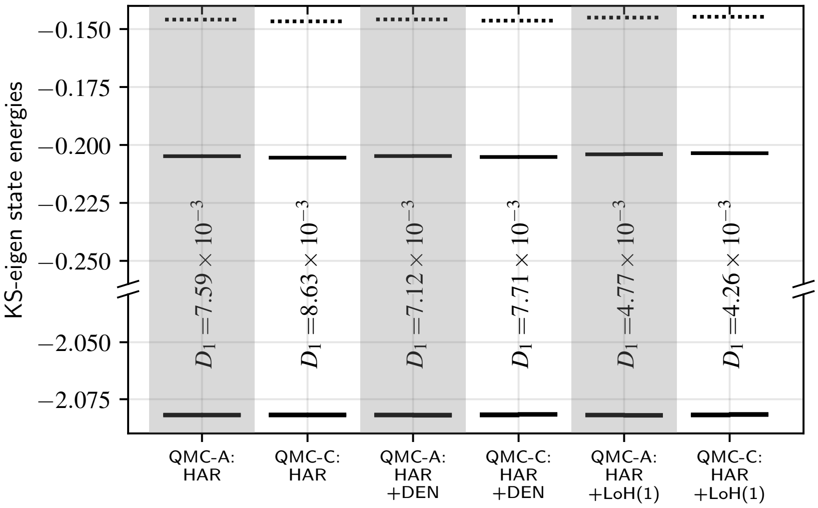


Figure 10.7.: The KS energy eigensystem for the same items as in Fig. 10.6. While originating from different inverted xc potentials, the energy spectrum is remarkably similar. The marginal differences in the lower eigenvalues at about −2.07 a.u. stem from the finite accuracy of the KS-solver. The update rules with a direct local feedback introduce a slight symmetry breaking of the spin channels. Nevertheless, note that the relatively noisy xc potentials from the HAR+LoH(1) update result in spectra visually almost indistinguishable from the smoother xc potentials of HAR or HAR+DEN. The eigenvalues have been shifted to align in the lowest up-eigenvalue due to the invariance of the KS system to a constant offset in the potential. Note the different scale of the plot for the two lowest eigenvalues, which is close to the accuracy to which the forward solver computes eigenvalues. The lowest two eigenvalues are almost degenerate and match within their line width.

11. Second order QMC

The toolset presented so far allows for generating accurate xc potentials from QMC densities. To this end, a standard DFT calculation is carried out to yield orbitals and a KS Slater determinant. The orbitals are used in VMC and DMC calculations to estimate the quasi-exact QMC density by extrapolating the mixed estimator error. Several inversions are carried out from the quasi-exact density to arrive at multiple candidate xc potentials. The reliability of the xc potential or its specific characteristics can be assessed from the multitude of potentials. However, the inverted xc potential corresponds to an inverted KS system, i.e., a density and a set of orbitals. These orbitals are accurate approximations of the orbitals of the true KS system and thus incorporate the physics of the true density. Additionally, they are expected to outcompete DFT orbitals from any a priori density functional approximation.

This makes the inverted orbitals ideal candidates for performing another QMC calculation with them again. Intuitively, the superior orbitals, enriched with the information of the quasi-exact density, should result in a better description of the system and also a lower DMC energy. The QMC calculation utilizing inverted orbitals is called second-order QMC in the following. The outcome of a second-order QMC calculation is given below. Higher orders, which were not calculated in the scope of this thesis, result from repeated cycles of QMC calculations and inversions, cf. Fig. 11.3.

Note that this procedure is an extraordinarily subtle task. The various hyper-parameters in a QMC calculation have to be chosen from experience and specifically for the system under study. The inversion of the resulting QMC density is a challenge on its own. The ambiguity in the inverted xc potential also manifests in the orbitals, even though in a less severe and obvious way. However, even calculating a single higher-order cycle requires a large amount of human and computer time. For the DMC method, the nodes of the trial wave function alone determine the DMC energy. In the case of a single-determinant trial wave function, the orbitals alone determine the nodes. While the single determinant has an incorrect topological nodal structure, cf. Sec. 5.6, there is nevertheless a determinat with the optimal nodal structure, i.e. single-determinant nodes, that minimize the DMC energy. The inverted orbitals are a promising candidate for approximating the optimal single-determinant nodes.

11.1. QMC results from inverted orbitals

The process to a second order of QMC depends on many hyper-parameters and qualitative choices. The procedure to arrive at a second order is therefore outlined below.

With the results and experience gathered from the previous sections, the systems Li_2 and N_2 are chosen for the second-order QMC calculation. First, with the DARSEC code a standard DFT calculation in the xKLI approximation is performed on a 181×181 grid. The resulting orbitals are used in the QMC calculations. In VMC, the Jastrow parameter is optimized. The subsequent DMC calculation uses the optimized trial wave function. The resulting QMC density, estimated by the extrapolation estimator based on the VMC and DMC densities, is then subject to the refinement process described in Sec. G. To this point, the procedure represents a standard QMC calculation. Next, the refined density on the 181×181 grid is reduced to a grid of extent 103×103, which is necessary to achieve the desired accuracy in the inversion process since larger grids slow down the inversion significantly. To this end, the density on the larger grid is

Table 11.1.: Second order DMC results based on the initial xKLI calculation and the update rule HAR for the systems Li_2 and N_2. The second-order DMC results are given in the last row. The other results are already reported in Tab. 9.1.

	Li_2		N_2	
	energy	variance Σ^2	energy	variance Σ^2
DFT@xKLI	$-14.870\,580$		$-108.985\,178$	
VMC@xKLI	$-14.930\,64(3)$	$0.406\,71(4)$	$-109.103\,16(5)$	$4.7855(2)$
DMC@xKLI	$-14.991\,43(4)$	$0.382\,04(4)$	$-109.516\,86(3)$	$4.3839(1)$
2nd order DMC	$-14.991\,724(6)$	$0.383\,453(9)$	$-109.516\,77(4)$	$4.4015(1)$

interpolated and subsequently evaluated at the grid points of the smaller grid. The reduction of the grid is an additional step in the regularization scheme used in inversions. The inversion uses the update rule HAR since this update rule is found to be the most reliable and stable and yields smooth xc potentials. The inverted xc potential is represented on the 103 × 103 grid and has to be extrapolated to the more accurate grid of 181 × 181 grid points, for the reasons described in Sec. E.4 and Sec. F. This is again accomplished by interpolation. The inverted xc potential on the 181 × 181 grid is again used in a DFT calculation to self-consistently calculate the orbitals, i.e., the xc potential is kept fixed while the Hartree potential and the orbitals are converged to self-consistency. This step might seem irrelevant since the inverted orbitals could be extrapolated from the inversion grid of size 103 × 103 to the accurate 181 × 181 grid directly. However, the extra DFT calculation ensures a consistent KS system, with correctly normalized orbitals and potentials represented on the same grid. At this point, the inversion of the QMC density yields an accurate KS system that can be studied on its own. With the resulting DFT orbitals, the second-order DMC calculation is finally carried out. From the second-order DMC calculation, energies and densities for Li_2 and N_2 are available. The evolution of the total energy over the course of the second-order QMC calculation is given in Tab. 11.1. The evolution of the density is given by Fig. 11.1 and Fig. 11.2.

Note that the Jastrow parameter is not reoptimized with the new orbitals of the second order. Instead, the Jastrow parameter is the same as in the original QMC calculation. This should not influence the DMC results, since the Jastrow factor does not alter the nodes.

For Li_2, the DMC energy in second order improves over the first-order DMC energy, while for N_2 the second-order DMC energy is slightly worse than in first order. However, the differences are in the range of the time-step error or the range of the DMC energies of different orbitals, cf. the extrapolated results for DMC based on LDA orbitals from Tab. 9.1. The variances for the second-order wave function increase slightly for both systems, indicating less eigenstate character. This is likely related to the fixed Jastrow parameter, which, in second order, is not optimized for the new orbitals. The second order therefore is not significantly improving the QMC results, and hence, neither the nodes nor the remaining aspects of the trial wave function. This contradicts the intuitive assumption that the inverted orbitals are well-suited for representing the nodes.

A remarkable fact concerning DFT orbitals is that for reasonable density functional approximations as different as LDA and xKLI, but also other density functional approximations like PBE [149] or SCAN [150], their orbitals are very similar (not shown). The orbitals only slightly differ quantitatively but show the same qualitative behavior. Since intricate quantitative details in the orbitals can nevertheless significantly affect the Slater determinant's nodes, the orbitals could have a severe impact on the DMC energy. From the results in Tab. 11.1, however, this is not apparent. Neither the first order DMC energies depend significantly on the used density functional, cf. Tab. 9.1, nor the second-order DMC energies differ from the first order energies.

In Fig. 11.1 and Fig. 11.2, the densities for first and second order are quite aligned. There is no significant qualitative difference in the densities. This is an expected result, as densities for the same system tend to be very similar, for example, cf. Fig. 9.3 or Fig. H.3. However, some slight changes in the density can be identified. For example, the density at the inter-atomic bond of Li_2 is increasing in the second order. For the nitrogen molecule, the inter-atomic density is slightly flattened in the second order. The figures 11.1 and 11.2 show DFT and DMC densities, the latter being biased by the mixed-estimator bias. The VMC densities or extrapolated densities are not shown.

The inverted xc potential can be compared to the original xc potential in the lower panels of Fig. 11.1 and Fig. 11.2. The difference in the potentials is significant compared to the difference in the densities. For both systems, inter-shell bumps arise. Note that the original and inverted xc potentials are not connected to the second-order DMC results but are only shown for completeness.

Whether the reported differences in energies and densities of the second order are self-consistent and can be confirmed in higher orders of the feedback loop between DFT and QMC or vanish, change, or oscillate with higher orders, requires the realization of further iterations of the feedback cycle.

11.2. Outlook

Although the second-order DMC energies do not consistently improve upon the first-order energies, the second-order densities do change, and therefore the underlying orbitals and the KS Slater determinant. The examination of third and higher orders is hence an interesting outlook. It is common for a self-consistent iteration not to show monotonous improvements in its first steps. Fig. 11.3 puts the higher orders in an iterative scheme, where inverted orbitals from a QMC density are fed back into the QMC method to calculate the next order. Higher orders could reveal if the self-consistent cycle is only oscillating around the results of the first order, or if a significant improvement is reached before self-consistency is achieved.

Iterating higher orders of the DFT-QMC cycle to self-consistency generates a unique KS system whose properties are yet to be determined. Two leading principles govern the self-consistent cycle. First, the QMC method aims to find a state of minimal energy constrained by the given nodes, which directly originate from the orbitals. Second, the inversion conserves the quasi-exact QMC density with some KS orbitals. If the QMC methods recover the true ground state density in the self-consistent limit, then the corresponding inverted KS system must also resemble the true density. A KS system that reproduces the true density is the true KS system. Hence, the orbitals must play a double role in generating the true density and, at the same time, should yield optimal nodes for the DMC wave function. Only if the true KS system coincidentally also provides the optimal single-determinantal nodes can the iterative scheme reach both goals simultaneously. This hope is questioned by Ref. [151], at least for a model system with reduced dimensionality and some model interaction. Without this coincidence, the iterative scheme might converge to some intermediate Slater determinant, deviating from the true KS system and also deviating from the optimal single-determinant energy.

Nevertheless, this KS system might be of interest since the DMC enhances the given orbitals via a refined density in every iteration. Interesting cases to further analyze by higher order QMC are the Li-atom or the Be-atom, for which the exact nodes are analytically known [86] and available for comparison. Additionally, the limit of vanishing electron-electron interaction strength might give further insights. In this limit, the non-interacting system with vanishing effective potential, $v_{\mathrm{KS}} = 0$, yields the true wave function. It also exhibits the true nodes and, at the same time, the true density. Deviating from this limit slightly, for example, with the help of the Hellmann-Feynman theorem, might allow for assessing the role of the effective potential.

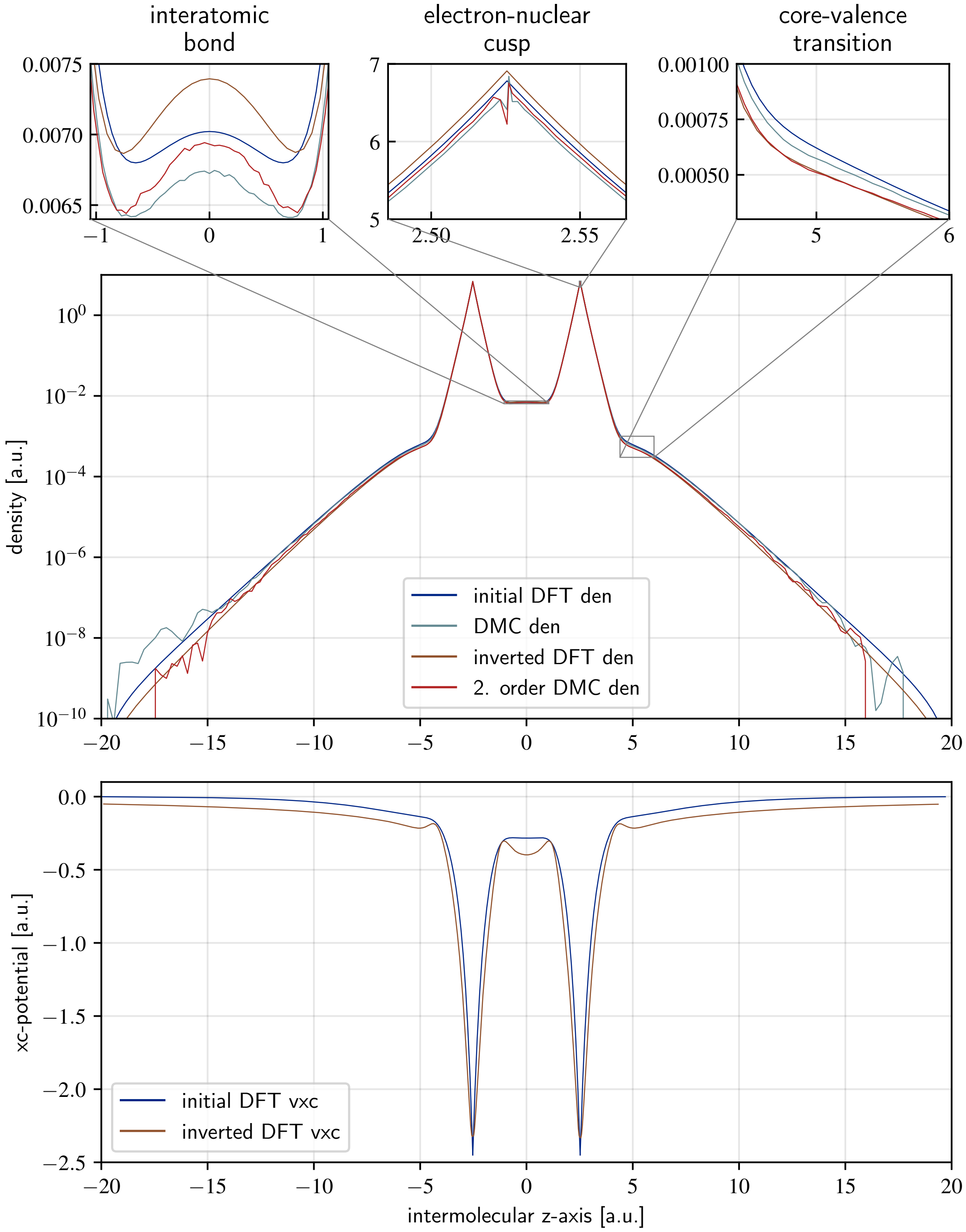


Figure 11.1.: Upper panel: Densities of the diatomic lithium Li_2 system starting from a standard DFT calculation based on the xKLI functional approximation to second-order QMC. Lower panel: The original xc potential from the self-consistent DFT calculation and the xc potential inverted from the first-order QMC density.

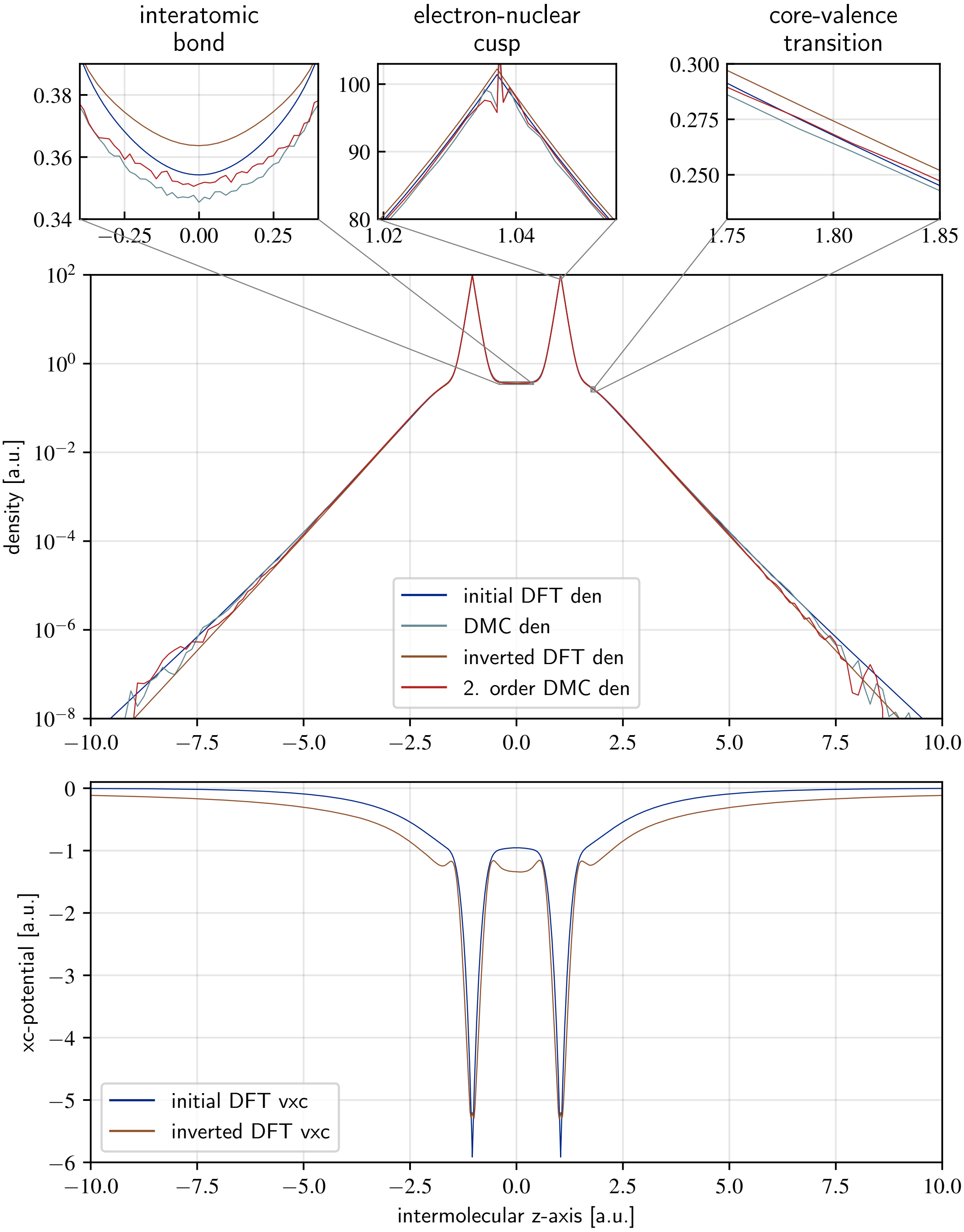


Figure 11.2.: Upper panel: Densities of the nitrogen molecule N_2 system starting from a standard DFT calculation based on the xKLI functional approximation to second-order QMC. Lower panel: The original xc potential from the self-consistent DFT calculation and the xc potential inverted from the first-order QMC density.

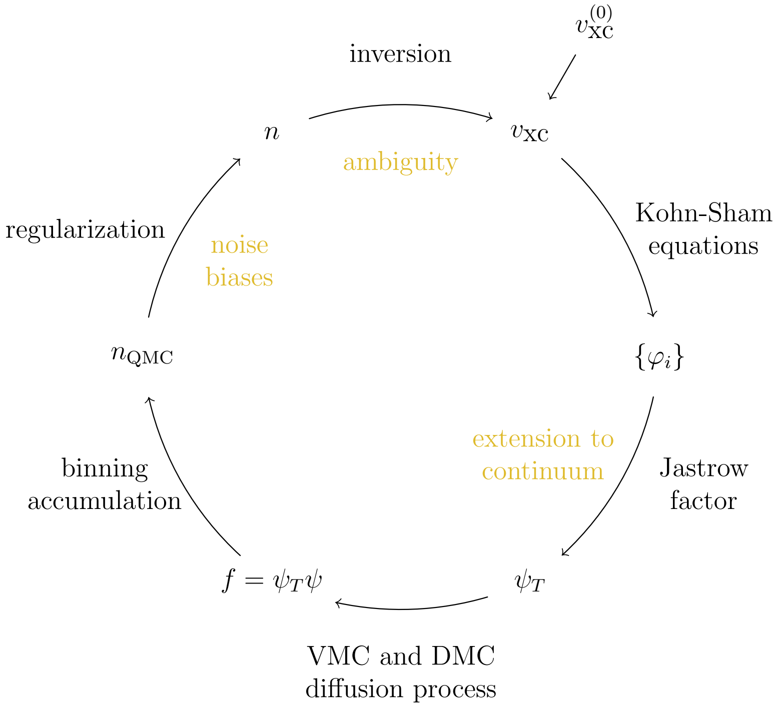


Figure 11.3.: The sketch of an iterative scheme to couple the QMC approach and density-inversion. From some initial orbitals a QMC calculation generates a quasi-exact density. The density is inverted to yield a KS system with inverted orbitals and therefore an inverted Slater determinant. Then the QMC calculation can be restarted with the inverted quantities. Iteration of the scheme defines a feedback loop, which might converge to some self-consistent solution.

12. Conclusion

This thesis provides a detailed self-contained introduction to the variational and diffusion Monte Carlo methods and their underlying concepts. The presented QMC methods have been implemented in a new and efficient QMC code that allows for exploring new computational strategies. The trial wave function used in QMC is deliberately kept simple in this first study. Although more sophisticated forms of trial wave functions, mainly multi-determinantial wave functions including sums over Slater determinants, are readily available and might be implemented in the future, this work focuses on the orbitals' representation on a specialized real-space grid. In a coordinate system such as the prolate spheroidal coordinates, with singular contributions to the gradient and Laplacian, converting the finite difference representation of the orbitals to a continuous representation is challenging. The interpolation from the real-space grid has been tailored thoroughly to accurately describe the orbitals in the vicinity of the nuclei and let the wave function's cusp cancel the Coulomb singularity. The resulting energies are highly accurate, recovering 90 to 97% of all correlation effects.

The corresponding QMC densities are estimated on a histogram that resembles the densities' characteristics to reduce the variance significantly. With this specialized binning method, quasi-exact densities are readily available. However, the proposed propagating estimator promises to enhance the quality of the resulting density further. The fluctuations of the QMC densities are analyzed to assess the statistical error of the densities. The regions of large density fluctuations, i.e., in the vicinity of the nuclei and the asymptotic domain, are remedied by a regularization technique that exploits the exact exponential behavior of the density in these regions of space. Thus, densities with minimal fluctuations can be constructed.

With such densities available, inversions of the KS map are carried out, i.e., an xc potential is deduced from a given density. Inversion problems are ill-posed and notoriously unstable in their results for the xc potential as well as their computational procedure. The pathological characteristics of inversions are studied in detail, and the general limitations of density inversions are elaborated. The inverse problem is shown to be bound to a significant level of ambiguity in the xc potential, and the chosen inversion algorithm already determines some properties of the resulting potentials. A range of iterative inversion procedures have been developed that are capable of inverting densities with statistical noise. The presence of noise in the density reveals how the inversion overemphasizes numerical artifacts. With this general assessment of the inversion process, accurate xc potentials are constructed and their properties are determined reliably. The inverse problem is contrasted with the exceptionally well-behaved forward KS map and its extensive linear regime for the density response to variations in the potential. A new inversion procedure based on potential fitting is proposed from the forward map's linearity.

Finally, a concept to improve QMC results with the aid of the exact KS orbitals is realized by combining the central topics of this thesis: accurate QMC results, estimation of the density, and reliable inversions. The exact KS orbitals construct a presumably superior trial wave function. However, the results based on this second-order wave function resemble the original QMC results calculated with approximate DFT orbitals. An outlook of the second-order approach discusses improving the second-order scheme.

A. Slater determinants

This chapter discusses properties of Slater determinants and numerical approaches for their efficient evaluation. The Slater determinant handles the anti-symmetry requirement for fermionic wave functions and thus introduces great redundancy in the description of the state, and potentially also in practical calculations. To avoid numerical redundancies and to arrive at efficient computations, the representation of the wave function is slightly altered in Sec. A.1. In Sec. A.2, the evaluation of Slater determinants and their derivatives is accelerated by the matrix determinant lemma.

A.1. Product representation

This section shows that a partially anti-symmetrized form of the wave function suffices to describe the state and that full anti-symmetry is not required in QMC.

With the antisymmetrizer of Eq. (5.2), the standard form of a Slater determinant of Eq. (5.3) can be rewritten in a simple but lengthy calculation as

$$\Phi(\mathbf{X}) = \tilde{N}\hat{\mathrm{A}}\left(\prod_{i=1}^{N_\alpha}\alpha(s_i)\right)\left(\prod_{i=N_\alpha+1}^{N}\beta(s_i)\right)\phi(\mathbf{R}). \tag{A.1}$$

The partially anti-symmetric form

$$\phi(\mathbf{R}) = \phi_\alpha(\mathbf{R}_\alpha)\phi_\beta(\mathbf{R}_\beta) \tag{A.2}$$

is the product of spin-free, normalized, and anti-symmetric determinants, $\phi_{\alpha,\beta}$, of the α- and β-channel orbitals, reading

$$\phi_\sigma(\mathbf{R}_\sigma) = \sqrt{N_\sigma!}\hat{\mathrm{A}}\prod_{i=1}^{N_\sigma}\varphi_{\sigma i}(\mathbf{R}_{\sigma,i}),\ \ \sigma = \alpha,\beta. \tag{A.3}$$

The product ϕ is anti-symmetric for each spin channel separately, but has no definite symmetry for permutations exchanging coordinates corresponding to different spins. The normalization factor

$$\tilde{N} = (N!/N_\alpha!N_\beta!)^{\frac{1}{2}} \tag{A.4}$$

accounts for and corrects the double countings due to multiple anti-symmetrizers (one in Eq. (A.1) and one for each spin channel in Eq. (A.3)).

Omitting the normalization factor and anti-symmetrizer in front of Eq. (A.1) results in a wave function

$$\tilde{\Phi}(\mathbf{X}) = \left(\prod_{i=1}^{N_\alpha}\alpha(s_i)\right)\left(\prod_{i=N_\alpha+1}^{N}\beta(s_i)\right)\phi(\mathbf{R}) \tag{A.5}$$

that is not fully anti-symmetric in its coordinates but nevertheless normalized. This form yields expectation values and overlap integrals identical to those from the fully anti-symmetrized form Φ. The proof follows below. Additionally, this form attaches the first N_α coordinates to the α channel and the remaining ones to the β channel, meaning the first N_α coordinates are the arguments to the Slater determinant of the α-orbitals, and vice versa. $\tilde{\Phi}$ is only anti-symmetric in the spatial coordinates of each spin channel separately but asymmetric across spin channels. The advantage of expression (A.5) is its product form of functions of spin and spatial coordinates, cf. Sec. 5.1.

To prove the equivalence of $\tilde{\Phi}$ and Φ regarding overlap integrals, suppose $\tilde{\Phi}'$ and $\tilde{\Phi}$ are of the form of Eq. (A.5) with the same spin configuration but different spatial parts, ϕ' and ϕ, respectively, and $\hat{\mathrm{O}}$ is a spin-independent and symmetric observable. Then

$$\begin{aligned}\langle\tilde{\Phi}'|\hat{\mathrm{O}}|\tilde{\Phi}\rangle &= \sum_{s_1\dots s_N}\left(\prod_{i=1}^{N_\alpha}|\alpha(s_i)|^2\right)\left(\prod_{i=N_\alpha+1}^{N}|\beta(s_i)|^2\right)\int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\mathbf{R})\,\mathrm{d}\mathbf{R} &\text{(A.6)}\\ &= \int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\mathbf{R})\,\mathrm{d}\mathbf{R}. &\text{(A.7)}\end{aligned}$$

Note that the spin functions α, β form an orthonormal basis,

$$\sum_s \sigma(s)\sigma'(s) = \delta_{\sigma,\sigma'} \text{ with } \sigma, \sigma' \in \{\alpha, \beta\}. \tag{A.8}$$

Next, the same calculation is written down for the corresponding fully symmetrized determinant of Eq. (A.1).

$$\begin{aligned}\langle\Phi'|\hat{\mathrm{O}}|\Phi\rangle =&\frac{\tilde{N}^2}{N!^2}\sum_{s_1\dots s_N}\int\sum_{\pi,\tau\in\mathbb{S}(N)}\mathrm{sign}(\pi)\mathrm{sign}(\tau)\\ &\times\alpha^*(s_{\pi(1)})\alpha(s_{\tau(1)})\dots\beta^*(s_{\pi(N)})\beta(s_{\tau(N)})\phi'^*(\pi\mathbf{R})\hat{\mathrm{O}}\phi(\tau\mathbf{R})\,\mathrm{d}\mathbf{R}\\ =&\frac{\tilde{N}^2}{N!^2}\sum_{\pi,\gamma\in\mathbb{S}(N)}\mathrm{sign}(\pi\gamma\pi)\\ &\times\left[\sum_{s_1\dots s_N}\alpha^*(s_{\pi(1)})\alpha(s_{\gamma\pi(1)})\dots\beta^*(s_{\pi(N)})\beta(s_{\gamma\pi(N)})\right]\\ &\times\int\phi'^*(\pi\mathbf{R})\hat{\mathrm{O}}\phi(\gamma\pi\mathbf{R})\,\mathrm{d}\mathbf{R}\\ =&\frac{\tilde{N}^2}{N!}\sum_{\gamma\in\mathbb{S}(N)}\mathrm{sign}(\gamma)\\ &\times\left[\sum_{s_1\dots s_N}\alpha^*(s_1)\alpha(s_{\gamma(1)})\dots\beta^*(s_N)\beta(s_{\gamma(N)})\right]\int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\gamma\mathbf{R})\,\mathrm{d}\mathbf{R}\end{aligned}\tag{A.9}$$

The permutation τ was substituted with $\gamma\pi$, and the reordering of the coordinates in $s_1, \dots s_N$ and $\mathbf{R}$ subsequently eliminates any action of the permutation π, making the sum over π count all $N!$ permutations in the symmetric group $\mathbb{S}(N)$. Note that the last step reveals $\langle\Phi'|\hat{\mathrm{O}}|\Phi\rangle = \tilde{N}\langle\tilde{\Phi}'|\hat{\mathrm{O}}|\Phi\rangle$.

Due to orthonormality of the spin functions, the expression in the square brackets is either zero or one. If γ maps any index from $\{1\dots N_\alpha\}$ to $\{N_\alpha+1\dots N\}$ the square bracket is zero, and the corresponding summand does not contribute to the summation. The remaining permutations with non-zero summands are restricted to

$$\begin{aligned}\gamma(\{1\dots N_\alpha\}) &= \{1\dots N_\alpha\} &\text{(A.10)}\\ \gamma(\{N_\alpha+1\dots N\}) &= \{N_\alpha+1\dots N\} &\text{(A.11)}\end{aligned}$$

Therefore, every permutation γ can be split into $\gamma = \gamma_\alpha\gamma_\beta$ with $\gamma_\alpha \in \mathbb{S}(\{1\dots N_\alpha\})$ and $\gamma_\beta \in \mathbb{S}(\{N_\alpha+1,\dots N\})$. A permutation such as $\gamma_\alpha\gamma_\beta$ acting on the spatial Slater determinant ϕ is readily evaluated as

$$\begin{aligned}
\phi(\gamma_\alpha\gamma_\beta\mathbf{R}) &= \phi_\alpha(\gamma_\alpha\gamma_\beta\mathbf{R}_\alpha)\phi_\beta(\gamma_\alpha\gamma_\beta\mathbf{R}_\beta) = \phi_\alpha(\gamma_\alpha\mathbf{R}_\alpha)\phi_\beta(\gamma_\beta\mathbf{R}_\beta) \\
&= \operatorname{sign}(\gamma_\alpha)\phi_\alpha(\mathbf{R}_\alpha)\operatorname{sign}(\gamma_\beta)\phi_\beta(\mathbf{R}_\beta) \\
&= \operatorname{sign}(\gamma_\alpha\gamma_\beta)\phi(\mathbf{R}).
\end{aligned} \tag{A.12}$$

Together this yields

$$\begin{aligned}
\langle\Phi'|\hat{\mathrm{O}}|\Phi\rangle =& \frac{\tilde{N}^2}{N!}\sum_{\gamma_\alpha\in\mathbb{S}(\{1\ldots N_\alpha\})}\ \sum_{\gamma_\beta\in\mathbb{S}(\{N_\alpha+1,\ldots N\})}\operatorname{sign}(\gamma_\alpha\gamma_\beta)\int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\gamma\mathbf{R})\,\mathrm{d}\mathbf{R} \\
=& \frac{\tilde{N}^2}{N!}\sum_{\gamma_\alpha\in\mathbb{S}(\{1\ldots N_\alpha\})}\ \sum_{\gamma_\beta\in\mathbb{S}(\{N_\alpha+1,\ldots N\})}\operatorname{sign}(\gamma_\alpha\gamma_\beta)^2\int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\mathbf{R})\,\mathrm{d}\mathbf{R} \\
=& \tilde{N}^2\frac{N_\alpha!N_\beta!}{N!}\int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\mathbf{R})\,\mathrm{d}\mathbf{R} \\
=& \int\phi'^*(\mathbf{R})\hat{\mathrm{O}}\phi(\mathbf{R})\,\mathrm{d}\mathbf{R} = \langle\tilde{\Phi}'|\hat{\mathrm{O}}|\tilde{\Phi}\rangle.
\end{aligned} \tag{A.13}$$

This concludes the proof of $\langle\Phi'|\hat{\mathrm{O}}|\Phi\rangle = \langle\tilde{\Phi}'|\hat{\mathrm{O}}|\tilde{\Phi}\rangle$.

A few notes are appropriate: First, the reduced symmetry of $\tilde{\Phi}$ can always be restored to full anti-symmetry by $\Phi = \tilde{N}\hat{\mathrm{A}}\tilde{\Phi}$. However, $\tilde{N}\hat{\mathrm{A}}$ is only norm-conserving when acting on wave functions with the symmetry of $\tilde{\Phi}$ of Eq. (A.5). Further $\tilde{N}\hat{\mathrm{A}}$ commutes with any symmetric observable, e.g., the Hamiltonian, and therefore $\tilde{\Phi}$ has the same eigenvalue structure as Φ with symmetric operators. As a major consequence of the above, the partially anti-symmetrized Slater determinants of a non-interacting Hamiltonian build an orthonormal basis equivalent to the basis of fully anti-symmetrized Slater determinants. Thus, the form of reduced symmetry also exists for correlated wave functions, as stated in Sec. 5.1.

A.2. Fast determinant evaluation

In a QMC calculation, a heavily repeated and computationally demanding task is the evaluation of Slater determinants. Often, not only is the value of the determinant required but its gradient is also used in the drift vector and the Laplacian in the local energy. The standard evaluation of the determinant of a $n \times n$ matrix is a numerically expensive calculation of complexity $\mathcal{O}(n^3)$. To evaluate probability, drift, and local energy of a Slater determinant ϕ at some coordinate $\mathbf{R}$ therefore requires $1 + 3 \times n + n$ determinant evaluations, resulting in a complexity of $\mathcal{O}(n^4)$. Alternatives exist to evaluating every determinant itself and instead using the properties of Slater matrices

$$\forall i, j: \quad M_{ij}(\mathbf{R}) = \varphi_i(\mathbf{r}_j). \tag{A.14}$$

Derivatives of a Slater determinant $\phi = \det\mathbf{M}$ with respect to a single electron coordinate correspond to single-row or -column change of the Slater matrix. Let L_k be a differential operator that only acts on the variable $\mathbf{r}_k$. Then, by Laplace expansion of the determinant, its derivative is

$$
\begin{aligned}
L_k\phi(\mathbf{R}) &= L_k \sum_{i=1}^{n} (-1)^{k+i} \varphi_i(\mathbf{r}_k)\phi^{\cancel{ik}}(\mathbf{R}) = \\
&= \sum_{i=1}^{n} (-1)^{k+i} \phi^{\cancel{ik}}(\mathbf{R}) L_k \varphi_i(\mathbf{r}_k) = \\
&= \det \begin{pmatrix} \varphi_1(\mathbf{r}_1) & \dots & L\varphi_1(\mathbf{r}_k) & \dots & \varphi_1(\mathbf{r}_n) \\ \vdots & & \vdots & & \vdots \\ \varphi_k(\mathbf{r}_1) & \dots & L\varphi_k(\mathbf{r}_k) & \dots & \varphi_k(\mathbf{r}_n) \\ \vdots & & \vdots & & \vdots \\ \varphi_n(\mathbf{r}_1) & \dots & L\varphi_n(\mathbf{r}_k) & \dots & \varphi_n(\mathbf{r}_n) \end{pmatrix} .
\end{aligned}
\tag{A.15}
$$

$\phi^{\cancel{ik}}$ is the sub-determinant, without orbital i and without coordinate k. For Slater determinants, constructing the determinant and taking the derivative commutes. The derivative of the Slater matrix is different from the Slater matrix only in a single column. This property allows the application of the matrix determinant lemma, which is shortly stated here for completeness. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be column-vectors and $\mathbf{M} \in \mathbb{R}^{n\times n}$ a non-singular matrix with determinant $\phi = \det \mathbf{M}$. The matrix

$$
\mathbf{M}' = \mathbf{M} + \mathbf{u}\mathbf{v}^{\mathrm{T}} = \begin{pmatrix} M_{11} + u_1 v_1 & \dots & M_{1n} + u_1 v_n \\ M_{21} + u_2 v_1 & \dots & M_{2n} + u_2 v_n \\ \vdots & \ddots & \vdots \\ M_{n1} + u_n v_1 & \dots & M_{nn} + u_n v_n \end{pmatrix}
\tag{A.16}
$$

originates from $\mathbf{M}$ by addition of a dyadic product only. The determinant of $\mathbf{M}'$ is given by the matrix determinant lemma as

$$
\phi' = \det \mathbf{M}' = \left(1 + \mathbf{v}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{u}\right) \phi .
\tag{A.17}
$$

If the inverse $\mathbf{M}^{-1}$ is known, the new determinant can be calculated with $\mathcal{O}(n^2)$ complexity in general.

Comparing Eq. (A.15) and Eq. (A.16), the derivative of a Slater matrix can be cast as a dyadic update with vectors

$$
\forall i: \quad u_i = L\varphi_i(\mathbf{r}_k) - \varphi_i(\mathbf{r}_k) \tag{A.18}
$$

$$
\forall j: \quad v_j = \delta_{kj} \tag{A.19}
$$

The specific form of $\mathbf{u}$ with the subtraction of the original value $\varphi_i(\mathbf{r}_k) = \mathbf{M}_{ik}$ results in the special formula

$$
\begin{aligned}
L_k\phi &= \phi \left(1 + \sum_i M^{-1}_{ki} \left(L\varphi_i(\mathbf{r}_k) - \varphi_i(\mathbf{r}_k)\right)\right) = \\
&= \phi \left(1 + \sum_i M^{-1}_{ki} L\varphi_i(\mathbf{r}_k) - \sum_i M^{-1}_{ki} M_{ik}\right) = \\
&= \phi \sum_i M^{-1}_{ki} L\varphi_i(\mathbf{r}_k)
\end{aligned}
\tag{A.20}
$$

which has even less complexity than the general case, namely the complexity $\mathcal{O}(n)$ of a scalar product. The gradient in x-, y- and z-direction and the Laplacian result from replacing L by ∂_x,

∂_y, ∂_z and ∇^2, respectively. For every coordinate $\mathbf{R}$, one thus computes the inverse matrix in $\mathcal{O}(n^3)$ runtime first from a LU decomposition. Then, the determinant is generated trivially from the LU matrices, and the gradients and Laplacians of the determinant follow with $\mathcal{O}(n)$ runtime. Overall, the evaluation reduces to a computational cost of $\mathcal{O}(n^3)$.

B. Critical characteristics of wave functions

Wave functions exhibit special characteristics near certain points or hypersurfaces in configuration space, like particle coincidences or wave function nodes. In the following, the characteristics of the exact and the single-reference Slater-Jastrow wave function are derived and compared.

To first assess the general behavior of the wave function when two particles meet, consider the particles A and B with charges q_A and q_B and with masses m_A and m_B. Both particles are assumed to interact with each other and are accounted for in the Hamiltonian. Let the separation of the two coinciding particles be $\boldsymbol{\varepsilon} = \mathbf{r}_A - \mathbf{r}_B$, defining a relative coordinate. The corresponding center-of-mass coordinate is $\mathbf{s} = (m_A\mathbf{r}_A + m_B\mathbf{r}_B)/(m_A + m_B)$ with the reduced mass $\mu = m_A m_B/(m_A + m_B)$. The Hamiltonian can be expressed in the new coordinates accounting for the transformed kinetic contributions $\nabla_A^2/2m_A + \nabla_B^2/2m_B$,

$$\hat{\mathrm{H}} = \hat{\mathrm{H}}_s(\mathbf{s}, \tilde{\mathbf{R}}) + \frac{1}{2(m_A + m_B)}\nabla_{\mathbf{s}}^2 + \frac{1}{2\mu}\nabla_{\boldsymbol{\varepsilon}}^2 + \frac{q_A q_B}{|\boldsymbol{\varepsilon}|}. \tag{B.1}$$

This renders the singularity to be placed at $\boldsymbol{\varepsilon} \to 0$, irrespective of $\mathbf{s}$. The operator $\hat{\mathrm{H}}_s$ unites all remaining parts of the Hamiltonian that are smooth, with $\tilde{\mathbf{R}}$ representing all particle coordinates except those of particles A and B. Note that any pair of particles, other than particles A and B, is assumed to be well separated with some sufficient lower bound on their distance, such that their potential and kinetic contributions in the Hamiltonian can be neglected compared to the singular contributions of the meeting particles. The coordinate-transformed Schrödinger equation in the vicinity of the singularity $\boldsymbol{\varepsilon} \to 0$ with a non-vanishing wave function value was shown to impose a cusp in the wave function by Kato [69]. Ref. [152] gives a comprehensive review. The wave function has the asymptotic form

$$\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[1 + q_A q_B \mu |\boldsymbol{\varepsilon}| + \mathbf{w} \cdot \boldsymbol{\varepsilon}\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \tag{B.2}$$

with some vector $\mathbf{w}$, that depends on the location of the other particles and is usually unknown, $\mathbf{w} = \mathbf{w}(\tilde{\mathbf{R}})$. This form is realized in the case of electron-nucleus coincidences and electron-electron coincidences for electrons of opposite spin. This particular form for the non-vanishing case is often written using the spherically averaged wave function

$$\hat{\psi}(|\boldsymbol{\varepsilon}|) = \frac{1}{4\pi}\int \psi(\boldsymbol{\varepsilon})\,\mathrm{d}\Omega_{\boldsymbol{\varepsilon}}, \tag{B.3}$$

where the integration is over the spherical coordinates $\Omega_{\boldsymbol{\varepsilon}}$ of the relative coordinate $\boldsymbol{\varepsilon}$ around the coincident point $\boldsymbol{\varepsilon} = 0$. With $\hat{\psi}$ the cusp condition takes the form

$$\lim_{|\boldsymbol{\varepsilon}| \to 0} \partial_{|\boldsymbol{\varepsilon}|} \ln \hat{\psi}(|\boldsymbol{\varepsilon}|) = q_A q_B \mu, \tag{B.4}$$

in which the scalar product or dipole contribution $\mathbf{w} \cdot \boldsymbol{\varepsilon}$ was averaged out. Although the averaged form is often used in the literature, omitting the dipole contribution neglects a characteristic of the cusp, and the non-averaged form is preferred in the following.

The case of a particle incident with vanishing wave function is not covered by Eq. (B.2). In Ref. [152], this case is also analyzed for two meeting electrons. Applied to the electron-electron coalescence, the cusp condition reads

$$\psi(\boldsymbol{\varepsilon}) = \mathbf{w}\cdot\boldsymbol{\varepsilon}\left[1+\frac{1}{4}|\boldsymbol{\varepsilon}|\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2). \tag{B.5}$$

B.1. Electron-nucleus coalescence

Applying the cusp condition to the case of an electron approaching a nucleus of charge Z and reduced mass μ, its value, gradient, and Laplacian read

$$\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[1 - Z\mu|\boldsymbol{\varepsilon}| + \mathbf{w}\cdot\boldsymbol{\varepsilon}\right] + \frac{1}{2}\boldsymbol{\varepsilon}\cdot\mathbf{t}\boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^3), \tag{B.6}$$

$$\nabla_{\boldsymbol{\varepsilon}}\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[-Z\mu\hat{\boldsymbol{\varepsilon}} + \mathbf{w}\right] + \mathbf{t}_s\boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \tag{B.7}$$

$$\nabla^2_{\boldsymbol{\varepsilon}}\psi(\boldsymbol{\varepsilon}) = \psi(0)\left[-\frac{2Z\mu}{|\boldsymbol{\varepsilon}|}\right] + \operatorname{tr}(\mathbf{t}) + \mathcal{O}(|\boldsymbol{\varepsilon}|^1). \tag{B.8}$$

Here, $\mathbf{t}$ is some matrix with symmetric part $\mathbf{t}_s = (\mathbf{t}+\mathbf{t}^T)/2$. The reduced mass $\mu \approx 1$ is neglected due to the large nuclear mass compared to the electron mass. For an exact wave function $\psi = \psi_0$, the local energy E_L is constant, and can be written as

$$E_L(\boldsymbol{\varepsilon}) = E_s - \frac{Z}{|\boldsymbol{\varepsilon}|} - D\frac{\nabla^2_{\boldsymbol{\varepsilon}}\psi_0(\boldsymbol{\varepsilon})}{\psi_0(\boldsymbol{\varepsilon})} = E_s - \operatorname{tr}(\mathbf{t}). \tag{B.9}$$

The energy E_s corresponds to $\hat{\mathrm{H}}_s$ and incorporates the bound kinetic and potential energy of all other particles. The Coulomb singularity of the potential energy is canceled exactly by the singularity of the kinetic energy.

The drift velocity $\mathbf{u}_k$ of the exact wave function for electron k is governed by the limiting behavior of

$$\mathbf{u}_k(\boldsymbol{\varepsilon}) = 2D\frac{\nabla_{\boldsymbol{\varepsilon}}\psi_0(\boldsymbol{\varepsilon})}{\psi_0(\boldsymbol{\varepsilon})} = 2D\frac{-Z\hat{\boldsymbol{\varepsilon}} + \mathbf{w} + \mathbf{t}_s\boldsymbol{\varepsilon}}{1 - Z|\boldsymbol{\varepsilon}| + \boldsymbol{\varepsilon}\cdot\mathbf{w}} \to 2D\left(-Z\hat{\boldsymbol{\varepsilon}} + \mathbf{w}\right). \tag{B.10}$$

Thus, the drift velocity points to the nucleus, up to a constant vector, showing a cusp but neither a discontinuity nor a divergence.

Next, the trial wave function is examined. The Jastrow factor and Slater determinant both contribute to the cusp condition. The Jastrow factor, which might explicitly incorporate a nuclear cusp in J_{en}, readily splits into a smooth and a cusped part

$$J = J_s(\mathbf{s}, \tilde{\mathbf{R}}) + \beta u(|\boldsymbol{\varepsilon}|), \tag{B.11}$$

with a Jastrow parameter β. Expanded to second order in $\boldsymbol{\varepsilon}$, the Jastrow factor reads

$$J - J_s = \beta u(|\boldsymbol{\varepsilon}|) = \beta\left(|\boldsymbol{\varepsilon}| + \frac{u''(0)}{2}|\boldsymbol{\varepsilon}|^2\right) + \mathcal{O}(|\boldsymbol{\varepsilon}|^3), \tag{B.12}$$

with $u(0) = 0$ and $u'(0) = 1$ from Eq. (5.19). To access the cusp in the Slater determinant, assume any orbital φ_i in ϕ has a cusp at the nuclear position of magnitude α_i (and a dipole contribution $\mathbf{w}_i$) and treat the orbital as a function of $\boldsymbol{\varepsilon}$.

$$\varphi_i(\boldsymbol{\varepsilon}) = \varphi_i(0)\left[1 + \alpha_i|\boldsymbol{\varepsilon}| + \mathbf{w}_i\cdot\boldsymbol{\varepsilon}\right] + \mathcal{O}(|\boldsymbol{\varepsilon}|^2) \tag{B.13}$$

The spin dependence of the orbital and determinant is suppressed for clarity. Define, for the brevity of notation, a Slater determinant $\phi^{\cancel{ik}}$, omitting the kth coordinate and the ith orbital,

$$\phi^{\cancel{ik}}(\mathbf{R}_{1:N\setminus k}) = \det \varphi_{1:N\setminus i}(\mathbf{R}_{1:N\setminus k}). \tag{B.14}$$

Then, expanding the determinant at the nucleus in a Laplace expansion yields

$$\phi(\mathbf{R}) = \sum_i (-1)^{k+i} \phi^{\cancel{ik}}(\mathbf{R}_{1:N\setminus k}) \varphi_i(\mathbf{r}_k) = \hat{\mathrm{S}}_{ik} \varphi_i(\mathbf{r}_k), \tag{B.15}$$

where the last term defines the operator $\hat{\mathrm{S}}_{ik}$ for simpler notation. Note that any choice for $k \in \{1, \dots N\}$ is legitimate. Then, both expansions can be inserted into the Eq. (5.21) for the gradient of the trial wave function.

$$\nabla_{\varepsilon} \psi_T = e^{J_s} \hat{\mathrm{S}}_{ik} \varphi_i(0) \Big[(\alpha_i + \beta)\, \hat{\varepsilon} + \mathbf{w}_i \Big] + \mathcal{O}(|\varepsilon|) \tag{B.16}$$

This shows that the trial wave function's cusp is inherited from the cusp of the Jastrow factor and the Slater determinant's cusp.

The cusp condition is fulfilled only if the term $(\alpha_i + \beta)\hat{\varepsilon}$ is independent of the orbital index i since only in this case can the expression reduce to the exact cusp condition. Additionally, the magnitude of the cusp is fixed to

$$\forall i: \quad \alpha_i + \beta = -Z. \tag{B.17}$$

Hence, every orbital has to individually fulfill the cusp condition with the same cusp. If all orbitals simultaneously fail to reproduce the cusp but have a common, identical cusp, the Jastrow factor can compensate for the global error. In the standard case, either the orbitals or the Jastrow satisfy the cusp condition.

The summation in $\hat{\mathrm{S}}_{ik}$ averages over the dipole vectors $\mathbf{w}_i$ to yield some new vector $\bar{\mathbf{w}}$. The drift velocity at the nucleus is then given by

$$\mathbf{u}_k(\varepsilon) = 2D\,[(\alpha + \beta)\hat{\varepsilon} + \bar{\mathbf{w}}] = 2D\,[-Z\hat{\varepsilon} + \bar{\mathbf{w}}] \tag{B.18}$$

Determining the behavior of the local energy requires an longer but elementary calculation. Taking the transformed Hamiltonian into account and applying it to the trial wave function yields

$$E_L(\varepsilon) - E_s = -\frac{Z}{|\varepsilon|} - \frac{1}{2}\frac{\alpha + \beta}{|\varepsilon|} + \beta \hat{\varepsilon} \cdot \bar{\mathbf{w}} + \mathcal{O}(0) = \beta \hat{\varepsilon} \cdot \bar{\mathbf{w}} + \mathcal{O}(0). \tag{B.19}$$

The cusp condition of Eq. (B.17) is enforced in the last equal sign to cancel the singularities. For a trial wave function that violates the cusp condition, its local energy exhibits a Coulombic divergence at the nucleus. For exact cusps in the trial wave function, there still is a discontinuity in E_L due to $\hat{\varepsilon} \cdot \bar{\mathbf{w}}$.

B.2. Electron-electron coalescence

The same analysis can be applied to the electron-electron coalescence points with the distinction of the triplet and singlet cases. When two electrons approach another with opposite spin (singlet) or parallel spin (triplet), from Eq. (B.2) and Eq. (B.5) the exact wave function at the incident point reads

$$\psi(\varepsilon) = \begin{cases} \psi(0)\left(1 + \frac{1}{2}|\varepsilon|\right) + \mathcal{O}(|\varepsilon|^2), & \text{singlet} \\ \varepsilon \cdot \mathbf{w}\left(1 + \frac{1}{4}|\varepsilon|\right) + \mathcal{O}(|\varepsilon|^3), & \text{triplet.} \end{cases} \tag{B.20}$$

To arrive here, $q_A q_B = 1$ and $\mu = 1/2$ are used. The gradient

$$\nabla_{\boldsymbol{\varepsilon}} \psi(\boldsymbol{\varepsilon}) = \begin{cases} \frac{1}{2}\psi(0)\hat{\boldsymbol{\varepsilon}} + \mathcal{O}(|\boldsymbol{\varepsilon}|), & \text{singlet} \\ \left(1 + \frac{|\boldsymbol{\varepsilon}|}{4}\right) \mathbf{w} + \frac{1}{4}(\boldsymbol{\varepsilon} \cdot \mathbf{w})\hat{\boldsymbol{\varepsilon}} + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), & \text{triplet} \end{cases} \tag{B.21}$$

is bound at the coalescence point, while the drift vector

$$\mathbf{u}(\boldsymbol{\varepsilon}) = 2D \begin{cases} \frac{1}{2}\hat{\boldsymbol{\varepsilon}}, & \text{singlet} \\ \frac{1}{4}\hat{\boldsymbol{\varepsilon}} + \frac{\mathbf{w}}{\boldsymbol{\varepsilon}\cdot\mathbf{w}}, & \text{triplet} \end{cases} \tag{B.22}$$

shows a discontinuity in singlet and triplet cases. The triplet case even features a singularity, which is entirely due to the node in the wave function and not caused by the Coulomb singularity. For the trial wave function of Eq. (5.14), the partially anti-symmetric Slater determinant is not analytically accessible. A generic expansion reads

$$\phi(\boldsymbol{\varepsilon}) = \phi(0) + \mathbf{g} \cdot \boldsymbol{\varepsilon} + \frac{1}{2}\boldsymbol{\varepsilon} \cdot \mathbf{t}\boldsymbol{\varepsilon} + \frac{1}{6}\mathbf{k} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^4), \tag{B.23}$$

or in index notation with Einstein summation,

$$\phi(\boldsymbol{\varepsilon}) = \phi(0) + \mathbf{g}_i \varepsilon_i + \frac{1}{2}\mathbf{t}_{ij}\varepsilon_i\varepsilon_j + \frac{1}{6}\mathbf{k}_{ijk}\varepsilon_i\varepsilon_j\varepsilon_k + \mathcal{O}(|\boldsymbol{\varepsilon}|^4). \tag{B.24}$$

In the partial anti-symmetrized form of Eq. (5.5) and for the triplet case, the anti-symmetry imposes $\phi(-\boldsymbol{\varepsilon}) = -\phi(\boldsymbol{\varepsilon})$. The determinant will show a node of first order in the triplet case, $\phi(0) = 0$, and $\mathbf{t}$ must either be zero or at least skew-symmetric. The higher orders are the gradient $\mathbf{g}_i = \partial_i \phi(0)$, Hessian $\mathbf{t}_{ij} = \partial_i \partial_j \phi(0)$ and the triple derivative tensor $\mathbf{k}_{ijk} = \partial_i \partial_j \partial_k \phi(0)$. The expansions gradient and Laplacian are given by

$$\begin{aligned} \nabla\phi(\boldsymbol{\varepsilon}) &= \mathbf{g} + \mathbf{t}^S\boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \\ \nabla^2\phi(\boldsymbol{\varepsilon}) &= \operatorname{tr}(\mathbf{t}) + \mathbf{k}^{**} \cdot \boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \end{aligned} \tag{B.25}$$

with $\mathbf{t}^S = (\mathbf{t} + \mathbf{t}^T)/2$ and $\mathbf{k}^{**} = \hat{\mathbf{e}}_i(k_{ijj} + k_{jij} + k_{jji})/3$.

The Jastrow factor J_{ee}, on the other hand, is tailored to the electron-electron coalescence. It can, again, be split into a smooth and a cusped part,

$$J - J_s = \beta_{ee} u(|\boldsymbol{\varepsilon}|) = \beta_{ee}\left(|\boldsymbol{\varepsilon}| + \frac{u''(0)}{2}|\boldsymbol{\varepsilon}|^2\right) + \mathcal{O}(|\boldsymbol{\varepsilon}|^3). \tag{B.26}$$

The cusp parameter β_{ee} is either $\beta_{ee}^S = 1/2$ or $\beta_{ee}^T = 1/4$. The full trial wave function's drift vector from Eq. (5.22) is

$$\mathbf{u}_k(\boldsymbol{\varepsilon})/2D = \beta_{ee}\hat{\boldsymbol{\varepsilon}} + \frac{\mathbf{g} + \mathbf{t}^S\boldsymbol{\varepsilon}}{\phi(0) + \mathbf{g}\cdot\boldsymbol{\varepsilon}} = \beta_{ee}\hat{\boldsymbol{\varepsilon}} + \begin{cases} \phi(0)^{-1}\mathbf{g} & \text{singlet} \\ \frac{\mathbf{g}}{\mathbf{g}\cdot\boldsymbol{\varepsilon}} & \text{triplet.} \end{cases} \tag{B.27}$$

Both spin scenarios show a cusp of the form $\beta_{ee}\hat{\boldsymbol{\varepsilon}}$, and at the triplet coalescence, the drift vector diverges. The local Laplacian of the trial wave function in the singlet case reads

$$\frac{\nabla_{\boldsymbol{\varepsilon}}^2 \psi_T(\boldsymbol{\varepsilon})}{\psi_T(\boldsymbol{\varepsilon})} = \beta_{ee}^2 + \begin{cases} \frac{2\beta_{ee}}{|\boldsymbol{\varepsilon}|} + 3\beta_{ee}u''(0) + \phi(0)^{-1}\left(\operatorname{tr}(\mathbf{t}) + 2\beta_{ee}\mathbf{g}\cdot\hat{\boldsymbol{\varepsilon}}\right) & \text{singlet} \\ \frac{4\beta_{ee}}{|\boldsymbol{\varepsilon}|} + 5\beta_{ee}u''(0) + \frac{\mathbf{k}^{**}\cdot\boldsymbol{\varepsilon}}{\mathbf{g}\cdot\boldsymbol{\varepsilon}} & \text{triplet.} \end{cases} \tag{B.28}$$

Inserting the above kinetic contributions into the full Hamiltonian and separating the smooth contributions yields a local energy of the form

$$E_L - E_s = \frac{1}{|\boldsymbol{\varepsilon}|} - \begin{cases} \frac{2\beta_{ee}}{|\boldsymbol{\varepsilon}|} + 2\phi(0)^{-1}\beta_{ee}\mathbf{g}\cdot\hat{\boldsymbol{\varepsilon}} & \text{singlet} \\ \frac{4\beta_{ee}}{|\boldsymbol{\varepsilon}|} + \frac{\mathbf{k}^{**}\cdot\boldsymbol{\varepsilon}}{\mathbf{g}\cdot\boldsymbol{\varepsilon}} & \text{triplet} \end{cases} = -\begin{cases} 2\phi(0)^{-1}\beta_{ee}\mathbf{g}\cdot\hat{\boldsymbol{\varepsilon}} & \text{singlet} \\ \frac{\mathbf{k}^{**}\cdot\boldsymbol{\varepsilon}}{\mathbf{g}\cdot\boldsymbol{\varepsilon}} & \text{triplet} \end{cases}. \tag{B.29}$$

The last equal sign imposes the correct values for the electron-electron cusps β_{ee}, and the Coulomb singularity cancels for the triplet and singlet case. Nevertheless, for both multiplets, the local energy is discontinuous due to the scalar products in the remaining terms. For the triplet case, $\mathbf{k}^{**} \nparallel \mathbf{g}$ in general, and the local energy diverges for the direction $\hat{\boldsymbol{\varepsilon}} \perp \mathbf{g}$ even for $|\boldsymbol{\varepsilon}| > 0$. Note that this divergence is not due to the Coulomb potential, since the Coulomb singularity cancels out, but due to the node and the definition of the local energy with the inverse wave function value.

B.3. Nodes

The exact and trial wave functions, fermionic states, have nodes. The nodes originating from the anti-symmetry are predictable. They give rise to the nodal hyperplanes, $\mathbf{r}_i = \mathbf{r}_j$ for all $i \neq j$, also called Pauli planes. The nodes not related to anti-symmetry are not predictable in almost all relevant cases since the Slater determinant or the expansion thereof vanishes by a complicated interplay of all involved orbitals. In the vicinity of a non-trivial node, $\psi(\mathbf{R}_n) = 0$, the wave function usually is regular without cusps. An expansion of the coordinate vector $\mathbf{R} = \mathbf{R}_n + \boldsymbol{\varepsilon}$ allows casting the wave function as

$$\begin{aligned}\psi(\mathbf{R}) &= \psi(\mathbf{R}_n) + \mathbf{g}\cdot\boldsymbol{\varepsilon} + \frac{1}{2}\boldsymbol{\varepsilon}\cdot\mathbf{t}\boldsymbol{\varepsilon} + \frac{1}{6}\mathbf{k}\otimes\boldsymbol{\varepsilon}\otimes\boldsymbol{\varepsilon}\otimes\boldsymbol{\varepsilon} + \mathcal{O}(|\varepsilon|^4)\\ &= |\mathbf{g}|\,|\boldsymbol{\varepsilon}_\perp| + \frac{1}{2}\boldsymbol{\varepsilon}\cdot\mathbf{t}\boldsymbol{\varepsilon} + \frac{1}{6}\mathbf{k}\otimes\boldsymbol{\varepsilon}\otimes\boldsymbol{\varepsilon}\otimes\boldsymbol{\varepsilon} + \mathcal{O}(|\varepsilon|^4)\end{aligned}\ . \tag{B.30}$$

The expansion coefficients, $\mathbf{g}$, $\mathbf{t}$ and $\mathbf{k}$, are analogous to Sec. B.2. Additionally, $\mathbf{g}$ is a vector normal to the nodal hypersurface at $\mathbf{R}_n$. Since the gradient $\mathbf{g} = \nabla\psi(\mathbf{R}_n)$ at the node points away from the nodal surface, $\mathbf{g}\cdot\boldsymbol{\varepsilon} = |\mathbf{g}|\,|\boldsymbol{\varepsilon}_\perp|$, while $\boldsymbol{\varepsilon}_\perp$ is the vector component of $\boldsymbol{\varepsilon}$ parallel to the gradient and perpendicular to the nodal surface. Further, $|\boldsymbol{\varepsilon}_\perp|$ is the shortest distance from $\mathbf{R}$ to the tangent plane of the nodal surface at $\mathbf{R}_n$.

The drift vector then takes the form

$$\mathbf{u}(\mathbf{R}) = 2D\left(\frac{\mathbf{g}}{|\mathbf{g}|\,|\boldsymbol{\varepsilon}_\perp|} + \frac{\mathbf{t}^S\boldsymbol{\varepsilon}}{|\mathbf{g}|\,|\boldsymbol{\varepsilon}_\perp|}\right) + \mathcal{O}(|\boldsymbol{\varepsilon}|) \tag{B.31}$$

with symmetric part $\mathbf{t}^S$ as in Sec. B.2. Without the cusp condition, as in the triplet case, the drift vector exhibits directionally singular behavior only.

The Laplacian reduces to

$$\nabla^2_{\boldsymbol{\varepsilon}}\psi(\boldsymbol{\varepsilon}) = \operatorname{tr}(\mathbf{t}) + \mathbf{k}^{**}\boldsymbol{\varepsilon} + \mathcal{O}(|\boldsymbol{\varepsilon}|^2), \tag{B.32}$$

with $\mathbf{k}^{**}$ as above. This renders the local energy

$$E_L - E_s = -D\frac{\nabla^2_{\boldsymbol{\varepsilon}}\psi(\boldsymbol{\varepsilon})}{\psi(\boldsymbol{\varepsilon})} = -D\frac{\operatorname{tr}(\mathbf{t}) + \mathbf{k}^{**}\cdot\boldsymbol{\varepsilon}}{|\mathbf{g}|\,|\boldsymbol{\varepsilon}_\perp|} + \mathcal{O}(|\boldsymbol{\varepsilon}|). \tag{B.33}$$

For an exact ground state, the local energy is a constant, $E_L = E_0$. Thus, the exact wave function must satisfy $\operatorname{tr}(\mathbf{t}) = 0$ and $\mathbf{k}^{**} \parallel \mathbf{g}$ to prevent singular behavior in the local energy. For a trial wave function, i.e., not an eigenfunction of the Hamiltonian in practice, directional divergencies of the two kinds $\operatorname{tr}(\mathbf{t})/\mathbf{g}\cdot\boldsymbol{\varepsilon}$ and $\mathbf{k}^{**}\cdot\boldsymbol{\varepsilon}/\mathbf{g}\cdot\boldsymbol{\varepsilon}$ are present at the nodes if $\operatorname{tr}(\mathbf{t}) \neq 0$. Although the local energy of a trial wave function diverges at the nodes, the contribution to the total energy originating from the point $\mathbf{R}_n$ on the nodal surface is bound since the total energy integrates over the product of the trial wave function and local energy.

C. Probability theory

This chapter aims to briefly overview the probability theory affiliated with Monte Calo methods. It is not a deep dive into the mathematical intricacies but a shallow summary of the tools used in this thesis. For a more precise introduction to probability theory the reader is referred to Ref. [61, 153]. The main topics are the Metropolis-Hastings algorithm, the estimation of random variables, and statistical correlation.

C.1. Definitions

Here, a random variable X on some space Ω is defined via its corresponding probability density function $f : \Omega \to \mathbb{R}_+$. The function f defines a probability density $f(x)$ for every element $x \in \Omega$. The random variable X describes the generic outcome of a random experiment in which some element of Ω is realized with probabilities according to f. This relation is denoted $X \sim f$ and X is said to be distributed according to f. In the following, the space Ω is always continuous and may be represented by $\mathbb{R}^n$. The random variable X encapsulates the probabilistic character in a single variable, giving rise to an elegant algebraic formulation of probability theory. The distribution f carries the equivalent information, often more accessible regarding computation as the random variable X, but is not as compelling regarding algebraic relations of random variables. The distribution f maps each $x \in \mathbb{R}^n$ to some positive probability density with the normalization $\int_\Omega f(x)\,\mathrm{d}x = 1$.

From this shorthand definition of a random variable, the expectation value of some function $g(X)$ can be calculated by

$$\operatorname{Exp}\left[g(X)\right] = \int g(x) f(x)\,\mathrm{d}x. \tag{C.1}$$

The above formula links the random variable with its probability distribution function in defining an algebraic function $\operatorname{Exp}\left[\cdot\right]$ for random variables. Another key quantity is the variance of a random variable

$$\operatorname{Var}\left[g(X)\right] = \int f(x)\left(g(X) - \operatorname{Exp}\left[g(X)\right]\right)^2\,\mathrm{d}x = \operatorname{Exp}\left[g(X)^2\right] - \operatorname{Exp}\left[g(X)\right]^2 \tag{C.2}$$

While the expectation value gives a notion of what to expect from realizations of $g(X)$, the variance quantifies the probability of how close a number of realizations is to the expectation value on average. Whenever the variance is available, the standard deviation can also be calculated.

$$\mathrm{s}\left[g(X)\right] = \sqrt{\operatorname{Var}\left[g(X)\right]} \tag{C.3}$$

It is an error measure, quantifying the same information as the variance but in a more accessible way.

For two random variables X and Y with joint probability distribution f, i.e., a probability distribution that takes x and y as arguments and returns the probability for the tuple (x, y), the covariance of X and Y is defined as

$$\begin{aligned}\operatorname{Cov}\left[X, Y\right] &= \int \left(X - \operatorname{Exp}\left[X\right]\right)\left(Y - \operatorname{Exp}\left[Y\right]\right) f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \operatorname{Exp}\left[XY\right] - \operatorname{Exp}\left[X\right]\operatorname{Exp}\left[Y\right]\end{aligned} \tag{C.4}$$

With the definitions of the expectation value, the variance, and the covariance, the theoretical tools for the following statistical analysis are set.

C.2. Monte Carlo integration

In this thesis, the probability distribution under study is, in the majority of cases, a high-dimensional one. Key target quantities to calculate are averages over the distribution of f in the form of Eq. (C.1). The integration over the high-dimensional domain Ω of f cannot be carried out by standard numerical means, and the central theorem to tackle the integration is the law of large numbers [61]. It allows for the estimation of the value of an integral of the form

$$I = \int g(x) f(x)\, \mathrm{d}x = \mathrm{Exp}\left[g(X)\right]. \tag{C.5}$$

To this end, m random variables (x_i) with $x_i \sim f, \forall i = 1, \dots m$, are required. The generation of these random variables is discussed in Sec. C.4. Then, the estimate from the law of large numbers reads,

$$I_m = \frac{1}{m} \sum_{i=1}^{m} g(x_i) \tag{C.6}$$

and in the limit $m \to \infty$, the estimate is guaranteed to converge to the true value of the integral,

$$I = \lim_{m \to \infty} I_m. \tag{C.7}$$

In the case of independent and identically distributed (i.i.d.) variables (x_i) with finite variances $\mathrm{Var}\left[g(X)\right]$ and $\mathrm{Var}\left[x_i\right]$ for each i, the variance of the estimate I_m is given by

$$\mathrm{Var}\left[I_m\right] = \frac{\mathrm{Var}\left[g(x_i)\right]}{m} = \frac{\mathrm{Var}\left[g(X)\right]}{m} \tag{C.8}$$

The mathematical nature of the convergence is omitted here for the sake of simplicity. The estimation of I by a sample, i.e., a series of random variates (x_i), is called Monte Carlo integration. The uncertainty of the estimate, its standard deviation, scales as

$$\mathrm{s}\left[I_m\right] = \sqrt{\mathrm{Var}\left[I_m\right]} = \frac{\sqrt{\mathrm{Var}\left[g(X)\right]}}{\sqrt{m}} \tag{C.9}$$

The great advantage of this integration method is the scaling of its uncertainty as $m^{-1/2}$ independent of dimension of Ω. Although the scaling of the uncertainty is not fast in an intuitive sense, it easily outcompetes numerical schemes in higher dimensions.

C.3. Estimators

Applying the law of large numbers to the expectation value yields an estimator

$$\check{\mu}\left[g(X)\right] = \frac{1}{m} \sum_{i=1}^{m} g(x_i). \tag{C.10}$$

The estimate's accuracy increases with sample size m and converges by the law of large numbers in the limiting case

$$\mathrm{Exp}\left[g(X)\right] = \lim_{m \to \infty} \check{\mu}\left[g(X)\right] \tag{C.11}$$

In the following i.i.d. samples, i.e., $\mathrm{Cov}\left[g(x_i), g(x_k)\right] = \delta_{ik}\mathrm{Var}\left[g(x_k)\right]$, are assumed. Applying the same estimator scheme to the variance $\mathrm{Var}\left[g(X)\right]$, i.e., applying the law of large numbers to the expectation values in the definition of the variance, gives an estimator for the variance

$$\check{\sigma}_b^2\left[g(X)\right] = \frac{1}{m}\sum_{i=1}^{m}\left(g(x_i) - \check{\mu}\left[g(X)\right]\right)^2 . \tag{C.12}$$

While the above variance estimator $\check{\sigma}_b^2\left[g(X)\right]$ converges in the limit of infinite m, it is biased for any finite m. This is apparent from its expectation value

$$\begin{aligned}\mathrm{Exp}\left[\check{\sigma}_b^2\left[g(X)\right]\right] &= \mathrm{Exp}\left[\left(g(x_i) - \check{\mu}[g(X)]\right)^2\right] = \mathrm{Var}\left[g(x_i) - \check{\mu}[g(X)]\right] \\ &= \mathrm{Var}\left[g(x_i)\right] + \mathrm{Var}\left[\check{\mu}\left[g(X)\right]\right] - 2\mathrm{Cov}\left[g(x_i), \check{\mu}\left[g(X)\right]\right] = \\ &= \mathrm{Var}\left[g(X)\right]\left(1 + \frac{1}{m}\right) - \frac{2}{m}\sum_{k=1}^{m}\mathrm{Cov}\left[g(x_i), g(x_k)\right] \\ &= \left(1 - \frac{1}{m}\right)\mathrm{Var}\left[g(X)\right] \neq \mathrm{Var}\left[g(X)\right] .\end{aligned} \tag{C.13}$$

This example shows that the estimation of random variables can become cumbersome, even if the estimators have a defined limit in which their target value is recovered. An unbiased estimator for the variance employs the so-called Bessel correction $m/(m-1)$ and reads

$$\check{\sigma}^2\left[g(X)\right] = \frac{1}{m-1}\sum_{i=1}^{m}\left(g(x_i) - \check{\mu}\left[g(X)\right]\right)^2 . \tag{C.14}$$

To estimate integrals via the law of large numbers, the main quantity of interest is, of course, the expectation value. However, to assess the statistical certainty of the estimated value, its corresponding statistical error, the standard deviation, is required. The variance itself is also important in some cases. All quantities have to be estimated by an appropriate estimator. While $\check{\mu}\left[g(X)\right]$ qualifies as an estimator of the expectation value, the estimation of the variance of either $g(X)$ or $\check{\mu}\left[g(X)\right]$ is achieved by $\check{\sigma}^2\left[g(X)\right]$ and $\check{\sigma}^2\left[\check{\mu}\left[g(X)\right]\right]$, respectively. From the variance of the estimator of the mean, an uncertainty for the estimated mean can be defined via the standard deviation.

$$\check{s}\left[\check{\mu}\left[g(X)\right]\right] = \sqrt{\check{\sigma}^2\left[\check{\mu}\left[g(X)\right]\right]} \tag{C.15}$$

However, estimating the standard deviation from an estimated variance is biased since the square root is a non-linear operator. Without the assumption of an i.i.d. sample and thus a serially correlated sample, the calculation of the variance is much more complicated as covariances $\mathrm{Cov}\left[x_i, x_j\right]$ have to be taken into account.

C.4. Metropolis-Hastings algorithm

To use the law of large numbers, that is, to estimate integrals of a probability distribution f by averaging realizations of that distribution, an efficient method is required to generate random variates distributed according to f, i.e., to sample the function. Random variables of simple distributions, like the uniform, normal, or Poisson distributions can be sampled directly. Unfortunately, it is impossible for most probability distributions of real applications to efficiently and directly draw a variate from it. In many cases, the probability distribution even is a computationally expensive function to evaluate. However, multiple implicit schemes allow for sampling such distribution. A successful approach is to start with some initial random guess for a random variate and generate new variates from the information of the previous variates. The result of such a process is called a Markov chain. Formally, it is defined as a random process in which every new variate only depends on its (direct) ancestor. The corresponding Markov process can be tailored to generate the desired set of variates for the law of large numbers, i.e. ensure $(\mathbf{x}_0, \mathbf{x}_1, \dots)$ with $\mathbf{x}_i \sim f$ for all i. However, in almost any case, the Markov chain shows serial correlation, i.e. statistical dependence between chain elements. The correlation complicates the estimation of the uncertainty of the mean estimator, the variance, and higher moments of the distribution, cf. the following Sec. C.5. Since the elements of the Markov chain $\mathbf{x}_i$ belong to some $\mathbb{R}^m$ and resemble stochastic trajectories, bold symbols highlight their multi-dimensional character and relation to classical mechanics. Random variables from a discrete space, or state space, are associated with states. For random variates from a continuous space, the term configuration references the variates.

The popular Metropolis-Hastings algorithm [62, 63] provides a universal method to construct such a series by defining the Markov process. Any time-independent Markov process without memory can be represented by a single transition probability density kernel $T(\mathbf{y} \leftarrow \mathbf{x})$, that specifies the probability to transition from configuration $\mathbf{x}$ to configuration $\mathbf{y}$ in a single step. A transition kernel has to satisfy the normalization

$$\int T(\mathbf{y} \leftarrow \mathbf{x})\,\mathrm{d}\mathbf{y} = 1 \tag{C.16}$$

and is non-negative. The normalization ensures that the probability to arrive at any configuration $\mathbf{y}$, starting from an arbitrary configuration $\mathbf{x}$, is unity. Every ergodic transition kernel (see below) is guaranteed to have a stationary distribution

$$f(\mathbf{y}) = \int T(\mathbf{y} \leftarrow \mathbf{x}) f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \hat{\mathrm{T}}(\mathbf{y}) f. \tag{C.17}$$

And an initial probability distribution f_0 can be converged to the stationary distribution f by repeated application of the kernel.

$$f = \lim_{n \to \infty} \hat{\mathrm{T}}^n f_0 \tag{C.18}$$

A transition kernel is ergodic, or equivalently irreducible, if it allows the transition from any configuration $\mathbf{x}$ to any other configuration $\mathbf{y}$ with positive probability in a finite amount of steps, i.e., if a finite number n exists such that

$$\int \cdots \int T(\mathbf{y} \leftarrow \mathbf{x}_{n-1}) \dots T(\mathbf{x}_1 \leftarrow \mathbf{x})\,\mathrm{d}\mathbf{x}_{n-1} \dots \mathbf{x}_1 > 0. \tag{C.19}$$

This requirement ensures that the complete configuration space is accessible by transitions irrespective of the starting state and does not separate into subspaces between which no transitions are possible. Ergodicity is met for all cases present in this work.

Assume the target distribution to draw from is f. The Markov process defined by T should be tailored to have an equilibrium state identical to f. There are infinitely many transition kernels with equilibrium distribution f, i.e. there is no unique map between transition kernel and equilibrium distribution. A sufficient condition for the transition kernel T to have the equilibrium distribution f is detailed balance or equivalently reversibility

$$T(\mathbf{y} \leftarrow \mathbf{x}) f(\mathbf{x}) = T(\mathbf{x} \leftarrow \mathbf{y}) f(\mathbf{y}). \tag{C.20}$$

Detailed balance ensures that the transported probability $T(\mathbf{y} \leftarrow \mathbf{x}) f(\mathbf{x})$ for the transition $\mathbf{y} \leftarrow \mathbf{x}$ is identical to the transported probability $T(\mathbf{x} \leftarrow \mathbf{y}) f(\mathbf{y})$ of the reverse process $\mathbf{x} \leftarrow \mathbf{y}$.

The transported probability $T(\mathbf{y} \leftarrow \mathbf{x})f(\mathbf{x})$ is given by the probability density $f(\mathbf{x})$ for the configuration $\mathbf{x}$ multiplied by the probability $T(\mathbf{y} \leftarrow \mathbf{x})$ of the transition from configuration $\mathbf{x}$ to $\mathbf{y}$. Note that detailed balance is not a necessary condition, and there are transition kernels that might violate it but nevertheless converge to f in the long-time limit. However, detailed balance is a convenient choice. If the kernel T satisfies detailed balance with the distribution f, then it automatically yields this distribution as its limiting distribution, since

$$\int T(\mathbf{y} \leftarrow \mathbf{x})f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int T(\mathbf{x} \leftarrow \mathbf{y})f(\mathbf{y})\,\mathrm{d}\mathbf{x} = f(\mathbf{y}), \tag{C.21}$$

where the last step follows from the normalization of Eq. (C.16). Detailed balance ensures the existence of an equilibrium distribution. The uniqueness follows from the two additional criteria of aperiodicity and positive recurrence. Aperiodic transition kernels prevent periodic flow in configuration space, i.e. that a configuration is not deterministically revisited after a fixed number of steps. The transition kernel is positive recurrent if the number of steps to revisit an arbitrary configuration has a finite expectation value.

Now that the formal connection from the target distribution f to some transition kernel T is established, a constructive representation of the transition kernel is required that satisfies the detailed balance. Therefore, T is cast as a conditional transition, i.e. T is split into a proposal transition kernel G and an accetpance probability A,

$$T(\mathbf{y} \leftarrow \mathbf{x}) = A(\mathbf{y} \leftarrow \mathbf{x})G(\mathbf{y} \leftarrow \mathbf{x}) + R(\mathbf{x})\delta(\mathbf{y} - \mathbf{x}) \tag{C.22}$$

While $G(\mathbf{y} \leftarrow \mathbf{x})$ proposes a transition, the acceptance probability $A(\mathbf{y} \leftarrow \mathbf{x})$ reweights the probability of the proposed transition. The rejection probability $R(\mathbf{x})$ accounts for the probability of the transition $\mathbf{x} \leftarrow \mathbf{x}$, that is, to stay in the configuration $\mathbf{x}$.

$$R(\mathbf{x}) = 1 - \int A(\mathbf{y} \leftarrow \mathbf{x})G(\mathbf{y} \leftarrow \mathbf{x})\,\mathrm{d}\mathbf{y} \tag{C.23}$$

In this form, the kernel satisfies the normalization of Eq. (C.16) without any constraint on A or G, but not necessarily the detailed balance Eq. (C.20). Imposing detailed balance leads to a condition on A.

$$A(\mathbf{y} \leftarrow \mathbf{x}) = \frac{f(\mathbf{y})G(\mathbf{x} \leftarrow \mathbf{y})}{f(\mathbf{x})G(\mathbf{y} \leftarrow \mathbf{x})}A(\mathbf{x} \leftarrow \mathbf{y}) \tag{C.24}$$

There are many possible forms for A that satisfy this condition, leading to Gibbs sampling or Hamiltonian Monte Carlo. However, the most popular one is the Metropolis-Hastings acceptance probability [62, 63]:

$$A(\mathbf{y} \leftarrow \mathbf{x}) = \min\left[1, \frac{f(\mathbf{y})G(\mathbf{x} \leftarrow \mathbf{y})}{f(\mathbf{x})G(\mathbf{y} \leftarrow \mathbf{x})}\right] \tag{C.25}$$

With this choice for A, the Markov process is fully defined. Despite the multiple choices already taken (imposing detailed balance and choosing the MH acceptance probability), there is still great freedom in the choice of G, as long as G renders the total transition kernel ergodic. There is a soft but important condition on G regarding the numerical implementation: The proposal transition should allow for numerically efficient sampling.

The optimal choice without any rejections ($A = 1$) implies $G(\mathbf{y} \leftarrow \mathbf{x}) = f(\mathbf{y})$, which contradicts the assumption that f is computationally expensive to sample and $G(\cdot \leftarrow \mathbf{x})$ can be sampled efficiently. Nevertheless, this points to the fact that a weak dependence of $G(\mathbf{y} \leftarrow \mathbf{x})$ on $\mathbf{x}$ and a big overlap with $f(\mathbf{y})$ is beneficial.

Next, the algorithm is outlined, that follows from the MH choice for the acceptance probability. The algorithm utilizes repeated application of the transition kernel to some initial distribution to arrive at the equilibrium distribution.

With some choice for G, a target distribution f and some initial choice for the Markov chain $\mathbf{x}_0$, a new configuration $\mathbf{y}_1$ is drawn from the proposal kernel $\mathbf{y}_1 \sim G(\cdot \leftarrow \mathbf{x}_0)$. For the proposed transition from $\mathbf{x}_0$ to $\mathbf{y}_1$ the acceptance probability $p = A(\mathbf{y}_1 \leftarrow \mathbf{x}_0)$ is calculated. With probability p, the proposed new configuration $\mathbf{y}_1$ is accepted to be the next configuration in the Markov chain, $\mathbf{x}_1 = \mathbf{y}_1$. With the reverse probability $1-p$ the Markov chain's next configuration is $\mathbf{x}_1 = \mathbf{x}_0$, i.e. the Markov process remains in the current configuration $\mathbf{x}_0$. Iteration of the scheme,

$$\mathbf{y}_{i+1} \sim G(\cdot \leftarrow \mathbf{x}_i) \tag{C.26}$$

$$\mathbf{x}_{i+1} = \begin{cases} \mathbf{y}_{i+1} & \text{with probability } A(\mathbf{y}_{i+1} \leftarrow \mathbf{x}_i) \\ \mathbf{x}_i, & \text{with probability } 1 - A(\mathbf{y}_{i+1} \leftarrow \mathbf{x}_i), \end{cases} \tag{C.27}$$

yields the Markov chain $(\mathbf{x}_0, \mathbf{x}_1, \dots)$. Note that the Markov chain requires a certain and previously unknown number of iterations n_{eq} to become independent of the initial configuration $\mathbf{x}_0$ and resemble the equilibrium distribution $f(\mathbf{x})$, since detailed balance only holds in equilibrium. Applying $\hat{\mathrm{T}}$ to some initial distribution does not yield the equilibrium distribution immediately. However, it drives the distribution towards the equilibrium distribution by the usual argument of a power method. In this regard, it is worthwhile to note that the eigenvalues of a transition kernel have the upper bound of one.

Many proposal transition probabilities are compatible with a parameter that scales the broadness, or variance, of the distribution. Thus, it is possible to introduce the notion of a time step Δt. Then the transition kernel is interpreted to propagate a configuration $\mathbf{x}_i$ to some stochastically chosen configuration $\mathbf{x}_{i+1} \sim T(\cdot \leftarrow \mathbf{x}, \Delta t)$ in the time Δt. The distance $|\mathbf{x}_{i+1} - \mathbf{x}_i|$ is interpreted as a distance traveled in the time step Δt and the passed time $t = i\Delta t$ scales with the number of iterations i. This allows for interpreting the Markov chain as a dynamic process in time $\mathbf{x}(t)$, although the time is not a continuous variable and stays discrete in this picture. This view on the MH algorithm aligns with the method of Green functions, see Sec. 4.3 or Sec. 6.2, where different proposal kernels are discussed. Another advantage of the MH algorithm is its genuine support for parallelization. Assume, not only a single configuration, $\mathbf{x}(t)$, is propagated, but an whole ensemble, $(\mathbf{x}_k(t))$. The propagation of the individual configurations is completely independent of the remaining ensemble. The ensemble propagation can be spread out across different parallel units, enabling straightforward parallelization.

C.5. Serial correlation

This section reviews the effects of serial correlation, or autocorrelation, in the Markov chain. The correlation quantifies any statistical dependence between two random variables X and Y. The correlation is given by the algebraic relation

$$\rho\left[X, Y\right] = \frac{\mathrm{Cov}\left[X, Y\right]}{\sqrt{\mathrm{Var}\left[X\right]\mathrm{Var}\left[Y\right]}}. \tag{C.28}$$

The elements of the Markov chain $\mathbf{x}(t)$ exhibit a natural correlation. Note that in equilibrium, the Markov chain is invariant under translation in time. Thus, the autocorrelation A, i.e., the correlation of the time series $\mathbf{x}(t)$ with itself after some time interval, depends only on the time difference only

$$A(t) = \rho\left[\mathbf{x}(t' + t), \mathbf{x}(t')\right] = \rho\left[\mathbf{x}(t), \mathbf{x}(0)\right]. \tag{C.29}$$

Different Markov chains generated with the MH algorithm differ in their autocorrelation. In the case of an optimal proposal kernel $G(\mathbf{y} \leftarrow \mathbf{x}) = f(\mathbf{y})$ the autocorrelation vanishes, since every element of the Markov chain is independent from its ancestor elements. For any non-optimal proposal, the autocorrelation is non-vanishing and ranges in the interval $[-1, 1]$ depending on the proposal transition probability. However, according to Reynolds [154], the serial correlation of a Markov chain at equilibrium decays multi-exponentially. This limits the magnitude of the autocorrelation for large times. In the following, the correlation is assumed to decay with a single exponent,

$$A(t) = \exp(-t/T_c) = \chi^t. \tag{C.30}$$

This approximation simplifies analytic considerations and focuses on the autocorrelation in the long-time limit. Therefore, neighbouring elements of the Markov chain are strongly correlated, and the correlation vanishes exponentially with the time interval between elements with some decay constant $\chi = \exp(-1/T_c)$. High correlation times T_c correspond to χ close to one, while a vanishing correlation time $T_c = 0$ results in $\chi = 0$. The correlation time T_c is the typical time scale on which the correlation decays. Note that this is not necessarily a reasonable approximation for short to intermediate times, where the correlation also can become negative.

C.6. Estimators with reduced correlation

Estimating expectation values by the law of large numbers in Eq. (C.10) depends heavily on how well the generated Markov chain describes the underlying probability distribution. If a Markov chain has a large serial correlation, the exploration of the configuration space is limited. While the expectation value is recovered in the long-time limit, for finite times the correlation introduces a potentially large uncertainty to it. For a chain $x(t)$ with $t = 1, \dots T$, the uncertainty manifests in the variance of the mean-estimator

$$\begin{aligned}
\mathrm{Var}\left[\check{\mu}\left[g(\mathbf{x})\right]\right] &= \frac{1}{T^2}\mathrm{Cov}\left[\sum_{t=1}^{T} g(\mathbf{x}(t)), \sum_{t'=1}^{T} g(\mathbf{x}(t'))\right] \\
&= \frac{1}{T^2}\sum_{t,t'=1}^{T} \mathrm{Cov}\left[g(\mathbf{x}(t)), g(\mathbf{x}(t'))\right] \\
&= \frac{1}{T^2}\left(\sum_{t=1}^{T} \mathrm{Cov}\left[g(\mathbf{x}(t)), g(\mathbf{x}(t))\right] + \sum_{t\neq t'} \mathrm{Cov}\left[g(\mathbf{x}(t)), g(\mathbf{x}(t'))\right]\right) \\
&= \frac{\mathrm{Var}\left[g(\mathbf{x})\right]}{T}\left(1 + \frac{1}{T}\sum_{t\neq t'} \rho\left[g(\mathbf{x}(t)), g(\mathbf{x}(t'))\right]\right)
\end{aligned} \tag{C.31}$$

Discarding the assumption of an i.i.d. sample, the above is a refinement of Eq. (C.14) to the correlated case. Thus, Eq. (C.14) disqualifies as an estimator for the variance in the presence of correlation. While the last term in Eq. (C.31) could be estimated, it is a more convenient and reliable approach to generate an uncorrelated sample from the correlated Markov chain. The correlation can be reduced with the introduction of new variables that represent an averaging over the variables $\mathbf{x}_k(t)$. First, the effect of parallel samples is discussed. Then, a method to transform the original Markov chain into a less correlated one is reviewed.

C.6.1. Parallel sampling

As mentioned above, the parallel generation of multiple independent Markov chains poses a computational advantage. In addition, averaging over the parallelly propagated configurations reduces the uncertainty of estimates and the covariance. Assume m independent Markov processes $\mathbf{x}_1(t), \ldots \mathbf{x}_m(t)$ are computed in parallel. Then, the covariances between the configurations are given by $\rho\left[\mathbf{x}_k(t), \mathbf{x}_{k'}(0)\right] = \delta_{kk'}\chi^t$. The arithmetic mean over the m configurations at time t is given by the sample average

$$\mathbf{Z}(t) = \frac{1}{m}\sum_{k=1}^{m} \mathbf{x}_k(t), \tag{C.32}$$

and, of course, conserves the expectation value. The variance of the sample at time t can be determined with the variance of an i.i.d. sample and takes the form

$$\mathrm{Var}\left[\mathbf{Z}\right] = \frac{\mathrm{Var}\left[\mathbf{x}\right]}{m}. \tag{C.33}$$

It monotonously reduces with the size of the sample. The covariance of two sample averages at different times becomes

$$\mathrm{Cov}\left[\mathbf{Z}(t), \mathbf{Z}(0)\right] = \frac{1}{m^2}\sum_{k,k'=1}^{m} \mathrm{Cov}\left[\mathbf{x}_k(t), \mathbf{x}_{k'}(0)\right] = \frac{1}{m^2}\sum_{k=1}^{m}\mathrm{Cov}\left[\mathbf{x}_k(t), \mathbf{x}_{k'}(0)\right] = \frac{1}{m}\mathrm{Cov}\left[\mathbf{x}(t), \mathbf{x}(0)\right]. \tag{C.34}$$

Thus, variance and covariance are affected in the same way by the parallelization. In the correlation coefficient, the reduction factors $1/m$ cancel exactly, and the correlation is not affected by the sample-averaging,

$$\rho\left[\mathbf{Z}(t), \mathbf{Z}(0)\right] = \rho\left[\mathbf{x}_1(t), \mathbf{x}_1(0)\right]. \tag{C.35}$$

Nevertheless, parallel sampling significantly reduces statistical uncertainty. Each Markov chain, independent of the other chains, explores the configuration space along its trajectory. It is not unusual to parallelly propagate hundreds or even beyond thousands of Markov processes, rendering the increase in exploration and statistical certainty a very attractive option. With, for example, a 10^4 parallelly propagated Markov chains, the statistical error bars can be lowered by a factor of 100 compared to a single Markov chain.

C.6.2. Data blocking

A second method exists, reformulating the sampling within a single Markov chain. This method is called data blocking or binning analysis [84]. A serially correlated time series $\mathbf{x}(t)$ with $t = 1, \ldots T = nB$ is chunked in blocks of length B and averaged to get the blocked time series $\mathbf{y}(b)$, where the new index b accounts for the respective block.

$$\mathbf{y}(b) = \frac{1}{B}\sum_{t=bB}^{(b+1)B-1} \mathbf{x}(t) = \frac{1}{B}\sum_{t=0}^{B-1}\mathbf{x}(t+bB) \tag{C.36}$$

Instead of the series $x(t)$ of length $T = nB$, the chunked sample $(\mathbf{y}(b))$ has lengths n. The blocking transformation conserves the expectation value. To calculate the correlation, first rewrite the covariance of $\mathbf{y}$ for $b > 0$

$$\begin{aligned}\mathrm{Cov}\left[\mathbf{y}(b), \mathbf{y}(0)\right] &= \frac{1}{B^2}\sum_{t,t'=0}^{B-1} \mathrm{Cov}\left[\mathbf{x}(t+bB), \mathbf{x}(t')\right] = \\ &= \frac{\mathrm{Var}\left[\mathbf{x}\right]}{B^2}\sum_{t,t'=0}^{B-1} \chi^{|bB+t-t'|} = \\ &= \frac{\mathrm{Var}\left[\mathbf{x}\right]}{B^2}\chi^{(b-1)B+1}\left(\frac{1-\chi^B}{1-\chi}\right)^2,\end{aligned} \tag{C.37}$$

where the last equality follows from the formula Eq. (C.65) derived in Sec. C.10. The variance, on the other hand, follows from the covariance for the case $b = 0$, which is not covered by the above formula. Instead Eq. (C.61), also derived in Sec. C.10, is used to arrive at

$$\begin{aligned}\mathrm{Var}\left[\mathbf{y}\right] &= \frac{\mathrm{Var}\left[\mathbf{x}\right]}{B^2}\sum_{t,t'=0}^{B-1} \chi^{|t-t'|} = \\ &= \frac{\mathrm{Var}\left[\mathbf{x}\right]}{B^2}\left(B\frac{1+\chi}{1-\chi} - 2\chi\frac{1-\chi^B}{(1-\chi)^2}\right) = \\ &= \frac{\mathrm{Var}\left[\mathbf{x}\right]}{B}\left(\frac{1+\chi}{1-\chi} - \frac{2\chi}{B}\frac{1-\chi^B}{(1-\chi)^2}\right).\end{aligned} \tag{C.38}$$

The correlation for $b > 0$ is hence

$$\rho\left[\mathbf{y}(b), \mathbf{y}(0)\right] = \chi^{(b-1)B+1}\frac{(1-\chi^B)^2}{(1-\chi^2)B - 2\chi(1-\chi^B)} \tag{C.39}$$

Clearly, the correlation of $\mathbf{y}(b)$ decays exponentially faster than the correlation of $\mathbf{x}(t)$. Comparing the correlation between two consecutive chunked variables $\mathbf{y}(1)$ and $\mathbf{y}(0)$ and two consecutive variables of the original Markov chain yields

$$r = \frac{\rho\left[\mathbf{y}(1), \mathbf{y}(0)\right]}{\rho\left[\mathbf{x}(1), \mathbf{x}(0)\right]} = \frac{(1-\chi^B)^2}{(1-\chi^2)B - 2\chi(1-\chi^B)} \leq 1 \tag{C.40}$$

The positive ratio can only lower the correlation or does not alter it in the limiting case $B = 1$. The above ratio sets a scale for the choice of the block length B, given the correlation coefficient χ, which is a constant of the Markov chain or transition kernel under study. With large enough B, the Markov chain can be decorrelated, and its variance can be calculated using the formula for the i.i.d. case. The autocorrelation formula with an assumed autocorrelation time for the blocked series T_c^B can be invoked to recast the correlation reduction.

$$r = \frac{\exp(-t/T_c^B)}{\exp(-t/T_c)} \tag{C.41}$$

Solving for the blocked autocorrelation time yields

$$T_c^B = T_c\frac{1}{1 - T_c \ln r}. \tag{C.42}$$

The correlation time of the blocked versus the original Markov chain is shown in Fig. C.1. For increased correlation, a larger block size effectively reduces the correlation. For example, to arrive at a correlation time of about $T_c^B = 1$, with an unblocked correlation time of $T_c = 8$, a block length $B = 10$ is sufficient. A small block length is already able to reduce the correlation time significantly. The block lengths $B = 50$ and $B = 100$ reduce the blocked correlation time to below one, even for large correlations.

However, the reduced correlation comes at the cost of fewer observations since B elements of the original Markov chain are collapsed into a single element of the chunked Markov chain. This loss in the number of observations results in a larger statistical uncertainty. An optimal B therefore is large enough to significantly decorrelate the Markov chain but at the same time small enough to prevent unnecessary elimination of data. The block length B is a hyper-parameter of the MC integration and has to be chosen for each system individually from experience.

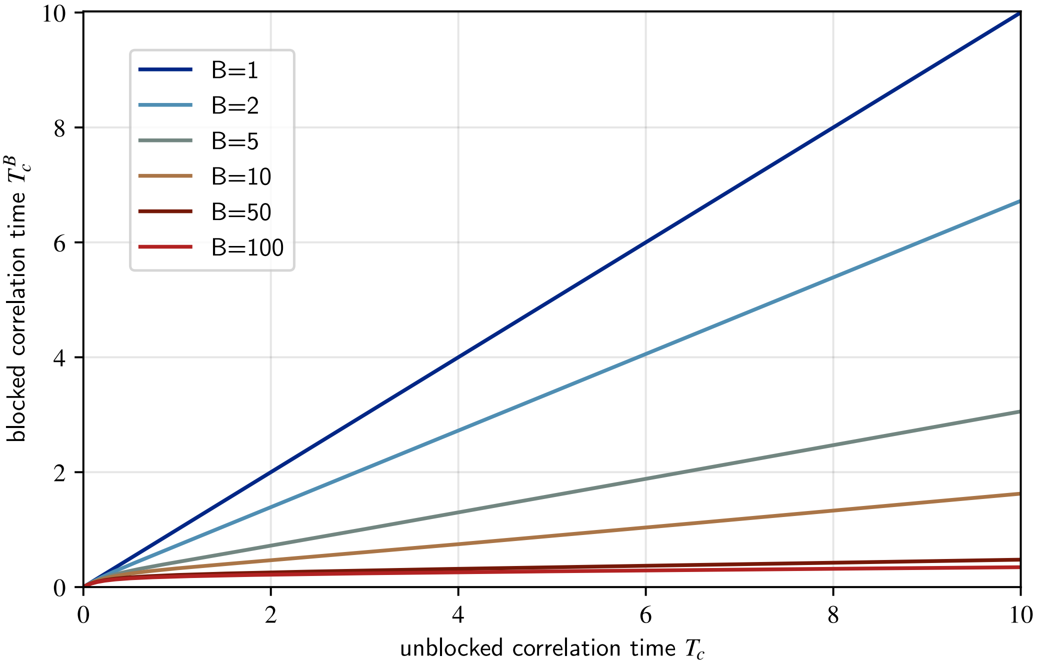


Figure C.1.: A plot of Eq. (C.42) for various values of the block length B. With a high correlation time, a larger block length is required to decorrelate the Markov chain. The shown graphs become linear relations for large correlation times. Non-linear behavior is only present for small correlation.

C.7. Weighted estimators

In a DMC calculation, the observables have to be evaluated from a weighted ensemble, cf. Sec. 6.11. That is a Markov process for combined weights and configurations, $(\lambda(t), \mathbf{x}(t))$. The weight is readily incorporated in the stochastic process since it can be seen as one component of the multi-dimensional vector of the Markov process alongside the configuration. The weight is naturally a random variable, subject to fluctuations, induced by the random walk of the walker's coordinate vector.

However, in the following the weights are treated as deterministic variables, with a fully deterministic time evolution, to simplify the calculations. Without this assumption, i.e. accounting for the random character of the weights, analytical expressions for the estimators are simply not accessible for analytic calculations.

In this section the estimators of a sample $(\lambda(t), \mathbf{x}(t))$ of length $t = 1, \ldots T$ are rewritten to account for the weight. First, assume an i.i.d. sample. The estimator of the expectation value is

$$\breve{\mu}\left[\mathbf{x}\right] = \frac{1}{\Lambda}\sum_{t=1}^{T}\lambda(t)\mathbf{x}(t), \tag{C.43}$$

with a normalization of

$$\Lambda = \sum_{t=1}^{T}\lambda(t) \quad \text{and} \quad \Lambda_2 = \sum_{t=1}^{T}\lambda(t)^2. \tag{C.44}$$

For clarity and to prevent an overcomplicated notation, the weighted estimators are not presented for some function $g(\mathbf{x})$ but for the expectation value of $\mathbf{x}$ alone. The results of the following analysis, of course, apply to the more general case of a function wrapped around $\mathbf{x}$, since every function of a random varibale is again a random variable.

The uncertainty of the expectation estimator for i.i.d. samples follows from the corresponding variance.

$$\mathrm{Var}\left[\breve{\mu}\left[\mathbf{x}\right]\right] = \frac{1}{\Lambda^2}\sum_{t=1}^{T}\lambda(t)^2\mathrm{Var}\left[\mathbf{x}(t)\right] = \frac{\Lambda_2}{\Lambda^2}\mathrm{Var}\left[\mathbf{x}\right] \tag{C.45}$$

The reduction of the variance due to a larger sample is, hence, Λ_2/Λ^2, which reduces to the known $1/T$ for uniform weights from Eq. (C.12). From the estimator for the expectation value above, the variance estimator

$$\breve{\sigma}_b^2\left[\mathbf{x}\right] = \frac{1}{\Lambda}\sum_{t=1}^{T}\lambda(t)\left(\mathbf{x}(t) - \breve{\mu}\left[\mathbf{x}\right]\right)^2 \tag{C.46}$$

follows. However, this estimator is again biased and a extended notion of the previously mentioned Bessel correction is required to correct it. To derive the correction for the weighted case, the correction from Eq. (C.14) for a uniform-weight sample must be extended. The corrected, unbiased variance estimator is given by

$$\breve{\sigma}^2\left[\mathbf{x}\right] = \frac{\Lambda^2}{\Lambda^2 - \Lambda_2}\breve{\sigma}_b^2\left[\mathbf{x}\right] = \frac{\Lambda}{\Lambda^2 - \Lambda_2}\sum_{t=1}^{T}\lambda(t)\left(\mathbf{x}(t) - \breve{\mu}\left[\mathbf{x}\right]\right)^2. \tag{C.47}$$

From the unbiased estimator the weighted Bessel correction is identified as $\Lambda^2/(\Lambda^2 - \Lambda_2)$, which reduces to the known $T/(T-1)$ for uniform weights. The proof is straightforward but lengthy. First, the biased estimator is expanded

$$\begin{aligned}
\Lambda\breve{\sigma}_b^2\left[\mathbf{x}\right] &= \sum_{t=1}^{T}\lambda(t)\left(\mathbf{x}(t) - \breve{\mu}\left[\mathbf{x}\right]\right)^2 \\
&= \sum_{t=1}^{T}\lambda(t)\left[(\mathbf{x}(t) - \mathrm{Exp}\left[\mathbf{x}\right]) - (\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right])\right]^2 \\
&= \sum_{t=1}^{T}\lambda(t)\left[(\mathbf{x}(t) - \mathrm{Exp}\left[\mathbf{x}\right])^2 - 2(\mathbf{x}(t) - \mathrm{Exp}\left[\mathbf{x}\right])(\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right]) + (\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right])^2\right] \\
&= \sum_{t=1}^{T}\lambda(t)(\mathbf{x}(t) - \mathrm{Exp}\left[\mathbf{x}\right])^2 - 2\Lambda(\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right])^2 + \Lambda(\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right])^2 \\
&= \Lambda\mathrm{Var}\left[\mathbf{x}\right] - \Lambda(\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\mathbf{x}\right])^2 = \Lambda\mathrm{Var}\left[\mathbf{x}\right] - \Lambda(\breve{\mu}\left[\mathbf{x}\right] - \mathrm{Exp}\left[\breve{\mu}\left[\mathbf{x}\right]\right])^2.
\end{aligned} \tag{C.48}$$

Then, the expectation value of the unbiased estimator

$$\begin{aligned}\mathrm{Exp}\left[\check{\sigma}^2\left[\mathbf{x}\right]\right] &= \mathrm{Exp}\left[\frac{\Lambda^2}{\Lambda^2-\Lambda_2}\left[\mathrm{Var}\left[\mathbf{x}\right]-\left(\check{\mu}\left[\mathbf{x}\right]-\mathrm{Exp}\left[\check{\mu}\left[\mathbf{x}\right]\right]\right)^2\right]\right] = \\ &= \frac{\Lambda^2}{\Lambda^2-\Lambda_2}\left[\mathrm{Exp}\left[\mathrm{Var}\left[\mathbf{x}\right]\right]-\mathrm{Exp}\left[\left(\check{\mu}\left[\mathbf{x}\right]-\mathrm{Exp}\left[\check{\mu}\left[\mathbf{x}\right]\right]\right)^2\right]\right] = \\ &= \frac{\Lambda^2}{\Lambda^2-\Lambda_2}\left[\mathrm{Var}\left[\mathbf{x}\right]-\mathrm{Var}\left[\check{\mu}\left[\mathbf{x}\right]\right]\right]\end{aligned} \tag{C.49}$$

is combined with Eq. (C.45) to result in the Bessel correction.

$$\mathrm{Exp}\left[\check{\sigma}^2\left[\mathbf{x}\right]\right] = \frac{\Lambda^2}{\Lambda^2-\Lambda_2}\left[1-\frac{\Lambda_2}{\Lambda^2}\right]\mathrm{Var}\left[\mathbf{x}\right] = \mathrm{Var}\left[\mathbf{x}\right] \tag{C.50}$$

The analytic formula for the variance of the expectation estimator, including serial correlation and weights, is intricate.

$$\begin{aligned}\mathrm{Var}\left[\check{\mu}\left[\mathbf{x}\right]\right] &= \frac{1}{\Lambda^2}\sum_{t,t'=1}^{T}\lambda(t)\lambda(t')\mathrm{Cov}\left[\mathbf{x}(t),\mathbf{x}(t')\right] = \\ &= \frac{1}{\Lambda^2}\left(\sum_{t=1}^{T}\lambda(t)^2\mathrm{Var}\left[\mathbf{x}(t)\right]+\sum_{t\neq t'}\lambda(t)\lambda(t')\mathrm{Cov}\left[\mathbf{x}(t),\mathbf{x}(t')\right]\right)\end{aligned} \tag{C.51}$$

As in the unweighted case, it poses a dead end for analytic considerations, and the only alternative to calculate the variance is again decorrelation. For a decorrelated Markov chain, the variance is estimated by the above $\check{\sigma}^2\left[\mathbf{x}\right]$.

C.8. Numerical evaluation of the variance

The numerical calculation of the variance has some fundamental difficulties. Suppose multiple i.i.d. Markov chains are combined into a single sample $(\mathbf{x}_i)$ of large size $i = 1, \ldots n$, irrespective of how many Markov chains are merged and how long each Markov chain is. Its variance is given by Eq. (C.14). For the sake of clarity, an unweighted sample is assumed without the Bessel correction and the special case $g = id$. The simplified estimator reads

$$\check{\sigma}^2\left[\mathbf{x}\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i-\check{\mu}\left[\mathbf{x}\right]\right)^2 . \tag{C.52}$$

To calculate the variance, however, first an estimated value for the mean $\check{\mu}\left[\mathbf{x}\right]$ is required. With n ranging into the orders of beyond 10^{10} and $\mathbf{x}$ being high-dimensional, it is impossible, and also computationally inefficient, to store the full sample $(\mathbf{x}_i)$ first and evaluate the estimators after the generation of the sample. The estimators have to be recast to allow for an iterative updating, where new elements of the series are accumulated “on-the-fly” to the estimators. A first attempt to rewrite Eq. (C.52) is the well-known alternate representation of the variance

$$\check{\sigma}^2\left[\mathbf{x}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\right)^2 . \tag{C.53}$$

A straightforward implementation leads to the two estimators

$$\begin{aligned}S_1(n) &= \sum_{i=1}^{n}x_i \\ S_2(n) &= \sum_{i=1}^{n}x_i^2,\end{aligned} \tag{C.54}$$

which recover the expectation value by $\check{\mu}[\mathbf{x}] = S_1(n)/n$ and the variance by $\check{\sigma}^2[\mathbf{x}] = \big(S_2(n) - S_1^2(n)/n\big) / n$ and allow for iterative accumulation by

$$\begin{aligned} S_1(n+1) &= S_1(n) + x_{n+1} \\ S_2(n+1) &= S_2(n) + x_{n+1}^2. \end{aligned} \tag{C.55}$$

However, the great disadvantage of this scheme is its numerical volatility. The two sums reach large values over the course of the accumulation. The variance is given by the difference of these large numbers and is thus subject to critical cancelation. This can be circumvented since the variance is invariant under a shift, $\mathrm{Var}[\mathbf{x} - \mathbf{a}] = \mathrm{Var}[\mathbf{x}]$ for any constant $\mathbf{a}$. Applied to the sums S_1 and S_2, this yields a new scheme:

$$\begin{aligned} S_1(n) &= \sum_{i=1}^{n}(x_i - a) = S_1(n-1) + (x_n - a) \\ S_2(n) &= \sum_{i=1}^{n}(x_i - a)^2 = S_2(n-1) + (x_n - a)^2 \\ \check{\mu}[\mathbf{x}] &= S_1(n)/n + \mathbf{a} \\ \check{\sigma}^2[\mathbf{x}] &= \frac{S_2(n)}{n} - \left(\frac{S_1(n)}{n}\right)^2 \end{aligned} \tag{C.56}$$

The critical cancellation is alleviated if $\mathbf{a}$ is close to the true expectation value $\mathrm{Exp}[\mathbf{x}]$. In the limit $\mathbf{a} \to \mathrm{Exp}[\mathbf{x}]$ the variance is govered by $S_2(n)$ since $S_1(n)$ averages to zero. The shift requires an a priori estimated value for $\mathrm{Exp}[a]$, and in an actual implementation, the shift is iteratively adapted. It is also possible to change the shift from one value to another during the accumulation with the proper corresponding modifications to S_1 and S_2, which are straightforward to derive but not presented here.

With the above scheme, the variance can be calculated on-the-fly. As more elements of the Markov chain are considered and the accumulators S_1 and S_2 are growing, the shift can be refined. The accumulators thus never exceed the threshold for critical cancellation. While this scheme was derived with the above-mentioned simplifications, it is readily extended to weighted samples, a general observable $g(\mathbf{x})$ instead of $\mathbf{x}$, and able to take into account the Bessel correction.

C.9. Multivariant exponential distribution

The Green function from Sec. 6.13, tailored to more accurately reproduce the true Green function in the vicinity of nuclei and nodes of the trial wave function, requires drawing from an exponential distribution in three dimensions. The distribution of interest is known, for example, from the hydrogen-like atom, and reads

$$f(\mathbf{r}) = \frac{\gamma^3}{\pi} e^{-2\gamma|\mathbf{r}|} \tag{C.57}$$

with a scale parameter -2γ, that resembles the density of an atom of charge γ. For the sake of completeness, this section is dedicated to how to draw from this distribution since the literature on this specific distribution appears to be scattered. To draw a vector from this distribution, it is convenient to change to spherical coordinates since the distribution does not depend on the direction of $\mathbf{r}$. Note that since f does not factorize into marginal distributions for the coordinates x, y, and z, one cannot draw an exponential random variable for every coordinate and combine them into a vector. With $\mathbf{r} = r\hat{\mathbf{e}}_r(\theta, \varphi)$ the angles $\theta \sim \mathcal{U}(0, \pi)$ and $\varphi \sim \mathcal{U}(0, 2\pi)$ are readily drawn uniformly from their respective intervals. The radius itself is not distributed with a one-dimensional exponential distribution. However, the radial distribution is readily identified by marginalizing the angular variables. For some function g of the radius, the expectation value can be written as

$$\operatorname{Exp}\left[g(r)\right] = \operatorname{Exp}\left[g(|\mathbf{r}|)\right] = 4\gamma^3 \int_0^\infty g(r) e^{-2\gamma r} r^2 \, \mathrm{d}r. \tag{C.58}$$

This reveals the distribution of the radius as $4\gamma^3 r^2 e^{-2\gamma r}$, which can be found to be the Gamma distribution [61] with shape parameter $\alpha = 3$ and scale parameter $\lambda = 2\gamma$. More precisely, it is the Erlang distribution, since α is a positive integer.

$$f_\Gamma(r) = \frac{\lambda^\alpha}{\Gamma(\alpha)} r^{\alpha-1} e^{-\lambda r} \tag{C.59}$$

Therefore, to draw a vector distributed with f, two angles must be drawn from the uniform distribution and a radius from the Erlang distribution, $r \sim \operatorname{Erlang}(\alpha = 3, \lambda = 2\gamma)$. Since the Erlang distribution represents a random variable, that is the sum of α one-dimensional exponential random variables, and exponential variables can be generated from uniform variables, this chain of transformation can be used to arrive at

$$r = -\frac{1}{\lambda} \sum_{i=1}^{\alpha} \ln u_i \sim \operatorname{Erlang}(\alpha, \lambda) \tag{C.60}$$

where $u_i \sim \mathcal{U}(0, 1)$. In the present case for a single vector $\mathbf{r} \sim f$, five uniform variables are required, two for the angle and three for the radius.

C.10. Covariance-related formulas

This section covers the tedious derivation of covariance-related formulas, that are given here to relief Sec. C.6 from it. Let $n \in \mathbb{N}$, $a \in \mathbb{R}$ and $0 < a < 1$, then

$$S_0 := \sum_{i,j=1}^{n} a^{|i-j|} = n\frac{1+a}{1-a} - 2a\frac{(1-a^n)}{(1-a)^2}. \tag{C.61}$$

Proof:

$$\begin{aligned}
S_0 :&= \sum_{i,j=1}^{n} a^{|i-j|} = n + 2\sum_{k=1}^{n-1}(n-k)a^k = n + 2\left[n\left(\sum_{k=0}^{n-1} a^k - 1\right) - \sum_{k=1}^{n-1} k a^k\right] \\
&= n + 2\left[n\left(\frac{1-a^n}{1-a} - 1\right) - \frac{(n-1)a^{n+1} - na^n + a}{(1-a)^2}\right] \\
&= n + 2\left[n\left(\frac{(1-a^n)(1-a) - (1-a)^2}{(1-a)^2}\right) - \frac{(n-1)a^{n+1} - na^n + a}{(1-a)^2}\right] \\
&= n + 2\frac{-na^n - na + na^{n+1} - na^2 - (n-1)a^{n+1} + na^n - a}{(1-a)^2} \\
&= n + 2\frac{-na + a^{n+1} - na^2 - a}{(1-a)^2} = n + 2a\frac{n(1-a) + a^n - 1}{(1-a)^2} \\
&= n + 2a\frac{n(1-a) - (1-a^n)}{(1-a)^2} = n\left[1 + \frac{2a}{1-a}\right] - \frac{2a(1-a^n)}{(1-a)^2} \\
&= n\frac{1+a}{1-a} - 2a\frac{(1-a^n)}{(1-a)^2}
\end{aligned} \tag{C.62}$$

Here the two formulas

$$\sum_{i,j=1,\ i\neq j}^{N} a^{|i-j|} = 2\sum_{\delta=1}^{N-1}(N-\delta)f(0,\delta) \tag{C.63}$$

and

$$\sum_{k=1}^{n-1}(n-k)a^{k} = a\ \frac{a^{n}+n(1-a)-1}{(1-a)^{2}} \tag{C.64}$$

are required. The first originates from a reordering of the summation to count the number of summands with the same exponent. The second is a well-known equation related to the geometric series.

With the same assumptions on a, and $n-1 < m \in \mathbb{N}$, the following holds.

$$S_m := \sum_{i,j=1}^{n} a^{|i-j+m|} = a^{m-n+1}\frac{(1-a^{n})^{2}}{(1-a)^{2}} \tag{C.65}$$

Proof: The integer m is bigger than any difference $i-j$ in the sum and the abolute value in the exponent can be omitted. Then

$$\begin{aligned} S_m :&= \sum_{i,j=1}^{n} a^{|i-j+m|} = \sum_{i,j=0}^{n-1} a^{i-j+m} = a^{m}\sum_{i=0}^{n-1}a^{i}\sum_{j=0}^{n-1}a^{-j} \\ &= a^{m}\frac{1-a^{n}}{1-a}\frac{1-a^{-n}}{1-a^{-1}} = a^{m}\frac{1-a^{n}}{1-a}\frac{1-a^{-n}}{1-a^{-1}} \\ &= -a^{m}\frac{(1-a^{n})(a-a^{-n+1})}{(1-a)^{2}} = -\frac{a^{m+1}}{(1-a)^{2}}(1-a^{n})(1-a^{-n}) \\ &= \frac{a^{m+1}}{(1-a)^{2}}(a^{n}+a^{-n}-2) = a^{m-n+1}\frac{(1-a^{n})^{2}}{(1-a)^{2}}. \end{aligned} \tag{C.66}$$

D. Drift-diffusion equation

The diffusion equation (6.17) is the central equation for DMC. With a vanishing sink-source term, it also plays a vital role in efficient VMC. This chapter is dedicated to understanding the basic properties of this PDE, devising first approximations to its Green function, and recasting it in a framework compatible with MC.

A suitable starting point is the Fokker-Planck equation [64, 65]

$$\partial_t p(\mathbf{r},t) = -\nabla\left[\mathbf{u}(\mathbf{r},t)p(\mathbf{r},t)\right] + \frac{1}{2}\mathrm{tr}\left(\nabla\nabla\left[\mathbf{D}(\mathbf{r},t)p(\mathbf{r},t)\right]\right), \tag{D.1}$$

that specifies the time evolution of a probability distribution $p(\mathbf{r},t)$ with an initial condition $p(\mathbf{r},0)$ in the presence of a drift $\mathbf{u}$ and a diffusion matrix $\mathbf{D}$. This work's drift and diffusion are time-independent, allowing for a stationary state of p. The time evolution reaches this steady state after sufficient propagation time when the dependence on the initial condition is weak enough. The above Fokker-Planck equation does not include a source term, that is required for the DMC method. With a source term added and the other properties inserted, the diffusion equation under study reads

$$-\partial_t p(\mathbf{r},t) = -D\nabla^2 p(\mathbf{r},t) + \nabla\cdot\left[\mathbf{u}(\mathbf{r})p(\mathbf{r},t)\right] - S(\mathbf{r},t)p(\mathbf{r},t). \tag{D.2}$$

The source term S and the drift vector $\mathbf{u}$ show intricate behavior in real DMC applications. Their spatial dependence on the coordinate vector $\mathbf{r}$ even possesses singularities. For the sake of simplicity, both quantities are assumed to be continuous here.

D.1. Fundamental solution

Consider a constant drift $\mathbf{u}$ without a source term $S = 0$ in the diffusion equation on the domain with $\mathbf{r} \in \mathbb{R}^n$. The initial value problem can be written

$$-\partial_t p(\mathbf{r},t) = -D\nabla^2 p(\mathbf{r},t) + \mathbf{u}\cdot\boldsymbol{\nabla} p(\mathbf{r},t) \tag{D.3}$$

$$p(\mathbf{r},0) = p_0(\mathbf{r}) \tag{D.4}$$

$$\lim_{|r|\to\infty} p(\mathbf{r},t) = 0. \tag{D.5}$$

To arrive at an analytic solution, define the Fourier transformation in symmetric form,

$$\hat{p}(\mathbf{k},t) = \mathcal{FT}[p(\mathbf{r},t)] = \frac{1}{(2\pi)^{\frac{n}{2}}}\int p(\mathbf{r},t)e^{-i\mathbf{r}\cdot\mathbf{k}}\,\mathrm{d}\mathbf{r}. \tag{D.6}$$

$$p(\mathbf{r},t) = \mathcal{FT}^{-1}[\hat{p}(\mathbf{k},t)] = \frac{1}{(2\pi)^{\frac{n}{2}}}\int \hat{p}(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}\mathbf{k} \tag{D.7}$$

The diffusion equation is recast in Fourier space by substituting $p(\mathbf{r},t)$ by the backward transformation Eq. (D.7).

$$\begin{aligned}
0 &= -\partial_t p(\mathbf{r},t) + D\nabla^2 p(\mathbf{r},t) - \mathbf{u}\cdot\boldsymbol{\nabla} p(\mathbf{r},t) \\
\Leftrightarrow 0 &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int \left[-\partial_t \hat{p}(\mathbf{k},t) + D\nabla^2 \hat{p}(\mathbf{k},t) - \mathbf{u}\cdot\boldsymbol{\nabla}\hat{p}(\mathbf{k},t)\right] e^{i\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}\mathbf{k} \\
\Leftrightarrow 0 &= \int \left[-\partial_t \hat{p}(\mathbf{k},t) - D|\mathbf{k}|^2 \hat{p}(\mathbf{k},t) - i\mathbf{u}\cdot\mathbf{k}\hat{p}(\mathbf{k},t)\right] e^{i\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}\mathbf{k} \\
\Leftrightarrow 0 &= -\partial_t \hat{p}(\mathbf{k},t) - D|\mathbf{k}|^2 \hat{p}(\mathbf{k},t) - i\mathbf{u}\cdot\mathbf{k}\hat{p}(\mathbf{k},t).
\end{aligned} \tag{D.8}$$

The resulting equation can be solved by separation of variables

$$\hat{p}(\mathbf{k},t) = \hat{p}(\mathbf{k},0)\exp\left[\left(-D|\mathbf{k}|^2 - i\mathbf{u}\cdot\mathbf{k}\right)t\right] = \hat{p}_0(\mathbf{k})\hat{g}(\mathbf{k},t) \tag{D.9}$$

with

$$\hat{g}(\mathbf{k},t) = \exp\left[\left(-D|\mathbf{k}|^2 - i\mathbf{u}\cdot\mathbf{k}\right)t\right]. \tag{D.10}$$

Now that the problem is solved in Fourier representation, the convolution theorem can be applied to the product $\hat{p}_0\hat{g}$.

$$p(\mathbf{r},t) = \mathcal{FT}^{-1}[\hat{p}_0(\mathbf{k})\hat{g}(\mathbf{k},t)] = \frac{1}{(2\pi)^{\frac{n}{2}}}(p_0 * g)(\mathbf{r},t) \tag{D.11}$$

To resolve the convolution with g, the backward Fourier transformation must be applied to $\hat{g}$.

$$\begin{aligned}
g(\mathbf{r},t) &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int e^{-D|\mathbf{k}|^2 t - i(\mathbf{u}t-\mathbf{r})\cdot\mathbf{k}}\,\mathrm{d}\mathbf{k} \\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\int e^{-Dk_i^2 t - i(u_i t - r_i)\cdot k_i}\,\mathrm{d}k_i \\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{2Dt}} e^{\frac{-(u_i t - r_i)^2}{4Dt}} \\
&= \frac{1}{(2Dt)^{\frac{n}{2}}} e^{\frac{-|\mathbf{r}-\mathbf{u}t|^2}{4Dt}}
\end{aligned} \tag{D.12}$$

This removes the imaginary unit from the equations. Finally, the solution to Eq. (D.3) is given by

$$\begin{aligned}
p(\mathbf{r},t) &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int p_0(\mathbf{r}')g(\mathbf{r}-\mathbf{r}',t)\,\mathrm{d}\mathbf{r}' = \\
&= \frac{1}{(4\pi Dt)^{\frac{n}{2}}}\int p_0(\mathbf{r}')\exp\left[-\frac{|\mathbf{r}-\mathbf{r}'-\mathbf{u}t|^2}{4Dt}\right]\mathrm{d}\mathbf{r}'.
\end{aligned} \tag{D.13}$$

The corresponding Green function is readily read from this equation to be the normal distribution

$$G_D(\mathbf{r},\mathbf{r}',t) = \frac{1}{(4\pi Dt)^{\frac{n}{2}}}\exp\left[-\frac{|\mathbf{r}-\mathbf{r}'-\mathbf{u}t|^2}{4Dt}\right]. \tag{D.14}$$

This is the drift-diffusion Green function for constant drift in n dimensions. It propagates any initial probability distribution $p_0(\mathbf{r})$ for the given time t by the kernel integral

$$p(\mathbf{r}',t) = \int G_D(\mathbf{r}',\mathbf{r},t)p_0(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{D.15}$$

The exact Green function of the diffusion equation with coordinate-dependent drift and sink-source terms of the DMC Eq. (6.17) clearly will differ from the Green function found here. Nevertheless, the above Green function poses a significant part of much more elaborate Green functions and can be utilized as a short-time approximation, as will become apparent in the next sections.

D.2. Stochastic trajectories

The Green function found in Eq. (D.14) is relatively simple and only solves a diffusion equation with constant drift. It does not describe more complex diffusion equations exactly that, for example, show coordinate-dependent drift. For this section, however, assume the Green function G_D, that solves some diffusion equation under study, is already found and not bound to be of the form (D.14). Nevertheless, for illustration of the concepts and instructive insights, the Green function for constant drift will be used as an example in the following.

Due to the high number of dimensions, the n-dimensional integral in Eq. (D.15) can neither be calculated analytically nor numerically with standard integration schemes for some arbitrary initial probability distribution. Thus, even with the full Green function at hand, the time evolution $p(\mathbf{r}', t)$ is not accessible, and a different approach is required to reach the steady state.

Recall that the solution of the diffusion equation, $p(\mathbf{r}, t)$, is a probability distribution. As such, it can either be interpreted as an ensemble representation of infinitely many particles in the absence of uncertainty or describe a single or more particles with uncertainty. To begin with, assume the probability distribution p describes the location of a single particle in the presence of uncertainty. If the drifting and diffusing particle has the definite position $\mathbf{r}(t)$ at time t, the probability distribution at this time is given by a deterministic Dirac distribution

$$p(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}(t)). \tag{D.16}$$

Next, the particle is subject to some general drift-diffusion process described by G_D for some time interval Δt. After this time, the distribution has broadened due to diffusion, its mean might be shifted, and its shape may has changed compared to the original distribution. The new distribution is, of course, given by propagation with the Green function.

$$p(\mathbf{r}', t + \Delta t) = \int G_D(\mathbf{r}', \mathbf{r}, \Delta t)\delta(\mathbf{r} - \mathbf{r}(t))\,\mathrm{d}\mathbf{r} = G_D(\mathbf{r}', \mathbf{r}(t), \Delta t) \tag{D.17}$$

In the simple case of Eq. (D.14) of constant drift,

$$p(\mathbf{r}', t + \Delta t) = \frac{1}{(4\pi D\Delta t)^{\frac{n}{2}}} \exp\left[-\frac{|\mathbf{r}' - \mathbf{r}(t) - \mathbf{u}\Delta t|^2}{4D\Delta t}\right], \tag{D.18}$$

with a displacement $\mathbf{u}\Delta t$ given by the drift vector. The broadening of the distribution can be described by the variance of the normal distribution $\sigma = \sqrt{2D\Delta t}$ and is increasing with time.

If the particle described by $p(\mathbf{r}, \Delta t)$ drift-diffuses for an additional time interval Δt, the resulting probability is

$$\begin{aligned} p(\mathbf{r}, t + 2\Delta t) &= \int G_D(\mathbf{r} - \mathbf{r}', \Delta t)p(\mathbf{r}', t + \Delta t)\,\mathrm{d}\mathbf{r}' \\ &= \int G_D(\mathbf{r} - \mathbf{r}', \Delta t)G_D(\mathbf{r}' - \mathbf{r}(t))\,\mathrm{d}\mathbf{r}'. \end{aligned} \tag{D.19}$$

The integral with a double Green function can be solved analytically for a constant drift since integrals over two Gaussian kernels can be evaluated analytically. However, this is not possible for more realistic Green functions of different shapes, and therefore, $p(\mathbf{r}, t + 2\Delta t)$ is inaccessible. One might question why the calculation of $p(\mathbf{r}, t + 2\Delta t)$ should be written as a two-step process instead of direct propagation from time zero to time $2\Delta t$. This will become clear later when it is established that the Green function above only stays a reasonable approximation to more complex problems when used with small time steps Δt, which motivates splitting a large time interval into multiple smaller steps. Instead of repetitively calculating a new distribution from an ever-increasing series of Green function integrals

$$p(\mathbf{r}, t + n\Delta t) = \int \cdots \int G_D(\mathbf{r}, \mathbf{r}_{n-1}) \ldots G_D(\mathbf{r}_1, \mathbf{r}_0) p(\mathbf{r}_0, t) \, \mathrm{d}\mathbf{r}_0 \ldots \, \mathrm{d}\mathbf{r}_{n-1} \tag{D.20}$$

the focus can be put on

$$\mathbf{r}(t + \Delta t) \sim p(\cdot, t + \Delta t) = G_D(\cdot, \mathbf{r}(t)), \tag{D.21}$$

recalling its nature as a random variable. Then, the key idea to solve the diffusion process is to collapse the propagated probability distribution $p(\mathbf{r}, t + \Delta t)$, or, equivalently, its random variable $\mathbf{r}(t + \Delta t)$. This is accomplished by drawing a random variate from it, i.e., generating a random realization. For the simple case of the Green function Eq. (D.14) with constant drift, such a realization would be

$$\mathbf{r}(t + \Delta t) \sim \mathcal{N}(\mathbf{r}(t) + \mathbf{u}\Delta t, 2D\Delta t), \tag{D.22}$$

where $\mathcal{N}(\boldsymbol{\mu}, \sigma^2)$ denotes the normal distribution with expectation value $\boldsymbol{\mu}$ and variance σ^2. This distribution is easily drawn from utilizing standard random number generators and, for example, the Box-Muller transformation. The probability distribution $\tilde{p}$ that describes the distribution after the collapse of the uncertainty is again a Dirac distribution,

$$\tilde{p}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}(t + \Delta t)). \tag{D.23}$$

The expectation value of $\tilde{p}$ is $\mathbf{r}(t + \Delta t)$, and since $\mathbf{r}(t + \Delta t)$ is by construction the expectation value of $p(\mathbf{r}, t + \Delta t)$, the collapse conserves the expectation value. The variance of $\tilde{p}$ of course vanishes, compared to the uncollapsed distribution. Thus, the essential approximation of this approach is that the collapsed distribution replaces the uncollapsed,

$$p(\mathbf{r}, t + \Delta t) = \tilde{p}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}(t + \Delta t)). \tag{D.24}$$

This yields an iterative scheme, where at any time t from the deterministic position $\mathbf{r}(t)$ a new distribution $p(\mathbf{r}, t + \Delta t)$ is generated from propagation of $\delta(\mathbf{r} - \mathbf{r}(t))$, and is subsequently collapsed to $\mathbf{r}(t + \Delta t)$.

$$\mathbf{r}(t) \overset{\text{propagation}}{\mapsto} G_D(\mathbf{r}, \mathbf{r}(t)) \overset{\text{collapse}}{\sim} \mathbf{r}(t + \Delta t) \overset{\text{propagation}}{\mapsto} \cdots \tag{D.25}$$

The series $(\mathbf{r}(\Delta t), \mathbf{r}(2\Delta t), \ldots)$ is a Markov chain since the generation of any next step $\mathbf{r}(t + \Delta t)$ depends only on the current state $\mathbf{r}(t)$. Clearly, the collapse of the probability distribution to a Dirac distribution eliminates information about the full distribution. A large number of time steps are required to restore this information. However, most of the eliminated information is not required for the accurate evaluation of expectation values. Another view of the approximation is the law of large numbers. The distribution $G_D(\cdot, \mathbf{r}(t))$ is approximated by a single random variate of itself, instead of a whole set of variates. The error that is introduced by the application of the law of large numbers with only a single variate has to be remedied with a large propagation time. After an initial phase of convergence to the steady state, which eliminates the dependence on the initial position $\mathbf{r}(0)$, the probability distribution $p(\mathbf{r}, t)$ is independent of time, and all random variates resemble the same distribution in the long-time limit.

In the limit of small time steps Δt, the above process might be interpreted as a continuous process and can be reformulated into a classical (ordinary) but stochastic differential equation

$$\dot{\mathbf{r}} = \boldsymbol{\xi} \quad \text{with} \quad \boldsymbol{\xi} \sim G_D(\cdot, \mathbf{r}). \tag{D.26}$$

For the case of constant drift, examination of the equation of motion yields

$$\dot{\mathbf{r}} = \mathbf{u} + \boldsymbol{\xi} \quad \text{with} \quad \boldsymbol{\xi} \sim \mathcal{N}(0, 2D). \tag{D.27}$$

This suggests the notion of a continuous stochastic trajectory $\mathbf{r}(t)$ with a stochastic equation of motion or a Langevin-type equation. Integrating the equations of motion yields

$$\mathbf{r}(t) = \mathbf{r}(0) + \mathbf{u}t + \int_0^t \boldsymbol{\xi}(t')\,\mathrm{d}t'. \tag{D.28}$$

The expectation value $\mathrm{Exp}\left[\mathbf{r}(t)\right] = \mathbf{r}(0) + \mathbf{u}t$ averages out the random displacement. The drift guides the particle, or equivalently its corresponding distribution, in the direction $\mathbf{u}$. However, its variance is ever-increasing and reveals the random character of the trajectory.

$$\begin{aligned} \mathrm{Var}\left[|\mathbf{r}(\mathbf{t})|\right] &= \int_0^t \int_0^t \mathrm{Cov}\left[\boldsymbol{\xi}(t'), \boldsymbol{\xi}(t'')\right]\,\mathrm{d}t'\,\mathrm{d}t'' \\ &= \mathrm{Var}\left[\boldsymbol{\xi}\right] t = 2Dt \end{aligned} \tag{D.29}$$

The random displacement $\boldsymbol{\xi}(t)$ is an uncorrelated and independent random variable for each time step, $\mathrm{Cov}\left[\boldsymbol{\xi}(t), \boldsymbol{\xi}(t')\right] = \mathrm{Var}\left[\boldsymbol{\xi}\right]\delta(t - t')$.

Returning to a general drift-diffusion equation, Eq. (D.26) is still valid. A variate for $\boldsymbol{\xi}$ drawn from an arbitrary Green function, $\boldsymbol{\xi} \sim G_D(\cdot, \mathbf{r})$, can still be described by some vector $\mathbf{u}$ and some diffusive random displacement $\boldsymbol{\xi}'$ with zero mean, $\mathrm{Exp}\left[\boldsymbol{\xi}'\right] = 0$. The drift vector is in general allowed to depend on the coordinate $\mathbf{u}(\mathbf{r})$. Simultaneously, the diffusive displacement $\boldsymbol{\xi}'$ is drawn from a zero-mean distribution that might depend on $\mathbf{r}$. The equations of motion can be integrated numerically, discretizing the timeline into time steps Δt with a discrete time evolution

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \big(\mathbf{u}(\mathbf{r}) + \boldsymbol{\xi}'(t)\big)\,\Delta t \tag{D.30}$$

with some i.i.d. realization of $\boldsymbol{\xi}'(t)$ for every single time step.

The notable conceptual advantage of the classical stochastic trajectory of Eq. (D.26) compared to the PDE formulation of Eq. (D.3) cannot be underestimated. Instead of solving a PDE, which becomes exponentially infeasible with increasing dimensions, a classical trajectory, even though stochastic, can be solved with a minimal computational cost, even in high-dimensional spaces.

D.3. Sink-source term

In Eq. (D.2) an additional sink-source term, S, is present. It modulates the normalization constant of p. Assuming both a constant source term S and constant drift, the diffusive process reads

$$-\partial_t p(\mathbf{r}, t) = -D\nabla^2 p(\mathbf{r}, t) + \mathbf{u} \cdot \boldsymbol{\nabla} p(\mathbf{r}, t) - Sp(\mathbf{r}, t), \tag{D.31}$$

This equation is not norm-conserving since S adds or removes probability. The interpretation of p as a probability distribution can therefore no longer hold. A reinterpretation of the solution p to a concentration profile with undetermined total mass is required. Equivalently, the process can be generalized to consist of a probability distribution f and a normalization constant λ that together solve the Eq. (D.31) as a tuple (f, λ) with $p = \lambda f$. Hence, $f(\cdot, t)$ represents the spatial probability distribution as a normalized probability density function subject to a time evolution, while $\lambda(t)$ holds a global (positive) normalization that also changes with time.

Repeating the derivations made in Sec. D.1 with an additional constant source S, the corresponding Green function reads

$$G(\mathbf{r}', \mathbf{r}, \Delta t) = \frac{\exp\left[S\Delta t\right]}{(4\pi D\Delta t)^{\frac{n}{2}}} \exp\left[-\frac{|\mathbf{r}' - \mathbf{r} - \mathbf{u}\Delta t|^2}{4D\Delta t}\right]. \tag{D.32}$$

The Green function splits into two conceptually different factors, $G = G_D \times G_B$. While the drift-diffusion part G_D is a valid, i.e., normalized, transition probability, the branching Green function G_B accounts for the increase or decay of total normalization. In the simple case above, both factors can be easily identified. The drift-diffusion Green function is the same as in Sec. D.1. The branching Green function, however, takes the form $G_B = \exp\left[S\Delta t\right]$. In this simple formulation, the effect of the sink-source term is trivial, increasing or decreasing the total normalization at a constant rate. The branching term prevents the interpretation of Eq. (D.31) as a single stochastic trajectory since usually a particle moving along a trajectory has a normalized probability. However, the introduction of a weight term alongside the coordinate in a tuple, $(\lambda(t), \mathbf{r}(t))$, allows one to stay very close to the trajectory interpretation Eq. (D.26). A simple and illustrative example of such a process is that of a raindrop falling to the ground (drift) while experiencing random collisions (diffusion) with other raindrops and an increase of mass due to the collisions or condensation and evaporation in more humid or more dry air.

The tuple $(\lambda(t), \mathbf{r}(t))$ has coupled equations of motion. For every update of the coordinates, there is a corresponding update to the weight of the form

$$\lambda(t + \Delta t) = G_B(\Delta t)\lambda(t) = \exp\left[S\Delta t\right]\lambda(t) \tag{D.33}$$

In differential form, the equations of motion read

$$\dot{\mathbf{r}} = \mathbf{u} + \boldsymbol{\xi} \tag{D.34}$$

$$\dot{\lambda} = S\lambda, \tag{D.35}$$

with $\lambda > 0$ for all times.

D.4. Short-time approximation

To better approximate a solution to the full DMC diffusion process of Eq. (D.2), the coordinate dependency of the drift vector $\mathbf{u}(\mathbf{r})$ and sink-source term $S(\mathbf{r})$ has to be respected. Assuming both terms have only small deviations on a small length scale compared to the displacement $\dot{\mathbf{r}}\Delta t$, the simple Green function of Eq. (D.32) can be used to propagate the probability distribution for small enough time steps. Insertion of the local values of $\mathbf{u}(\mathbf{r})$ and $S(\mathbf{r})$ into the equation of motion yields

$$\begin{aligned} \dot{\mathbf{r}} &= \mathbf{u}(\mathbf{r}) + \boldsymbol{\xi} \\ \dot{\lambda} &= S(\mathbf{r})\lambda \end{aligned} \tag{D.36}$$

with $\boldsymbol{\xi} \sim \mathcal{N}(0, 2D)$. The assumption is particularly relevant for small time steps Δt. Integration of the equations of motion along the trajectory $\mathbf{r}(t)$ results in

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{r}(0) + \int_0^t \big(\mathbf{u}(\mathbf{r}(t')) + \boldsymbol{\xi}(t')\big)\, \mathrm{d}t' \\ \lambda(t) &= \exp\left[\int_0^t S(\mathbf{r}(t'))\, \mathrm{d}t'\right] \lambda(0) \end{aligned} \tag{D.37}$$

With the spatial dependence inserted into the equations of motion, the previously mostly trivial equations of motion become much more useful to approximate Eq. (D.2) immediately. Instead of a particle drifting with a constant vector, the drift vector now changes depending on the coordinate of the particle. Furthermore, depending on its location in space, the particle can gain or lose normalization. This allows the time evolution to exhibit much more details. Note that the source term S only influences the normalization as a side effect and has no influence on the trajectory.

With some numerical discretization scheme for the time steps Δt, a numerical integration yields the trajectory. A rather simple but sufficient approach is the Euler method,

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \left[\mathbf{u}(\mathbf{r}(t)) + \boldsymbol{\xi}(t)\right]\Delta t \tag{D.38}$$

$$\lambda(t+\Delta t) = \exp\left[S(\mathbf{r}(t))\Delta t\right]\lambda(t). \tag{D.39}$$

An improved numerical integration scheme to the classical problem is to approximate the integrals in Eq. (D.37) to a higher order. Recalling that the evaluation of $\mathbf{u}(\mathbf{r})$ and $S(\mathbf{r})$ might be computationally expensive, higher orders might also slow down the propagation. However, for the weight a simple yet considerably improved integration compared to Eq. (D.39), a simple Crank-Nicolson scheme, uses the information of the location of $\mathbf{r}(t+\Delta t)$ for integration of S:

$$\lambda(t+\Delta t) = \exp\left[\frac{\Delta t}{2}\left(S(\mathbf{r}(t+\Delta t)) + S(\mathbf{r}(t))\right)\right]\lambda(t) \tag{D.40}$$

Since the source term S has to be evaluated at $\mathbf{r}(t+\Delta t)$ in any case for the next iterations, it poses no additional computational cost to include the information available in $S(\mathbf{r}(t+\Delta t))$ already in the present iteration. From this weight update, a Green function for the branching can be inferred.

$$G_B(\mathbf{r}',\mathbf{r},\Delta t) = \exp\left[\frac{\Delta t}{2}\left(S(\mathbf{r}') + S(\mathbf{r})\right)\right] \tag{D.41}$$

This completes the derivation of the Green function in its short-time approximation. It takes the final form,

$$G(\mathbf{r}',\mathbf{r},\Delta t) = \frac{1}{(4\pi D\Delta t)^{\frac{n}{2}}}\exp\left[-\frac{|\mathbf{r}'-\mathbf{r}-\mathbf{u}\Delta t|^2}{4D\Delta t}\right]\exp\left[\frac{\Delta t}{2}\left(S(\mathbf{r}') + S(\mathbf{r})\right)\right]. \tag{D.42}$$

Note that the same Green function can be derived following Feynman et al. [20] using his path integral formalism. Furthermore, using the Baker-Campbell-Hausdorff formula [79] to expand the operator $\hat{\mathrm{G}}$ from Chapt. 6 in second order (cf. Eq. (6.8)) also corresponds to the above Green function of Eq. (D.42). The expansion to first order (cf. Eq. (6.7)) corresponds to the Green function with branching term from Eq. (D.33), where the contributions from configuration $\mathbf{r}'$ in the branching term are omitted.

The short-time approximation works best for slowly varying drift and source terms. However, in the DMC use case with both terms showing intricate behavior, including discontinuities and singularities, the Green function does not accurately capture the distribution characteristics near those irregularities. Although significant improvement to the Green function can be made in this case, discussed in Sec. 6.13, the short-time approximation remains at the core of DMC.

Note that the transition from constant to spatially varying drift and source was accomplished by simply adding the spatial dependence in the equations of motion (D.36) that originally were derived for constant drift and source. This shortcoming will be addressed in the next section.

D.5. Revealing approximations

This section reviews the stochastic realization of the short-time approximation to the Green function from the previous section, revealing the approximations made therein.

Although various authors refer to the numerical scheme described above, the given derivations lack certain details and do not show all approximations involved [53, 96, 155].

The connection between the equilibrium distribution p and a classical stochastic trajectory or random walk was already highlighted in Sec. D.2. Compared to Sec. D.2, the Green function of Eq. (D.2) is not normalized anymore, which introduced the notion of a weight in addition to the random walk in sec. D.3. The weighted random walk still assumes to collapse the propagated probability distribution into the deterministic Dirac distribution. Assume, at some time t, the distribution is given by

$$p(\mathbf{r},t) = \lambda(t)\delta(\mathbf{r}-\mathbf{r}(t)). \tag{D.43}$$

Propagating p with an arbitrary Green function G results in

$$p(\mathbf{r}',t+1) = \lambda(t)G(\mathbf{r}',\mathbf{r}(t)) \tag{D.44}$$

and a new weight and position have to be generated from $p(\mathbf{r}',t+1)$. The core idea in Sec. D.2, namely to collapse the probability distribution to a single position, remains the same. However, in the presence of a spatially varying source term $S(\mathbf{r})$ and therefore the spatially varying branching Green function $G_B(\mathbf{r}',\mathbf{r},\Delta t)$ from Eq. (D.41), the differentiation between the probability distribution $G_D(\mathbf{r}',\mathbf{r},\Delta t)$ to draw a new coordinate from and the weight adjustment from the branching Green function $G_B(\mathbf{r}',\mathbf{r},\Delta t)$ is not possible anymore. Instead of the drift-diffusion distribution alone, now the product of drift-diffusion and branching Green function is the full unnormalized distribution $\lambda(t)G(\cdot,\mathbf{r}(t))$ to draw from. The product distribution has a normalization constant, which simultaneously is the new weight. It can be calculated from

$$\lambda(t+1) = \int p(\mathbf{r}',t+1)\,\mathrm{d}\mathbf{r}' = \lambda(t)\int G(\mathbf{r}',\mathbf{r}(t))\,\mathrm{d}\mathbf{r}'. \tag{D.45}$$

With the known normalization constant $\lambda(t+1)$, the normalized Green function to draw the displacement from is readily given by

$$\mathbf{r}(t+1) \sim \frac{p(\mathbf{r}',t+1)}{\lambda(t+1)} = \frac{\lambda(t)}{\lambda(t+1)}G(\mathbf{r}',\mathbf{r}(t)) \tag{D.46}$$

In practice, however, $\lambda(t+1)$ cannot be calculated since it includes the high-dimensional integration in Eq. (D.45). Hence, the update is handled differently in practice: Due to the lack of alternatives, the new configuration is simply drawn from the drift-diffusion part G_D of the full Green function $G = G_D \times G_B$ only,

$$\tilde{\mathbf{r}}(t+1) \sim G_D(\mathbf{r}',\mathbf{r}(t)). \tag{D.47}$$

Whereas in Eq. (D.23), the replacement conserved the expectation value exactly, the above certainly is an approximation and therefore introduces some error.

The error in the expectation value of $\mathbf{r}(t+1)$ can be understood from the mean-value theorem extended to vector-valued functions of several variables applied to the correct distribution from

Eq. (D.46).

$$\begin{aligned}\mathrm{Exp}\left[\mathbf{r}(t+1)\right] &= \frac{\lambda(t)}{\lambda(t+1)}\int \mathbf{r}' G(\mathbf{r}',\mathbf{r}(t))\,\mathrm{d}\mathbf{r}' \\ &= \frac{\lambda(t)}{\lambda(t+1)}\int \mathbf{r}' G_D(\mathbf{r}',\mathbf{r}(t)) G_B(\mathbf{r}',\mathbf{r}(t))\,\mathrm{d}\mathbf{r}' \\ &= \frac{\lambda(t)}{\lambda(t+1)}\left[\hat{\mathbf{e}}_i G_B(\boldsymbol{\xi}_i,\mathbf{r}(t))\int \mathbf{r}'_i G_D(\mathbf{r}',\mathbf{r}(t))\,\mathrm{d}\mathbf{r}'\right] \\ &= \mathbf{b}\odot \mathrm{Exp}\left[\tilde{\mathbf{r}}(t+1)\right]\end{aligned} \tag{D.48}$$

with the vector

$$\mathbf{b} = \frac{\lambda(t)}{\lambda(t+1)}\begin{pmatrix} G_B(\boldsymbol{\xi}_1,\mathbf{r}(t)) \\ G_B(\boldsymbol{\xi}_2,\mathbf{r}(t)) \\ \dots \\ G_B(\boldsymbol{\xi}_n,\mathbf{r}(t))\end{pmatrix} \tag{D.49}$$

and where $\odot$ denotes the Hadamard or element-wise product. With the right choice for all $\boldsymbol{\xi}_i$, the above relation is guaranteed to hold. From the mean-value theorem, every $\boldsymbol{\xi}_i$ is formally bound only within the support (the domain of non-vanishing values) of $G_D(\mathbf{r}',\mathbf{r}(t))$. Suppose G_D is given by a normal distribution, then its support is the whole configuration space. However, since G_D is a rapidly decaying function centered around $\mathbf{r}(t)+\Delta t\mathbf{u}(\mathbf{r}(t))$, every $\boldsymbol{\xi}_i$ is limited with high probability to a small volume, decreasing in size with smaller time steps. Thus, $\mathbf{b}$ approaches element-wise unity in the limit of zero time step, i.e. $\mathbf{b}=(1,1,\dots)$. For finite time steps, drawing from G_D instead of $G_D G_B$ poses an error to the expectation value of the random walk of linear order in the deviation from the correct choices for $\boldsymbol{\xi}_i$.

From this limit with $\lambda(t+1)=\lambda(t)G_B(\boldsymbol{\xi},\mathbf{r}(t))$ also an approximation for the new weight can be deferred as

$$\lambda(t+1)=\lambda(t)G_B(\mathbf{r}(t+1),\mathbf{r}(t)) \tag{D.50}$$

where $\mathbf{r}(t+1)$ was inserted for $\boldsymbol{\xi}$. Then, this weight does not require the full integration of Eq. (D.45) but only requires S to be evaluated at a single point, $\mathbf{r}(t+1)$. This choice for $\boldsymbol{\xi}$ is somewhat arbitrary and introduces a linear error but stays within the probable constraints of the mean-value theorem. Nevertheless, this arbitrariness also poses the opportunity to improve the weight update, see Sec. 6.5, where another approximation is introduced that is more along the line of the analytical form of $\lambda(t+1)$ in Eq. (D.45).

With all these approximations in mind, the probability distribution realized in practice becomes

$$p(\mathbf{r}',t+1)=\lambda(t+1)\delta(\mathbf{r}'-\tilde{\mathbf{r}}(t+1)) \tag{D.51}$$

with $\tilde{\mathbf{r}}(t+1)$ from Eq. (D.47) and $\lambda(t+1)$ following from Eq. (D.50).

E. DARSEC

In this thesis, all DFT calculations have been carried out with the DARSEC code [146]. The computational framework in which DARSEC solves the DFT calculations is presented below. DARSEC calculations are restricted to systems of axial symmetry to remove the axial angle from the numerical equations. The remaining two-dimensional problem is solved with grid methods. The DARSEC grid is additionally tailored to the geometry of atoms and diatomic molecules. The resulting non-Euclidean grid allows for a very accurate representation of the orbitals, enabling accurate all-electron calculations for diatomic molecules.

The non-interacting Schrödinger equation of a diatomic system is given by the Hamiltonian

$$\hat{\mathrm{H}} = -\frac{1}{2}\nabla^2 + v(\mathbf{r}), \tag{E.1}$$

$$v(\mathbf{r}) = v_i(\mathbf{r}) + v_e(\mathbf{r}) = -\frac{Z_A}{|\mathbf{r}-\boldsymbol{\xi}_A|} - \frac{Z_B}{|\mathbf{r}-\boldsymbol{\xi}_B|} + v_e(\mathbf{r}), \tag{E.2}$$

with ion A of charge Z_A at $\boldsymbol{\xi}_A$ and ion B of charge Z_B at $\boldsymbol{\xi}_B$. The ions A and B comprise the diatomic molecule or, in the case of $Z_A = 0$, the atom. Without loss of generality, the ions can be placed at $\xi_\mathbf{A} = (0, 0, -R/2)$ and $\xi_\mathbf{B} = (0, 0, R/2)$, to align the inter-atomic axis with the z-axis of the coordinate system. Their separation or bond distance is given by $R = |\boldsymbol{\xi}_B - \boldsymbol{\xi}_A|$. An additional potential v_e, which could model electric fields, harmonics potentials, potential steps, etc. , might enter the Hamiltonian if $v_e(\mathbf{r})$ is invariant to rotations around the z-axis or, equivalently, if v_e is a function of the variables z and $\sqrt{x^2+y^2}$ alone. Then, the Hamiltonian has cylindrical symmetry and could be expressed in a cylindrical coordinate system with coordinates $\rho = \sqrt{x^2+y^2}$, φ, and z. While a cylindrical coordinate system would allow the separation of the Hamiltonian, the use of an ellipsoidal coordinate system offers a more numerically efficient approach. The corresponding coordinates are the prolate spheroidal (PS) coordinates.

In cylindrical as well as PS coordinates, the Hamiltonian is independent of the azimuthal angle, $\partial\hat{\mathrm{H}}/\partial\varphi = 0$. Thus, the solution of the Schrödinger equation can be cast as a product of an angular solution and a solution in the ρ-z plane or the μ-ν-plane of the PS coordinates. In the following, the PS coordinate system and its grid are introduced together with details of the numerical scheme for the DFT calculation relevant to this work. The sections below do not cover every detail and intricacy of the DARSEC approach; they only review the aspects that are of special interest in this work. For a deeper insight into the DARSEC approach, refer to Ref. [146].

E.1. Prolate spheroidal coordinates

Grids suitable to describe diatomic molecules in an all-electron approach require high resolution at the nuclei. For computational efficiency, the regions far away from any nuclei, where orbitals show small variations, should be represented with lower resolution. A uniform grid in the natural coordinates of the PS coordinate system meets these requirements. It features two focal points along the z-axis, where the nuclei are located, that show a fine-grain resolution. In the asymptotic domains, the grid's resolution is lower.

Any Euclidean point $\mathbf{r} \in \mathbb{R}^3$ is represented in PS coordinates by the coordinate vector $\mathbf{q} = (\mu, \nu, \varphi) \in P$, with $P = [0, \infty] \times [0, \pi] \times [0, 2\pi]$. The transformation formula is given by

$$T : P \to \mathbb{R}^3, (\mu, \nu, \varphi) \to \mathbf{r} = a \begin{pmatrix} \sinh(\mu)\sin(\nu)\cos(\varphi) \\ \sinh(\mu)\sin(\nu)\sin(\varphi) \\ \cosh(\mu)\cos(\nu) \end{pmatrix} \tag{E.3}$$

The positive constant $2a$ defines the distance between the focal points $(0, 0, \pm a)$. For simplicity, denote the focal point at $(0, 0, -a)$ with A and the focal point at $(0, 0, +a)$ with B, as depicted in Fig. E.1.

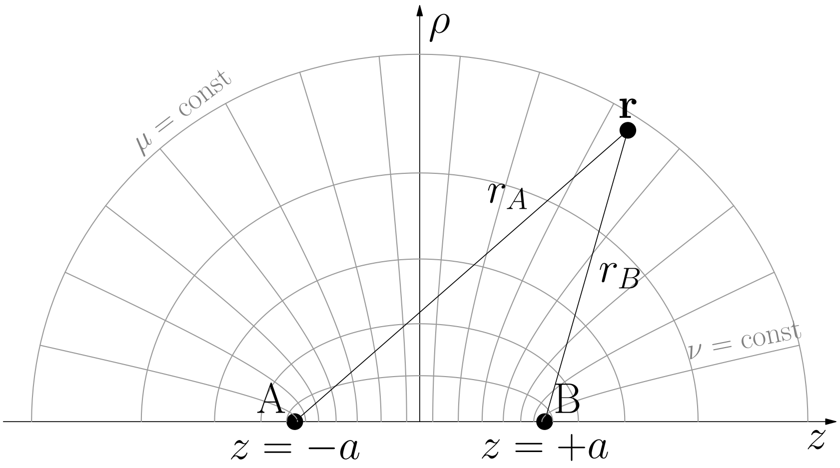


Figure E.1.: A sketch of the prolate spheroidal coordinate system with distances to the foci and isolines for the coordinates native variables μ and ν. By rotation of this plane with the azimuthal angle φ, the whole three-dimensional space is covered.

The distances from any point $\mathbf{r}$ to either focus are given by

$$r_A(x, y, z) = \sqrt{x^2 + y^2 + (z - a)^2} \tag{E.4}$$

$$r_B(x, y, z) = \sqrt{x^2 + y^2 + (z + a)^2}. \tag{E.5}$$

Together with the angle φ, the distances r_A and r_B are a unique representation of the point $\mathbf{r}$ and the reverse transformation reads

$$T^{-1} : \mathbb{R}^3 \to P, (x, y, z) \to \mathbf{q} = \begin{pmatrix} \operatorname{arcosh}\left(\frac{r_A + r_B}{2a}\right) \\ \arccos\left(\frac{r_A - r_B}{2a}\right) \\ \arctan\left(\frac{y}{x}\right) \end{pmatrix}. \tag{E.6}$$

Scale factors are required to evaluate quantities from differential or integral calculus.

$$h_\mu = h_\nu = a\sqrt{\sinh(\mu)^2 + \sin(\nu)^2} \tag{E.7}$$

$$h_\varphi = a|\sinh(\mu)\sin(\nu)| \tag{E.8}$$

With them, integration in Euclidean space is possible in the PS coordinates with the corresponding Jacobi determinant

$$\det J(\mu, \nu, \varphi) = h_\mu h_\nu h_\varphi = a^3 \sinh(\mu)\sin(\nu)\left(\sinh^2(\mu) + \sin^2(\nu)\right). \tag{E.9}$$

For example, the integral over the Euclidean region $T(p)$ with a corresponding rectangular PS region $p = [\mu_0, \mu_1] \times [\nu_0, \nu_1] \times [\varphi_0, \varphi_1] \in P$, the is accessible by by change of coordinates.

$$\int\limits_{T(p)} f(\mathbf{r})\,\mathrm{d}\mathbf{r} = \int\limits_{\mu_0}^{\mu_1}\int\limits_{\nu_0}^{\nu_1}\int\limits_{\varphi_0}^{\varphi_1} f(T(\mathbf{q}))\det J(\mathbf{q})\,\mathrm{d}\varphi\,\mathrm{d}\nu\,\mathrm{d}\mu \tag{E.10}$$

For calculations, including vector calculus, the gradient operator takes the form

$$\begin{aligned}\nabla = \sum_i \frac{\hat{\mathbf{e}}_i}{h_i}\partial_i = \quad & a h_\mu^{-2}\begin{pmatrix}\cosh(\mu)\sin(\nu)\cos(\varphi)\\ \cosh(\mu)\sin(\nu)\sin(\varphi)\\ \sinh(\mu)\cos(\nu)\end{pmatrix}\partial_\mu\\ &+ a h_\nu^{-2}\begin{pmatrix}\sinh(\mu)\cos(\nu)\cos(\varphi)\\ \sinh(\mu)\cos(\nu)\sin(\varphi)\\ -\cosh(\mu)\sin(\nu)\end{pmatrix}\partial_\nu\\ &+ a h_\varphi^{-2}\begin{pmatrix}-\sinh(\mu)\sin(\nu)\sin(\varphi)\\ \sinh(\mu)\sin(\nu)\cos(\varphi)\\ 0\end{pmatrix}\partial_\varphi,\end{aligned} \tag{E.11}$$

from which also the Laplacian

$$\nabla^2 = h_\mu^{-2}\left(\partial_\mu^2 + \partial_\nu^2 + \coth(\mu)\partial_\mu + \cot(\nu)\partial_\nu\right) + h_\varphi^{-2}\partial_\varphi^2 = \nabla_{\mu\nu}^2 + h_\varphi^{-2}\partial_\varphi^2 \tag{E.12}$$

follows. Here $\nabla_{\mu\nu}^2$ denotes the Laplacian acting on the coordinates μ and ν only. This particular form allows the Schrödinger equation to be separated into a two-dimensional μ-ν solution and an angle-dependent part.

E.2. Representation of the orbitals

The solutions of the non-interacting Schrödinger equation from the Hamiltonian of Eq. (E.1) admit orbitals to be of the product form

$$\varphi_m(\mu, \nu, \varphi) = \tilde{\varphi}_m(\mu, \nu)\phi_m(\varphi), \tag{E.13}$$

with a seperated system of equations for $\tilde{\varphi}_m$ and ϕ_m. Both functions have symmetry constraints and boundary conditions. Firstly, the angular equation simplifies to $\phi_m'' = -m^2\phi_m$ and ϕ_m requires 2π-periodicity due to axial symmetry. Additionally, the functions $\{\phi_m\}$ need to span an orthonormal basis.

$$\int_0^{2\pi} \phi_m^*\phi_{m'} = \delta_{mm'} \tag{E.14}$$

The defining properties of ϕ_m are independent of representation. However, some representation must be chosen in practice. A common choice for ϕ_m is the complex representation

$$\phi_m = \frac{e^{im\varphi}}{\sqrt{2\pi}}, \tag{E.15}$$

which is most suited for analytic calculations. However, for the numerical implementation with a real ground state, a real representation, e.g.,

$$\phi_m = \begin{cases}\sqrt{2\pi} & m = 0\\ \frac{1}{\sqrt{\pi}}\cos(m\varphi) & m > 0\\ \frac{1}{\sqrt{\pi}}\sin(m\varphi) & m < 0\end{cases}, \tag{E.16}$$

is mandatory. Some further relations required in the following are

$$-\int_0^{2\pi} |\phi'_m|^2 = \int_0^{2\pi} \phi^*_m \phi''_m = -m^2. \tag{E.17}$$

Inserting the Laplacian term $-m^2$ back into the to equations for $\tilde{\varphi}_m$ and focusing on the μ-ν coordinates, the orbitals obey

$$\left(-\frac{1}{2}\nabla^2_{\mu\nu} + \frac{m^2}{2h^2_\varphi} + v_i(\mu,\nu) + v_e(\mu,\nu)\right)\tilde{\varphi}_m(\mu,\nu) = \epsilon_m \tilde{\varphi}_m(\mu,\nu). \tag{E.18}$$

This PDE is solved numerically with a uniform grid in the μ-ν-plain. The boundary conditions dictate vanishing $\tilde{\varphi}_m$ at infinity, introducing the boundary condition $\tilde{\varphi}_m \to 0$ for $\mu \to \infty$. At the z-axis, given by either $\mu = 0$ or $\nu = 0, \pi$ (exclusive), the orbitals must have a continuous derivative, whereas at the focal positions, $\mu = 0 \wedge \nu = 0$ and $\mu = 0 \wedge \nu = \pi$, they show a cusp in the presence of an ion.

E.3. The prolate spheroidal grid

The PS grid is not only used in the DARSEC code to perform DFT calculations, where it is a mandatory tool to achieve the desired accuracy at low computational cost. It also supports the accumulation of the QMC densities in defining the bins for the density histogram of Sec. 7. A sketch of the grid in Euclidean space is given in Fig. E.2. In the coordinates μ and ν, the grid is regular and uniform and thus not shown explicitly. The uniform grid for the two-dimensional PS space is the direct product of two one-dimensional grids for μ and ν. It is given by only three parameters: the number of grid points for both coordinates, N_μ and N_ν, and the radial extent $\mu_{\max}$. The grid spacing for each coordinate is given by

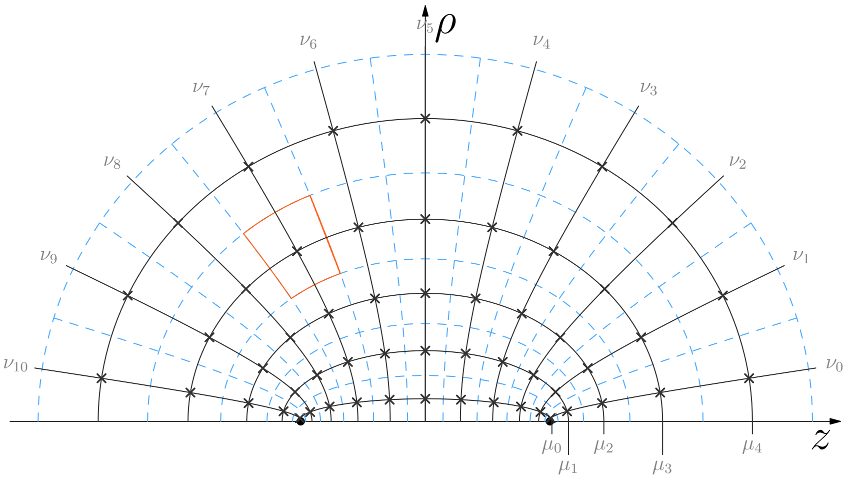


Figure E.2.: A sketch of the PS grid with a resolution of $N_\nu = 11$ and $N_\mu = 5$. The grid points are given by the crosses. The solid lines indicate the isolines running through the grid points, while the dashed lines envelop each grid point's corresponding grid cell. A single grid cell is highlighted for the grid index $(i, j) = (3, 7)$.

$$\Delta_\mu = \frac{\mu_{\max}}{N_\mu}, \tag{E.19}$$

$$\Delta_\nu = \frac{\pi}{N_\nu}. \tag{E.20}$$

Grid points for μ and ν are found at the coordinates

$$\mu(i) = \Delta_\mu \left(i - \frac{1}{2}\right), \qquad \forall i \in \{1, \dots, N_\mu\}, \tag{E.21}$$

$$\nu(j) = \Delta_\nu \left(j - \frac{1}{2}\right), \qquad \forall j \in \{1, \dots, N_\nu\}. \tag{E.22}$$

Thus, the entire set of grid points is given by

$$G = \{(\mu(i), \nu(j)) | i = 1, \dots, N_\mu \wedge j = 1, \dots, N_\nu\}. \tag{E.23}$$

For given coordinates (μ, ν), in the respective bounds, the closest grid point can be calculated by

$$i(\mu) = \left\lfloor \frac{1}{2} + \frac{\mu}{\Delta_\mu} \right\rfloor, \tag{E.24}$$

$$j(\nu) = \left\lfloor \frac{1}{2} + \frac{\nu}{\Delta_\nu} \right\rfloor. \tag{E.25}$$

Note the difference between a grid point $(\mu(i), \nu(j))$ and the corresponding grid cell P_{ij}, which is the set of all coordinates that are closer to the grid point of index (i, j) than to any other grid point. The Euclidean volume of the grid cell P_{ij} is covered by the integral of Eq. (E.10), which in this case reads

$$\begin{aligned} \mathrm{vol}(P_{ij}) = \frac{2\pi a^3}{12} \Big[&- (\cosh(3\mu) - 9\cosh(\mu)) \Big|_{\mu_i - \Delta\mu/2}^{\mu_i + \Delta\mu/2} \cos(\nu) \Big|_{\nu_i - \Delta\nu/2}^{\nu_i + \Delta\nu/2} \\ &+ (\cos(3\nu) - 9\cos(\nu)) \Big|_{\nu_i - \Delta\nu/2}^{\nu_i + \Delta\nu/2} \cosh(\mu) \Big|_{\mu_i - \Delta\mu/2}^{\mu_i + \Delta\mu/2} \Big]. \end{aligned} \tag{E.26}$$

The smallest volumes in the grid are those of indices $(0, 0)$ and $(0, N_\nu)$ with a volume of approximately

$$\mathrm{vol}(P_{00}) = \mathrm{vol}(P_{0N_\nu}) \approx \frac{\pi a^3}{4} \Delta\mu^2 \Delta\nu^2 \left(\Delta\mu^2 + \Delta\nu^2\right). \tag{E.27}$$

Compared to the volume of the grid cells in an Euclidean uniform grid in two dimensions, e.g., in a cylindrical grid for the variables (ρ, z) with cell volume $\Delta\rho\Delta z$, the volume in the PS grid depends on the sixth power of the grid spacings. This scaling certainly allows for much finer grid resolution near the focal points or ion locations.

E.4. Evaluation of orbital derivatives

In a standard DFT calculation, the most relevant derivative is the Laplacian. The orbitals' gradients are rarely calculated explicitly. In a QMC calculation, however, the gradient of the orbitals enters the local energy and the drift term. To utilize finite-difference orbitals in QMC, the orbitals' values, gradients, and Laplacians have to be available at any point in continuous space to calculate the value, gradient, and Laplacian of the corresponding Slater determinant, the key ingredient of the QMC trial wave function. The obvious approach to evaluate orbitals in the continuum is the interpolation of the orbitals from their values at the grid points. Since the PS coordinates allow the separation of the orbitals, the interpolation is only required for the two-dimensional μ-ν-space. The angular solution can be calculated analytically. The interpolation of the orbitals includes an intrinsic but rather small error for orbitals with little variations. The derivatives evaluated in PS coordinates, however, include multiple singular terms in the gradient and Laplacian due to the singularities in the inverse scale factors, $h_\mu^{-2} = h_\nu^{-2}$ and h_φ^{-2}. These singularities are located at the foci, $(\mu = 0) \wedge (\nu = 0, \pi)$, for $h_\mu = h_\nu$ and along the entire the z-axis, $(\mu = 0) \vee (\nu = 0, \pi)$, for h_φ.

Within the DARSEC code, the divergencies cancel out with high accuracy since the numerical treatment is consistent, e.g., the finite-difference approach, orders of expansions, etc. For interpolations that yield a continuous function, however, the DARSEC representation is not suitable since Lagrange interpolations corresponding to the finite-difference approach of DARSEC are not continuous. Thus, in the QMC method, B-Splines are used to interpolate the orbitals' grid values with continuity across grid cells. With the B-Spline representation, a different numerical framework is used, and the divergencies of the gradient and Laplacian become a major obstacle in QMC since the cancellation of singularities happens with far less accuracy. While some cancellations will remain difficult, for example, those of the kinetic energy with a Coulomb singularity near a nucleus, others can be tailored to lower the interpolation error.

A large source of interpolation error is introduced if an orbital is represented by a single B-Spline and its derivatives are calculated from the B-Spline representation. While this seems most consistent, the polynomial B-Spline cannot exhibit the correct singular behavior of the orbital at the z-axis. Therefore, derivatives of the orbital must be calculated separately with a separate B-Spline representation. The following discusses optimal representations for the Laplacian and the gradient.

At the core of the (non-interacting) kinetic energy is the Laplacian of the orbitals. In PS coordinates and applied to the DARSEC orbitals of Eq. (E.13), it reads

$$\nabla^2 \varphi_m = \phi_m (\nabla^2_{\mu\nu} - m^2 h_\varphi^{-2}) \tilde{\varphi}_m \tag{E.28}$$

One might be tempted to separate the terms including $\nabla^2_{\mu\nu}$ and m^2, evaluate each term independently and rearrange them to yield the Laplacian. With such a separation, however, both terms become divergent, and the rearrangement cannot fully cancel the singularities to arrive at regular behavior. Thus, the Laplacian is best approximated if a distinct B-Spline is constructed for the full eq. (E.28).

To evaluate the gradient, the form

$$\nabla \varphi_m = \phi_m \begin{pmatrix} g_r \cos(\varphi) \\ g_r \sin(\varphi) \\ g_z \end{pmatrix} + g_t \phi'_m \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \\ 0 \end{pmatrix} \tag{E.29}$$

has proven to have fewer numerical errors at the singular sites, with the constituents

$$g_r(\mu, \nu) = a h_\mu^{-2} \left(\cosh(\mu) \sin(\nu) \partial_\mu \tilde{\varphi}_m + \sinh(\mu) \cos(\nu) \partial_\nu \tilde{\varphi}_m\right) \tag{E.30}$$

$$g_z(\mu, \nu) = a h_\mu^{-2} \left(\sinh(\mu) \cos(\nu) \partial_\mu \tilde{\varphi}_m - \cosh(\mu) \sin(\nu) \partial_\nu \tilde{\varphi}_m\right) \tag{E.31}$$

$$g_t(\mu, \nu) = a h_\varphi^{-2} \tilde{\varphi}_m \sinh(\mu) \sin(\nu) = h_\varphi^{-1} \tilde{\varphi}_m. \tag{E.32}$$

With B-Splines for the expressions g_r, g_z and g_t, no divergent terms are present in the formula for the gradient besides the intrinsic singularities present "inside" g_r, g_z and g_t.

A simple quality check on the orbitals' derivatives can be carried out with the aid of Green's first identity.

$$\int \varphi_m^* \nabla^2 \varphi_m = -\int |\nabla \varphi_m|^2 \tag{E.33}$$

Written out in terms of the above-defined expressions, this yields

$$\begin{aligned}\int \varphi_m^* \nabla^2 \varphi_m &= \int |\phi_m|^2 \tilde{\varphi}_m^* \left(\nabla_{\mu\nu}^2 \tilde{\varphi}_m - m^2 h_\varphi^{-2} \tilde{\varphi}_m\right) \\ &= \int \tilde{\varphi}_m^* \left(\nabla_{\mu\nu}^2 \tilde{\varphi}_m - m^2 h_\varphi^{-2} \tilde{\varphi}_m\right) \\ &= \int \tilde{\varphi}_m^* \nabla_{\mu\nu}^2 \tilde{\varphi}_m - m^2 \int |g_t|^2\end{aligned} \tag{E.34}$$

for the Laplacian and

$$\begin{aligned}-\int |\nabla \varphi_m|^2 &= -\int |\phi_m|^2 (g_r^2 + g_z^2) + |\phi_m'|^2 g_t^2 \\ &= -\int \left(|g_r|^2 + |g_z|^2\right) - m^2 \int |g_t|^2\end{aligned} \tag{E.35}$$

for the gradient. The last equation holds because the gradient norm separates since its constituents are orthogonal. These identities can be evaluated numerically to assess the quality of the orbitals' numerical representation.

F. Influence of the orbitals' grid on QMC

> The following section is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

Here we aim to analyze the dependence of the QMC results on the grid and converge them with respect to our grid parameters, as the quality and consistency of the orbitals has a great impact on QMC – as already pointed out in Ref. [86] in the context of basis-set represented orbitals. Within our grid there are three parameters to adjust: The number of grid points along the two PS axes, N_ν and N_μ, and the total extent of the grid by $R_{\max}$, or equivalently $\mu_{\max}$. We chose $R_{\max}$ to yield densities of the order 10^{-8} to 10^{-10} at the border of the simulation box. This might seem unnecessary, since densities of this magnitude do not contribute to any physical effects, but we deliberately sampled this regions to test our numerical methods. To determine the required accuracy for the number of grid points we examined the grid of $N_\nu \times N_\mu$ in the range of 55×55, 79×79, 103×103, 121×121, 151×151, 181×181 and 211×211 for Li_2 based on LDA orbitals. For each of those grids we calculated the DFT, VMC and DMC energies. The results are shown in Fig. F.1. While a grid of around 103×103 cells suffices to converge the DFT energy, for VMC and DMC the energies converge only at around 151×151 and beyond. Note that our notion of convergence stems from the behaviour of the graphs shown in Fig. F.1, while convergence within an error range of chemical accuracy (0.0016 a.u.) would be reached with a less strict criterion. The variance of the DMC ground state reaches a plateau at this grid parameters, reflecting the improvement in grid accuracy. While the DFT energies are decreasing with increasing grid accuracy, the QMC energies have lower energies for insufficient grid resolution. This is particularly important to note, since an unconverged grid would yield "variationally better" results.

Considering this convergence analysis, we performed all QMC calculation with a grid of 181×181 grid cells to yield accurate energies. Note that densities would already be accurate with a smaller grid as long as one is not interested in the direct vicinity of the nuclei.

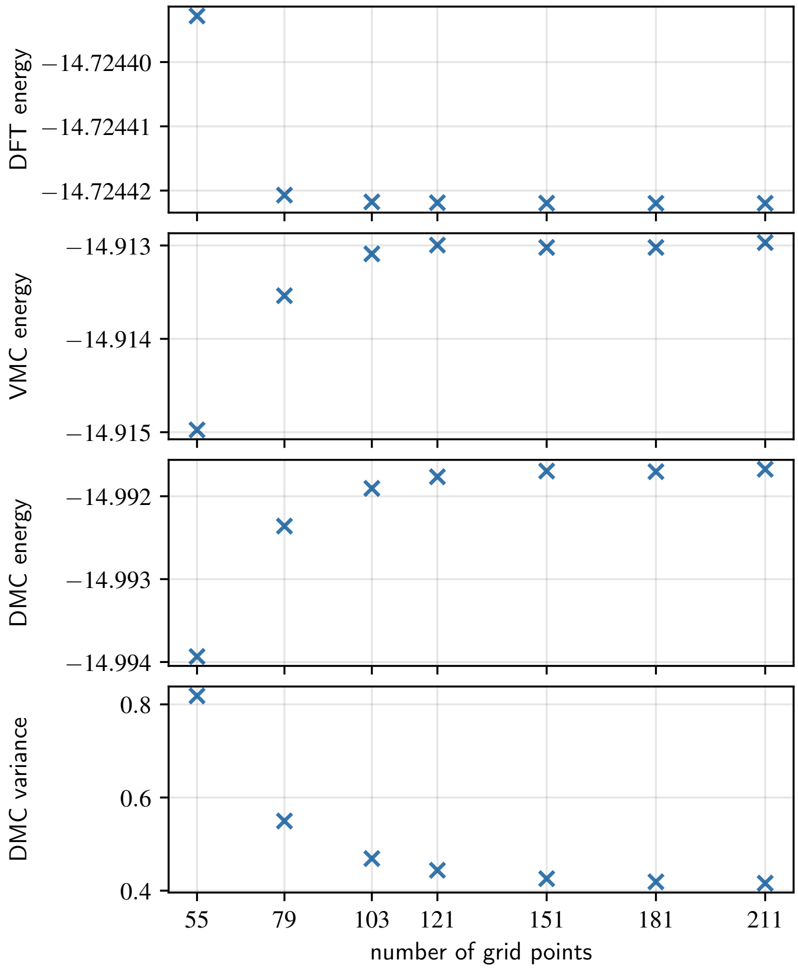


Figure F.1.: Effects of the grid resolution on energies from DFT (LDA), VMC and DMC as well as the ground state variance for DMC in Li_2 based on LDA orbitals in Hartree units. VMC and DMC energies converge at about 151 × 151 grid cells, while the LDA energies converge with less grid points at about 103 × 103 cells. The DMC variance is strictly decreasing with more accurate grids. This indicates that the QMC convergence is determined by the accuracy of the conversion of the derivatives from discrete orbitals to the continuous representation in the QMC code.

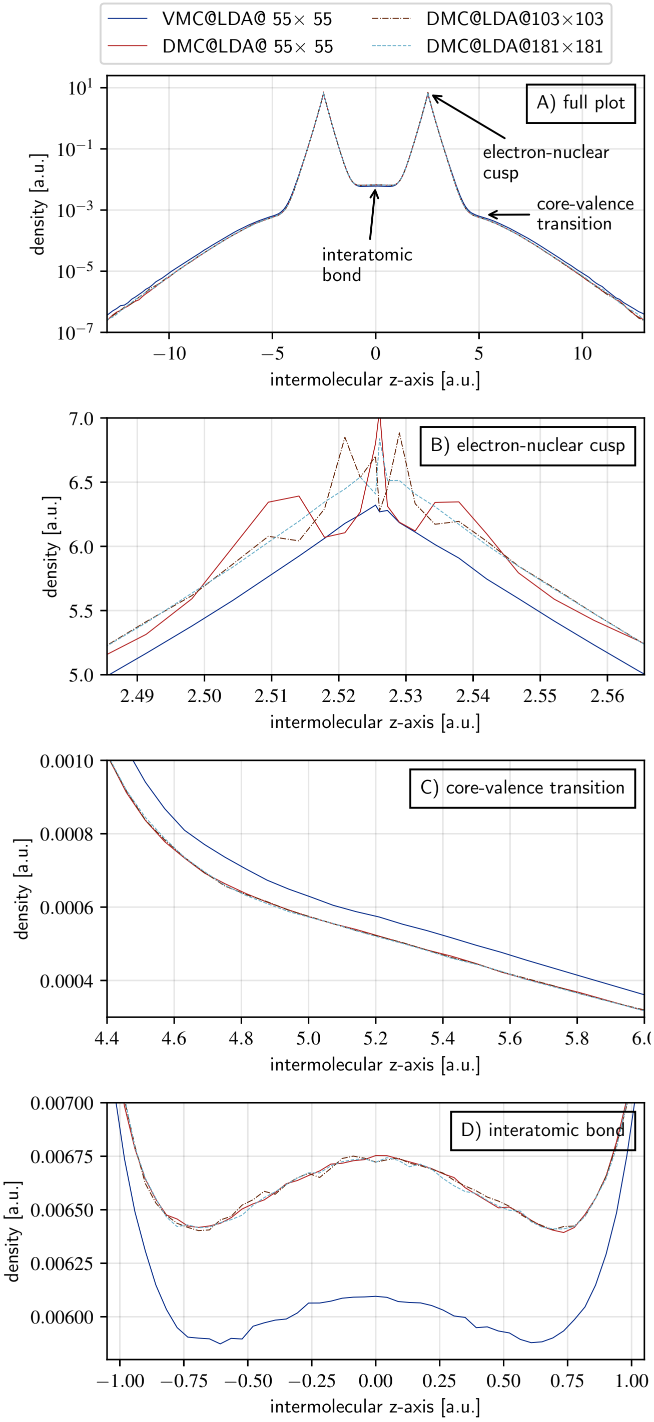


Figure F.2.: Densities for the grid parameters 55 × 55, 103 × 103 and 181 × 181 for DMC and 55 × 55 for VMC. The density along the z-axis is plotted in panel A. Panels B, C and D zoom in on the nuclear cusp, the region of the core-valence transition, and the bond region, respectively. The DMC densities align perfectly anywhere except the nuclei. At the focal points numerical artifacts govern the density in DMC, increasing in magnitude with coarser grids. To the contrary the VMC density does not show numerical issues beyond statistical noise, even for the rather low grid resolution of 55 × 55.

G. Regularization of noisy densities

> The following section is an excerpt from the author's publication [45] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

Naturally, the density has large statistical errors in the asymptotic regions. In our approach the core regions in addition show large fluctuations due to the high resolution of the grid at its focal points and the decreasing grid cell volume. As can be seen from Eq. (7.5), the fluctuations of the density estimation in a bin depend on the bin's volume, leading to infinite variance in the limit of vanishing bin size.

These fluctuations, either due to the lack of density or small bins, occur on a scale that does not influence the quality of the over-all density: the low-density regime beyond density values of 10^{-8} in the asymptotic region has little physical meaning, while the volume around the focal points with large statistical noise is of negligible extent. To nevertheless obtain a smooth density on the entire domain, we invoke an extrapolation procedure that replaces the density in vicinity of the PS foci (e.g., the nuclei in the case of Li_2) with an extrapolation from the close-by grid points. This procedure is also applied in the asymptotic domain and explained in detail in the following.

First the whole domain Ω of the density is split in four sectors: Two sectors around each of the two focal points, one sector for the asymptotic domain, and the remainder. This is visualized in Fig. G.1 for the Li_2 molecule. To define these domains, we introduce distance measures

$$\mathrm{dist}_{\mathrm{A}}(\mathbf{r}) = \min\left(|\mathbf{r} + a\hat{\mathbf{e}}_z|, |\mathbf{r} - a\hat{\mathbf{e}}_z|\right), \tag{G.1}$$

$$\mathrm{dist}_{\pm\mathrm{F}}(\mathbf{r}) = |\mathbf{r} \mp a\hat{\mathbf{e}}_z|. \tag{G.2}$$

The asymptotic domain is then given by $\mathrm{dist}_{\mathrm{A}}(\mathbf{r}) > r_{\mathrm{A}}^{\mathrm{min}}$, whereas the focal sites are identified via $\mathrm{dist}_{\pm\mathrm{F}}(\mathbf{r}) < r_{\mathrm{F}}^{\mathrm{max}}$. The parameters $r_{\mathrm{F}}^{\mathrm{max}}$, $r_{\mathrm{A}}^{\mathrm{min}}$ are chosen such that the noisy areas are confined within the domains, but in addition also a reasonable amount of trustworthy data are contained within the domains.

Next, we define two subdomains in the focal and asymptotic domains: The first subdomain Ω^{src} is chosen such that its density values are trustworthy, i.e. have low noise. This source-subdomain's data is then used to optimized the parameters of a fitting function by least squares. The second subdomain, Ω^{fit}, selects the density that is dominated by noise. Its density is replaced completely by data from the fit. Thus we extrapolate the local behaviour of the density in the source region to the fit region. The resulting six sub-domains are

$$\Omega_A^{\mathrm{src}} = \left\{\mathbf{r} \in \Omega | r_{\mathrm{A}}^{\mathrm{min}} < \mathrm{dist}_{\mathrm{A}}(\mathbf{r}) < r_{\mathrm{A}}^{\mathrm{max}}\right\} \tag{G.3}$$

$$\Omega_A^{\mathrm{fit}} = \left\{\mathbf{r} \in \Omega | r_{\mathrm{A}}^{\mathrm{max}} < \mathrm{dist}_{\mathrm{A}}(\mathbf{r})\right\} \tag{G.4}$$

$$\Omega_{\pm\mathrm{F}}^{\mathrm{src}} = \left\{\mathbf{r} \in \Omega | r_{\mathrm{F}}^{\mathrm{min}} < \mathrm{dist}_{\pm\mathrm{F}}(\mathbf{r}) < r_{\mathrm{F}}^{\mathrm{max}}\right\} \tag{G.5}$$

$$\Omega_{\pm\mathrm{F}}^{\mathrm{fit}} = \left\{\mathbf{r} \in \Omega | \mathrm{dist}_{\pm\mathrm{F}}(\mathbf{r}) < r_{\mathrm{F}}^{\mathrm{min}}\right\}, \tag{G.6}$$

while the remainder is

$$\Omega^{\mathrm{D}} = \Omega \backslash \cup \{\Omega_A^{\mathrm{fit}}, \Omega_A^{\mathrm{src}}, \Omega_{+\mathrm{F}}^{\mathrm{fit}}, \Omega_{-\mathrm{F}}^{\mathrm{fit}}, \Omega_{+\mathrm{F}}^{\mathrm{src}}, \Omega_{-\mathrm{F}}^{\mathrm{src}}\}. \tag{G.7}$$

Around both foci and the asymptotic we now have isolated a trust region with valid density values and the regions with notorious fluctuations that we want to replace. Note that the density in the remainder domain is not altered by our procedure. Thus, the scale on which we manipulate the over-all density is tiny and the remainder domain covers almost entirely the domain of physical importance. The number of grid points subject to refinement shown in Fig. G.1 is only a small fraction of the total number of grid points and is visually over-represented in the Euclidean space.

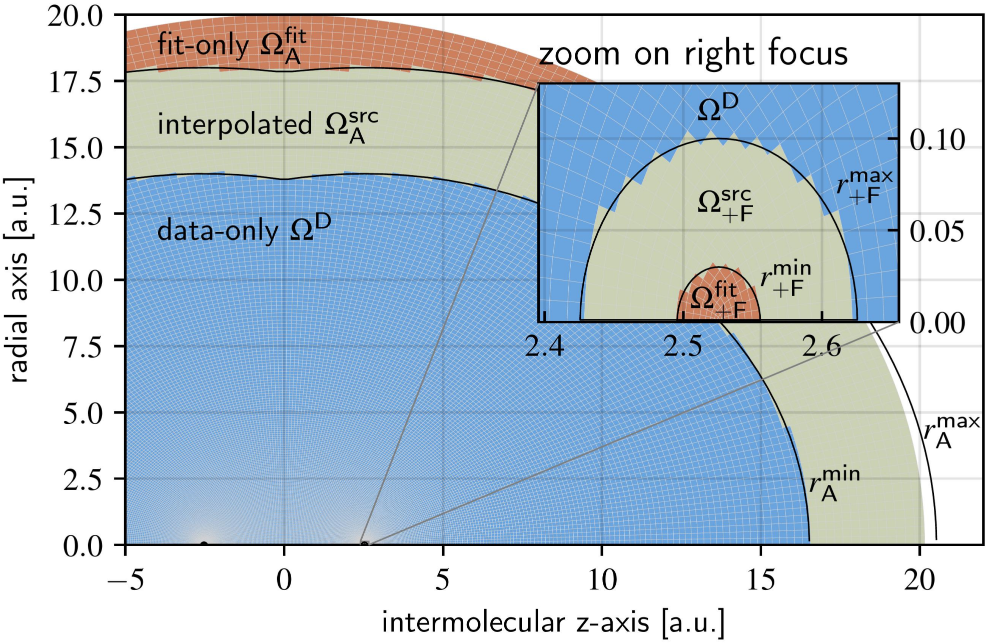


Figure G.1.: Partitioning of the density's support in the refinement procedure, here shown for the Li_2 molecule. In the (beige) source sub-domains an appropriate extrapolating fit is optimized. The fit is applied to the fit-only sub-domains (orange) replacing their original density values. A smooth transition between data and fit is used in the beige domains to avoid discontinuities at the boundaries to the data-only domain (blue). The QMC density is not manipulated in the data-only domain, which covers almost all of the physically relevant space.

To restore the density in the subdomains of high statistical fluctuations, an appropriate fit function has to be chosen. We use a cusp-form fit (including the often spherically averaged dipole contribution) for the focal points occupied with ions, and an exponential polynomial form for vacant focal points (viz. in the carbon atom),

$$n^{\text{cusp}} = n_0 \exp\left(-(\gamma + p\cos\vartheta)r\right), \tag{G.8}$$

$$n^{\text{smooth}} = n_0 \exp\left(-\gamma r - \sigma^2 r^2 - \kappa r^3\right), \tag{G.9}$$

where r and ϑ are the polar coordinates centered at the closest ion (in the case of a vacant focus, the closest ion is the one on the other focus). For the asymptotic decay there are multiple reasonable fits available. The simplest is a s-type exponential decay, but to better represent the density in the near-asymptotic regions, any LCAO-type fit can be used. For Li_2 and N_2 the decay is straightforward to capture, e.g., using s-orbitals centered at each atom of the molecule. The carbon atom shows an-isotropic exponential decay, which is best represented with s-type orbital for the non-HOMO part of the density and an additional $2p_{\pm 1}$-type orbital for the HOMO-density. The fits are optimized with data from the trustworthy source domain.

Combining the bare QMC density values in the source region with the density values from the fit in the neighboring fit-only region would introduce a small discontinuity at the boundary, especially if the fit or the density have poor quality, or if the spatial boundaries are inappropriately chosen. Therefore, we mix the bare QMC density with the fitted density by introducing weights that smoothly interpolate both densities in the source sub-domain, $\mathbf{r} \in \Omega^{\mathrm{src}}$,

$$n^{\mathrm{refined}}(\mathbf{r}) = n^{\mathrm{fit}}(\mathbf{r}) \times W(\mathbf{r}) + n^{\mathrm{QMC}}(\mathbf{r}) \times (1 - W(\mathbf{r})), \tag{G.10}$$

where the weights are

$$W(\mathbf{r}) = I\left(\frac{\mathrm{dist}_{\pm\mathrm{F,A}}(\mathbf{r}) - r^{\mathrm{min}}_{\pm\mathrm{F,A}}}{r^{\mathrm{max}}_{\pm\mathrm{F,A}} - r^{\mathrm{min}}_{\pm\mathrm{F,A}}}\right). \tag{G.11}$$

I is a smooth transitioning function with $I(0) = I'(0) = I'(1) = 0$ and $I(1) = 1$. Here, we chose $I(x) = (3 - 2x)x^2$, but other choices are equally justified.

This concludes the refinement procedure. The results can be inspected in Fig. G.2. The four spatial boundary parameters $r^{\mathrm{min}}_{\mathrm{F}}, r^{\mathrm{max}}_{\mathrm{F}}, r^{\mathrm{min}}_{\mathrm{A}}$ and $r^{\mathrm{max}}_{\mathrm{A}}$ are adapted for every system to reflect the densities' behavior. While this procedure might seem to introduce a significant dependence on the user-defined parameters and fit functions, varying the parameters in a reasonable range does only marginally influence the resulting refined density. This is a direct consequence of the electron-nuclear cusp condition and the exponential asymptotic decay of the density, governing the behavior of the density in the domains we extrapolate in. For the refinement of the carbon atom shown in Fig. G.2 and G.3 we use $r^{\mathrm{min}}_{+\mathrm{F}} = 0.03$ a.u. and $r^{\mathrm{max}}_{+\mathrm{F}} = 0.10$ a.u. for the focus with the C atom and $r^{\mathrm{min}}_{-\mathrm{F}} = 0.06$ a.u. and $r^{\mathrm{max}}_{-\mathrm{F}} = 0.20$ a.u. for the vacant focus, while $r^{\mathrm{min}}_{\mathrm{A}} = 5$ a.u. and $r^{\mathrm{max}}_{\mathrm{A}} = 7$ a.u. in the asymptotic region. The fit at the carbon atom is Eq. (G.8), while the fit at the unoccupied focus is (G.9). The fit in the asymptotic domain could be chosen from a broader variety of fitting functions and spatial parameters, but we restrain to the LCAO fit of type $1s2p(\pm 1)$ with variable coefficients and decay rates.

The resulting density meets high numerical standards in terms of smoothness with little remaining statistical noise and conserves the original features of the density.

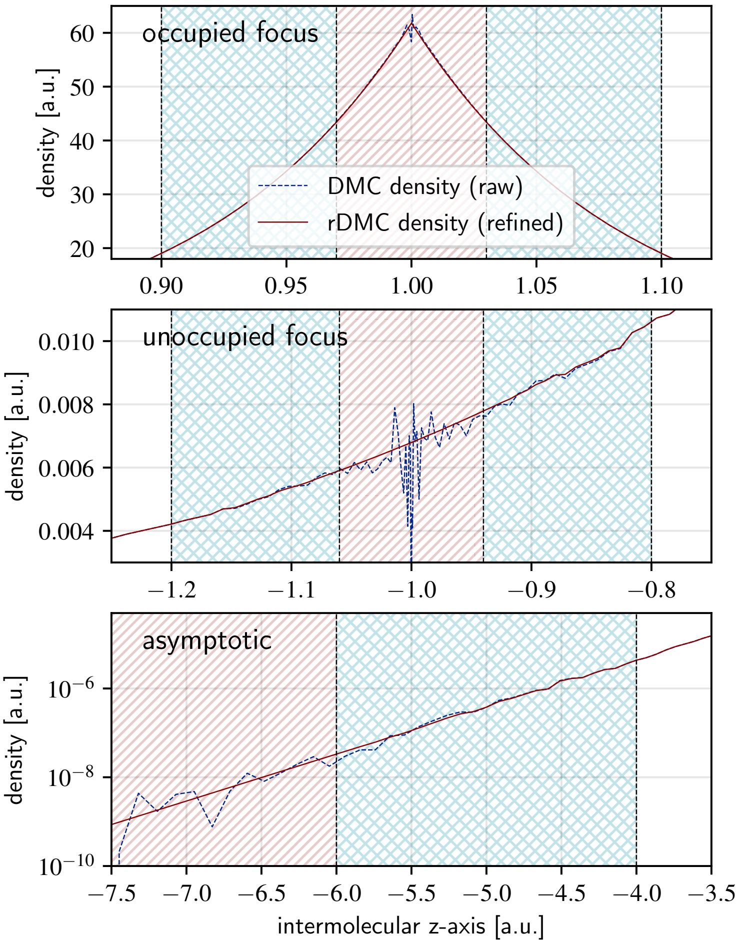


Figure G.2.: The refinement improves the density in those domains where QMC shows large fluctuations. At the PS focus occupied by the carbon atom at $z = 1$, the density is higher (thus has less noise) compared to the unoccupied focus at $z = -1$, where larger fluctuations arise due to the lack of walkers visiting the corresponding bins around the empty focus. Fit-only subdomains are textured with a light-red diagonal pattern, while source subdomains use a light-blue crisscross texture.

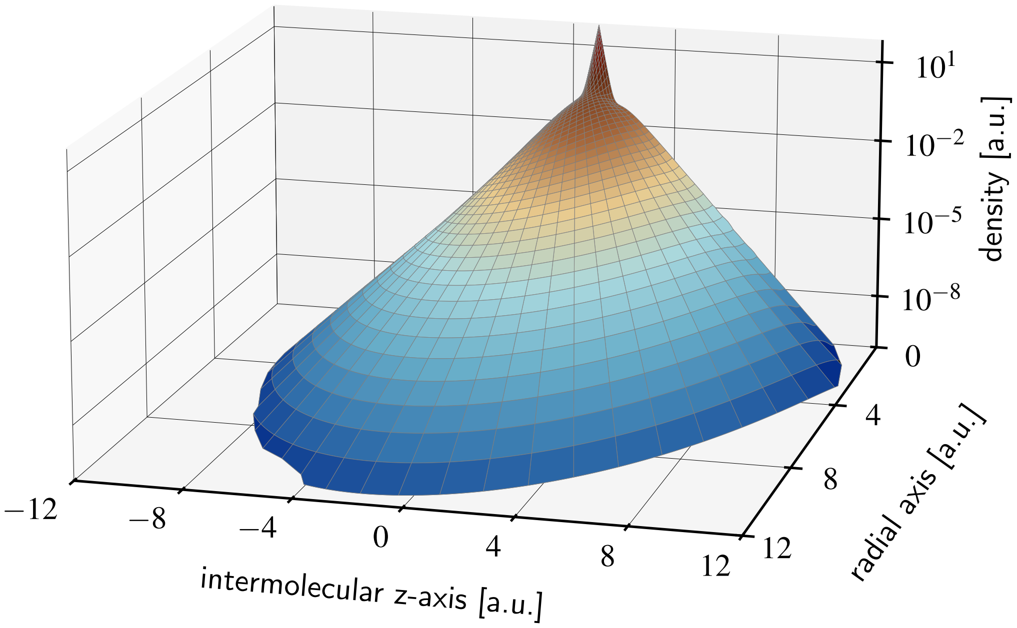


Figure G.3.: With the regularization of the refinement procedure the statistical noise can be eliminated. The $1s2p(\pm1)$ fit in the asymptotic domain decays with different decay rates depending on the direction, thus including the nodal line of the carbon atom, extrapolating the density to values beyond 10^{-11}.

H. Preparation of densities for inversion

> The following section is an excerpt from the author's publication [116] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

We use the QMC densities from our recent article Ref. [45]. These highly accurate ab initio densities were calculated with the FNDMC method and thus are subject to statistical fluctuations and some biases, discussed in Sec. 7.2. The FNDMC method requires a set of single particle orbitals, for which we chose the LDA and xKLI orbitals. These should not be confused with the DFT reference densities or initial xc potentials of this work. In Sec. G we further devised a regularization method to remedy most of the statistical fluctuations close to any nuclei as well as in the asymptotic domain of the system. While it is possible to eliminate the fluctuations in the density's asymptotic domain, here we restricted ourselves to regularization of the cusp. Regularization in the asymptotic domain did only marginally influence our inversions and we want to avoid any superfluous regularization. To arrive at a very high accuracy in the QMC calculation we calculated the QMC densities on a 181×181 grid in prolate spheroidal coordinates. To perform an inversion with a grid of such an extent is tedious and unnecessary for the accuracy of the inversion. Therefore we reduce the grid to size 103×103 to accelerate the inversion. The PS grid does not allow to simply merge neighboring bins, and we explain our reduction procedure in the following:
First we interpolate the density on the fine grid. Next, we split every grid cell in the coarse grid into a small subgrid of size $m \times m$ and evaluate the interpolation at those subgrid points to represent the interpolated density from the fine grid. The value of the density in the coarse grid is the integral of the values on the subgrid. We found $m = \lceil N^{\text{fine}}/N^{\text{coarse}} \rceil = 2$ to be sufficient for all our cases. The densities resulting from the reduction process are almost indistinguishable from the densities on the finer grid. Since our systems N_2 and Li_2 are symmetric with respect to the z-axis, we could further symmetrize the density to lower the magnitude of numerical artifacts, but we deliberately allowed the asymmetries to investigate their consequences in the resulting inverted xc potentials. The figures H.1, H.2 and H.3 show the reference densities we used.

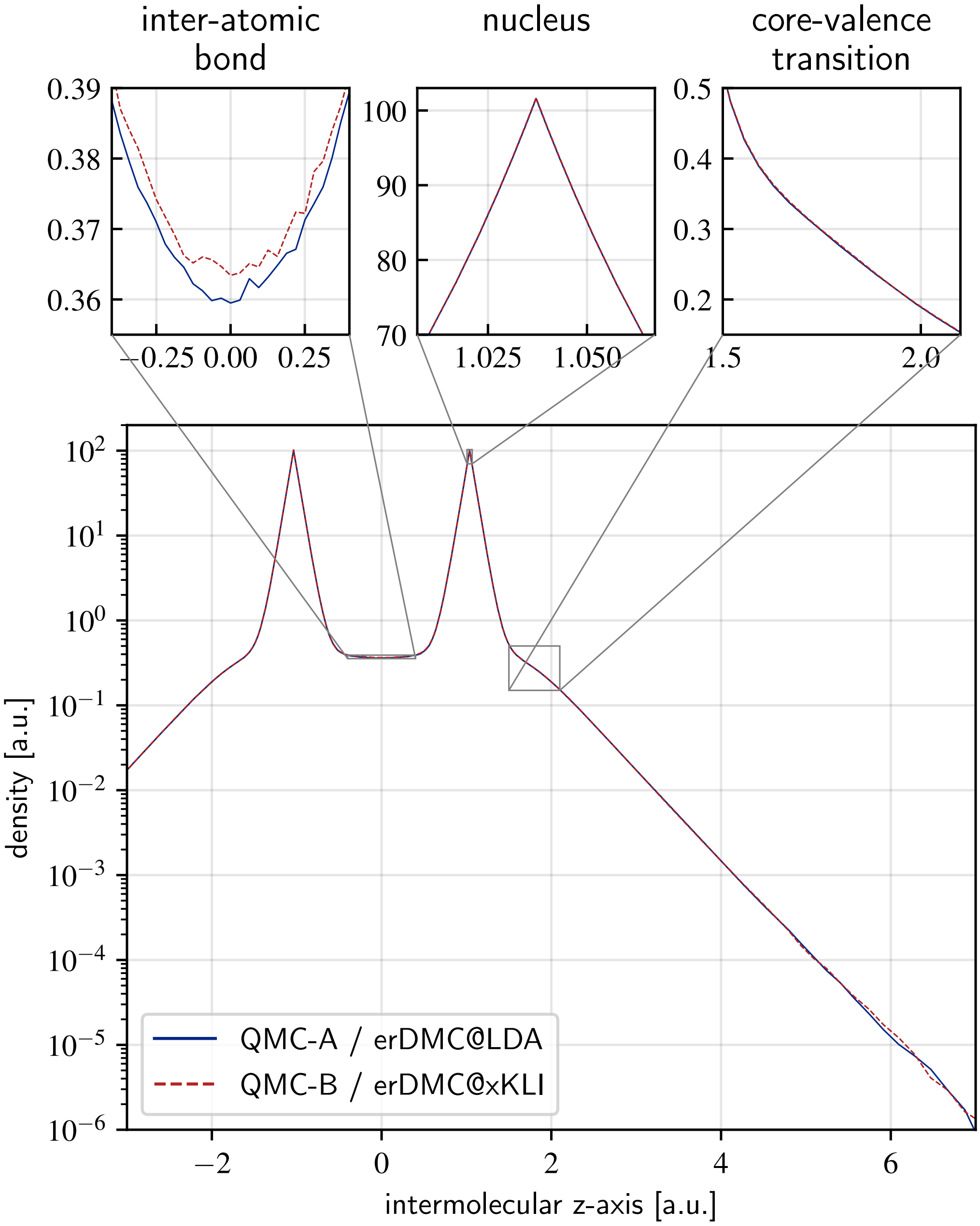


Figure H.1.: Reference densities for the nitrogen molecule N_2 used for the inversion. The DMC densities are extrapolated to reduce the mixed estimator bias and are regularized at the cusp. The QMC calculations are based on LDA and xKLI orbitals.

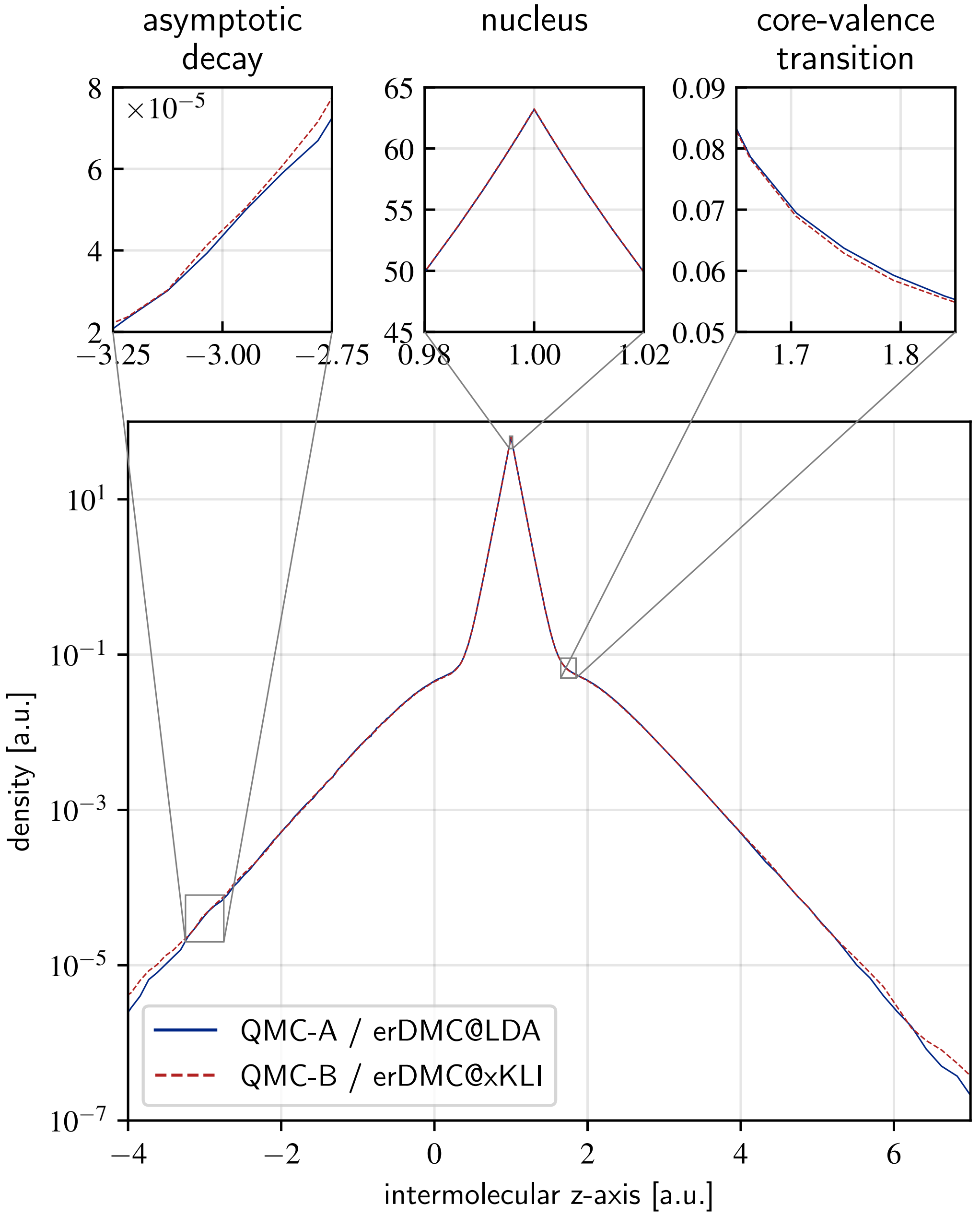


Figure H.2.: Reference densities for the carbon atom C used for the inversion. The DMC densities are extrapolated to reduce the mixed estimator bias and are regularized at the cusp. The QMC calculations are based on LDA and xKLI orbitals.

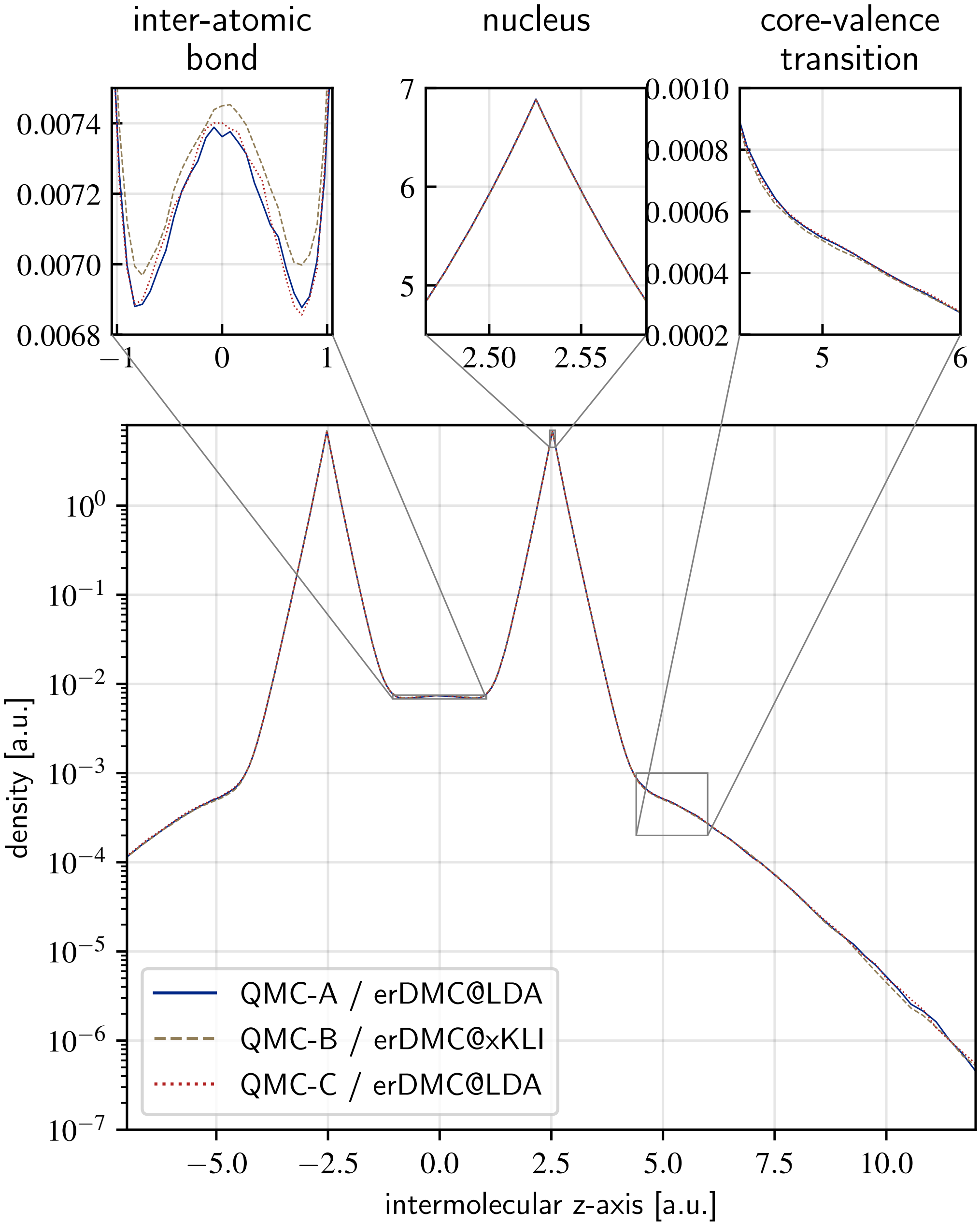


Figure H.3.: Reference densities for the lithium dimer Li_2 used for the inversion. The DMC densities are extrapolated to reduce the mixed estimator bias and are regularized at the cusp. The QMC calculations are based on LDA and xKLI orbitals.

I. Iterative inversion scheme

The iterative process of updating the potential with feedback from the density can suffer from instabilities inherent to feedback loops. A common strategy to avoid instabilities is the adjustment of the feedback strength ϵ. Ideally ϵ is lowered before instabilities arise and increased if the feedback has too little effects on the density. While the latter does affect performance, the former decides whether the inversion is successful or fails. A common situation, in which the update cycles could diverge, is an excessive feedback strength. A mismatch in the density is then mapped to a potential update that overshoots the mismatch, leading to a new mismatch of the densities but with changed sign. Thus oscillations arise with potentially increasing magnitude. Such oscillations often start at the nuclei, where the density takes large values. Additionally, changes in the density anywhere in space affect the density non-locally due to normalization, contributing to an initial large mismatch at the nucleus that could trigger divergence by oscillation. The oscillatory feedback is not bound to the nuclei and possibly propagates into a larger environment. Update rules like DEN, that do not mediate the wide range of density values to the range of reasonable potential values, heavily suffer from such oscillations. If oscillations are detected early, the feedback strength can be lowered. This prevents a divergence but also lowers the feedback at other locations, and the updates might become effectively unable to apply the required changes to reconstruct the density. Therefore it is vital to use update rules that map density mismatches to reasonable potential updates.

Another question is how to detect feedback oscillations early. While the spatially resolved auto-correlation of the density, $A(\mathbf{r}, m) = \mathrm{Corr}[n^{(i+m)}(\mathbf{r}), n^{(i)}(\mathbf{r})]$, gives a hint to local oscillations, it is memory-inefficient to calculate and tedious to analyze. While one could reduce the spatial dimensionality of $A(\mathbf{r}, m)$ to for example only monitor the auto-correlation of the density at the nuclear position, we found it sufficient to lower the feedback strength by a constant factor (usually close to but lower than one) after some fixed number of iterations. We choose to reduce the feedback to 90% every 30 iteration, $\epsilon \to 0.9 \times \epsilon$, which works with our choice for the initial feedback strength.

For some inversions of the QMC-B reference density for the nitrogen molecule based on an initial xKLI potential, we plot the metric D_1 with a log-log scale in Fig. I.1. First of all, the inversion greatly improves the density only in the first hundreds of iterations, while subsequent updates typically yield only small improvements to the density. The graphs do not reach a stable plateau in any inversion we considered, and therefore further updates provide further improvements to the density. Despite small incremental improvements of the density the required iterations become exponentially larger, e.g., the first 10^3 iterations improve the inverted density from $D_1 \approx 0.4$ to $D_1 \approx 3 \times 10^{-3}$, but the last 10^3 only further to not even $D_1 \approx 2 \times 10^{-3}$ for HAR+LoH(1), such that the inversion is beyond practical improvement at some iteration.

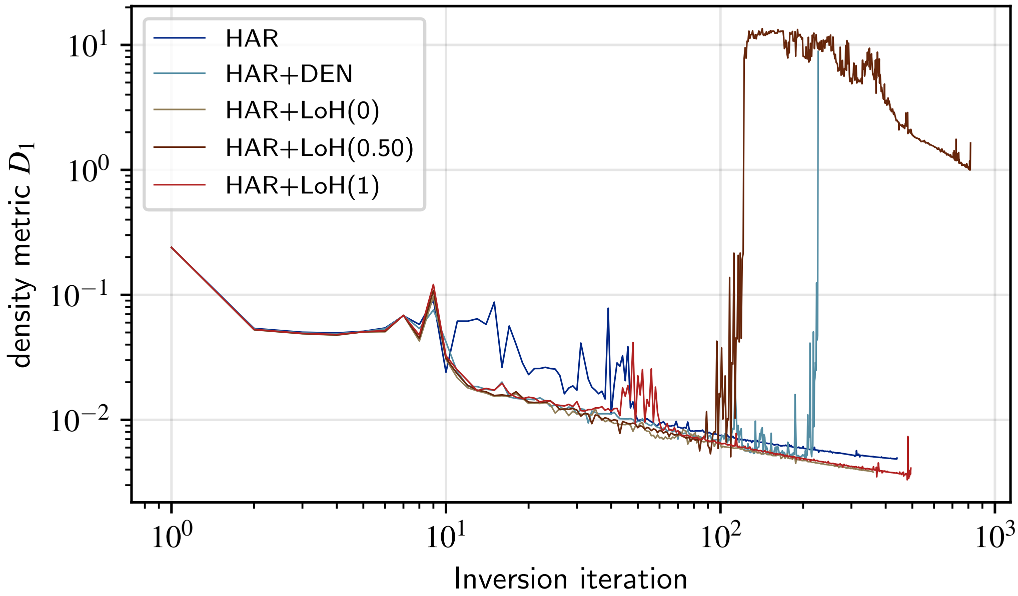


Figure I.1.: Evolution of the inversion process for the nitrogen molecule (with the reference density QMC-B and xKLI initial potential) in a log-log-plot. The iteratively updated xc potentials do not produce monotonically improving densities, but improve logarithmically with rather large fluctuations and outliers. Despite the noise the converging metrics follow a distinct average graph, even in the case of intermediate to large deviations as present in the evolution of the rule HAR+LoH(1) for at about 50 iterations. The inversion HAR+DEN failed at about 200 iterations, and its optimal inverted potential is determined from the preceding iterations. The rule HAR+LoH(0.5) almost diverged at about the 100th iteration, but recovered subsequently. Inversions of this type most often show extremely large artifacts, e.g., wiggles in the asymptotic domain, broken symmetries, etc., and cannot be expected to yield physical insights, even if the metrics converge after a divergence-like event in the iteration.

J. The forward KS-map in detail

> The following section is an excerpt from the author's publication [116] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

This section explains the method and presents the results of the analysis of the forward KS map summarized in Sec. 10.1. To start with the variation we apply to the xc potential, we utilize a linear mixture, i.e. an interpolation or extrapolation, of two xc potentials $v_{\mathrm{xc}}^{(0)}$ and $v_{\mathrm{xc}}^{(1)}$

$$v_{\mathrm{xc}}^{(\lambda)} = (1-\lambda)v_{\mathrm{xc}}^{(0)} + \lambda v_{\mathrm{xc}}^{(1)} \tag{J.1}$$

Alternatively we perturbe $v_{\mathrm{xc}}^{(0)}$ by $v_{\mathrm{xc}}^{(1)} - v_{\mathrm{xc}}^{(0)}$. Note that mixing potentials differs from mixing functionals. Mixing two functionals results in a new xc functional, which has some self-consistent xc potential and valid xc and total energies. Mixing two xc potentials and self-consistently obtaining the corresponding density (which essentially also converges the a priori unknown Hartree potential), however, results in a KS potential, that does not come with a corresponding energy functional and a corresponding energy functional might not even exist. As already mentioned in Sec. 10.1, the linear variation in the potential results in a linear response in the eigenvalues and total energies due to the Hellmann-Feynman theorem. On the other had, the density is assumed to respond in an intricate way. Approximations to the KS potential are designed to map to the exact interacting wave function and therefore the potential and the density aim to compress the full information of the exact wave function. The complexity of the compression originates from the complexity of the Hohenberg-Kohn map and even small variations of the KS potential should therefore manifest in the density in a subtle way. The formal connection of an infinitesimal variation of the xc potential and to its infinitesimal density response is given by

$$\frac{\mathrm{d}n_\sigma(\mathbf{r})}{\mathrm{d}\lambda} = \sum_{i\neq a} \frac{\varphi_{\sigma i}^*(\mathbf{r})\varphi_{\sigma a}(\mathbf{r})}{\epsilon_{\sigma i} - \epsilon_{\sigma a}} \left\langle \varphi_{\sigma a} \middle| v_{\mathrm{xc}}^{(1)} - v_{\mathrm{xc}}^{(0)} \middle| \varphi_{\sigma i} \right\rangle . \tag{J.2}$$

and is thus only valid for vanishing magnitudes for λ. The Taylor expansion of the density inherits this constraint, being valid only for small $|\lambda - \lambda_0|$, intuitively.

$$n^{(\lambda)} = n^{(\lambda_0)} + \left.\frac{\mathrm{d}n_\sigma}{\mathrm{d}\lambda}\right|_{\lambda_0} (\lambda - \lambda_0) \tag{J.3}$$

Assume the density to only respond in the linear regime for the moment. From the densities at $\lambda = 0$ and $\lambda = 1$ the linear map is fully determined. The linear prediction reads

$$n_{\mathrm{lin}}^{(\lambda)} = (1-\lambda)n^{(0)} + \lambda n^{(1)}. \tag{J.4}$$

In reality the density cannot respond purely linear. The values of λ for which the density responds linearly might be very limited to the reference potentials at $\lambda = 0$ and $\lambda = 1$ and higher order terms could become important even with small deviations in λ. We will clarify the questions of how the density response in first order and to which extend the linear regime is valid in the following.

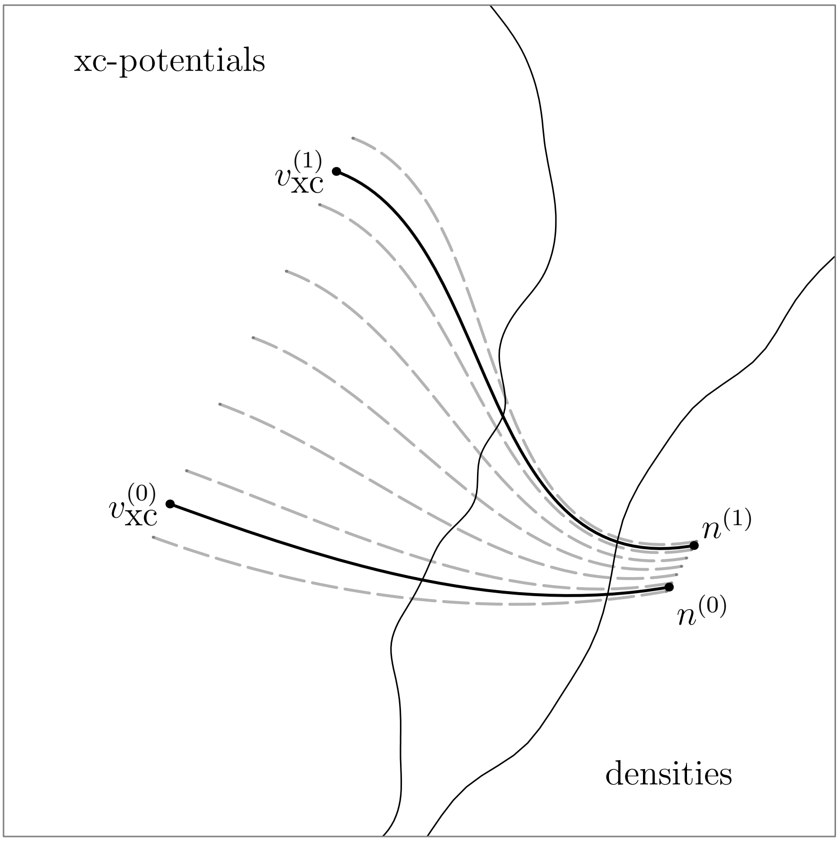


Figure J.1.: Sketch of the interpolation or extrapolation between two xc potentials $v_{\mathrm{xc}}^{(0)}$ and $v_{\mathrm{xc}}^{(1)}$, for which the corresponding densities, $n^{(0)}$ and $n^{(1)}$, are known. Since the forward KS map F transports the full information of the ground state, it could return rather unpredictable and artificial densities for the interpolation and especially extrapolation. However, we show evidence for a regular and even strongly linear response in the density. While our sketch suggestively shows a linear response, this cannot be inferred to the KS map a priori.

To start with, we chose $v_{\mathrm{xc}}^{(0)} = v_{\mathrm{xc}}^{\mathrm{xKLI}}$ and $v_{\mathrm{xc}}^{(1)} = v_{\mathrm{xc}}^{\mathrm{LDA}}$ in the Perdew-Wang parametrization [47] with their corresponding densities $n^{(0)} = n^{\mathrm{xKLI}}$ and $n^{(1)} = n^{\mathrm{LDA}}$ for diatomic lithium Li_2. With $\lambda \in [-6, 7]$ we interpolate and also extrapolate far beyond $\lambda \in [0, 1]$, exploring the xc potentials along the straight line passing through $v_{\mathrm{xc}}^{\mathrm{xKLI}}$ and $v_{\mathrm{xc}}^{\mathrm{LDA}}$ in xc potential space, see Fig. J.1 Next, we calculate the densities $n^{(\lambda)} = F[v_{\mathrm{xc}}^{(\lambda)}]$ for numerous values of λ in the interval. F is the forward KS map, cf. (8.1). To this end, we fix the xc potential in a DFT calculation and self-consistently solve for the Hartree potential. Thereafter the densities for different values of λ can be compared. Fig. J.2 shows the results of this analysis. The interpolated potentials at the bounds of the interval for λ, shown in the upper left and upper right panels, have a strong artificial character. Thus, the extrapolation is taken to the extreme at these values. Nevertheless the graphs of the metrics are uniform over the whole range. The density metric D_1 evaluated at the xKLI density (where $\lambda = 0$) and the density $n^{(\lambda)}$ for some other λ stays in the linear regime to very good accuracy over the large extend for λ. To the right (positive λ) the graph is only slightly convex, and concave to the left (negative λ). This is an indication, that the density itself responds in a mostly linear way. More importantly, the linear response is entirely given by the corresponding linear interpolation of the densities, $n_{\mathrm{lin}}^{(\lambda)}$, at least in the integrated metric $D_1[n^{(\lambda)}, n_{\mathrm{lin}}^{(\lambda)}]$. As expected J_{xc} is quadratic in λ.

While Fig. J.2 grants insights to the the extend of the linear regime for the integrated density metric D_1, it does not spatially resolve the density response. To get a better understanding of how the density is affected by the potential variation, we plot the density response against its potential variation in Fig. J.3, in the following way: In our prolate spheroidal grid, we first choose a set of random coordinates $\{\mathbf{r}_i\}$. For each coordinate $\mathbf{r}_i$ we evaluate the density $n^{(\lambda)}(\mathbf{r}_i)$ and the potential $v_{\mathrm{xc}}^{(\lambda)}(\mathbf{r}_i)$ for every λ. We plot all potential-density pairs $(n^{\lambda}(\mathbf{r}_i), v_{\mathrm{xc}}^{\lambda}(\mathbf{r}_i))$ as points (where the potential value corresponds to the values along the horizontal axis and the density corresponds to the vertical axis). The points are colored to discriminate the respective value of λ. Starting from a single random coordinate $\mathbf{r}_0$ and a single value $\lambda = 0$ (corresponding to the xKLI approximation) we draw the potential-density pair point $(n^{(0)}(\mathbf{r}_0), v_{\mathrm{xc}}^{(0)}(\mathbf{r}_0))$. Step by step drawing the density-potential pair for different values of λ with a different color, the pair-point will slightly shift, since the values of potential and density change. This perturbation by λ is therefore represented by a series of points or a string of beads. For $\lambda = 0$ (xKLI) we colored the point black as a guide for the eye, whereas at $\lambda = 1$ (LDA) we increased the points size slightly. Repeating the process for many randomly chosen coordinates $\mathbf{r}_i$, we can represent the density response by a collection of strings of beads. Note that this representation is not invariant with respect to a constant offset in the potential, causing stretched or compressed strings of beads. We have aligned the xc potentials to be zero at the boundary of the simulation ellipsoid to construct Fig. J.3.

Fig. J.3 covers many aspects of the KS map. First, focusing on the black points of the xKLI density, as expected in molecular systems mostly low density values correspond to potentials close to zero, whereas high densities map to low potential. For the LDA, the density-potential map is directly given by the functional construction, resulting in a predefined graph. For the xKLI functional, this mostly remains true, with the exception for densities of the order of about 10^{-2} or xc potentials in the range -0.4 to -0.2, where the inter-shell bumps are located, bifurcating the graph slightly. The LDA potential is less negative than the xKLI potential, therefore higher values of λ (which mean higher contribution from LDA) are mostly shifted to a more positive potential. As a result the strings of bead are mostly red on their right end. This is not true for the regions of inter-shell bump, where the xc potential of xKLI characteristically has larger potential than LDA.

Next, focus on the orientation of the strings of beads, i.e. the first order density response. In the high-density regime, we can directly see that large changes in the xc potential affect the

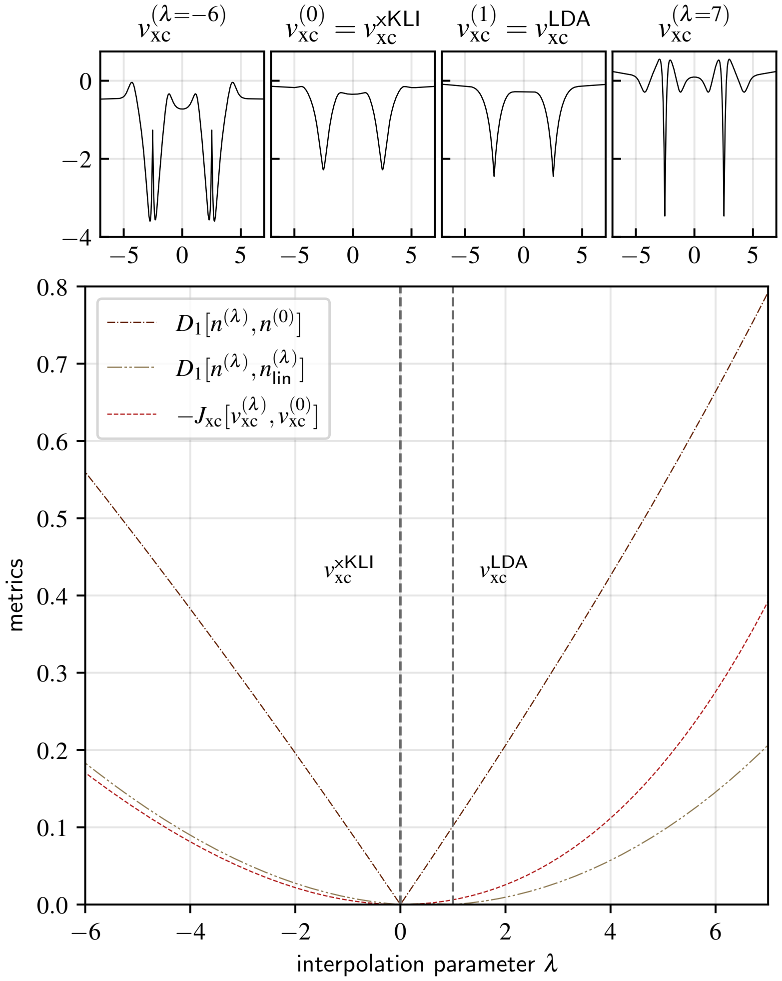


Figure J.2.: The xc potential interpolation from self-consistent xKLI to LDA with large extrapolation for Li_2. The upper insets shows the interpolated xc potential $v_{\text{xc}}^{(\lambda)}$ along the inter-atomic axis for different values of λ. The corresponding densities $n^{(\lambda)}$ deviate from the xKLI density (at $\lambda = 0$) almost entirely linear in the metric $D_1[n^{(\lambda)}, n^{(0)}]$. Furthermore, the linear prediction for the density $n_{\text{lin}}^{(\lambda)}$ covers the entire linear response, since its deviation from the actual density is quadratic, c.f. $D_1[n^{(\lambda)}, n_{\text{lin}}^{(\lambda)}]$. Note that at extreme extrapolation the potentials have strong artificial character, but are not far outside the linear regime, apparently. The behavior of J_{xc} is clearly quadratic as expected.

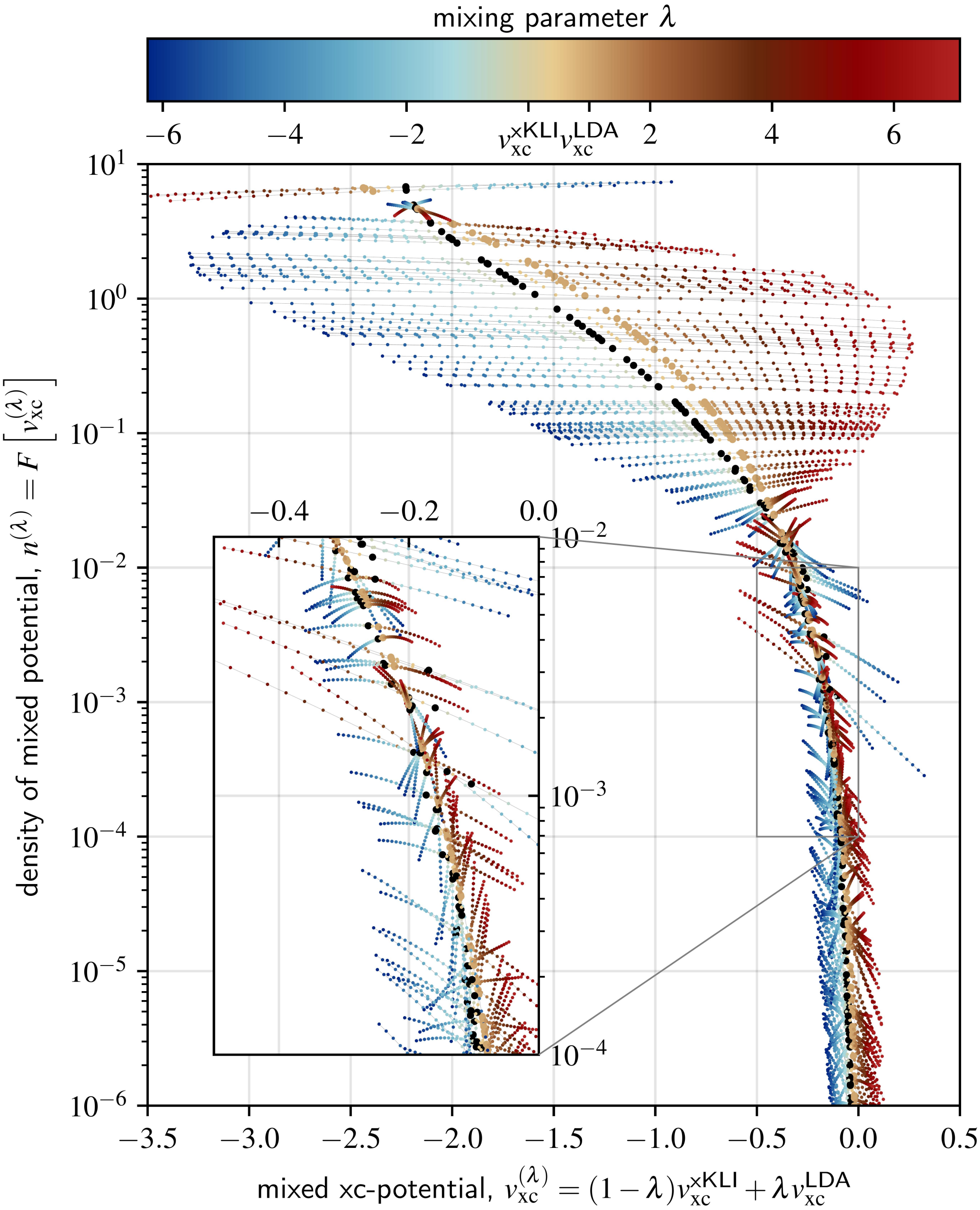


Figure J.3.: The response of the density to a linear perturbation in the potential for Li_2. For a number of random coordinates we plot the density value versus the potential value at these coordinates for different values of the mixing parameter λ. The color indicates the value of the mixing parameter. The xKLI calculation ($\lambda = 0$) is marked with black circles, the LDA calculation ($\lambda = 1$) with enlarged circles. Most of the density responds in the linear regime, significant higher orders almost exclusively appear at low density. The inset zooms in on the region of potential and density, where the non-linear behavior of the KS increases.

density only marginally indicated by a almost horizontal string of beads, as expected with the linear behavior of D_1 in Fig. J.2. This behavior is a clear trend only at high densities and breaks down at the inter-shell bumps, where the density responds much more detailed. The astonishing fact is, that the density responds mostly linear to changes in the potential, indicated by straight strings of beads. The linearity persists to high accuracy from the high density regime to density values of approximately 10^{-2}, where slight curvature of the strings of beads is visible, and breaks down only for very low densities and therefore regions of physically less relevance, which have almost negligible contribution to the total energy. The increasing quadratic contributions can be seen for density values of about 10^{-4} and lower. The common assumption [132] that an increased potential dictates a reduced density and vice versa, would forbid strings of beads to extend in the direction of the first and third quadrant. The assumption holds in the case of moderately high densities from about 10 to 10^{-1}. However, at lower densities, starting at about 0.1, and also the highest densities above about 5, this directional restriction is not met. For densities from 10^{-2} to 10^{-1} an increased potential even corresponds to an increased density. This again highlights the non-local character of the KS map.

The linear prediction $n_{\text{lin}}^{(\lambda)}$ is the dominating contribution to the density response. Thus within the limitations discussed above (e.g., a non-linear response in the low densities regime) a reasonable estimate of the resulting density at any λ is available without evaluation of the KS map. We have found the same behavior for interpolations from xKLI to PBE and SCAN potentials. The above implies linearity of the KS map only along a the straight line, but in conjunction with interpolations in different directions (e.g., PBE or SCAN) is indicates linearity to some extend in a larger scope.

$$F\left[(1-\lambda)v_{\text{xc}}^{(0)} + \lambda v_{\text{xc}}^{(1)}\right] = (1-\lambda)F\left[v_{\text{xc}}^{(0)}\right] + \lambda F\left[v_{\text{xc}}^{(1)}\right] \tag{J.5}$$

K. An efficient inversion procedure based on KS linearity

> The following section is an excerpt from the author's publication [116] including minor adjustments, see the Statement of authorship for an explanation of the authorship.

Here we propose an iterative inversion scheme, not based on update rules, but solely on minimization. We have not yet carried out any such inversions, due to time constraints. In Sec. 10.1 we illustrated the linearity of the KS map for interpolation/extrapolation of two xc potentials, $v_{\mathrm{xc}}^{(0)}$ and $v_{\mathrm{xc}}^{(1)}$. Assume that some reference density n, whose xc potential v_{xc} we are interested in, is "close" to the densities on the line of inter-/extrapolation, $n^{(\lambda)}$. Then, the inverted xc potential can readily be calculated by minimizing the cost $K = \int |n - n^{(\lambda)}|^2$ with respect to λ to yield some optimal λ^*. For reference densities that lie on the line of interpolation, K will vanish. The true xc potential is then given by $v_{\mathrm{xc}} = (1-\lambda^*)v_{\mathrm{xc}}^{(0)} + \lambda^* v_{\mathrm{xc}}^{(1)}$. Let us generalize this idea to a mixture of multiple xc potentials (and their corresponding densities), with the set of potential-density pairs $(v_{\mathrm{xc}}^{(i)}, n^{(i)})$ for $i = 1, \dots m$ and mixing parameters $\lambda_1, \dots \lambda_m$.

$$v_{\mathrm{xc}}(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^{m} \lambda_i v_{\mathrm{xc}}^{(i)} \tag{K.1}$$

Those potential-density pairs could for example be generated from common density functional approximations. We still assume linearity for the mixing even in the multidimensional case since interpolation in many variables can be executed step by step for each constituent xc potential. Thus we know a reasonable linear prediction for any given set of $(\lambda_1, \dots, \lambda_m)$,

$$n_{\mathrm{lin}}(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^{m} \lambda_i n^{(i)} \tag{K.2}$$

We can optimize the linear prediction to resemble the reference density n by minimization of some cost function

$$K(\lambda_1, \dots, \lambda_m) = \|n_{\mathrm{lin}}(\lambda_1, \dots, \lambda_m) - n\|, \tag{K.3}$$

where any reasonable metric $\|\cdot\|$ can be used. The minimizing parameters $(\lambda_1^*, \dots, \lambda_m^*)$ correspond to a minimizing density n^* and a candidate minimizing xc potential $v_{\mathrm{xc}}^* = \sum_{i=1}^{m} \lambda_i^* v_{\mathrm{xc}}^{(i)}$. If the reference density lies within the span of the densities $\{n^{(i)}\}_{i=1,\dots m}$, the cost vanishes, the optimal density is the reference density $n^* = n$ and with linearity the true xc potential is found by linearity as $v_{\mathrm{xc}}^* = v_{\mathrm{xc}}$. With any non-linearity in the KS map the optimal xc potential will deviate from the true xc potential. This is also the case if the reference density is not in this span, indicated by a non-vanishing cost. In this case we nevertheless get some optimal mixing parameters, density and candidate potential v_{xc}^*. Note that up to this point the forward map F was only evaluated m times to generate the initial potential-density pairs. We can now evaluate the forward map with the candidate potential v_{xc}^* to get its actual density $F[v_{\mathrm{xc}}^*]$. Let us suggestively denote this new potential-density pair with $(v_{\mathrm{xc}}^*, F[v_{\mathrm{xc}}^*]) = (v_{\mathrm{xc}}^{m+1}, n^{m+1})$. While this might not be an accurate estimate of the true xc potential and the reference density (due to nonlinearity or an insufficient span), we now can restart the minimization procedure with a new set of potential-density pairs, $(v_{\mathrm{xc}}^{(i)}, n^{(i)})$ for $i = 1, \dots m+1$, to more accurately recover the reference density by an increased span. Iterating this scheme will produce new potential-density pairs that increase the span and closer and closer resemble the true pair. Eventually the reference density will lie in the span over which the minimization is taken, such that the reference density is reconstructed. Additionally, after the density has become a good approximation to the reference density, new pairs will probe the close environment to the reference density, where linearity eventually must hold. Note that the deviation of the optimal density n^* to the actual density of the predicted potential $F[v_{\mathrm{xc}}^*]$ is key to the process, since it enriches the span of densities with linearly independent densities. Thus, far from the linear regime the algorithm mainly enlarges the span, while close to the reference density it solidifies the accuracy of the recovered density. With a reasonable linear prediction of the optimal density n^*, the evaluation of the forward map is significantly easier than a common self-consistent DFT calculation, since the xc potential is not self-consistently updated but fixed and the Hartree potential $v_{\mathrm{H}}[n^*]$ is already close to its self-consistent final form. The implementation of the algorithm is straightforward and rather simple. It only requires a solver for the forward problem and some minimization procedure. From the ever increasing span of densities and potentials, those pairs, that contribute little to the optimal pair can be excluded, to retain a reasonable number of parameters or dimensions for the minimization. Furthermore, casting the inversion as a minimization problem allows to add constraints or costs to the minimization. Two necessary constraints, that we omitted so far for simplicity, are $\sum_i \lambda_i = 1$ and $n_{\mathrm{lin}} \geq 0$ to ensure a normalized and non-negative density. Cost functions could for example incorporate weights to the density, include integral kernels or penalize the lack of exact characteristics of the density, such as the cusp condition.

Bibliography

[1]E. Romero, V. I. Novoderezhkin, and R. van Grondelle, "Quantum design of photosynthesis for bio-inspired solar-energy conversion", Nature **543**, 355 (2017).

[2]B. Mennucci and S. Corni, "Multiscale modelling of photoinduced processes in composite systems", Nat Rev Chem **3**, 315 (2019).

[3]M. Brütting, J. M. Foerster, and S. Kümmel, "Understanding Primary Charge Separation in the Heliobacterial Reaction Center", J. Phys. Chem. Lett. **14**, 3092 (2023).

[4]D. Elsing, A. Pálffy, and Y. Wu, "Quantum effects on plasma screening for thermonuclear reactions in laser-generated plasmas", Phys. Rev. Res. **4**, L022004 (2022).

[5]D. J. Hoarty et al., "Observations of the Effect of Ionization-Potential Depression in Hot Dense Plasma", Phys. Rev. Lett. **110**, 265003 (2013).

[6]L. B. Fletcher et al., "Observations of Continuum Depression in Warm Dense Matter with X-Ray Thomson Scattering", Phys. Rev. Lett. **112**, 145004 (2014).

[7]O. Ciricosta et al., "Measurements of continuum lowering in solid-density plasmas created from elements and compounds", Nat Commun **7**, 11713 (2016).

[8]S. M. Vinko, O. Ciricosta, and J. S. Wark, "Density functional theory calculations of continuum lowering in strongly coupled plasmas", Nat Commun **5**, 3533 (2014).

[9]R. Jin, M. M. Abdullah, Z. Jurek, R. Santra, and S.-K. Son, "Transient ionization potential depression in nonthermal dense plasmas at high x-ray intensity", Phys. Rev. E **103**, 023203 (2021).

[10]D. Kraus et al., "Characterizing the ionization potential depression in dense carbon plasmas with high-precision spectrally resolved x-ray scattering", Plasma Phys. Control. Fusion **61**, 014015 (2018).

[11]J. Zeng, Y. Li, C. Gao, and J. Yuan, "Screening potential and continuum lowering in a dense plasma under solar-interior conditions", A&A **634**, A117 (2020).

[12]S. Redner, "Citation Statistics from 110 Years of Physical Review", Physics Today **58**, 49 (2005).

[13]W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects", Phys. Rev. **140**, A1133 (1965).

[14]P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas", Phys. Rev. **136**, B864 (1964).

[15]J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems", Phys. Rev. B **23**, 5048 (1981).

[16]D. M. Ceperley and B. J. Alder, "Ground State of the Electron Gas by a Stochastic Method", Phys. Rev. Lett. **45**, 566 (1980).

[17]M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln", Annalen der Physik **389**, 457 (1927).

[18]E. Schrödinger, "An Undulatory Theory of the Mechanics of Atoms and Molecules", Phys. Rev. **28**, 1049 (1926).

[19] W. Pauli, "The Connection Between Spin and Statistics", Phys. Rev. **58**, 716 (1940).

[20] R. P. Feynman, R. B. Leighton, and M. L. Sands, *The Feynman Lectures on Physics* (Addison-Wesley Publishing Company, 1963).

[21] 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 **76**, 6062 (1979).

[22] E. H. Lieb, "Density functionals for coulomb systems", International Journal of Quantum Chemistry **24**, 243 (1983).

[23] L. D. Site, "Levy-Lieb constrained-search formulation as a minimization of the correlation functional", J. Phys. A: Math. Theor. **40**, 2787 (2007).

[24] E. Sim, S. Song, and K. Burke, "Quantifying Density Errors in DFT", The Journal of Physical Chemistry Letters **9**, 6385 (2018).

[25] E. Runge and E. K. U. Gross, "Density-Functional Theory for Time-Dependent Systems", Phys. Rev. Lett. **52**, 997 (1984).

[26] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer, Berlin, Heidelberg, 1990).

[27] A. Görling and M. Levy, "Hybrid schemes combining the Hartree-Fock method and density-functional theory: Underlying formalism and properties of correlation functionals", The Journal of Chemical Physics **106**, 2675 (1997).

[28] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, "Generalized Kohn-Sham schemes and the band-gap problem", Phys. Rev. B **53**, 3764 (1996).

[29] A. K. Rajagopal, "Inhomogeneous relativistic electron gas", J. Phys. C: Solid State Phys. **11**, L943 (1978).

[30] A. H. MacDonald and S. H. Vosko, "A relativistic density functional formalism", J. Phys. C: Solid State Phys. **12**, 2977 (1979).

[31] U. v. Barth and L. Hedin, "A local exchange-correlation potential for the spin polarized case. i", J. Phys. C: Solid State Phys. **5**, 1629 (1972).

[32] A. K. Rajagopal and J. Callaway, "Inhomogeneous Electron Gas", Phys. Rev. B **7**, 1912 (1973).

[33] E. Engel, *Density functional theory* (Springer, 2011).

[34] J. B. Krieger, Y. Li, and G. J. Iafrate, "Systematic approximations to the optimized effective potential: Application to orbital-density-functional theory", Phys. Rev. A **46**, 5453 (1992).

[35] J. B. Krieger, Y. Li, and G. J. Iafrate, "Construction and application of an accurate local spin-polarized Kohn-Sham potential with integer discontinuity: Exchange-only theory", Phys. Rev. A **45**, 101 (1992).

[36] M. Grüning, O. V. Gritsenko, and E. J. Baerends, "Exchange potential from the common energy denominator approximation for the Kohn-Sham Green's function: Application to (hyper)polarizabilities of molecular chains", The Journal of Chemical Physics **116**, 6435 (2002).

[37] O. V. Gritsenko and E. J. Baerends, "Orbital structure of the Kohn-Sham exchange potential and exchange kernel and the field-counteracting potential for molecules in an electric field", Phys. Rev. A **64**, 042506 (2001).

[38] F. Della Sala and A. Görling, "Efficient localized Hartree-Fock methods as effective exact-exchange Kohn-Sham methods for molecules", The Journal of Chemical Physics **115**, 5718 (2001).

[39] V. Sahni, J. Gruenebaum, and J. P. Perdew, “Study of the density-gradient expansion for the exchange energy”, Phys. Rev. B **26**, 4371 (1982).

[40] J. D. Talman and W. F. Shadwick, “Optimized effective atomic central potential”, Phys. Rev. A **14**, 36 (1976).

[41] R. T. Sharp and G. K. Horton, “A Variational Approach to the Unipotential Many-Electron Problem”, Phys. Rev. **90**, 317 (1953).

[42] S. Kümmel and J. P. Perdew, “Simple Iterative Construction of the Optimized Effective Potential for Orbital Functionals, Including Exact Exchange”, Phys. Rev. Lett. **90**, 043004 (2003).

[43] S. Kümmel and J. P. Perdew, “Optimized effective potential made simple: Orbital functionals, orbital shifts, and the exact Kohn-Sham exchange potential”, Phys. Rev. B **68**, 035103 (2003).

[44] C.-O. Almbladh and U. von Barth, “Exact results for the charge and spin densities, exchange-correlation potentials, and density-functional eigenvalues”, Phys. Rev. B **31**, 3231 (1985).

[45] A. Kaiser and S. Kümmel, “Accurate electron densities from Quantum Monte Carlo calculations using real-space grids”, The Journal of Chemical Physics **162**, 134108 (2025).

[46] K. Capelle, “A bird’s-eye view of density-functional theory”, Braz. J. Phys. **36**, 1318 (2006).

[47] J. P. Perdew and Y. Wang, “Accurate and simple analytic representation of the electron-gas correlation energy”, Phys. Rev. B **45**, 13244 (1992).

[48] V. Bach and L. Delle Site, *Many-electron approaches in physics, chemistry and mathematics* (Springer, 2014).

[49] W. Kohn, “Nobel Lecture: Electronic structure of matter—wave functions and density functionals”, Rev. Mod. Phys. **71**, 1253 (1999).

[50] J. Toulouse and C. J. Umrigar, “Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules”, The Journal of Chemical Physics **128**, 174101 (2008).

[51] J. Toulouse and C. J. Umrigar, “Optimization of quantum Monte Carlo wave functions by energy minimization”, The Journal of Chemical Physics **126**, 084102 (2007).

[52] P. Seth, P. L. Ríos, and R. J. Needs, “Quantum Monte Carlo study of the first-row atoms and ions”, The Journal of Chemical Physics **134**, 084105 (2011).

[53] 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 **99**, 2865 (1993).

[54] Y. Kwon, D. M. Ceperley, and R. M. Martin, “Effects of three-body and backflow correlations in the two-dimensional electron gas”, Phys. Rev. B **48**, 12037 (1993).

[55] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, “Inhomogeneous backflow transformations in quantum Monte Carlo calculations”, Phys. Rev. E **74**, 066701 (2006).

[56] C. Filippi and C. J. Umrigar, “Multiconfiguration wave functions for quantum Monte Carlo calculations of first-row diatomic molecules”, The Journal of Chemical Physics **105**, 213 (1996).

[57] E. Buendía, F. J. Gálvez, P. Maldonado, and A. Sarsa, “Quantum Monte Carlo ground state energies for the atoms Li through Ar”, The Journal of Chemical Physics **131**, 044115 (2009).

[58] J. Toulouse, R. Assaraf, and C. J. Umrigar, “Zero-variance zero-bias quantum Monte Carlo estimators of the spherically and system-averaged pair density”, The Journal of Chemical Physics **126**, 244112 (2007).

[59] M. P. Nightingale and V. Melik-Alaverdian, “Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters”, Phys. Rev. Lett. **87**, 043401 (2001).

[60] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, “Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions”, Phys. Rev. Lett. **98**, 110201 (2007).

[61] R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics*, 4th ed. (Macmillan, London, England, 1978).

[62] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines”, The Journal of Chemical Physics **21**, 1087 (1953).

[63] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications”, Biometrika **57**, 97 (1970).

[64] A. D. Fokker, “Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld”, Annalen der Physik **348**, 810 (1914).

[65] V. M. Planck, “Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie”, Sitzungberichte der (1917).

[66] R. M. Lee, G. J. Conduit, N. Nemec, P. López Ríos, and N. D. Drummond, “Strategies for improving the efficiency of quantum Monte Carlo calculations”, Phys. Rev. E **83**, 066706 (2011).

[67] J. Sherman and W. J. Morrison, “Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix”, The Annals of Mathematical Statistics **21**, 124 (1950).

[68] C. J. Umrigar, “Two aspects of quantum monte carlo: Determination of accurate wavefunctions and determination of potential energy surfaces of molecules”, International Journal of Quantum Chemistry **36**, 217 (1989).

[69] T. Kato, “On the eigenfunctions of many-particle systems in quantum mechanics”, Communications on Pure and Applied Mathematics **10**, 151 (1957).

[70] N. D. Drummond, M. D. Towler, and R. J. Needs, “Jastrow correlation factor for atoms, molecules, and solids”, Phys. Rev. B **70**, 235119 (2004).

[71] C.-J. Huang, C. J. Umrigar, and M. P. Nightingale, “Accuracy of electronic wave functions in quantum Monte Carlo: The effect of high-order correlations”, The Journal of Chemical Physics **107**, 3007 (1997).

[72] S. A. Alexander and R. L. Coldwell, “Atomic wave function forms”, International Journal of Quantum Chemistry **63**, 1001 (1997).

[73] A. D. Güçlü, G. S. Jeon, C. J. Umrigar, and J. K. Jain, “Quantum Monte Carlo study of composite fermions in quantum dots: The effect of Landau-level mixing”, Phys. Rev. B **72**, 205327 (2005).

[74] C.-J. Huang, C. Filippi, and C. J. Umrigar, “Spin contamination in quantum Monte Carlo wave functions”, The Journal of Chemical Physics **108**, 8838 (1998).

[75] N. Metropolis and S. Ulam, “The Monte Carlo Method”, Journal of the American Statistical Association **44**, 335 (1949).

[76] M. H. Kalos, “Monte Carlo Calculations of the Ground State of Three- and Four-Body Nuclei”, Phys. Rev. **128**, 1791 (1962).

[77] J. B. Anderson, “A random-walk simulation of the Schrödinger equation: H+3”, J. Chem. Phys. **63**, 1499 (1975).

[78] R. P. Feynman, A. R. Hibbs, and G. H. Weiss, “Quantum Mechanics and Path Integrals”, Physics Today **19**, 89 (1966).

[79] B. C. Hall, “Lie Groups, Lie Algebras, and Representations”, in *Quantum Theory for Mathematicians*, edited by B. C. Hall (Springer, New York, NY, 2013), pp. 333–366.

[80] H. F. Trotter, “On the product of semi-groups of operators”, Proc. Amer. Math. Soc. **10**, 545 (1959).

[81] M. Suzuki, “Generalized Trotter’s formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems”, Commun.Math. Phys. **51**, 183 (1976).

[82] I. Kosztin, B. Faber, and K. Schulten, “Introduction to the diffusion Monte Carlo method”, American Journal of Physics **64**, 633 (1996).

[83] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester Jr., “Fixed-node quantum Monte Carlo for moleculesa) b)”, The Journal of Chemical Physics **77**, 5593 (1982).

[84] J. Toulouse, R. Assaraf, and C. J. Umrigar, *Introduction to the variational and diffusion Monte Carlo methods*, Aug. 2015.

[85] M. F. DePasquale, S. M. Rothstein, and J. Vrbik, “Reliable diffusion quantum Monte Carlo”, The Journal of Chemical Physics **89**, 3629 (1988).

[86] D. Bressanini, G. Morosi, and S. Tarasco, “An investigation of nodal structures and the construction of trial wave functions”, The Journal of Chemical Physics **123**, 204109 (2005).

[87] D. Bressanini and G. Morosi, “On the nodal structure of single-particle approximation based atomic wave functions”, The Journal of Chemical Physics **129**, 054103 (2008).

[88] D. M. Ceperley and B. J. Alder, “Quantum Monte Carlo for molecules: Green’s function and nodal release”, The Journal of Chemical Physics **81**, 5833 (1984).

[89] F. Pederiva and M. H. Kalos, “Fermion Monte Carlo”, Computer Physics Communications, Proceedings of the Europhysics Conference on Computational Physics CCP 1998 **121-122**, 440 (1999).

[90] D. M. Arnow, M. H. Kalos, M. A. Lee, and K. E. Schmidt, “Green’s function Monte Carlo for few fermion problems”, The Journal of Chemical Physics **77**, 5562 (1982).

[91] A. Aspuru-Guzik and W. A. L. Jr, *Quantum Monte Carlo methods for the solution of the Schroedinger equation for molecular systems*, June 2002.

[92] D. M. Ceperley, “Fermion nodes”, J Stat Phys **63**, 1237 (1991).

[93] D. Bressanini, “Implications of the two nodal domains conjecture for ground state fermionic wave functions”, Phys. Rev. B **86**, 115120 (2012).

[94] P.-F. Loos and D. Bressanini, “Nodal surfaces and interdimensional degeneracies”, The Journal of Chemical Physics **142**, 214112 (2015).

[95] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).

[96] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, “Quantum Monte Carlo simulations of solids”, Rev. Mod. Phys. **73**, 33 (2001).

[97] P. H. Acioli, “Review of quantum Monte Carlo methods and their applications”, Journal of Molecular Structure: THEOCHEM, Proceedings of the Eighth Brazilian Symposium of Theoretical Chemistry **394**, 75 (1997).

[98] R. Assaraf, M. Caffarel, and A. Scemama, “Improved Monte Carlo estimators for the one-body density”, Phys. Rev. E **75**, 035701 (2007).

[99] Y. Wang and R. G. Parr, "Construction of exact Kohn-Sham orbitals from a given electron density", Phys. Rev. A **47**, R1591 (1993).

[100] C. J. Umrigar and X. Gonze, "Accurate exchange-correlation potentials and total-energy components for the helium isoelectronic series", Phys. Rev. A **50**, 3827 (1994).

[101] Q. Zhao, R. C. Morrison, and R. G. Parr, "From electron densities to Kohn-Sham kinetic energies, orbital energies, exchange-correlation potentials, and exchange-correlation energies", Phys. Rev. A **50**, 2138 (1994).

[102] O. V. Gritsenko, R. van Leeuwen, and E. J. Baerends, "Molecular Kohn-Sham exchange-correlation potential from the correlated ab initio electron density", Phys. Rev. A **52**, 1870 (1995).

[103] C. Filippi, C. J. Umrigar, and X. Gonze, "Excitation energies from density functional perturbation theory", The Journal of Chemical Physics **107**, 9994 (1997).

[104] M. J. Allen and D. J. Tozer, "Eigenvalues, integer discontinuities and NMR shielding constants in Kohn-Sham theory", Molecular Physics **100**, 433 (2002).

[105] D. P. Chong, O. V. Gritsenko, and E. J. Baerends, "Interpretation of the Kohn-Sham orbital energies as approximate vertical ionization potentials", The Journal of Chemical Physics **116**, 1760 (2002).

[106] M. Lein and S. Kümmel, "Exact Time-Dependent Exchange-Correlation Potentials for Strong-Field Electron Dynamics", Phys. Rev. Lett. **94**, 143003 (2005).

[107] A. M. Teale, F. De Proft, and D. J. Tozer, "Orbital energies and negative electron affinities from density functional theory: Insight from the integer discontinuity", The Journal of Chemical Physics **129**, 044110 (2008).

[108] M. Thiele, E. K. U. Gross, and S. Kümmel, "Adiabatic Approximation in Nonperturbative Time-Dependent Density-Functional Theory", Phys. Rev. Lett. **100**, 153004 (2008).

[109] A. S. d. Wijn, M. Lein, and S. Kümmel, "Strong-field ionization in time-dependent density functional theory", EPL **84**, 43001 (2008).

[110] M. Thiele and S. Kümmel, "Frequency Dependence of the Exact Exchange-Correlation Kernel of Time-Dependent Density-Functional Theory", Phys. Rev. Lett. **112**, 083001 (2014).

[111] O. V. Gritsenko, Ł. M. Mentel, and E. J. Baerends, "On the errors of local density (LDA) and generalized gradient (GGA) approximations to the Kohn-Sham potential and orbital energies", The Journal of Chemical Physics **144**, 204114 (2016).

[112] A. Kaiser and S. Kümmel, "Revealing the field-counteracting term in the exact Kohn-Sham correlation potential", Phys. Rev. A **98**, 052505 (2018).

[113] E. Fabiano et al., "Investigation of the Exchange-Correlation Potentials of Functionals Based on the Adiabatic Connection Interpolation", J. Chem. Theory Comput. **15**, 1006 (2019).

[114] R. Garrick, A. Natan, T. Gould, and L. Kronik, "Exact Generalized Kohn-Sham Theory for Hybrid Functionals", Phys. Rev. X **10**, 021040 (2020).

[115] J. D. Ramsden and R. W. Godby, "Exact Density-Functional Potentials for Time-Dependent Quasiparticles", Phys. Rev. Lett. **109**, 036402 (2012).

[116] A. Kaiser and S. Kümmel, "Rethinking the Kohn-Sham Inverse Problem", The Journal of Chemical Physics **163**, 104101 (2025).

[117] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Vol. 120, Applied Mathematical Sciences (Springer, New York, NY, 2011).

[118] "Sur les problèmes aux dérivées partielles et leur signification physique", Princeton university bulletin, 49 (1902).

[119] A. K. Theophilou, "The energy density functional formalism for excited states", J. Phys. C: Solid State Phys. **12**, 5419 (1979).

[120] O. V. Gritsenko and E. J. Baerends, "The spin-unrestricted molecular Kohn-Sham solution and the analogue of Koopmans's theorem for open-shell molecules", The Journal of Chemical Physics **120**, 8364 (2004).

[121] P. Strange, *Relativistic Quantum Mechanics: with applications in condensed matter and atomic physics* (Cambridge University Press, 1998).

[122] Q. Zhao and R. G. Parr, "Quantities T_s[n] and T_c[n] in density-functional theory", Phys. Rev. A **46**, 2337 (1992).

[123] T. W. Hollins, S. J. Clark, K. Refson, and N. I. Gidopoulos, "A local Fock-exchange potential in Kohn-Sham equations", J. Phys.: Condens. Matter **29**, 04LT01 (2017).

[124] Q. Wu and W. Yang, "A direct optimization method for calculating density functionals and exchange-correlation potentials from electron densities", The Journal of Chemical Physics **118**, 2498 (2003).

[125] M. Levy, J. P. Perdew, and V. Sahni, "Exact differential equation for the density and ionization energy of a many-particle system", Phys. Rev. A **30**, 2745 (1984).

[126] J. Erhard, E. Trushin, and A. Görling, "Numerically stable inversion approach to construct Kohn-Sham potentials for given electron densities within a Gaussian basis set framework", The Journal of Chemical Physics **156**, 204124 (2022).

[127] Y. Shi and A. Wasserman, "Inverse Kohn-Sham Density Functional Theory: Progress and Challenges", J. Phys. Chem. Lett. **12**, 5308 (2021).

[128] M.-C. Kim, E. Sim, and K. Burke, "Understanding and Reducing Errors in Density Functional Calculations", Phys. Rev. Lett. **111**, 073003 (2013).

[129] I. G. Ryabinkin, S. V. Kohut, and V. N. Staroverov, "Reduction of Electronic Wave Functions to Kohn-Sham Effective Potentials", Phys. Rev. Lett. **115**, 083001 (2015).

[130] E. Ospadov, I. G. Ryabinkin, and V. N. Staroverov, "Improved method for generating exchange-correlation potentials from electronic wave functions", The Journal of Chemical Physics **146**, 084103 (2017).

[131] K. Peirs, D. Van Neck, and M. Waroquier, "Algorithm to derive exact exchange-correlation potentials from correlated densities in atoms", Phys. Rev. A **67**, 012505 (2003).

[132] A. Kumar, R. Singh, and M. K. Harbola, "Universal nature of different methods of obtaining the exact Kohn-Sham exchange-correlation potential for a given density", J. Phys. B: At. Mol. Opt. Phys. **52**, 075007 (2019).

[133] L. Boccardo and T. Gallouët, "Non-linear elliptic and parabolic equations involving measure data", Journal of Functional Analysis **87**, 149 (1989).

[134] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, "Density functional theory is straying from the path toward the exact functional", Science **355**, 49 (2017).

[135] A. Aouina, M. Gatti, S. Chen, S. Zhang, and L. Reining, "Accurate Kohn-Sham auxiliary system from the ground-state density of solids", Phys. Rev. B **107**, 195123 (2023).

[136] L. V. Kantorovich, "Mathematical Methods of Organizing and Planning Production", Management Science **6**, 366 (1960).

[137] L. V. Kantorovich, “On the Translocation of Masses”, J Math Sci **133**, 1381 (2006).

[138] K. P. Huber and G. Herzberg, “Constants of diatomic molecules”, in *Molecular Spectra and Molecular Structure: IV. Constants of Diatomic Molecules*, edited by K. P. Huber and G. Herzberg (Springer US, Boston, MA, 1979), pp. 8–689.

[139] M. V. Ivanov and P. Schmelcher, “Ground state of the carbon atom in strong magnetic fields”, Phys. Rev. A **60**, 3558 (1999).

[140] D. Sundholm, P. Pyykkö, and L. Laaksonen, “Two-dimensional, fully numerical molecular calculations”, Molecular Physics **56**, 1411 (1985).

[141] T. Schmidt, E. Kraisler, L. Kronik, and S. Kümmel, “One-electron self-interaction and the asymptotics of the Kohn-Sham potential: an impaired relation”, Phys. Chem. Chem. Phys. **16**, 14357 (2014).

[142] F. Della Sala and A. Görling, “Asymptotic Behavior of the Kohn-Sham Exchange Potential”, Phys. Rev. Lett. **89**, 033003 (2002).

[143] P. Gori-Giorgi, T. Gál, and E. J. Baerends, “Asymptotic behaviour of the electron density and the Kohn-Sham potential in case of a Kohn-Sham HOMO nodal plane”, Molecular Physics **114**, 1086 (2016).

[144] T. Aschebrock, R. Armiento, and S. Kümmel, “Orbital nodal surfaces: Topological challenges for density functionals”, Phys. Rev. B **95**, 245118 (2017).

[145] P. Gori-Giorgi and E. J. Baerends, “Asymptotic nodal planes in the electron density and the potential in the effective equation for the square root of the density”, Eur. Phys. J. B **91**, 160 (2018).

[146] A. Makmal, S. Kümmel, and L. Kronik, “Fully Numerical All-Electron Solutions of the Optimized Effective Potential Equation for Diatomic Molecules”, J. Chem. Theory Comput. **5**, 1731 (2009).

[147] R. P. Feynman, “Forces in Molecules”, Phys. Rev. **56**, 340 (1939).

[148] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Vol. 93, Applied Mathematical Sciences (Springer International Publishing, Cham, 2019).

[149] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, Phys. Rev. Lett. **77**, 3865 (1996).

[150] J. Sun, A. Ruzsinszky, and J. P. Perdew, “Strongly Constrained and Appropriately Normed Semilocal Density Functional”, Phys. Rev. Lett. **115**, 036402 (2015).

[151] F. A. Reboredo and P. R. C. Kent, “Density-density functionals and effective potentials in many-body electronic structure calculations”, Phys. Rev. B **77**, 245110 (2008).

[152] R. T. Pack and W. B. Brown, “Cusp Conditions for Molecular Wavefunctions”, The Journal of Chemical Physics **45**, 556 (1966).

[153] W. Härdle, L. Simar, et al., “Applied multivariate statistical analysis”, Springer, Berlin-Heidelberg **10**, 978 (2007).

[154] J. F. Reynolds, “Some theorems on the transient covariance of Markov chains”, Journal of Applied Probability **9**, 214 (1972).

[155] J. B. Anderson, “Fixed-node quantum Monte Carlo”, International Reviews in Physical Chemistry **14**, 85 (1995).

Statement of authorship

Embedded publications

This thesis integrates substantial portions of previously published work, especially in sections or chapters presenting research results. The previous work is published in Ref. [45] and Ref. [116]. Both publications are co-authored with the supervisor of this thesis, Stephan Kümmel. The author of this thesis was the primary author of both publications.

While some sections or chapters of this theses are direct excerpts from the publications, others have been modified slightly to adapt for consistency with this thesis in notation, language, and context. The following overview lists the sections or chapters of this thesis that originate from the above publications.

Table 1.: Sections and chapters of this thesis containing previoulsy published or soon to be submitted work of the author.

section/chapter name	reference	puplication
Persistent walkers	Sec. 6.14	Ref. [45]
QMC results	Chapt. 9	Ref. [45]
Inversion theory	Chapt. 8	Ref. [116]
Inversion results	Chapt. 10	Ref. [116]
Density fluctuations	Sec. 7.2	Ref. [45]
Influence of the orbitals' grid on QMC	Sec. F	Ref. [45]
Regularization of noisy densities	Sec. G	Ref. [45]
Binning technique	Sec. 7.1	Ref. [45]
Preparation of densities for inversion	Sec. H	Ref. [116]
The forward KS-map in detail	Sec. J	Ref. [116]
An efficient inversion procedure based on KS linearity	Sec. K	Ref. [116]

Auxiliary tools

The author of this thesis has used AI-tools (tools based on artificial intelligence) to accelerate the process of writing and enhance the linguistic quality of this thesis regarding clarity, coherence and comprehensibility. AI tools have been used to accelerate the generation of code. However, all AI suggestions have been thoroughly tested by the author. While the author of this thesis considered the output of large language models as a guide, no literal output, copy or unmodified content from any large language model is present in this thesis. The author has not used any AI tool to create or conceptualize intellectual content or to contribute to the research of this thesis. The following tools were used: grammarly, ChatGPT, claude, deepl

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.
Weiterhin erkläre ich, dass ich die Hilfe von gewerblichen Promotionsberatern bzw. –vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe, noch künftig in Anspruch nehmen werde.
Zusätzlich erkläre ich hiermit, dass ich keinerlei frühere Promotionsversuche unternommen habe.

Bayreuth, den

Unterschrift