Diagrammatic Methods for Strongly-Correlated Electrons in Realistic Systems

by

Jia Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics and Scientific Computing)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Emanuel Gull, Chair
Professor Lu Li
Associate Professor Xiaoming Mao
Associate Professor Kai Sun
Associate Professor Dominika Zgid

Jia Li
lijiax@umich.edu
ORCID iD: 0000-0001-7845-4748

Acknowledgments

First and foremost, I would like to give my sincere gratitude to my research advisor, Prof. Emanuel Gull, for his continuous support and guidance during the past five years. With great knowledge in both physics and computational technology, he always points me to the right direction in times of challenges and confusion, but also gives me much freedom to explore. He provided me with ample resources in my doctoral study, which has given me valuable opportunities to hold discussions with outstanding scientists in the field, to participate in international conferences and workshops, and to visit scientific institutions around the world. This thesis would not be possible without his help and encouragement.

I thank every previous and current members of Gull Group for a vibrant and fun environment. I would like to thank Dr. Sergei Iskakov for being a great mentor, colleague, and friend and for teaching me C++ programming as well as the art of debate; Dr. Markus Wallerberger for a fruitful collaboration in my main projects and for insightful discussions about science, culture, music, and society. I feel privileged to have worked with, and have learned a lot from, our great postdocs: Dr. Andrei Antipov, Dr. James LeBlanc, Prof. Hanna Terletska, and Dr. Igor Krivenko. I also thank my fellow PhD students for their friendship: Dr. Joseph Kleinhenz, Dr. Qiaoyuan Dong, Dr. Xi Chen, Dr. Joseph Paki, Chia-Nan Yeh, Xinyang Dong, Runxue Yu, and Thomas Blommel. Special thanks go to Yang Yu, who read the first draft of this thesis and provided me with valuable feedback.

Since a large part of my work focused on chemistry systems, I am grateful to Prof. Dominika Zgid for her helpful insights from a chemist's perspective. I also thank Dr. Avijit Shee, Dr. Tran Nguyen Lan from Zgid group for helpful discussions.

I feel fortunate that the Physics Department at the University of Michigan has provided great support to my study here. I thank Prof. Lu Li, Prof. Finn Larsen, Prof. David Lubensky for their mentorship during my first year in graduate school, Dr. Yuri Popov, Prof. Xiaoming Mao, and Prof. Christine Aidala for helping with my teaching, Prof. Henriette Elvang for the fantastic lectures on quantum field theory, and my PhD committee members: Prof. Lu

Li, Prof. Xiaoming Mao, Prof. Kai Sun, and Prof. Dominika Zgid, for their time and helpful suggests to my thesis.

My research projects benefitted a lot from interactions with researchers outside of Michigan. I would like to thank Prof. Hiroshi Shinaoka for his hospitality when I visited Saitama University and his enthusiastic support during the sparse sampling project. I thank Prof. Guy Cohen for generously sharing his ideas and experiences for the inchworm project and providing necessary help remotely from Tel Aviv. I am thankful to Dr. Igor Tupitsyn, Dr. Riccardo Rossi and Dr. Fedor Šimkovic for teaching me the basics of DiagMC and CDet as experts in the field. I also thank Prof. Shiwei Zhang, Prof. Hao Shi, and Prof. Lucas Wagner for their help during the benchmark projects.

My graduate school life was made enjoyable thanks to the support and companionship of my friends in Michigan. Although many of them have started new journeys around the world, I sincerely believe that our memories and friendship will transcend any dimensions of space and time.

Last but not least, I would like to thank my parents and my girlfriend Ellen for their unconditional love and support.

Table of Contents

List of Figures

List of Appendices

List of Acronyms

1PI one-particle irreducible.

2PI two-particle irreducible.

AO atomic orbital.

BDMC bold diagrammatic Monte Carlo.

CBS continuous basis set.

CDet connected determinant.

CDF cumulative distribution function.

CGF cumulant generating function.

CLT central limit theorem.

CT-INT interaction-expansion continuous time quantum Monte Carlo.

CT-QMC continuous time quantum Monte Carlo.

DFT density functional theory.

DiagMC diagrammatic Monte Carlo.

DMFT dynamical mean-field theory.

DMRG density matrix renormalization group.

ED exact diagonalization.

EOS equation of state.

FCI full configuration interaction.

GF2 self-consistent second-order Green's function perturbation theory.

GTO Gaussian-type orbital.

HF Hartree–Fock.

IR intermediate representation.

LW Luttinger–Ward.

MGF moment generating function.

PDF probability density function.

PEC potential energy curve.

SC-*GW* self-consistent *GW*.

SCF self-consistent field.

SEET self-energy embedding theory.

STO Slater-type orbital.

TDL thermodynamic limit.

Abstract

This thesis is devoted to the development of numerical methods for correlated electrons in realistic systems based on perturbation theories using Feynman diagrams. Electron correlations in realistic systems give rise to a wide range of interesting physical properties. Diagrammatic expansions and its low-order approximations are key building blocks of theoretical and numerical methods for studying correlated electrons. Recent developments of the diagrammatic quantum Monte Carlo (DiagMC) methods enable efficient stochastic evaluations of diagrams and does not suffer from numerical sign problem which worsens with increasing system size. It is therefore a promising direction to apply DiagMC techniques to realistic electron systems in the pursuit of *ab initio* calculations of correlated materials.

This thesis starts by reviewing basic concepts and definitions of realistic electron systems, the perturbation theory based on Feynman diagrams, Monte Carlo evaluations of the series, as well as self-consistent diagrammatic methods. Low-order diagrammatic methods are applied to realistic calculations and benchmarked against large collections of state-of-the-art many-body methods.

The thesis then proposes a diagrammatic Monte Carlo solver for realistic molecules and impurities based on a recursive evaluation of diagrams, which is tested on realistic molecules and impurities with varying temperature, system size, and interaction strength. We observe that this method is ideal for problems with many orbitals with moderate correlations, but may suffer from series divergence at strong correlations.

Next, we propose the inchworm Monte Carlo algorithm for interaction expansion, which overcomes limitations of divergent diagrammatic series by replacing the standard perturbation series with a sequence of incremental series expansions. A prototype implementation is presented for system with on-site Hubbard interactions.

Finally, we discuss technical aspects of numerical representations of the Green's function, which has significant impact on the efficiency of diagrammatic methods. We review several common numerical representations, and introduce our development of the sparse sampling

method, which provides compact representations of the Green's function in both imaginary time and frequency as well as fast transformation formula between the two.

Chapter 1

Introduction

Electron interactions in realistic solids give rise to many fascinating physical phenomena, such as metal–insulator transitions, high-temperature superconductors, colossal magnetoresistance, and so on. These phenomena are not only intellectually intriguing, but also impactful in advancing broader science, technology, and human society. For example, superconductors with higher transition temperatures make better magnets, which allows high-energy particle colliders to achieve higher collision energy; metal–insulator transitions lead to the development of Mott field effective transistors; colossal magnetoresistance effects can be applied to magnetic field sensors.

Due to the many-body nature of interacting electrons, explicit solution to the many-body Schrödinger equation becomes prohibitively difficult as one increase the system size. Approximations to the full electronic Hamiltonian are therefore necessary. In *ab initio* methods, such as the density functional theory (DFT) [1] and the Hartree–Fock approximation [2], electrons are treated on a mean-field level based on the quasi-particle picture. This makes numerical studies possible for realistic materials, but with limited capability of including electron correlations. On the other end of the spectrum, methods developed for strongly

correlated systems such as high temperature superconductors often treat the electron interactions explicitly, and therefore require approximating the system with model Hamiltonians.

Both types of approximations rely on physical "insights" to decide whether the systematic error as a result of the approximation is acceptable. However, as one moves toward intermediate interaction strength, the level of complexity of the many-body problem may become difficult to analyze intuitively. It is therefore beneficial to have controllable approximations, which are theoretically guaranteed to approach the exact solution by tuning some control parameters, and the systematic errors of the approximation can be estimated quantitatively.

Perturbation theory based on Feynman diagrams [3] is the cornerstone of modern quantum field theory in both high-energy physics and condensed matter physics. Diagrammatic expansions of the electronic system in terms of the electron interactions have become the foundation of a wide range of numerical applications. It is straightforward to construct controlled approximations based on finite-order diagrammatic approximations, as the expansion orders of the perturbation series naturally serve as a control parameter. For example, self-consistent methods based on low-order diagrams, such as self-consistent second-order Green's function perturbation theory (GF2) [4–7] and the *GW* approximation [8], are able to include electron correlation effects with relatively low computational cost. Continuous time quantum Monte Carlo (CT-QMC) methods [9–14] solve quantum impurity problems with unprecedented precision via stochastic sampling of the partition function diagrams, which significantly boost the power of the dynamical mean-field theory (DMFT) [15, 16]. The diagrammatic Monte Carlo (DiagMC) technique [17–21], formulated directly in the thermodynamic limit and stochastically samples individual Feynman diagrams, does not suffer from the fermionic sign problem that worsens exponentially with increasing system size and inverse temperature. Recent developments of DiagMC methods, such as the connected determinant (CDet) approach [22–26] for fast diagram summations, the inchworm Monte Carlo

method [27–33], and analytic resummation techniques [24, 34], have greatly improved the numerical efficiency of evaluating diagrams. Diagrammatic methods therefore show great potential in solving realistic Hamiltonians, either as a standalone solver, or as an "impurity" solver for quantum embedding methods, such as DMFT [15, 16] and self-energy embedding theory (SEET) [35–37].

My PhD study mainly focuses on the development of controlled diagrammatic methods which directly work with realistic Hamiltonians from first principles.

Structure of this thesis

The first few chapters summarizes basic concepts and notations for developing diagrammatic methods for realistic systems. Chapter 2 reviews basic definitions of the electronic Hamiltonian in realistic basis sets, and lays out notations that are used in later chapters. Chapter 3 derives the perturbation expansions of physical quantities in thermal equilibrium, which are often referred to as the "bare" diagrammatic expansion. Chapter 4 reviews the Monte Carlo integration method, and provides a general framework for Monte Carlo calculations of diagrammatic series. Chapter 5 reviews the "bold" diagrammatic series based on the Luttinger–Ward functional as well as the self-consistent approximations.

Chapter 6 summarizes two benchmark projects of many-body methods applied to realistic systems. During these projects, I applied self-consistent diagrammatic approximations to hydrogen chain systems as well as transition metal atoms and molecules, and the results are comprehensively analyzed in comparison with large collections of state-of-the-art many-body methods.

The next two chapters covers my contribution in developing two numerical algorithms based on DiagMC. Chapter 7 introduces the development of a DiagMC solver for realistic impurities with general interactions and hybridizations using recursive evaluations of Feyn-

man diagrams with determinants. The solver is tested on realistic molecules along various aspects of complexities, such as temperature, system size, basis size, and types of molecules. Chapter 8 introduces an "inchworm" algorithm for interaction expansion methods, which performs a series of incremental perturbation expansions in order to achieve better series convergence at each step. We show that the inchworm expansion converges in systems where the bare diagrammatic series diverges, which provides a more robust approach in evaluating diagrammatic expansions.

Finally, Chapter 9 focuses on a technical aspect when developing diagrammatic methods for realistic systems: efficient numerical representations of the electron Green's function. Common representation schemes, such as the high-frequency tails, orthonormal polynomials, and the intermediate representation (IR) basis are briefly reviewed. We then introduce the sparse sampling method, a framework for compact Green's function representations in both imaginary time and frequency with fast transformations, which I participated in the development and optimization.

Chapter 2

Realistic Electron Hamiltonian and Basis Sets

This chapter introduces the definition of the general electronic Hamiltonian in realistic materials as the main subject of study of this thesis, as well as the electronic basis sets which projects the continuous space into the discrete orbital space, allowing linear algebraic treatment of the Hamiltonian. Notations introduced in this chapter will serve as a foundation of most derivations in subsequent chapters of this thesis.

2.1 Many-body Hamiltonian of electrons

We study the quantum many-body problem of N_e electrons in the potential field of a realistic lattice or cluster of nuclei. Under the Born–Oppenheimer approximation, we separate the electronic motion from the nuclear degrees of freedom, and describe the electrons using the

following Hamiltonian [38]:

$$\begin{aligned}\hat{H} &= \sum_{i=1}^{N_e} \hat{h}(\mathbf{r}_i) + \sum_{i=1}^{N_e} \sum_{j=i+1}^{N_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \\ \hat{h}(\mathbf{r}_i) &= -\frac{\nabla_{\mathbf{r}_i}^2}{2m_e} - \sum_{I=1}^{N_n} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|},\end{aligned} \tag{2.1}$$

where $\mathbf{r}_i$ $(i = 1, \ldots, N_e)$ are spatial coordinates of electrons, $\mathbf{R}_I$ $(I = 1, \ldots, N_n)$ coordinates of N_n nuclei, Z_I the atomic numbers of the nuclei, and m_e the electronic mass. The one-particle Hamiltonian $\hat{h}$ contains the kinetic energy and the potential energy from the Coulomb potential of the nuclei. The existence of electron-electron Coulomb interactions in $\hat{H}$ prevents us from reducing the Hamiltonian to a single-particle problem, and requires many-body treatment of the system.

An electronic basis set is a set of functions in real space $\phi_\mu(\mathbf{r})$ which forms a basis for the single-particle electronic wave function. Common choices include plane wave, atomic orbitals (AOs), or combinations of the two. For some types of basis sets such as the AOs (usually represented either as Gaussian or Slater functions), it is possible that the basis functions are not orthogonal. The non-orthogonality is characterized by the overlap matrix $\boldsymbol{S}$, such that [38]

$$[\boldsymbol{S}]_{\mu\nu} = S_{\mu\nu} = \int \mathrm{d}^3\mathbf{r}\, \phi_\mu^*(\mathbf{r})\phi_\nu(\mathbf{r}). \tag{2.2}$$

Projected onto the basis functions [38, 39], the one-particle Hamiltonian becomes a matrix $\boldsymbol{h}$:

$$[\boldsymbol{h}]_{\mu\nu} = h_{\mu\nu} = \int \mathrm{d}^3\mathbf{r}\, \phi_\mu^*(\mathbf{r})\hat{h}(\mathbf{r})\phi_\nu(\mathbf{r}), \tag{2.3}$$

and the electron-electron interaction becomes a rank-4 tensor $\boldsymbol{\mathscr{V}}$:

$$[\boldsymbol{\mathscr{V}}]_{\mu\nu\lambda\rho} = V_{\mu\nu\lambda\rho} = \int \mathrm{d}^3\mathbf{r}\mathrm{d}^3\mathbf{r}'\ \phi_\mu^*(\mathbf{r})\phi_\nu(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\phi_\lambda^*(\mathbf{r}')\phi_\rho(\mathbf{r}'), \tag{2.4}$$

Here we have used bold italic symbols for matrices, bold calligraphic symbols for tensors, and regular italic symbols for matrix/tensor elements. These conventions will be assumed throughout this thesis unless otherwise stated. In quantum chemistry, $\boldsymbol{h}$ and $\boldsymbol{\mathscr{V}}$ are termed one- and two-electron integrals, respectively, and the following notations are commonly used [38]:

$$(\mu|h|\nu) = h_{\mu\nu}, \quad (\mu\nu|\lambda\rho) = V_{\mu\nu\lambda\rho}. \tag{2.5}$$

Introducing the electron creation operator $\hat{c}_{\mu\sigma}^\dagger$ and the annihilation operator $\hat{c}_{\mu\sigma}$ for an electron in orbital ϕ_μ and spin σ, the many-body Hamiltonian (2.1) can be rewritten in the second-quantized form [39]:

$$\hat{H} = \sum_{\mu\nu}\sum_{\sigma} h_{\mu\nu}\hat{c}_{\mu\sigma}^\dagger\hat{c}_{\nu\sigma} + \frac{1}{2}\sum_{\mu\nu\lambda\rho}\sum_{\sigma\sigma'} V_{\mu\nu\lambda\rho}\hat{c}_{\mu\sigma}^\dagger\hat{c}_{\lambda\sigma'}^\dagger\hat{c}_{\rho\sigma'}\hat{c}_{\nu\sigma}. \tag{2.6}$$

This effectively switched to a Fock space in which the electron number is variable, from the Hilbert space where Eq. (2.1) is defined. Note that since the basis can be non-orthogonal, the anti-commutation relations of the electronic operators are given by

$$\{\hat{c}_{\mu\sigma}^\dagger, \hat{c}_{\nu\sigma'}\} = \delta_{\sigma\sigma'}S_{\mu\nu}. \tag{2.7}$$

In (2.6), we do not assume any periodicity of the basis function ϕ, and each orbital index $\mu, \nu, \ldots$ represent a single AO at a specific nucleus location. In periodic systems such as lattice systems, different unit cells contain exactly the same atoms, and the orbital indices μ may

be split into the lattice indices $\mathbf{n} = (n_x, n_y, \ldots)$ and the orbital indices within the unit cell, i.e. $\mu \to (\mu, \mathbf{n})$. The basis function at lattice site $\mathbf{n}$ and orbital μ is then

$$\phi_{\mu\mathbf{n}}(\mathbf{r}) = \phi_\mu(\mathbf{r} - \mathbf{R}_\mathbf{n}). \tag{2.8}$$

2.2 Orthogonalization of the basis set

The most commonly used basis sets in quantum chemistry are constructed based on AOs, such as Slater-type orbitals (STOs) [40] and Gaussian-type orbitals (GTOs) [41, 42]. AOs are atom-centered functions motivated by the eigenstates of Schrödinger equation for hydrogen-like atoms, which provides an intuitive picture for interpreting calculation results. Different basis sets are optimized for each element with finite number of AOs. By systematically choosing larger basis sets, the system can be extrapolated to the continuous basis set (CBS) limit, where the basis is considered complete [43, 44]. See e.g. Ref. [45] for a review of atomic basis sets.

In molecular or lattice systems, AO basis sets usually yield non-orthogonal basis functions as a result of overlapping orbitals from different atoms, i.e. the overlap matrix $\boldsymbol{S}$ is not diagonal. However, many numerical methods, such as most diagrammatic methods studied in this thesis, are formulated in orthonormal basis sets. Is is therefore convenient to perform a linear transformation of the AO basis

$$\phi'_i(\mathbf{r}) = \sum_\mu \phi_\mu(\mathbf{r}) X_{\mu i} \tag{2.9}$$

via some transformation matrix $\boldsymbol{X}$, such that the resulting basis $\{\phi'_i\}$ is orthonormal. This

implies

$$\delta_{ij} \equiv \int \mathrm{d}^3\mathbf{r}\, \phi_i'^*(\mathbf{r})\phi_j'(\mathbf{r}) = \sum_{\mu\nu} X_{\mu i}^* X_{\nu j} \int \mathrm{d}^3\mathbf{r}\, \phi_\mu^*(\mathbf{r})\phi_\nu(\mathbf{r}) = [\boldsymbol{X}^\dagger \boldsymbol{S}\boldsymbol{X}]_{ij}, \tag{2.10}$$

or in matrix form

$$\boldsymbol{X}^\dagger \boldsymbol{S}\boldsymbol{X} = \boldsymbol{I}, \tag{2.11}$$

where $\boldsymbol{S}$ is the overlap matrix and $\boldsymbol{I}$ the identity matrix.

In order to find such a transformation $\boldsymbol{X}$, we perform an eigenvalue decomposition of $\boldsymbol{S}$:

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{s}\boldsymbol{U}^\dagger, \tag{2.12}$$

where $\boldsymbol{U}$ is unitary and $\boldsymbol{s}$ is the diagonal matrix of the eigenvalues. Equation 2.11 is reorganized as

$$(\boldsymbol{X}^\dagger \boldsymbol{U}\boldsymbol{s}^{1/2})(\boldsymbol{s}^{1/2}\boldsymbol{U}^\dagger \boldsymbol{X}) = \boldsymbol{I}. \tag{2.13}$$

Defining $\boldsymbol{Y} = \boldsymbol{s}^{1/2}\boldsymbol{U}^\dagger \boldsymbol{X}$, we have

$$\boldsymbol{Y}^\dagger \boldsymbol{Y} = \boldsymbol{I}, \tag{2.14}$$

i.e. $\boldsymbol{Y}$ is a unitary matrix. Solving for $\boldsymbol{X}$ yields

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{s}^{-1/2}\boldsymbol{Y}. \tag{2.15}$$

The choice of the unitary matrix $\boldsymbol{Y}$ is arbitrary, as the final basis sets ϕ' from different choices are only different by a unitary transformation.

In quantum chemistry, there are two common choices [38] for $\boldsymbol{Y}$. The **canonical orthogonalization** chooses $\boldsymbol{Y} = \boldsymbol{I}$, such that

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{s}^{-1/2}, \tag{2.16}$$

which is useful when $\boldsymbol{S}$ is ill-conditioned and truncations of the eigenvalues are needed[1]. Alternatively, the **symmetric orthogonalization** chooses $\boldsymbol{Y} = \boldsymbol{U}^\dagger$, such that the transformation matrix

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{s}^{-1/2}\boldsymbol{U}^\dagger \tag{2.17}$$

is symmetric. Once $\boldsymbol{X}$ is fixed, one can transform electronic integrals to the orthogonal basis:

$$h'_{ij} = \sum_{\mu\nu} h_{\mu\nu} X^*_{\mu i} X_{\nu j}, \quad V'_{ijkl} = \sum_{\mu\nu\lambda\rho} V_{\mu\nu\lambda\rho} X^*_{\mu i} X_{\nu j} X^*_{\lambda k} X_{\rho l}. \tag{2.18}$$

In the remainder of this thesis, we use latin indices $i, j, \ldots$ for orthogonal orbitals, greek indices $\mu, \nu, \ldots$ for non-orthogonal AOs, and omit the "primes" when there is no conflict of notation. The electronic Hamiltonian (2.6) can be rewritten using the orthogonalized basis as

$$\hat{H} = \sum_{ij} \sum_{\sigma} h_{ij} \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} + \frac{1}{2} \sum_{ijkl} \sum_{\sigma\sigma'} V_{ijkl} \hat{c}^\dagger_{i\sigma} \hat{c}^\dagger_{k\sigma'} \hat{c}_{l\sigma'} \hat{c}_{j\sigma}, \tag{2.19}$$

where the creation and annihilation operators now follow the canonical anti-commutation relation

$$\{\hat{c}^\dagger_{i\sigma}, \hat{c}_{j\sigma'}\} = \delta_{ij}\delta_{\sigma\sigma'}. \tag{2.20}$$

2.3 Spin-orbitals and anti-symmetrized interaction

Some applications benefit from combining the spin and orbital degrees of freedom together, which forms the so-called "spin-orbitals" [38]. We introduce the compound notation $\{a, b, \ldots\}$

[1] This yields a smaller basis after the transformation.

such that

$$h_{ab} = h_{(i\sigma)(j\sigma')} \equiv h_{ij}\delta_{\sigma\sigma'} \tag{2.21}$$

$$V_{abcd} = V_{(i\sigma)(j\sigma')(k\lambda)(l\lambda')} \equiv V_{ijkl}\delta_{\sigma\sigma'}\delta_{\lambda\lambda'}. \tag{2.22}$$

The fermionic anti-symmetry can be explicitly encoded in the anti-symmetrized interaction tensor $\mathscr{U}$ as

$$U_{abcd} = V_{abcd} - V_{adcb}, \tag{2.23}$$

such that Eq. (2.19) becomes

$$\hat{H} = \sum_{ab} h_{ab}\hat{c}_a^\dagger \hat{c}_b + \frac{1}{4}\sum_{abcd} U_{abcd}\hat{c}_a^\dagger \hat{c}_c^\dagger \hat{c}_d \hat{c}_b. \tag{2.24}$$

2.4 Thermal dynamic quantities

This section defines notations of the finite-temperature properties of electrons, the main physical quantities studied in this thesis. The grand partition function [39] of the Hamiltonian (2.19) writes

$$Z = \mathrm{Tr}\, e^{-\beta(\hat{H}-\mu\hat{N})}, \tag{2.25}$$

where $\beta = (k_\mathrm{B}T)^{-1}$ is the inverse temperature, μ is the chemical potential, and $\hat{N} = \sum_{i\sigma} \hat{c}_{i\sigma}^\dagger \hat{c}_{i\sigma}$ is the number operator. Grand canonical ensemble is necessary here as the electron number is not fixed in the Fock space. It is common to absorb the $\mu\hat{N}$ term into the Hamiltonian by shifting the one-electron Hamiltonian $\boldsymbol{h}$:

$$\boldsymbol{h}' = \boldsymbol{h} - \mu\boldsymbol{I}, \tag{2.26}$$

such that $Z = \mathrm{Tr}\, e^{-\beta \hat{H}'}$, and the primes can be dropped in the absence of confusion. The thermal expected value of an electronic operator $\hat{O}$ is given by

$$\langle \hat{O} \rangle = \frac{\mathrm{Tr}[e^{-\beta(\hat{H}-\mu\hat{N})}\hat{O}]}{Z}. \tag{2.27}$$

The central quantity for studying correlations of electrons is the many-body Green's function. An m-particle Green's function is defined as a correlation function of m creation operators and m annihilation operators in imaginary time [39]

$$G^{(m)}(x_1, x_1', x_2, x_2', \ldots, x_m, x_m') = (-1)^m \langle \mathscr{T}_\tau \hat{c}(x_1)\hat{c}^\dagger(x_1')\hat{c}(x_2)\hat{c}^\dagger(x_2') \cdots \hat{c}(x_m)\hat{c}^\dagger(x_m') \rangle, \tag{2.28}$$

where $\mathscr{T}_\tau$ is the time-ordering operator, and $x = (\tau, i, \sigma)$ is the compound index such that

$$\hat{c}^{(\dagger)}(x) = \hat{c}^{(\dagger)}_{i\sigma}(\tau) = e^{\tau(\hat{H}-\mu\hat{N})}\hat{c}^{(\dagger)}_{i\sigma}e^{-\tau(\hat{H}-\mu\hat{N})}. \tag{2.29}$$

At $m = 1$, we obtain the single-particle Green's function

$$G(x, x') = G_{i\sigma, j\sigma'}(\tau, \tau') = -\langle \mathscr{T}_\tau \hat{c}_{i\sigma}(\tau)\hat{c}^\dagger_{j\sigma'}(\tau') \rangle. \tag{2.30}$$

If $\hat{H}$ conserves spin, we have

$$G_{i\sigma, j\sigma'}(\tau, \tau') = G_{ij,\sigma}(\tau, \tau'). \tag{2.31}$$

Further simplification can be made due to the cyclic property of the trace:

$$G_{ij,\sigma}(\tau, \tau') = G_{ij,\sigma}(\tau - \tau'). \tag{2.32}$$

$G_{ij,\sigma}(\tau)$ is anti-periodic [39] in imaginary time, i.e. $G_{ij,\sigma}(\tau+\beta) = -G_{ij,\sigma}(\tau)$. At $\tau \to 0^-$, $G_{ij,\sigma}(\tau)$ gives the density matrix in one-particle basis

$$\rho_{ij,\sigma} = \langle \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} \rangle = G_{ji,\sigma}(0^-). \tag{2.33}$$

2.5 Coherent state path integral

Many derivations covered in this thesis work in the coherent state path integral formalism. This section briefly introduces the formalism and summarizes the notations closely following Ref. [39].

The Feynman path integral [3] provides an alternative formulation for quantum mechanics. For a single particle problem, it transforms the problem of solving the Schrödinger equation of a Hamiltonian, defined in terms of the quantum mechanical operators $\hat{x}$ and $\hat{p}$, to calculating a "path integral", i.e. summing up an infinite number of probability amplitudes over all possible paths in the configuration space of the canonical coordinates (x, p), which are exactly the eigenvalues of the operators $\hat{x}$ and $\hat{p}$. In analog, to apply path integral to our electronic Hamiltonian (2.19), we need to find the "coordinates" corresponding to the conjugate pair of operators $\hat{c}$ and $\hat{c}^\dagger$.

The coherent states are defined as the eigenstates of the electron annihilation operator:

$$\hat{c}_\alpha |\eta\rangle = \eta_\alpha |\eta\rangle , \tag{2.34}$$

where $\alpha = (i, \sigma)$ is the compound index for the single particle state. The eigenvalues η_α should maintain the anti-commutation properties of $\hat{c}$:

$$\hat{c}_\alpha \hat{c}_\beta |\eta\rangle = -\hat{c}_\beta \hat{c}_\alpha |\eta\rangle \Longrightarrow \eta_\alpha \eta_\beta |\eta\rangle = -\eta_\beta \eta_\alpha |\eta\rangle ,$$

therefore

$$\{\eta_\alpha, \eta_\beta\} = 0. \tag{2.35}$$

This indicates that the eigenvalues are not common complex numbers, but rather the anti-commuting Grassmann numbers[2]. A Grassmann number η has the following properties:

- Anti-commutation

$$\eta\eta' = -\eta'\eta, \quad \eta^2 = 0. \tag{2.36}$$

- Functions of analytic functions ("Taylor expansion")

$$f(\eta) = f_0 + f_1\eta, \quad f(\eta_1, \eta_2) = f_0 + f_1\eta_1 + f_2\eta_2 + f_{12}\eta_1\eta_2, \quad \ldots \tag{2.37}$$

- Derivatives

$$\frac{\partial}{\partial\eta}1 = 0, \quad \frac{\partial}{\partial\eta}\eta = 1. \tag{2.38}$$

- Integrals

$$\int \mathrm{d}\eta\, 1 = 0, \quad \int \mathrm{d}\eta\, \eta = 1. \tag{2.39}$$

For proper mixing with the creation and annihilation operators, each one-particle state α is associated with two Grassmann numbers η_α and $\bar{\eta}_\alpha$, which are defined to be conjugates of each other:

$$\overline{\bar{\eta}_\alpha} = \eta_\alpha, \quad \overline{\bar{\eta}_\alpha\bar{\eta}_\beta} = \bar{\eta}_\beta\bar{\eta}_\alpha. \tag{2.40}$$

The electronic operators are anti-commuting with the Grassmann numbers:

$$\{\eta, \hat{c}^{(\dagger)}\} = 0, \tag{2.41}$$

[2]Or more precisely, generators of the Grassmann algebra.

and the conjugation works on both the operators and the Grassmann numbers, e.g.

$$(\eta \hat{c}_\alpha)^\dagger = \hat{c}_\alpha^\dagger \bar{\eta}. \tag{2.42}$$

It can be shown that the fermionic coherent state takes the form

$$|\eta\rangle = e^{-\sum_\alpha \eta_\alpha \hat{c}_\alpha^\dagger} |0\rangle, \tag{2.43}$$

where $|0\rangle$ is the vacuum state in the Fock space. Note that $|\eta\rangle$ no longer lives in the fermion Fock space, but rather an extended space for linear combinations of states in the Fock space with Grassmann coefficients. The coherent states form an over-complete basis of the extended Fock space, with internal products

$$\langle \eta | \eta' \rangle = e^{\sum_\alpha \bar{\eta}_\alpha \eta_\alpha}, \tag{2.44}$$

and completeness relation

$$\int \left(\prod_\alpha \mathrm{d}\bar{\eta}_\alpha \mathrm{d}\eta_\alpha \right) e^{-\sum_\alpha \bar{\eta}_\alpha \eta_\alpha} |\eta\rangle \langle \eta| = 1. \tag{2.45}$$

Using the completeness relation (2.45) to compute the trace in the fermionic partition function (2.25), we have

$$Z = \mathrm{Tr}\, e^{-\beta(\hat{H}-\mu\hat{N})} = \int \prod_\alpha \mathrm{d}\bar{\eta}_\alpha \mathrm{d}\eta_\alpha e^{-\sum_\alpha \bar{\eta}_\alpha \eta_\alpha} \langle -\eta | e^{-\beta(\hat{H}-\mu\hat{N})} |\eta\rangle, \tag{2.46}$$

where the minus sign in $\langle -\eta|$ occurs when it is moved across $|\eta\rangle$ for the right. We now break

β up into M small intervals $\epsilon = \beta/M$, such that

$$\begin{aligned}\langle -\eta | \, e^{-\beta(\hat{H}-\mu\hat{N})} \, |\eta\rangle &= \langle -\eta | \, e^{-M\epsilon(\hat{H}-\mu\hat{N})} \, |\eta\rangle \\ &= \int \Big(\prod_{k=1}^{M-1} \prod_{\alpha} \mathrm{d}\bar{\eta}_\alpha^{(k)} \mathrm{d}\eta_\alpha^{(k)} \Big) e^{-\sum_{k=1}^{M-1} \sum_\alpha \bar{\eta}_\alpha^{(k)} \eta_\alpha^{(k)}} \prod_{k=1}^{M} \left\langle \eta^{(k)} \right| e^{-\epsilon(\hat{H}-\mu\hat{N})} \left| \eta^{(k-1)} \right\rangle .\end{aligned} \tag{2.47}$$

Here we have taken the boundary condition $\left|\eta^{(0)}\right\rangle = \left|-\eta^{(M)}\right\rangle = |\eta\rangle$. The "sandwich" factors can be evaluated to coherent state eigenvalues by writing the exponential in normal order:

$$e^{-\epsilon(\hat{H}-\mu\hat{N})} =: e^{-\epsilon(\hat{H}-\mu\hat{N})} : \, + \mathcal{O}(\epsilon^2), \tag{2.48}$$

and taking the $M \to \infty$ limit, such that

$$\lim_{M\to\infty} \left\langle \eta^{(k)} \right| e^{-\epsilon(\hat{H}-\mu\hat{N})} \left| \eta^{(k-1)} \right\rangle = \lim_{M\to\infty} e^{\sum_\alpha \bar{\eta}_\alpha^{(k)} \eta_\alpha^{(k-1)}} e^{-\epsilon[H(\bar{\eta}_\alpha^{(k)}, \eta_\alpha^{(k-1)}) - \mu \sum_\alpha \bar{\eta}_\alpha^{(k)} \eta_\alpha^{(k-1)}]}, \tag{2.49}$$

where $H(\bar{\eta}, \eta)$ is the Grassmann function obtained by replacing the $\hat{c}^\dagger$, $\hat{c}$ operators from $\hat{H}$ with the Grassmann numbers $\bar{\eta}, \eta$. Now we have

$$\begin{aligned} Z = \lim_{M\to\infty} \int \Big(\prod_{k=1}^{M} \prod_{\alpha} \mathrm{d}\bar{\eta}_\alpha^{(k)} \mathrm{d}\eta_\alpha^{(k)} \Big) \\ \times \exp\Big[\sum_{k=1}^{M} \epsilon \Big(- \sum_\alpha \bar{\eta}_\alpha^{(k)} \frac{\bar{\eta}_\alpha^{(k)} - \bar{\eta}_\alpha^{(k-1)}}{\epsilon} - H(\bar{\eta}_\alpha^{(k)}, \eta_\alpha^{(k-1)}) + \mu \sum_\alpha \bar{\eta}_\alpha^{(k)} \eta_\alpha^{(k-1)} \Big) \Big]. \end{aligned} \tag{2.50}$$

At the $M \to \infty$ limit, the imaginary time axis $[0, \beta]$ is split into an infinitely dense grid $\{\tau_k = k\epsilon\}$, which allows us to formally introduce a "path" of the coherent state eigenvalues $\eta_\alpha(\tau)$, such that

$$\bar{\eta}_\alpha^{(k)} \frac{\bar{\eta}_\alpha^{(k)} - \bar{\eta}_\alpha^{(k-1)}}{\epsilon} \to \bar{\eta}_\alpha(\tau) \frac{\partial}{\partial\tau} \eta_\alpha(\tau), \quad H(\bar{\eta}_\alpha^{(k)}, \eta_\alpha^{(k-1)}) \to H(\bar{\eta}_\alpha(\tau), \eta_\alpha(\tau)). \tag{2.51}$$

The exponent in (2.50) can now be formally written as

$$\begin{aligned}
&\sum_{k=1}^{M} \epsilon \Big[-\sum_{\alpha} \bar{\eta}_{\alpha}^{(k)} \frac{\bar{\eta}_{\alpha}^{(k)} - \bar{\eta}_{\alpha}^{(k-1)}}{\epsilon} - H(\bar{\eta}_{\alpha}^{(k)}, \eta_{\alpha}^{(k-1)}) + \mu \sum_{\alpha} \bar{\eta}_{\alpha}^{(k)} \eta_{\alpha}^{(k-1)} \Big] \\
&\to \int_0^{\beta} d\tau \Big[-\sum_{\alpha} \bar{\eta}_{\alpha}(\tau) \frac{\partial}{\partial \tau} \eta_{\alpha}(\tau) - H(\bar{\eta}_{\alpha}(\tau), \eta_{\alpha}(\tau)) + \mu \sum_{\alpha} \bar{\eta}_{\alpha}(\tau) \eta_{\alpha}(\tau) \Big] \\
&= -\int_0^{\beta} d\tau \Big[\sum_{\alpha} \bar{\eta}_{\alpha}(\tau)(\partial_{\tau} - \mu) \eta_{\alpha}(\tau) + H(\bar{\eta}_{\alpha}(\tau), \eta_{\alpha}(\tau)) \Big] =: -S[\bar{\eta}, \eta],
\end{aligned}$$

where we have defined the Euclidean action S.

In later chapters of this thesis, the Grassmann fields are usually written using the same letters as the corresponding operators but without the "hat", e.g. $(\bar{\eta}, \eta) \to (\bar{c}, c)$ for $(\hat{c}^{\dagger}, \hat{c})$. With this convention, the partition function is formulated as a coherent state path integral

$$Z = \int_{c(0) = -c(\beta)} \mathscr{D}[\bar{c}, c] e^{-S[\bar{c}, c]}, \tag{2.52}$$

where the integration measure is shorthand for

$$\mathscr{D}[\bar{c}, c] = \lim_{M \to \infty} \prod_{k=1}^{M} \prod_{\alpha} \mathrm{d}\bar{c}_{\alpha}^{(k)} \mathrm{d}c_{\alpha}^{(k)}. \tag{2.53}$$

By switching to the path integral formulation, one no longer deals with the Hamiltonian and operators in the Fock space, but rather a "scalar" functional of Grassmann fields $\bar{c}(\tau), c(\tau)$ — the Euclidean action. For example, the action corresponding to the electronic Hamiltonian (2.19) writes

$$S[\bar{c}, c] = \int_0^{\beta} \mathrm{d}\tau \Big\{ \sum_{ij} \sum_{\sigma} \bar{c}_{i\sigma}(\tau) [(\partial_{\tau} - \mu)\delta_{ij} + h_{ij}] c_{j\sigma}(\tau) + \frac{1}{2} \sum_{ijkl} \sum_{\sigma\sigma'} U_{ijkl} \bar{c}_{i\sigma}(\tau) \bar{c}_{k\sigma'}(\tau) c_{l\sigma'}(\tau) c_{j\sigma}(\tau) \Big\}. \tag{2.54}$$

This simplifies mathematical operations with a simpler commutation relations from the Grassmann numbers, and enables powerful mathematical techniques such as change of integration variables, Guassian integrals, the Legendre transformation, the Hubbard-Stratonovic transformation, and so on. Although a path integral such as (2.52) is formally written as an "integral", it is not a continuous integral and should be understood only as a limit of infinite slicing of the time axis (2.50).

The partition function (2.52) formally defines a "probability distribution"

$$p[\bar{c}, c] = \frac{e^{-S[\bar{c},c]}}{Z}, \quad \int \mathscr{D}[\bar{c}, c] p[\bar{c}, c] = 1, \tag{2.55}$$

for all functionals of the Grassmann fields $\bar{c}_\alpha(\tau), c_\alpha(\tau)$. We now overload the notation $\langle \cdot \rangle$ to also represent the expectation values under this "distribution":

$$\langle f[\bar{c}, c] \rangle = \int \mathscr{D}[\bar{c}, c] f[\bar{c}, c] p[\bar{c}, c] = \frac{1}{Z} \int \mathscr{D}[\bar{c}, c] f[\bar{c}, c] e^{-S[\bar{c},c]}, \tag{2.56}$$

which is implicitly distinguished from Eq. (2.27) unless otherwise stated: here the average is taken over Grassmann fields $\bar{c}, c$ whereas in (2.27) it is taken over operators.

The construction of the path integral relies on the sequential insertion of the coherent state completeness relation (2.45) as well as normal ordering of operators (2.48). If one plugs a function of creation and annihilation operators $\hat{c}^\dagger(\tau')$ and $\hat{c}(\tau)$ into a path-integral, time-ordering will be automatically applied to components at different times, and normal ordering will be applied to components at equal time, i.e.

$$\langle f[\bar{c}(\tau'), c(\tau)] \rangle = \frac{1}{Z} \int \mathscr{D}[\bar{c}, c] f[\bar{c}(\tau'), c(\tau)] e^{-S[\bar{c},c]} \equiv \langle \mathscr{T}_\tau \, : \hat{f}[\hat{c}^\dagger(\tau'), \hat{c}(\tau)] : \rangle. \tag{2.57}$$

For example, the path integral

$$-\int \mathscr{D}[\bar{c},c]\, c(\tau)\bar{c}(\tau')e^{-S[\bar{c},c]} = -\operatorname{Tr}[\mathscr{T}_\tau \hat{c}(\tau)\hat{c}^\dagger(\tau')e^{-\beta(\hat{H}-\mu\hat{N})}] \tag{2.58}$$

allows to define the one-particle Green's function as

$$\begin{aligned} G_{ij,\sigma}(\tau,\tau') &= -\langle \mathscr{T}_\tau \hat{c}_{i\sigma}(\tau)\hat{c}^\dagger_{j\sigma}(\tau')\rangle = -\frac{1}{Z}\operatorname{Tr}[\mathscr{T}_\tau \hat{c}_{i\sigma}(\tau)\hat{c}^\dagger_{j\sigma}(\tau')e^{-\beta(\hat{H}-\mu\hat{N})}] \\ &= -\frac{1}{Z}\int \mathscr{D}[\bar{c},c]\, c_{i\sigma}(\tau)\bar{c}_{j\sigma}(\tau')e^{-S[\bar{c},c]} =: -\langle c_{i\sigma}(\tau)\bar{c}_{j\sigma}(\tau')\rangle. \end{aligned} \tag{2.59}$$

Time-ordering operators are not needed as time-ordering is automatically enforced by the path integral.

A generating function for the correlations functions of the Grassmann fields, i.e. Green's functions, can be defined in analog to the moment generating function (MGF) in probability theory:

$$M[\bar{\eta},\eta] := \left\langle \exp\Big(-\int \mathrm{d}x[\bar{\eta}(x)c(x) + \bar{c}(x)\eta(x)]\Big)\right\rangle, \tag{2.60}$$

where $\bar{\eta},\eta$ are Grassmann "source" fields, x is the compound index such that

$$x = (\tau, i, \sigma), \qquad \int \mathrm{d}x := \int_0^\beta \mathrm{d}\tau \sum_i \sum_\sigma . \tag{2.61}$$

The many-body Green's function (2.28) can be generated via functional derivatives, in the same way moments of a probability distribution are generated from the MGF [39]:

$$\begin{aligned} G^{(m)}(x_1,x_1',x_2,x_2',\dots,x_m,x_m') &= (-1)^m\langle c(x_1)\bar{c}(x_1')c(x_2)\bar{c}(x_2')\cdots c(x_m)\bar{c}(x_m')\rangle \\ &= \langle \bar{c}(x_1')\bar{c}(x_2')\cdots\bar{c}(x_m')c(x_m)\cdots c(x_2)c(x_1)\rangle \\ &= \frac{\delta^{2m}M[\bar{\eta},\eta]}{\delta\bar{\eta}(x_1)\cdots\delta\bar{\eta}(x_m)\delta\eta(x_m')\cdots\delta\eta(x_1')}\bigg|_{\bar{\eta}=\eta=0}. \end{aligned} \tag{2.62}$$

Similar analogy can be made to the cumulant generating function (CGF) by taking a logarithm of M, which gives

$$K[\bar{\eta},\eta] = \log M[\bar{\eta},\eta] = \log\left\langle \exp\Big(-\int \mathrm{d}x[\bar{\eta}(x)c(x) + \bar{c}(x)\eta(x)]\Big)\right\rangle. \tag{2.63}$$

The same functional derivatives now generates the cumulants, called the "connected" Green's functions

$$G_c^{(m)}(x_1, x_1', x_2, x_2', \dots, x_m, x_m') := \left.\frac{\delta^{2m}K[\bar{\eta},\eta]}{\delta\bar{\eta}(x_1)\cdots\delta\bar{\eta}(x_m)\delta\eta(x_m')\cdots\delta\eta(x_1')}\right|_{\bar{\eta}=\eta=0}. \tag{2.64}$$

The physical meaning of $G_c^{(m)}$ can be found in standard textbooks such as Ref. [39].

Chapter 3

Diagrammatic Expansions in Thermal Equilibrium

The interaction term in the electronic Hamiltonian (2.1) couples motions of all electrons in the system together, resulting in a many-body problem whose complexity scales exponentially versus the number of electrons. It further enters the electronic action (2.54) as a quartic term, which prevents us from evaluating the path integral (2.52) analytically as a Gaussian Grassmann integral [39]. A perturbative treatment of the many-body problem by viewing the interaction as a "perturbation" to the non-interacting problem not only provides good approximations for weakly interacting systems, but also leads to the development of diagrammatic methods which have been applied to a wide range of applications.

In this chapter, we review the fundamentals of the perturbation theory for electron interactions and diagrammatic methods, which serves as the foundation for all diagrammatic methods studied in this thesis.

3.1 Perturbation expansion of the partition function

We partition the Hamiltonian with anti-symmetrized interactions (2.24) by treating the interaction term as the perturbation, such that

$$\begin{aligned} \hat{H} &= \hat{H}_0 + \hat{H}_I, \\ \hat{H}_0 &= \sum_{ab} h_{ab} \hat{c}_a^\dagger \hat{c}_b, \\ \hat{H}_I &= \frac{1}{4} \sum_{abcd} U_{abcd} \hat{c}_a^\dagger \hat{c}_c^\dagger \hat{c}_d \hat{c}_b. \end{aligned} \tag{3.1}$$

For a better analytical control over the perturbation series, we introduce a scalar "coupling constant" $\xi \in \mathbb{C}$, and define a parametrized Hamiltonian

$$\hat{H}_\xi = \hat{H}_0 + \xi \hat{H}_I, \tag{3.2}$$

which recovers the non-interacting Hamiltonian $\hat{H}_0$ at $\xi = 0$, and the full Hamiltonian $\hat{H}$ at $\xi = 1$. The Euclidean action (2.54) becomes

$$\begin{aligned} S_\xi[\bar{c}, c] &= S_0 + \xi S_I, \\ S_0 &= \int_0^\beta \mathrm{d}\tau \sum_{ab} \bar{c}_a(\tau)[(\partial_\tau - \mu)\delta_{ab} + h_{ab}]c_b(\tau), \\ S_I &= \int_0^\beta \mathrm{d}\tau \sum_{abcd} \frac{U_{abcd}}{4} \bar{c}_a(\tau)\bar{c}_c(\tau)c_d(\tau)c_b(\tau). \end{aligned} \tag{3.3}$$

The corresponding partition function is

$$Z_\xi = \mathrm{Tr}[e^{-\beta(\hat{H}_\xi - \mu\hat{N})}] = \int \mathscr{D}[\bar{c}, c] e^{-S_0 - \xi S_I}, \tag{3.4}$$

and the thermal expected values is denoted as

$$\langle f[\bar{c},c]\rangle_\xi = \frac{1}{Z_\xi}\int \mathscr{D}[\bar{c},c] f[\bar{c},c] e^{-S_\xi}. \tag{3.5}$$

Expanding Z_ξ as a power series of ξ, we have

$$Z_\xi = \int \mathscr{D}[\bar{c},c] e^{-S_0} \sum_{k=0}^{\infty} \frac{(-S_I)^k}{k!}\xi^k = Z_0 \sum_{k=0}^{\infty} \frac{\langle(-S_I)^k\rangle_0}{k!}\xi^k, \tag{3.6}$$

where $\langle\cdot\rangle_0$ is the non-interacting thermal average. The task of the perturbation theory is to evaluate the power series order by order, and specifically, to calculate

$$\begin{aligned}\langle(-S_I)^k\rangle_0 = \int_0^\beta \mathrm{d}\tau_1 \sum_{a_1 b_1 c_1 d_1} \cdots \int_0^\beta \mathrm{d}\tau_k \sum_{a_k b_k c_k d_k} (-1)^k \frac{U_{a_1 b_1 c_1 d_1}\cdots U_{a_k b_k c_k d_k}}{4^k} \times \\ \times \langle \bar{c}_{a_1}(\tau_1)\bar{c}_{c_1}(\tau_1) c_{d_1}(\tau_1) c_{b_1}(\tau_1) \cdots \bar{c}_{a_k}(\tau_k)\bar{c}_{c_k}(\tau_k) c_{d_k}(\tau_k) c_{b_k}(\tau_k)\rangle_0.\end{aligned} \tag{3.7}$$

The expansion can be done equivalently in the Hamiltonian formalism, which is briefly summarized here and detailed in standard textbooks such as Ref. [46]. Rewrite the partition function as

$$Z_\xi = \mathrm{Tr}[e^{-\beta(\hat{H}_0-\mu\hat{N})}\hat{U}_\xi(\beta)], \quad \hat{U}_\xi(\beta) = e^{\beta(\hat{H}_0-\mu\hat{N})}e^{-\beta(\hat{H}_\xi-\mu\hat{N})}. \tag{3.8}$$

$\hat{U}_\xi(\tau)$ is the time evolution operator in interaction picture and can be expanded as a Dyson series [46]

$$\hat{U}_\xi(\beta) = \sum_{k=0}^{\infty} \frac{(-\xi)^k}{k!} \int_0^\beta \mathrm{d}\tau_1 \cdots \int_k^\beta \mathrm{d}\tau_k \mathscr{T}_\tau\{\hat{H}_I(\tau_1)\cdots\hat{H}_I(\tau_k)\}, \tag{3.9}$$

where $\hat{H}_I(\tau)$ is the interaction Hamiltonian in interaction picture. The partition function is

therefore

$$\begin{aligned} Z_\xi &= \sum_{k=0}^{\infty} \frac{(-\xi)^k}{k!} \int_0^\beta \mathrm{d}\tau_1 \cdots \int_k^\beta \mathrm{d}\tau_k \, \mathrm{Tr}[\mathscr{T}_\tau \hat{H}_I(\tau_1) \cdots \hat{H}_I(\tau_k) e^{-\beta(\hat{H}_0 - \mu\hat{N})}] \\ &= Z_0 \sum_{k=0}^{\infty} \frac{(-\xi)^k}{k!} \int_0^\beta \mathrm{d}\tau_1 \cdots \int_k^\beta \mathrm{d}\tau_k \langle \mathscr{T}_\tau \hat{H}_I(\tau_1) \cdots \hat{H}_I(\tau_k) \rangle_0. \end{aligned} \tag{3.10}$$

This is equivalent to the same expansion in action formalism (3.6).

For a finite system, the exponent $\hat{H}_\xi - \mu\hat{N}$ is bounded, and Z_ξ is analytic in $\xi \in \mathbb{C}$. Therefore, the series expansion (3.6) has infinite convergence radius and thus always converges.

3.2 Wick's theorem and Feynman diagrams

The non-interacting correlation functions of $4k$ Grassmann fields in (3.7) can be evaluated using Wick's theorem [39, 46]. Here we take an intuitive approach of introducing the theorem by interpreting the correlator as a $2k$-particle Green's function of the non-interacting electrons:

$$\langle \bar{c}(x_1')\bar{c}(y_1')c(y_1)c(x_1) \cdots \bar{c}(x_k')\bar{c}(y_k')c(y_k)c(x_k) \rangle_0 = G_0^{(2k)}(x_1, x_1', y_1, y_1', \ldots, x_k, x_k', y_k, y_k'), \tag{3.11}$$

following the definition (2.28), where x, y are compound indices defined in (2.61). To compute $G_0^{(m)}$ as a functional derivative of the generating function following (2.62), we first derive the generating function for the non-interacting system:

$$\begin{aligned} M_0[\bar{\eta}, \eta] &= \left\langle \exp\Big(- \int \mathrm{d}x [\bar{\eta}(x)c(x) + \bar{c}(x)\eta(x)] \Big) \right\rangle_0 \\ &= \frac{1}{Z_0} \int \mathscr{D}[\bar{c}, c] \exp\Big(\int \mathrm{d}x \mathrm{d}x' \bar{c}(x') g^{-1}(x', x) c(x) - \int \mathrm{d}x [\bar{\eta}(x)c(x) + \bar{c}(x)\eta(x)] \Big). \end{aligned} \tag{3.12}$$

Here the non-interacting Green's function g is defined such that

$$-\sum_b [(\partial_\tau - \mu)\delta_{ab} + h_{ab}]g_{bc}(\tau, \tau') = \delta_{ac}\delta(\tau - \tau'). \tag{3.13}$$

The path integral in (3.12) is a Gaussian integral, and can be evaluated explicitly [39]

$$\int \mathscr{D}[\bar{c}, c] e^{\int \mathrm{d}x\mathrm{d}x' \bar{c}(x')g^{-1}(x',x)c(x) - \int \mathrm{d}x[\bar{\eta}(x)c(x) + \bar{c}(x)\eta(x)]} = \det[-g^{-1}] e^{-\int \mathrm{d}x\mathrm{d}x' \bar{\eta}(x)g(x,x')\eta(x')}, \tag{3.14}$$

where the determinant is taken in the combined linear space of $x = (\tau, i, \sigma)$. The non-interacting partition function Z_0 is obtained by simply taking $\bar{\eta} = \eta = 0$:

$$Z_0 = \int \mathscr{D}[\bar{c}, c] e^{\int \mathrm{d}x\mathrm{d}x' \bar{c}(x')g^{-1}(x',x)c(x)} = \det[-g^{-1}]. \tag{3.15}$$

Therefore,

$$M_0[\bar{\eta}, \eta] = e^{-\int \mathrm{d}x\mathrm{d}x' \bar{\eta}(x)g(x,x')\eta(x')}. \tag{3.16}$$

Plugging (3.16) into (2.62) gives Wick's theorem

$$
\begin{aligned}
G_0^{(m)}(x_1, x_1', \dots, x_m, x_m') &= \frac{\delta^{2m} e^{-\int \mathrm{d}y\mathrm{d}y' \bar{\eta}(y) g(y,y') \eta(y')}}{\delta\bar{\eta}(x_1)\cdots\delta\bar{\eta}(x_m)\delta\eta(x_m')\cdots\delta\eta(x_1')}\Big|_{\bar{\eta}=\eta=0} \\
&= \frac{\delta^m}{\delta\bar{\eta}(x_1)\cdots\delta\bar{\eta}(x_m)} \int \mathrm{d}y_m \bar{\eta}(y_m) g(y_m, x_m') \cdots \int \mathrm{d}y_1 \bar{\eta}(y_1) g(y_1, x_1') \times \\
&\quad \times e^{-\int \mathrm{d}y\mathrm{d}y' \bar{\eta}(y) g(y,y') \eta(y')}\Big|_{\bar{\eta}=\eta=0} \\
&= \int \mathrm{d}y_m \cdots \int \mathrm{d}y_1 \frac{\delta^m(\bar{\eta}(y_m)\cdots\bar{\eta}(y_1))}{\delta\bar{\eta}(x_1)\cdots\delta\bar{\eta}(x_m)} g(y_1, x_1') \cdots g(y_m, x_m')\Big|_{\bar{\eta}=\eta=0} \\
&= \int \mathrm{d}y_m \cdots \int \mathrm{d}y_1 \sum_{\pi\in\mathcal{S}_m} (-1)^\pi \delta(y_1 - x_{\pi(1)}) \cdots \delta(y_m - x_{\pi(m)}) g(y_1, x_1') \cdots g(y_m, x_m')\Big|_{\bar{\eta}=\eta=0} \\
&= \sum_{\pi\in\mathcal{S}_m} (-1)^\pi g(x_{\pi(1)}, x_1') \cdots g(x_{\pi(m)}, x_m').
\end{aligned}
\tag{3.17}
$$

Here $\mathcal{S}_m$ is the permutation group of order m, and $(-1)^\pi$ is the parity of the permutation, which is the result of the fermion sign when taking the $\bar{\eta}$ functional derivatives in different orders. Defining an $m \times m$ matrix $\boldsymbol{G}(x_1, x_1', \dots, x_m, x_m')$ such that

$$
[\boldsymbol{G}(x_1, x_1', \dots, x_m, x_m')]_{ij} = g(x_i, x_j'), \tag{3.18}
$$

we have

$$
G_0^{(m)}(x_1, x_1', \dots, x_m, x_m') = \det \boldsymbol{G}(x_1, x_1', \dots, x_m, x_m'), \tag{3.19}
$$

which gives all "contractions" of the Grassmann fields [39] $\overbrace{c(x_i)\bar{c}(x_j')} = \langle c(x_i)\bar{c}(x_j')\rangle_0 = -g(x_i, x_j')$. At $m = 1$, we have

$$
G_0(x, x') = g(x, x'). \tag{3.20}
$$

Applying Wick's theorem to (3.7), we have

$$\langle(-S_I)^k\rangle_0 = \int_0^\beta \mathrm{d}\tau_1 \sum_{a_1 b_1 c_1 d_1} \cdots \int_0^\beta \mathrm{d}\tau_k \sum_{a_k b_k c_k d_k} (-1)^k \frac{U_{a_1 b_1 c_1 d_1} \cdots U_{a_k b_k c_k d_k}}{4^k} \det \boldsymbol{G}, \tag{3.21}$$

where $\boldsymbol{G}$ is a $2k \times 2k$ matrix, which can be written in a block form as

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{g}_{11} & \boldsymbol{g}_{12} & \cdots & \boldsymbol{g}_{1k} \\ \boldsymbol{g}_{11} & \boldsymbol{g}_{12} & \cdots & \boldsymbol{g}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_{k1} & \boldsymbol{g}_{k2} & \cdots & \boldsymbol{g}_{kk} \end{bmatrix}, \quad \boldsymbol{g}_{mn} = \begin{bmatrix} g_{b_m a_n}(\tau_m, \tau_n) & g_{b_m c_n}(\tau_m, \tau_n) \\ g_{d_m a_n}(\tau_m, \tau_n) & g_{d_m c_n}(\tau_m, \tau_n) \end{bmatrix}. \tag{3.22}$$

Since the determinant in (3.21) effectively generates all possible contractions of Grassmann numbers, each of which carries indices (τ_n, a_n) that connect to an interaction tensor element $U_{a_n b_n c_n d_n}$, we can represent the integrand of (3.21) graphically as Feynman diagrams [39] where interaction "vertices" are connected by "propagator" lines. Defining compound "vertex indices" $\{\nu_1, \ldots, \nu_k\}$ such that

$$\nu_i = (\tau_i, a_i, b_i, c_i, d_i,), \quad U(\nu_i) = U_{a_i b_i c_i d_i}, \quad \int \mathrm{d}\nu_i = \int_0^\beta \mathrm{d}\tau_i \sum_{a_i b_i c_i d_i}, \tag{3.23}$$

we can rewrite (3.21) as

$$\langle(-S_I)^k\rangle_0 = \int \mathrm{d}\nu_1 \cdots \int \mathrm{d}\nu_k \, (-1)^k \frac{U(\nu_1) \cdots U(\nu_k)}{4^k} \det \boldsymbol{G}(\nu_1, \ldots, \nu_k), \tag{3.24}$$

where $\boldsymbol{G}(\nu_1, \ldots, \nu_k)$ is the same as $\boldsymbol{G}$ defined in (3.22). Diagrams contributing to $\langle(-S_I)^k\rangle_0$ are constructed according to the following Feynman rules [39] in graph theory terms:

1. Each diagram is a directed graph (i.e. a digraph) with k four-point vertices.

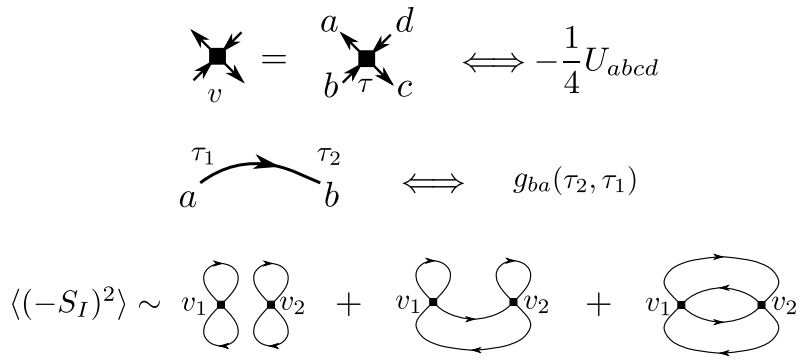

Figure 3.1: Example of Feynman diagrams contributing to Z at $k = 2$.

2. Each vertex carries a compound index ν_n and a value $-U(\nu_n)/4$. The indegree and outdegree of the vertex are both 2.

3. Vertices are connected by $2k$ propagator lines, each carries a value $g(x_m, x'_n)$ where x_m is the indices on the leg of the "target" vertex, and x'_n is the indices on the leg of the "source" vertex.

4. The whole diagram carries an overall sign of $(-1)^l$, where l is the number of fermion loops in the diagram.

5. Contribution of the diagram to $\langle(-S_I)^k\rangle_0$ is obtained by integrating over all internal indices $\nu_1, \ldots, \nu_k$.

Figure 3.1 shows examples of these diagrams.

At order k of the expansion (3.6), the determinant of the $2k \times 2k$ matrix $\boldsymbol{G}$ generates a total of $(2k)!$ different contractions, which in turn result in $(2k)!$ Feynman diagrams. Numerical enumeration of individual diagrams, either explicitly or stochastically, is expensive at high expansion orders due to such a factorial growth. In contrast, numerical evaluation of the determinant $\det \boldsymbol{G}$ can be performed in $\mathcal{O}(k^3)$ time via linear algebra decompositions, which significantly reduces the computational time and enables practical calculation of high-order

expansions. This approach is adopted in continuous time quantum Monte Carlo (CT-QMC) algorithms [9–12, 14], and serves as building blocks of the connected determinant (CDet) algorithm for diagrammatic Monte Carlo (DiagMC) [22–24, 47].

3.3 Expansion of intensive quantities

Since the partition function scales exponentially with system size, most physical observables that are intensive or extensive quantities directly relates to the logarithm of Z. For example, the grand potential

$$\Omega = -\frac{1}{\beta}\log Z, \tag{3.25}$$

and the internal energy

$$E = -\frac{\partial}{\partial\beta}\log Z. \tag{3.26}$$

By the linked cluster theorem [39], the logarithm effectively removes all disconnected Feynman diagrams, i.e.

$$\begin{aligned}
\frac{Z_\xi}{Z_0} &= \sum_{k=0}^{\infty}\frac{1}{k!}\sum\{\text{All closed Feynman diagrams with } k \text{ vertices}\},\\
\log\frac{Z_\xi}{Z_0} &= -\beta(\Omega_\xi - \Omega_0)\\
&= \sum_{k=1}^{\infty}\frac{1}{k!}\sum\{\text{All } \textbf{connected} \text{ closed Feynman diagrams with } k \text{ vertices}\},
\end{aligned} \tag{3.27}$$

where the diagrams are constructed following the rules described in the previous section.

The series expansion of the one-particle Green's function

$$G_\xi(x,x') = -\langle c(x)\bar{c}(x')\rangle_\xi = -\frac{1}{Z_\xi}\int\mathscr{D}[\bar{c},c]c(x)\bar{c}(x')e^{-S_\xi[\bar{c},c]} \tag{3.28}$$

is obtained by the ratio of two power series of ξ. The denominator is just (3.6) the expansion of Z_ξ. The numerator

$$-\int \mathscr{D}[\bar{c},c]c(x)\bar{c}(x')e^{-S_\xi[\bar{c},c]} = Z_0\sum_{k=0}^{\infty}\frac{\xi^k}{k!}\langle -c(x)\bar{c}(x')(-S_I)^k\rangle_0, \tag{3.29}$$

where the expected value can be evaluated with Wick's theorem:

$$\begin{aligned}\langle -c(x)\bar{c}(x')S_I^k\rangle_0 = &\int_0^\beta \mathrm{d}\tau_1 \sum_{a_1b_1c_1d_1}\cdots\int_0^\beta \mathrm{d}\tau_k\sum_{a_kb_kc_kd_k}\frac{U_{a_1b_1c_1d_1}\cdots U_{a_kb_kc_kd_k}}{4^k}\times\\ &\times\det\begin{bmatrix}\boldsymbol{G} & \boldsymbol{g}(x')\\ \boldsymbol{g}'(x)^T & g(x,x')\end{bmatrix},\end{aligned} \tag{3.30}$$

where $\boldsymbol{g}(x')$ and $\boldsymbol{g}'(x)$ are column vectors of dimension $2k$:

$$[\boldsymbol{g}(x')]_n = g(x_n,x'),\quad [\boldsymbol{g}'(x)]_n = g(x,x_n'). \tag{3.31}$$

Diagrammatically, the existence of the two "external" Grassmann fields $c(x)$ and $\bar{c}(x')$ adds two external "legs" to the diagrams. The determinant generates all possible connections of the external legs and vertices, resulting in both connected and disconnected diagram topologies. In consequence, the ratio between two diagrammatic series

$$G_\xi(x,x') = \frac{Z_0\displaystyle\sum_{k=0}^{\infty}\frac{(-\xi)^k}{k!}\sum\left\{\begin{array}{l}\text{All diagrams with } k \text{ vertices and 2 external legs}\\ \text{including disconnected "bubbles"}\end{array}\right\}}{Z_0\displaystyle\sum_{k=0}^{\infty}\frac{(-\xi)^k}{k!}\sum\{\text{All "bubble" diagrams with } k \text{ vertices}\}} \tag{3.32}$$

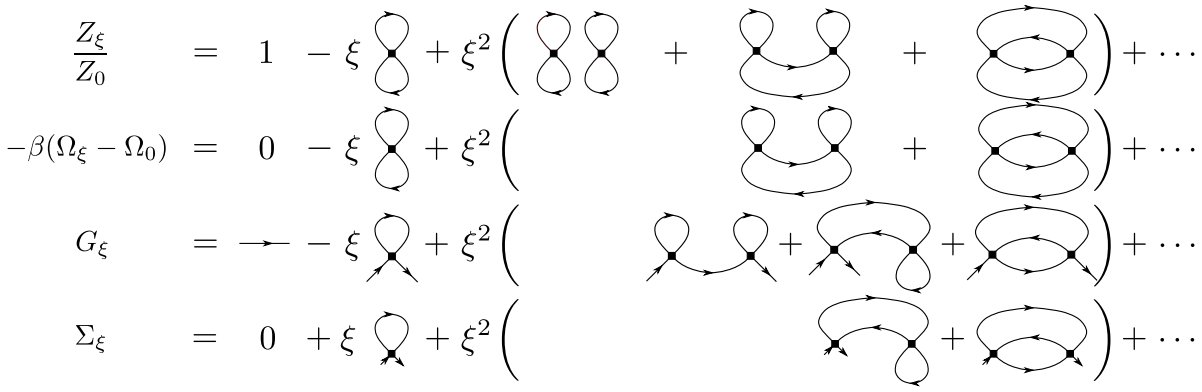

Figure 3.2: Low order diagram topologies in expansions for Z_ξ, Ω_ξ, G_ξ, and Σ_ξ.

effectively cancels all disconnected components, and thus

$$G_\xi(x, x') = \sum_{k=0}^{\infty} \frac{(-\xi)^k}{k!} \sum \{\text{All } \textbf{connected} \text{ diagrams with } k \text{ vertices and 2 external legs}\}. \tag{3.33}$$

Figure 3.2 includes examples of low-order diagram topologies in expansions of Z_ξ, Ω_ξ, and G_ξ.

The logarithm in (3.25) and the inverse Z_ξ in (3.28) will generally lead to poles in the complex plane of ξ. The series expansions of Ω_ξ and G_ξ therefore have finite convergence radii. Since the physical solution of the system is obtained at $\xi = 1$, a convergence radius smaller than 1 will lead to a divergent perturbation series for the physical quantities, which typically occurs when the electron-electron interaction is strong and $\hat{H}_0$ no longer serves as a good approximation for $\hat{H}$.

3.4 Self-energy and Dyson equation

The self-energy Σ describes the effect of electron-electron interaction on the single-particle properties [39, 46]. It is defined by the Dyson equation

$$G_{ab}(\mathrm{i}\omega_n) = g_{ab}(\mathrm{i}\omega_n) + \sum_{cd} g_{ac}(\mathrm{i}\omega_n)\Sigma_{cd}(\mathrm{i}\omega_n)G_{db}(\mathrm{i}\omega_n), \tag{3.34}$$

where we have transformed to the Matsubara frequency representation [39] via the Fourier transform[3]

$$f(\mathrm{i}\omega_n) = \int_0^\beta \mathrm{d}\tau\, f(\tau)e^{\mathrm{i}\omega_n\tau}. \tag{3.35}$$

When transformed back to the imaginary time representation, products in frequency become convolutions in time:

$$G(x, x') = g(x, x') + \int \mathrm{d}y\mathrm{d}y'\, g(x, y')\Sigma(y', y)G(y, x'). \tag{3.36}$$

One can formally insert the Dyson equation back to itself recursively and obtain a geometric series:

$$G = g + g\Sigma g + g\Sigma g\Sigma g + \cdots. \tag{3.37}$$

Diagrammatically, the self-energy can be expanded in terms of one-particle irreducible (1PI) diagrams with external legs amputated [39, 46]. Definitions for the interaction vertices and propagators are the same as in diagrams for Z or G. 1PI means that the diagram cannot be separated into disconnected components by removing any one propagator line, or in the graph theory language, that the edge connectivity of the graph is greater than one. The

[3] As a sanity check, $G(\tau) = -\langle c(\tau)\bar{c}(0)\rangle$ is a dimensionless quantity, the dimension of its Fourier transform $G(\mathrm{i}\omega_n)$ is inverse energy, therefore $\Sigma(\mathrm{i}\omega_n)$ has the dimension of energy, and $\Sigma(\tau)$ in turn has the dimension of squared energy.

bottom panel of Fig. 3.2 shows examples of Σ diagrams. Note that this is different from the skeleton diagrams of the self-energy, which is introduced in Chapter 5, where the propagator lines represent the full Green's function G. Diagrams described in this section is therefore also referred to as "bare" diagrams of Σ.

3.5 Shifted action formalism

In systems where electron interaction is strong, the perturbation series for G_ξ may converge very slowly or even diverge. One way of circumventing this problem is to change the partition of the "non-interacting" problem and the interacting "perturbation", such that the series is constructed around a starting point that is closer to the physical solution. This approach is often referred to as "shifted action" method in the context of DiagMC [48, 49], or the "α-shift" in the context of CT-QMC [9].

To change the perturbation starting point of the action (3.3), a "shift action" term can be added to S_0 and subtracted from S_I, such that

$$\tilde{S}_0 = S_0 + \Delta S, \quad \tilde{S}_I = S_I - \Delta S. \tag{3.38}$$

The perturbation theory is redefined by reintroducing the coupling constant ξ such that

$$\tilde{S}_\xi = \tilde{S}_0 + \xi \tilde{S}_I = (S_0 + \Delta S) + \xi(S_I - \Delta S), \tag{3.39}$$

which still recovers the physical action at $\xi = 1$.

So far the shift action ΔS is a generic functional of the Grassmann fields $\bar{c}, c$. For a perturbation theory to be practical, it is necessary that the new "non-interacting" problem is

solvable analytically, i.e. the path integral

$$\tilde{Z}_0 = \int \mathscr{D}[\bar{c}, c] e^{-\tilde{S}_0[\bar{c},c]} \tag{3.40}$$

is a Gaussian integral. This requires the ΔS to be defined in a bilinear form:

$$\Delta S = \int \mathrm{d}x\mathrm{d}x' \bar{c}(x')\alpha(x', x)c(x), \tag{3.41}$$

where α is a two point function whose explicit form writes

$$\alpha(x', x) = \alpha_{a',a}(\tau', \tau). \tag{3.42}$$

The shifted non-interacting action now also a bilinear form

$$\tilde{S}_0 = -\int \mathrm{d}x\mathrm{d}x' \bar{c}(x')[g^{-1}(x', x) - \alpha(x', x)]c(x). \tag{3.43}$$

Defining the shifted non-interacting propagator $\tilde{g}$ such that

$$\tilde{g}^{-1} = g^{-1} - \alpha, \tag{3.44}$$

we see that

$$\tilde{g} = (g^{-1} - \alpha)^{-1} \Rightarrow \tilde{g} = g + g\alpha\tilde{g}, \tag{3.45}$$

which implies that $\tilde{g}$ is obtained from the Dyson equation (3.34) with the self-energy defined as α.

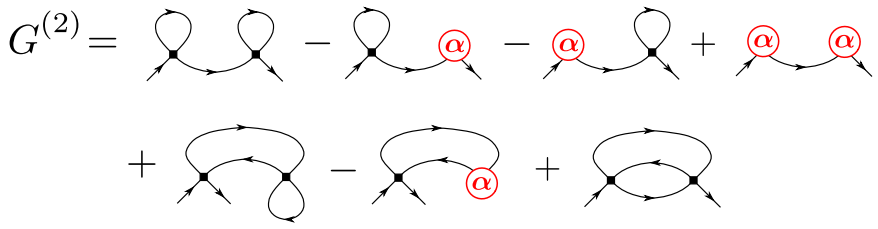

Figure 3.3: Second order Green's function diagrams from the shifted action expansion.

Perturbation expansion of $\tilde{Z}_\xi$ is obtained in the same way as in (3.6):

$$\tilde{Z}_\xi = \tilde{Z}_0 \sum_{k=0}^{\infty} \frac{\langle (-\tilde{S}_I)^k \rangle_0}{k!} \xi^k. \tag{3.46}$$

Note that the non-interacting expected value $\langle \cdot \rangle_0$ is now defined in terms of $\tilde{S}_0$. The shifted interacting action

$$\tilde{S}_I = \int_0^\beta \mathrm{d}\tau \sum_{abcd} \frac{U_{abcd}}{4} \bar{c}_a(\tau)\bar{c}_c(\tau)c_d(\tau)c_b(\tau) - \int_0^\beta \mathrm{d}\tau \mathrm{d}\tau' \sum_{ab} \bar{c}_a(\tau')\alpha_{a,b}(\tau',\tau)c_b(\tau) \tag{3.47}$$

now contains an additional bilinear term (often referred to as the counter term), which enters the diagrammatic series as a two-point vertex [50].

The shifted action leads to the following changes to the diagram rules in expansions of the partition function, grand potential, or Green's function:

1. All propagator lines now represent $\tilde{g}$.

2. At order k, each diagram still contain k vertices. Each vertex now can either be a four-point vertex representing $-U_{abcd}/4$, or a two-point vertex representing $\alpha(x', x)$.

All other rules and connectivity requirements are unchanged. Figure 3.3 shows an example of second order Green's function diagrams from a shifted action expansion, where red circles indicate the counter term α, each introduces a factor of -1.

Chapter 4

Monte Carlo Integration of Perturbation Series

Mathematically, evaluating a diagrammatic series is simply calculating a sum of some integrals over vertex indices: imaginary time, orbital, and spin. Direct numerical integration of diagrams using deterministic methods is only practical for low-order diagrams, and becomes prohibitively expensive at higher orders due to the "curse of dimensionality" [51].

Monte Carlo integration, thanks to the universal central limit theorem (CLT), converges in a $\mathcal{O}(N^{-1/2})$ scaling in terms of the number of samples N, regardless of the dimensionality of the integral. Hence, Monte Carlo methods have become a predominant choice for computing high-order Feynman diagrams.

In this chapter, we briefly review the Markov chain Monte Carlo method and introduce the basics of Monte Carlo integration of perturbation series introduced in Chapter 3. Concepts and notations introduced in this chapter is applied to methods and applications in subsequent chapters.

4.1 Monte Carlo integration and importance sampling

We start with the simplest integral of an arbitrary continuous function $f(x)$ in a interval $[x_0, x_1]$:

$$A = \int_{x_0}^{x_1} \mathrm{d}x f(x). \tag{4.1}$$

Suppose a probability density function (PDF) $p(x)$ is defined for a random variable X in the same interval, such that

$$\int_{x_0}^{x_1} \mathrm{d}x p(x) = 1, \quad p(x) > 0. \tag{4.2}$$

The integral A can then be understood as the expected value of the function $f(X)/p(X)$, i.e.

$$A = \mathrm{E}\Big[\frac{f(X)}{p(X)}\Big] = \int_{x_0}^{x_1} \mathrm{d}x \frac{f(x)}{p(x)} p(x) = \int_{x_0}^{x_1} \mathrm{d}x f(x). \tag{4.3}$$

Numerical Monte Carlo integration of A is carried out in the following steps:

1. Generate N Monte Carlo samples $X_1, \ldots, X_N$ which follows the same distribution $p(x)$.

2. For each Monte Carlo sample, calculate the function

$$a_i = a(X_i) = \frac{f(X_i)}{p(X_i)}. \tag{4.4}$$

3. Compute the Monte Carlo average as an estimate of the expected value I:

$$\langle a \rangle_p = \frac{1}{N} \sum_{i=1}^{N} a_i, \quad \mathrm{E}[\langle a \rangle_p] = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}[a_i] = A. \tag{4.5}$$

The variance of the Monte Carlo average $\langle a \rangle_p$ is related to $\mathrm{Var}[a]$ by a factor of N:

$$\mathrm{Var}[\langle a \rangle_p] = \frac{1}{N} \mathrm{Var}[a]. \tag{4.6}$$

Assuming the Monte Carlo samples are independent[4], the $\mathrm{Var}[a]$ can be estimated as the sample variance

$$\sigma_a^2 = \frac{1}{N-1} \sum_{i=1}^{N} (a_i - \langle a \rangle_p), \quad \mathrm{E}[\sigma_a^2] = \mathrm{Var}[I]. \tag{4.7}$$

The estimator for the root-mean-square error of $\langle a \rangle_p$ is therefore

$$\Delta \langle a \rangle_p = \sqrt{\frac{\sigma_a^2}{N}} \xrightarrow{N \to \infty} \sqrt{\frac{Var[a]}{N}}, \tag{4.8}$$

which has an asymptotic scaling of $\mathcal{O}(N^{-1/2})$ at large N.

To analyze the impact of the distribution $p(x)$ on the variance, we rewrite the variance of A as

$$\begin{aligned} \mathrm{Var}[a] &= \mathrm{E}[a^2] - \mathrm{E}[a]^2 = \mathrm{E}[|a|^2] - \mathrm{E}[a]^2 \\ &= \mathrm{Var}[|a|] + \mathrm{E}[|a|]^2 - \mathrm{E}[a]^2, \end{aligned} \tag{4.9}$$

in which $\mathrm{Var}[|a|]$ can be minimized by choosing $p(x)$ to be proportional to $|f(x)|$, i.e.

$$p^\star(x) = \frac{|f(x)|}{\int_{x_0}^{x_1} \mathrm{d}y |f(y)|} \tag{4.10}$$

such that

$$|a(x)| = \int_{x_0}^{x_1} \mathrm{d}y |f(y)| \tag{4.11}$$

[4]This is usually not true in practice, such as in Markov chain Monte Carlo, in which case on has to take into account the correlations between samples.

becomes a constant, and thus $\mathrm{Var}[|a|] = 0$. Since $\mathrm{E}[a] = A$ and $\mathrm{E}[|a|] = \int_{x_0}^{x_1} \mathrm{d}x|f(x)|$ are independent of $p(x)$, we have

$$\min_p \mathrm{Var}[a] = E[|a|]^2 - E[a]^2 = E[|a|]^2\left[1 - \left(\frac{E[a]}{E[|a|]}\right)^2\right], \tag{4.12}$$

where

$$\frac{E[a]}{E[|a|]} = \frac{\int_{x_0}^{x_1} \mathrm{d}x f(x)}{\int_{x_0}^{x_1} \mathrm{d}x|f(x)|} = \int_{x_0}^{x_1} \mathrm{d}x\, \mathrm{sgn}(f(x))p^\star(x) = \mathrm{E}[\mathrm{sgn}(f(X))]. \tag{4.13}$$

If $f(x)$ does not change sign in the whole interval, the variance is reduced to zero; otherwise, the variance effectively depends on how rapidly the function $f(x)$ changes sign.

From this simple 1-D example, we can summarize the following properties of the Monte Carlo integration method:

1. The integral is transformed to numerical sampling of the ratio $f(x)/p(x)$ for some probability distribution $p(x)$

2. The probability distribution $p(x)$ should be chosen such that

   a) $f(x)/p(x)$ can be computed numerically, and

   b) samples X_i can be generated following the distribution $p(x)$.

3. The stochastic error scales as $\mathcal{O}(N^{-1/2})$ for large N.

4. For the same N, the error is minimized by taking $p(x) \propto |f(x)|$.

5. The minimal error depends on the sign of the integrand $f(x)$: the error is larger if $f(x)$ changes sign more frequently.

By taking the Monte Carlo PDF to follow $|f(x)|$, more samples will be generated in regions where $f(x)$ contributes more significantly to the integral than those where $f(x)$ is almost

zero. This technique is referred to as **importance sampling** for Monte Carlo integration.

The Monte Carlo integration procedure can be generalized to integrate functions of elements ω in an arbitrary set Ω with a (discrete or continuous) integration measure $\mathrm{d}\mu(\omega)$, e.g.

$$A = \int_\Omega \mathrm{d}\mu(\omega) f(\omega). \tag{4.14}$$

A PDF $p(\omega)$ is similarly defined such that

$$\int_\Omega \mathrm{d}\mu(\omega) p(\omega) = 1, \quad p(\omega) > 0. \tag{4.15}$$

Monte Carlo integration for A is thus obtained by generating samples ω_i following the PDF $p(\omega)$, and measuring the average of the ratio f/p.

4.2 Markov chain Monte Carlo and Metropolis-Hastings algorithm

When computing the integral

$$\int_\Omega \mathrm{d}\mu(\omega) f(\omega) \tag{4.16}$$

using Monte Carlo integration with the optimal PDF

$$p(\omega) = \frac{|f(\omega)|}{W}, \quad W = \int_\Omega \mathrm{d}\mu(\omega) |f(\omega)|, \tag{4.17}$$

it is often difficult to generate samples following this distribution via a direct inversion of the cumulative distribution function (CDF) due to the complexity of the integrand $f(\omega)$. Moreover, the normalization factor W is often unknown, making explicit computations of

$p(\omega)$ impossible. Markov chain Monte Carlo provides a way to generate samples following a given distribution knowing only ratios of the PDF for two different samples, which is ideal for integrating functions such as high-order diagrammatic series.

A Markov chain is a sequence of random variables $X_1, X_2, \ldots$ where the distribution of each sample only depend on the previous one, i.e.

$$P(X_{i+1} = \omega) = \int_\Omega \mathrm{d}\mu(\omega') w(\omega|\omega') P(X_i = \omega'), \tag{4.18}$$

where the transition probability

$$w(\omega|\omega') = P(X_{i+1} = \omega | X_i = \omega') \tag{4.19}$$

has been chosen to be time-homogeneous, i.e. independent of the sample index i. The Markov chain is **ergodic** if any $\omega \in \Omega$ can be reached with finite probability in a finite number of steps starting from any other ω_0. A distribution $p(\omega)$ is the stationary distribution if it remains unchanged after a Markov step, i.e. the following balance equation is satisfied:

$$p(\omega) = \int_\Omega \mathrm{d}\mu(\omega') w(\omega|\omega') p(\omega'). \tag{4.20}$$

A stronger condition named **detailed balance** is commonly applied:

$$w(\omega|\omega') p(\omega') = w(\omega'|\omega) p(\omega). \tag{4.21}$$

For Monte Carlo integration, we always require that the Markov chain ensures ergodicity and detailed balance.

The Metropolis-Hastings algorithm allows us to design a Markov chain for a given sta-

tionary distribution $p(\omega)$. From each configuration ω_i, a new configuration ω_j is proposed following some proposal probability distribution $w^{\text{prop}}(\omega_j|\omega_i)$. To ensure detailed balance, an acceptance ratio R is calculated after each proposal as

$$R(\omega_j|\omega_i) = \frac{w^{\text{prop}}(\omega_i|\omega_j)p(\omega_j)}{w^{\text{prop}}(\omega_j|\omega_i)p(\omega_i)}. \tag{4.22}$$

The proposal $\omega_i \to \omega_j$ is accepted with probability

$$w^{\text{acc}}(\omega_j|\omega_i) = \min(1, R(\omega_j|\omega_i)). \tag{4.23}$$

This ensures the detailed balance of the Markov process, i.e.

$$w(\omega_j|\omega_i)p(\omega_i) = w(\omega_i|\omega_j)p(\omega_j), \tag{4.24}$$

where

$$w(\omega_j|\omega_i) = w^{\text{acc}}(\omega_j|\omega_i)w^{\text{prop}}(\omega_j|\omega_i), \tag{4.25}$$

which guarantees that samples obtain the equilibrium distribution $p(\omega)$ after thermalization.

Even after the equilibrium distribution is reached, samples generated from a Markov chain are not independent but have a finite autocorrelation time. A naïve variance estimate following (4.7) is thus biased. To obtain a correct error estimate, one should either measure the autocorrelation time of the Markov chain, or employ statistical resampling methods such as the jackknife or the bootstrap [51].

4.3 Monte Carlo integration of diagrammatic series

Diagrammatic expansions introduced in Chapter 3 share the following generic form:

$$Q = \sum_{k=0}^{\infty} \frac{1}{k!} \int \mathrm{d}\nu_1 \cdots \int \mathrm{d}\nu_k \sum_{t \in \mathscr{T}} d(\nu_1, \nu_2, \ldots, \nu_k; t), \tag{4.26}$$

where Q is the target physical quantity (Z, $G(x, x')$, etc.), $\nu_1, \ldots, \nu_k$ are compound indices associated to each vertex, $\mathscr{T}$ is the set of diagram topologies at given vertex configuration following specific diagram rules for Q, and d is the value of the diagram. To calculate the diagram using Monte Carlo integration, we first need to properly define the sampling space Ω and the integration measure $\mu(\omega)$. Traditionally, DiagMC methods perform Monte Carlo integrations over individual diagrams, such that each sample $\omega = (\nu_1, \nu_2, \ldots, \nu_k, t)$ includes the diagram topology t explicitly [17–21]. Alternatively, one can sum up all diagram topologies at a given vertex configuration explicitly:

$$D(\nu_1, \ldots, \nu_k) = \sum_{t \in \mathscr{T}} d(\nu_1, \nu_2, \ldots, \nu_k; t) \tag{4.27}$$

and compute the following series using Monte Carlo integration:

$$Q = \sum_{k=0}^{k_{\max}} \frac{1}{k!} \int \mathrm{d}\nu_1 \cdots \int \mathrm{d}\nu_k D(\nu_1, \nu_2, \ldots, \nu_k), \tag{4.28}$$

where a finite order truncation $k_{\max}$ is introduced. This approach is adopted in CT-QMC methods [9, 11, 12, 14] and the CDet approach for DiagMC [22–24, 47]. We will not discuss the first approach in this thesis and will only focus on the second approach.

To perform the Monte Carlo integration, one needs to define the sampling space Ω and the corresponding integration measure. For (4.28), two different definitions of the sampling

space are commonly used:

I. Each sample is a sequence of vertices, i.e. $\omega^{\mathrm{I}} = [v_1, v_2, \dots, v_k]$, such that exchanging two vertices leads to a different sample. The sampling space is formally given by

$$\Omega_{\mathrm{I}} = \bigoplus_{k=0}^{k_{\max}} \mathscr{V}^k, \tag{4.29}$$

where $\mathscr{V}$ is the space for the compound index v_i. The integration measure is given by

$$\int \mathrm{d}\mu(\omega^{\mathrm{I}}) = \sum_{k=0}^{k_{\max}} \int \mathrm{d}v_1 \cdots \int \mathrm{d}v_k. \tag{4.30}$$

II. Each sample is a set of vertices, i.e. $\omega^{\mathrm{II}} = (v_1, v_2, \dots, v_k)$, such that any permutations of the vertices result in the same sample. The sampling space is formally given by

$$\Omega_{\mathrm{II}} = \bigoplus_{k=0}^{k_{\max}} \frac{1}{k!} \mathscr{V}^k, \tag{4.31}$$

where the factor $1/k!$ accounts for the invariance under permutation. The integration measure is given by

$$\int \mathrm{d}\mu(\omega^{\mathrm{II}}) = \sum_{k=0}^{k_{\max}} \frac{1}{k!} \int \mathrm{d}v_1 \cdots \int \mathrm{d}v_k. \tag{4.32}$$

The integral (4.28) is thus rewritten under the two definitions as

$$\begin{aligned} Q &= \int_{\Omega_{\mathrm{I}}} \mathrm{d}\mu(\omega^{\mathrm{I}}) \, \frac{D(\omega^{\mathrm{I}})}{k!}, \quad (k = |\omega^{\mathrm{I}}|) \\ &= \int_{\Omega_{\mathrm{II}}} \mathrm{d}\mu(\omega^{\mathrm{II}}) \, D(\omega^{\mathrm{II}}). \quad (k = |\omega^{\mathrm{II}}|) \end{aligned} \tag{4.33}$$

Here $k = |\omega^{\mathrm{I/II}}|$ gives the number of vertices in each sample.

Following Eq. (4.10), the PDF of the stationary distribution is defined proportional to the absolute value of the integrand to minimize variance:

$$p(\omega^{\mathrm{I}}) = \frac{1}{W}\frac{|D(\omega^{\mathrm{I}})|}{|\omega^{\mathrm{I}}|!}, \quad p(\omega^{\mathrm{II}}) = \frac{1}{W}|D(\omega^{\mathrm{II}})|, \tag{4.34}$$

where the normalization factor

$$W = \int_{\Omega_{\mathrm{I}}} \mathrm{d}\mu(\omega^{\mathrm{I}})\,\frac{|D(\omega^{\mathrm{I}})|}{k!} = \int_{\Omega_{\mathrm{II}}} \mathrm{d}\mu(\omega^{\mathrm{II}})\,|D(\omega^{\mathrm{II}})| = \sum_{k=0}^{k_{\max}} \frac{(-1)^k}{k!} \int \mathrm{d}v_1 \cdots \int \mathrm{d}v_k |D(v_1, v_2, \ldots, v_k)| \tag{4.35}$$

is the same in both definitions. For each sample, the following quantities are measured to compute the final integral Q:

$$\begin{aligned} Q &= \left\langle \frac{D(\omega^{\mathrm{I}})/(k!)}{|D(\omega^{\mathrm{I}})|/(k!W)} \right\rangle_{\mathrm{I}} = W\langle \mathrm{sgn}(D(\omega^{\mathrm{I}}))\rangle_{\mathrm{I}}, \\ Q &= \left\langle \frac{D(\omega^{\mathrm{II}})}{|D(\omega^{\mathrm{II}})|/W} \right\rangle_{\mathrm{II}} = W\langle \mathrm{sgn}(D(\omega^{\mathrm{II}}))\rangle_{\mathrm{II}}. \end{aligned} \tag{4.36}$$

We see that the expression for measuring Q is the same from both definitions even though the sampling space are defined differently. In fact, the two definitions are equivalent in terms of Monte Carlo sampling. Each sample of definition II corresponds to $k!$ samples of definition I, which are compensated by a $1/k!$ prefactor in the PDF in definition I. The total probability of the $k!$ samples of definition I is therefore the same as the single sample of definition II.

After the stationary distribution p is determined, we can generate Monte Carlo samples in a Markov chain using Metropolis-Hastings algorithm. Monte Carlo updates are designed differently in different methods and applications to achieve better sampling efficiency. Here we discuss two conjugating updates that are commonly used: the insertion and the removal of a vertex.

We first focus on definition I of the sampling space. When proposing an insertion of a new vertex to an existing sample of k vertices in a sequence, the proposal probability is derived as follows. First propose a random vertex with uniform probability $|\mathscr{V}|^{-1}$, and then insert the vertex into the $k+1$ possible locations of the sequence with probability $1/(k+1)$. The proposal probability for insertion is therefore

$$w_{\mathrm{I}}^{\mathrm{insert}} = \frac{1}{|\mathscr{V}|(k+1)}. \tag{4.37}$$

The inverse update removes a random vertex from $k+1$ vertices with probability $1/(k+1)$, and the proposal probability is

$$w_{\mathrm{I}}^{\mathrm{remove}} = \frac{1}{(k+1)}. \tag{4.38}$$

The acceptance ratio (4.22) for the insertion update is thus

$$\begin{aligned} R_{\mathrm{I}}^{\mathrm{insert}} &= \frac{w_{\mathrm{I}}^{\mathrm{remove}} p(\omega_{\mathrm{to}}^{\mathrm{I}}, |\omega_{\mathrm{to}}^{\mathrm{I}}| = k+1)}{w_{\mathrm{I}}^{\mathrm{insert}} p(\omega_{\mathrm{from}}^{\mathrm{I}}, |\omega_{\mathrm{from}}^{\mathrm{I}}| = k)} = \frac{1/(k+1)}{1/[|\mathscr{V}|(k+1)]} \frac{|D(\omega_{\mathrm{to}}^{\mathrm{I}})|/((k+1)!W)}{|D(\omega_{\mathrm{from}}^{\mathrm{I}})|/(k!W)} \\ &= \frac{|\mathscr{V}|}{k+1} \frac{|D(\omega_{\mathrm{to}}^{\mathrm{I}})|}{|D(\omega_{\mathrm{from}}^{\mathrm{I}})|}. \end{aligned} \tag{4.39}$$

For definition II of the sampling space, the insertion to an existing set of k vertices only requires proposing a random vertex with probability $1/|\mathscr{V}|$. The inverse move randomly removes a vertex from $k+1$ vertices with probability $1/(1+k)$. The acceptance ratio is therefore

$$\begin{aligned} R_{\mathrm{II}}^{\mathrm{insert}} &= \frac{w_{\mathrm{II}}^{\mathrm{remove}} p(\omega_{\mathrm{to}}^{\mathrm{II}}, |\omega_{\mathrm{to}}^{\mathrm{II}}| = k+1)}{w_{\mathrm{II}}^{\mathrm{insert}} p(\omega_{\mathrm{from}}^{\mathrm{II}}, |\omega_{\mathrm{from}}^{\mathrm{II}}| = k)} = \frac{1/(k+1)}{1/|\mathscr{V}|} \frac{|D(\omega_{\mathrm{to}}^{\mathrm{II}})|/W}{|D(\omega_{\mathrm{from}}^{\mathrm{II}})|/(W)} \\ &= \frac{|\mathscr{V}|}{k+1} \frac{|D(\omega_{\mathrm{to}}^{\mathrm{II}})|}{|D(\omega_{\mathrm{from}}^{\mathrm{II}})|}. \end{aligned} \tag{4.40}$$

The acceptance ratio is the same for both definitions, which makes them equivalent in practical applications. One may choose to adopt either definition in designing Monte Carlo algorithms, but it would be dangerous to mix the two, e.g. take the probability distribution from one definition but design updates following the other. In the remainder of the thesis, we will always use definition II of the sampling space, unless otherwise stated.

4.4 Normalization

The normalization factor W, defined in (4.35) as a integral of the absolute value of the sum of diagrams at each configuration, occurs in measurements of physical quantities (4.36). However, the value for W is often unknown and difficult to compute explicitly. A common solution to this problem is to normalize Q using a different quantity Q_{norm}, defined as

$$Q_{\text{norm}} = \sum_{k=0}^{k_{\max}} \frac{1}{k!} \int \mathrm{d}\nu_1 \cdots \int \mathrm{d}\nu_k D_{\text{norm}}(\nu_1, \nu_2, \ldots, \nu_k) = \int_\Omega \mathrm{d}\mu(\omega)\, D_{\text{norm}}(\omega), \tag{4.41}$$

such that

$$Q_{\text{norm}} = W \left\langle \frac{D_{\text{norm}}(\omega)}{|D(\omega)|} \right\rangle_p. \tag{4.42}$$

If the value of Q_{norm} is available analytically, W can be solved as

$$W = \frac{Q_{\text{norm}}}{\langle D_{\text{norm}}(\omega)/|D(\omega)|\rangle_p}, \tag{4.43}$$

and thus the measurement for Q (4.36) becomes

$$Q = \frac{Q_{\text{norm}} \langle \operatorname{sgn}(D(\omega))\rangle_p}{\langle D_{\text{norm}}(\omega)/|D(\omega)|\rangle_p}. \tag{4.44}$$

Because the distribution $p(\omega)$ is optimized for the importance sampling of $D(\omega)$, it does not necessary have a good "overlap" with any choice of $D_{\text{norm}}(\omega)$. When choosing the normalization, it is therefore important to ensure the measurement for Q_{norm} does not result in excess statistical error. Here we list some common options for Q_{norm}.

4.4.1 Normalization against the volume of hyperspace

The easiest normalization to compute in analytics is the volume of the sampling space, i.e. by choosing

$$D_{\text{norm}}(\omega) = 1, \tag{4.45}$$

we have

$$Q_{\text{norm}} = \int_{\Omega} \mathrm{d}\mu(\omega) = |\Omega| = \sum_{k=0}^{k_{\max}} \frac{|\mathscr{V}|^k}{k!}. \tag{4.46}$$

The normalization for Q is obtained as

$$Q = \frac{Q_{\text{norm}} \langle \operatorname{sgn}(D(\omega)) \rangle_p}{\langle 1/|D(\omega)| \rangle_p}. \tag{4.47}$$

One major drawback of the hyperspace normalization occurs when the distribution $p(\omega)$ has "holes" in the sampling space, i.e. there are certain regions in the sample space Ω where $p(\omega)$ is very small or even zero. This will lead to a large error in the Monte Carlo estimation of the hyperspace.

4.4.2 Normalization against a subset of configurations

If certain low order diagrams can be computed analytically, one can choose

$$D_{\text{norm}}(\omega) = D(\omega)\, \mathbb{1}[\omega \in \Omega_{\text{norm}}], \tag{4.48}$$

where $\Omega_{\text{norm}} \subset \Omega$ is the subspace in which these low order diagrams are defined, and $\mathbb{1}[\textit{statement}]$ is the indicator function which gives 1 if *statement* is true and 0 otherwise. Integral (4.41) becomes

$$Q_{\text{norm}} = \int_{\Omega} \mathrm{d}\mu(\omega) D(\omega)\, \mathbb{1}[\omega \in \Omega_{\text{norm}}] = \int_{\Omega_{\text{norm}}} \mathrm{d}\mu(\omega) D(\omega), \tag{4.49}$$

which should be calculated analytically. The normalized estimate for Q is therefore

$$Q = \frac{Q_{\text{norm}} \langle \operatorname{sgn}(D(\omega)) \rangle_p}{\langle \operatorname{sgn}(D(\omega))\, \mathbb{1}[\omega \in \Omega_{\text{norm}}] \rangle_p}. \tag{4.50}$$

4.4.3 Normalization via extension of sampling space

Instead of normalizing in the original sampling space Ω, one could extend the sampling space by adding some other "virtual" space Ω_0, such that the new sampling space is

$$\Omega' = \Omega \oplus \Omega_0. \tag{4.51}$$

As an example, we assume that Ω_0 is a singleton set, containing only one additional "virtual" configuration. The integration measure now becomes

$$\int_{\Omega'} \mathrm{d}\mu'(\omega) = \int_{\Omega} \mathrm{d}\mu(\omega) + \int_{\Omega_0} \mathrm{d}\mu_0(\omega), \tag{4.52}$$

where $\mu_0(\omega)$ is the integration measure for the added sampling space. For a singleton set, $\mu_0(\omega)$ is simply the indicator function for Ω_0. The original integral (4.28) is rewritten as

$$Q = \int_{\Omega} \mathrm{d}\mu(\omega)\, D(\omega) + \int_{\Omega_0} \mathrm{d}\mu_0(\omega)\, 0 = \int_{\Omega'} \mathrm{d}\mu'(\omega)\, D(\omega)\, \mathbb{1}[\omega \in \Omega]. \tag{4.53}$$

The probability distribution needs to be updated to cover the extended space:

$$p'(\omega) = \begin{cases} \frac{|D(\omega)|}{W'} & \omega \in \Omega, \\ \frac{w_0}{W'} & \omega \in \Omega_0. \end{cases} \tag{4.54}$$

Here the $w_0 > 0$ is an arbitrary weight we assign to the sole configuration in Ω_0, and W' is the normalization factor:

$$W' = \int_\Omega \mathrm{d}\mu(\omega)\, |D(\omega)| + \int_{\Omega_0} \mathrm{d}\mu_0(\omega)\, w_0 = W + w_0. \tag{4.55}$$

Measurement for Q (4.36) thus becomes

$$Q = W' \langle \mathrm{sgn}(D(\omega)) \mathbb{1}[\omega \in \Omega] \rangle_{p'}. \tag{4.56}$$

We can now define the normalization using the virtual configuration

$$D_{\mathrm{norm}}(\omega) = w_0\, \mathbb{1}[\omega \in \Omega_0], \tag{4.57}$$

which gives

$$Q_{\mathrm{norm}} = \int_{\Omega'} \mathrm{d}\mu'(\omega)\, w_0\, \mathbb{1}[\omega \in \Omega_0] = \int_{\Omega_0} \mathrm{d}\mu_0(\omega)\, w_0 = w_0. \tag{4.58}$$

Monte Carlo measurement of the normalization factor is

$$Q_{\mathrm{norm}} = \langle D_{\mathrm{norm}}(\omega)/p'(\omega) \rangle_{p'} = W' \langle \mathbb{1}[\omega \in \Omega_0] \rangle_{p'}, \tag{4.59}$$

therefore

$$W' = \frac{w_0}{\langle \mathbb{1}[\omega \in \Omega_0] \rangle_{p'}}. \tag{4.60}$$

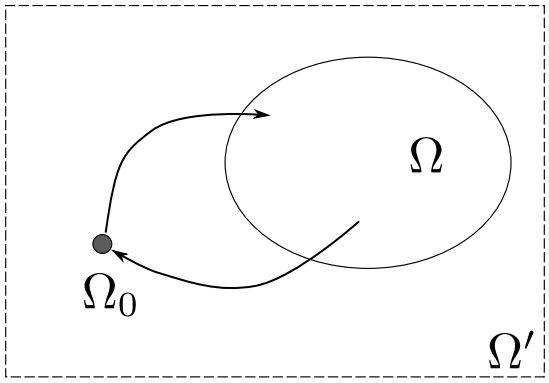

Figure 4.1: Schematic illustration for Monte Carlo normalization using extended sampling space.

The normalized estimate for Q is therefore

$$Q = \frac{w_0 \langle \mathrm{sgn}(D(\omega))\, \mathbb{1}[\omega \in \Omega] \rangle_{p'}}{\langle \mathbb{1}[\omega \in \Omega_0] \rangle_{p'}}. \tag{4.61}$$

When applying this normalization scheme to a Monte Carlo simulation using Metropolis–Hastings, it is crucial to also include the additional configuration from Ω_0 to the metropolis updates, as illustrated in Fig. 4.1.

Chapter 5

Self-Consistent Approximations

Self-consistent field (SCF) methods play a crucial role in the development of electronic structure theories. Mean-field methods such as the Hartree–Fock (HF) approximation and density functional theory (DFT) serve as the building blocks of modern quantum chemistry and material science. In the study of correlated electrons, self-consistency methods based on the Luttinger–Ward (LW) functional [52] and the corresponding Φ-derivable conserving approximations [53, 54] allow systematic inclusions of certain types of high-order diagrams with low computational cost, while conserving fundamental quantities such as charge, momentum, and angular momentum of the system. The LW functional framework also serves as the theoretical foundation of various self-consistent embedding methods, such as dynamical mean-field theory (DMFT) [15, 16], self-energy embedding theory (SEET) [35–37], etc.

This chapter is organized as follows. We first introduce the definition of the LW functional, and discuss its relation to self-consistent approximations. We then briefly introduce several examples of self-consistent methods, such as the HF approximation [2, 42], self-consistent second-order Green's function perturbation theory (GF2) [4–7], the *GW* approximation [8], and the bold diagrammatic Monte Carlo method [20]. As presented in Chapter 6, these

methods are applied to realistic calculations benchmarked against other many-body computational methods.

5.1 Definition of the Luttinger–Ward functional

In this section, we introduce the definition of the LW functional using a non-perturbative approach [39, 55]. The perturbative construction of the function is well described with the initial proposal [52], and Refs. [56, 57] provide rigorous mathematical definitions.

We start with the general Hamiltonian of interacting fermions based on (2.24)

$$\hat{H} = \sum_{ab} h_{ab} \hat{c}_a^\dagger \hat{c}_b + \frac{1}{4} \sum_{abcd} U_{abcd} \hat{c}_a^\dagger \hat{c}_c^\dagger \hat{c}_d \hat{c}_b, \tag{5.1}$$

where $a, b, \ldots$ are spin–orbital indices. The grand partition function is defined as a path integral following (2.52):

$$Z = \mathrm{Tr}\, e^{-\beta(\hat{H}-\mu\hat{N})} = \int \mathscr{D}[\bar{c}, c] e^{-S[\bar{c},c]}, \tag{5.2}$$

where the action is defined following (2.54):

$$S = \int_0^\beta \mathrm{d}\tau \left(\sum_{ab} \bar{c}_a(\tau)((\partial_\tau - \mu)\delta_{ab} + h_{ab}) c_b(\tau) + \frac{1}{4} \sum_{abcd} U_{abcd} \bar{c}_a(\tau) \bar{c}_c(\tau) c_d(\tau) c_b(\tau) \right). \tag{5.3}$$

We now add a bilinear source term to the action in the path integral for Z, such that

$$Z[J] = \int \mathscr{D}[\bar{c}, c] \exp\left(-S[\bar{c}, c] + \int_0^\beta \mathrm{d}\tau' \mathrm{d}\tau \sum_{ab} J_{ab}(\tau', \tau) \bar{c}_a(\tau') c_b(\tau) \right), \tag{5.4}$$

which serves as a MGF of particle–hole pairs, which generates the many-body Green's func-

tion in a similar way to the MGF of single particles (2.60). By the linked-cluster theorem [39], $W[J] = \log Z[J]$ is the CGF of particle–hole pairs, which generates the connected Green's functions. The one-particle Green's function can be written by a functional derivative

$$\begin{aligned} G_{ab}(\tau,\tau') &= \left.\frac{\delta W[J]}{\delta J_{ba}(\tau',\tau)}\right|_{J=0} = \frac{1}{Z[0]}\left.\frac{\delta Z[J]}{\delta J_{ba}(\tau',\tau)}\right|_{J=0} \\ &= \frac{1}{Z}\int \mathscr{D}[\bar{c},c]\bar{c}_b(\tau')c_a(\tau)e^{-S[\bar{c},c]} = \langle \mathscr{T}_\tau \hat{c}_b^\dagger(\tau')\hat{c}_a(\tau)\rangle \\ &= -\langle \mathscr{T}_\tau \hat{c}_a(\tau)\hat{c}_b^\dagger(\tau')\rangle. \end{aligned} \tag{5.5}$$

We perform a Legendre transform on the functional $W[J]$, which yields a functional of the one-particle Green's function $\tilde{G}[J] = \delta W[J]/\delta J$ (at this point $\tilde{G}$ is arbitrary because we have not set $J = 0$ yet) and is sometimes called the Baym-Kadanoff functional [53]

$$\Gamma[\tilde{G}] = -W[J] + \int_0^\beta \mathrm{d}\tau\mathrm{d}\tau' \sum_{ab} \tilde{G}_{ab}(\tau,\tau')J_{ba}(\tau',\tau) \tag{5.6}$$

where $J = J[\tilde{G}]$ is now a functional of $\tilde{G}$. One can show that (we absorb τ and τ' into indices and time integrals into index summations for simplicity)

$$\begin{aligned} \frac{\delta\Gamma[\tilde{G}]}{\delta\tilde{G}_{ab}} &= -\frac{\delta W[J]}{\delta\tilde{G}_{ab}} + J_{ba} + \sum_{a'b'}\tilde{G}_{a'b'}\frac{\delta J_{b'a'}}{\delta\tilde{G}_{ab}} \\ &= -\sum_{a'b'}\frac{\delta W[J]}{\delta J_{b'a'}}\frac{\delta J_{b'a'}}{\delta\tilde{G}_{ab}} + \sum_{a'b'}\frac{\delta W[J]}{\delta J_{b'a'}}\frac{\delta J_{b'a'}}{\delta\tilde{G}_{ab}} + J_{ba} \\ &= J_{ba} \end{aligned} \tag{5.7}$$

Recall that the "physical" Green's function G is obtained when we set $J = 0$, which is equivalent to say that G is the stationary point of functional $\Gamma[\tilde{G}]$.

For a non-interacting system, i.e. $U_{abcd} = 0$, since both the action S and the source term are Gaussian, we can calculate $Z[J]$ and $\Gamma[\tilde{G}]$ directly. Utilizing the following properties of

(positive-definite) matrices and Grassmann numbers [39]

$$\int \prod_a \mathrm{d}\bar{\eta}_a \mathrm{d}\eta_a \exp\left(-\sum_{\alpha\beta} \bar{\eta}_\alpha M_{\alpha\beta}\eta_\beta\right) = \det M, \quad e^{\mathrm{Tr}\log M} = \det M, \tag{5.8}$$

one can show

$$\begin{aligned}
&Z|_{U=0}[J] = \int \mathscr{D}[\bar{c}, c] \exp\left(-\sum_{ab} \bar{c}_a(-[g^{-1}]_{ab} - J_{ab})c_b\right) = \det(-g^{-1} - J) = e^{\mathrm{Tr}\log(-g^{-1}-J)}, \\
&W|_{U=0}[J] = \log Z|_{U=0}[J] = \mathrm{Tr}\log(-g^{-1} - J), \\
&[\tilde{G}|_{U=0}]_{ab} = \frac{\delta W|_{U=0}[J]}{\delta J_{ba}} = \sum_{a'b'}[(-g^{-1} - J)^{-1}]_{a'b'}\frac{\delta(-g^{-1} - J)_{b'a'}}{\delta J_{ba}} = [(g^{-1} + J)^{-1}]_{ab},
\end{aligned} \tag{5.9}$$

from which we can solve for J in terms of $\tilde{G}$, and perform the Legendre transform:

$$\begin{aligned}
&J|_{U=0} = \tilde{G}^{-1}|_{U=0} - g^{-1}, \\
&\Gamma|_{U=0}[\tilde{G}] = -W|_{U=0}[J] + \mathrm{Tr}\,\tilde{G}J|_{U=0} \\
&\quad = -\mathrm{Tr}\log(-g^{-1} - J|_{U=0}) + \mathrm{Tr}(\tilde{G}(\tilde{G}^{-1} - g^{-1})) \\
&\quad = -\mathrm{Tr}\log(-g^{-1} - (\tilde{G}^{-1} - g^{-1})) - \mathrm{Tr}(\tilde{G}(g^{-1} - \tilde{G}^{-1})) \\
&\quad = \mathrm{Tr}\log(-\tilde{G}) - \mathrm{Tr}(\tilde{G}g^{-1} - 1).
\end{aligned} \tag{5.10}$$

The LW functional is defined as the difference of $\Gamma[\tilde{G}]$ between the interacting system and the non-interacting system:

$$\Phi[\tilde{G}] = \Gamma[\tilde{G}] - \Gamma|_{U=0}[\tilde{G}] = \Gamma[\tilde{G}] - \mathrm{Tr}\log(-\tilde{G}) + \mathrm{Tr}(\tilde{G}g^{-1} - 1) \tag{5.11}$$

We can show that the LW functional has the following property:

$$\frac{\delta\Phi[\tilde{G}]}{\delta\tilde{G}} = \frac{\delta\Gamma[\tilde{G}]}{\delta\tilde{G}} - \tilde{G}^{-1} + \frac{\delta}{\delta\tilde{G}}\operatorname{Tr}(\tilde{G}g^{-1} - 1) = \frac{\delta\Gamma[\tilde{G}]}{\delta\tilde{G}} + g^{-1} - \tilde{G}^{-1}. \tag{5.12}$$

Recall that the physical Green's function is the stationary point of $\Gamma[\tilde{G}]$, we see that

$$\left.\frac{\delta\Phi[\tilde{G}]}{\delta\tilde{G}}\right|_{\tilde{G}=G} = g^{-1} - G^{-1} = \Sigma, \tag{5.13}$$

i.e. the functional derivative of $\Phi[\tilde{G}]$ gives the self-energy at $\tilde{G} = G$. This allows us to define a "self-energy functional"

$$\Sigma[\tilde{G}] = \frac{\delta\Phi[\tilde{G}]}{\delta\tilde{G}}, \tag{5.14}$$

which satisfies $\Sigma[\tilde{G} = G] = \Sigma$.

$\Phi[G]$ is directly related to the grand potential of the system at $\tilde{G} = G$, where we have $J = 0$, $W[0] = \log Z = -\beta\Omega$, $\Gamma[G] = -W[0] + \operatorname{Tr} G \cdot 0 = \beta\Omega$, and therefore

$$\begin{aligned}
\Phi[\tilde{G} = G] &= \beta\Omega - \operatorname{Tr}\log(-G) + \operatorname{Tr}(Gg^{-1} - 1) \\
&= \beta\Omega - \operatorname{Tr}\log(-G) + \operatorname{Tr}(G\Sigma), \\
\beta\Omega &= \Phi[\tilde{G} = G] + \operatorname{Tr}\log(-G) - \operatorname{Tr}(G\Sigma).
\end{aligned} \tag{5.15}$$

If the system is non-interacting, i.e. $U = 0$, the LW functional vanishes by definition (5.11).

5.2 Diagrammatic series of the Luttinger–Ward functional

The LW functional is parametrized by the electron interaction U, which we now write explicitly as $\Phi[G;U]$. The bold diagrammatic series is obtained as a series expansion of $\Phi[G;U]$ or $\Sigma[G;U]$ in terms of U [52, 56, 57]. Introducing an arbitrary scalar ξ as the "coupling

constant", the bold series are formally defined as

$$\Phi[G;\xi U] = \sum_{k=1}^{\infty} \xi^k \Phi^{(k)}[G;U], \quad \Sigma[G;\xi U] = \sum_{k=1}^{\infty} \xi^k \Sigma^{(k)}[G;U], \tag{5.16}$$

where $\Sigma^{(k)}$ is obtained as a sum of all "skeleton" diagrams of order k [52]. The skeleton diagrams follow a diagram rule similar to the bare diagrams introduced in Sec. 3.4, with the following main differences:

- Propagator lines are now "bold", representing G instead of g.
- Diagrams are two-particle irreducible (2PI), meaning that a diagram will not be broken into disconnected components by cutting any two propagator lines. In the graph theory language, the edge connectivity of the diagram is greater than two.

Similarly, $\Phi^{(k)}$ consists of 2PI closed diagrams of order k, each which yields the skeleton diagrams by removing one of the $2k$ propagator lines. The removal of propagator lines corresponds to the functional derivative (5.14). Figure 5.1 compares the skeleton diagrams and the "bare" diagram for Σ as well as the corresponding Φ diagrams. The topology corresponding to the second bare diagram is absent in the bold expansion because it can be reduced to the first bold diagram.

Relations between $\Phi[G;U]$ and $\Sigma[G;U]$ can be shown analytically following Ref. [57]. It can be shown that [57] for any arbitrary scalar $t > 0$,

$$\Phi[tG;U] = \Phi[G;t^2U]. \tag{5.17}$$

Consider

$$\Phi[tG;\xi U] = \Phi[G;t^2\xi U], \tag{5.18}$$

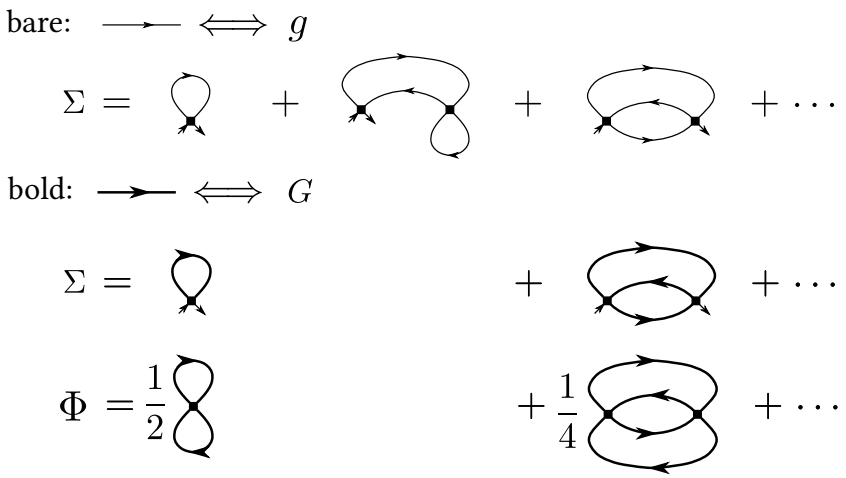

Figure 5.1: Comparison of bare and bold (skeleton) diagrams for the self-energy.

taking a functional derivable w.r.t. G on both sides, we have

$$t\Sigma[tG;\xi U] = \Sigma[G;t^2\xi U], \quad \Sigma[tG;\xi U] = \frac{1}{t}\Sigma[G;t^2\xi U]. \tag{5.19}$$

Since

$$\Phi[G;\xi U] = \int_0^1 \mathrm{d}t\, \frac{\mathrm{d}}{\mathrm{d}t}\Phi[tG;\xi U], \tag{5.20}$$

and the integrand evaluates to

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi[tG;\xi U] = \mathrm{Tr}\left\{G\frac{\delta\Phi[tG;\xi U]}{\delta(tG)}\right\} = \mathrm{Tr}\{G\Sigma[tG;\xi U]\} = \frac{1}{t}\,\mathrm{Tr}\{G\Sigma[G;t^2\xi U]\}, \tag{5.21}$$

which can be expanded following (5.16) as

$$\frac{1}{t}\,\mathrm{Tr}\{G\Sigma[G;t^2\xi U]\} = \sum_{k=1}^{\infty}\frac{1}{t}(t^2\xi)^k\,\mathrm{Tr}\{G\Sigma^{(k)}[G;U]\} = \sum_{k=1}^{\infty} t^{2k-1}\xi^k\,\mathrm{Tr}\{G\Sigma^{(k)}[G;U]\}, \tag{5.22}$$

we have

$$\Phi[G;\xi U] = \sum_{k=1}^{\infty} \xi^k \operatorname{Tr}\{G\Sigma^{(k)}[G;U]\} \int_0^1 \mathrm{d}t\, t^{2k-1} = \sum_{k=1}^{\infty} \frac{\xi^k}{2k} \operatorname{Tr}\{G\Sigma^{(k)}[G;U]\}. \tag{5.23}$$

Comparing to (5.16) gives

$$\Phi^{(k)}[G;U] = \frac{1}{2k} \operatorname{Tr}\{G\Sigma^{(k)}[G;U]\}, \tag{5.24}$$

which directly corresponds to the diagrammatic interpretation of "cutting" propagators.

5.3 Self-consistent approximations

As shown by Baym and Kadanoff [53, 54], any approximation to Φ by taking a subset of diagrams in Φ obeys the number, energy, momentum, and angular momentum conservation laws. Such approximations are called *Φ-derivable approximations*. Self-consistent methods are often constructed based on Φ-derivable approximations following a common pattern:

1. Construct an approximation to the LW functional $\Phi^{\text{approx}}[G]$, or equivalently, the self-energy functional $\Sigma^{\text{approx}}[G]$.

2. Start with an initial guess for G, and perform the following self-consistent loop:

   a) Compute $\Sigma^{\text{approx}}[G]$ using G from the previous iteration.

   b) Compute $G = (g^{-1} - \Sigma)^{-1}$ for the next iteration.

3. Iterate the self-consistent loop until convergence.

Since Φ is a universal functional of G [57], it is possible that the Φ functional converge to "unphysical" results. A self-consistent solution for the functional equation $\delta\Phi[G]/\delta G = \Sigma$

is equivalent to solving for the stationary point of the $\Gamma[\tilde{G}]$ functional, so multiple solutions might exist when the Γ functional has multiple stationary points. It is sometimes called the "multivaluedness" of the Φ functional [58–60].

This section lists common self-consistent methods formulated as Φ-derivable approximations.

5.3.1 Hartree-Fock

The HF approximation [2, 38, 42] is equivalent to truncating the Φ expansion at first order:

$$\Phi^{\mathrm{HF}}[G] = \Phi^{(1)} = \frac{1}{2} U_{abcd} G_{ba}(0^-) G_{dc}(0^-) = \frac{1}{2} U_{abcd} \rho_{ab} \rho_{cd}. \tag{5.25}$$

The HF self-energy is therefore

$$\Sigma^{\mathrm{HF}}_{ab}(\mathrm{i}\omega_n)[G] = U_{abcd} \rho_{cd}, \tag{5.26}$$

or equivalently by expanding $\mathscr{U}$ following (2.23):

$$\Sigma^{\mathrm{HF}}_{ij,\sigma}(\mathrm{i}\omega_n)[G] = \sum_{kl,\sigma'} [V_{ijkl} - V_{ilkj}\delta_{\sigma\sigma'}] \rho_{kl,\sigma'}. \tag{5.27}$$

The HF self-energy is frequency-independent, and only depends on the one-particle density matrix ρ of the electron, and Σ^{HF} can be viewed as a functional of the density.

In quantum chemistry, HF serves as the foundation for more advanced methods, which are often termed "post Hartree–Fock" methods [38]. The HF self-energy is often combined with the one-electron matrix elements $\boldsymbol{h}$ to a "Fock matrix":

$$F_{ij,\sigma} = h_{ij} + \Sigma^{\mathrm{HF}}_{ij,\sigma}, \tag{5.28}$$

such that

$$G^{\mathrm{HF}}(\mathrm{i}\omega_n) = [(\mathrm{i}\omega_n - \mu)\boldsymbol{I} - \boldsymbol{F}]^{-1}. \tag{5.29}$$

5.3.2 GF2

GF2 [4–7] adds the second order bold diagrams on top of HF, such that

$$\Sigma^{\mathrm{GF2}}_{ab}[G] = \Sigma^{(1)}_{ab}[G] + \Sigma^{(2)}_{ab}[G], \tag{5.30}$$

where $\Sigma^{(1)}[G]$ is just $\Sigma^{\mathrm{HF}}[G]$. The second-order self-energy is given by

$$\Sigma^{(2)}_{ab}(\tau)[G] = -\frac{1}{2}\sum_{b'cd}\sum_{a'c'd'} U_{ab'cd}U_{a'bc'd'}G_{b'a'}(\tau)G_{dc'}(\tau)G_{d'c}(-\tau), \tag{5.31}$$

or with explicit spin and orbital indices:

$$\begin{aligned}\Sigma^{(2)}_{ij,\sigma}(\tau)[G] = -\sum_{j'kl}\sum_{i'k'l'} V_{ij'kl}G_{j'i',\sigma}(\tau)\sum_{\sigma'} G_{lk',\sigma'}(\tau)G_{l'k,\sigma'}(-\tau)[V_{i'jk'l'} \\ - V_{k'ji'l'}\delta_{\sigma\sigma'}].\end{aligned} \tag{5.32}$$

5.3.3 *GW*

The *GW* formalism, introduced by Hedin [8], is an alternative self-consistent framework based on the Green's function G, the "screened" interaction W, and a three-point vertex function which gives the exact solution to the interacting electron problem when self-consistency is achieved. The *GW* approximation was originally formulated in terms of Hedin's equations by setting the three-point vertex to one, and thus only requires self-consistency at one-particle level. This approximation is Φ-derivable if full self-consistency in both G and W is enforced [61–63].

The GW self-energy is defined as

$$\Sigma^{GW}_{ij,\sigma}(\tau)[G] = \sum_{kl}\left[V_{ijkl}\sum_{\sigma'}\rho_{kl,\sigma'}\delta(\tau) - W_{ilkj}(\tau)G_{lk,\sigma}(\tau)\right], \tag{5.33}$$

where the screened interaction W is a functional of G, obtained from an RPA-like geometric series

$$W_{ijkl}(\mathrm{i}\omega_n)[G] = V_{ijkl} + \sum_{i'j'k'l'} V_{ijk'l'}\Pi_{l'k'j'i'}(\mathrm{i}\omega_n)W_{i'j'kl}(\mathrm{i}\omega_n), \tag{5.34}$$

in which Π is the bare polarization bubble

$$\Pi_{l'k'j'i'}(\tau,\tau') = -\sum_{\sigma} G_{j'k',\sigma}(\tau,\tau')G_{l'i',\sigma}(\tau',\tau). \tag{5.35}$$

Diagrammatically, $\Sigma^{GW}[G]$ consists of the first order skeleton diagram (HF diagram) plus all RPA-like bold diagrams of second order and above.

One can alternatively define a functional $\Psi[G,W]$ of both G and W, as introduced in Ref. [63], such that

$$\begin{aligned}\left(\frac{\delta\Psi[G,W]}{\delta G}\right)_W &= \Sigma \\ \left(\frac{\delta\Psi[G,W]}{\delta W}\right)_G &= -\frac{1}{2}\Pi.\end{aligned} \tag{5.36}$$

Diagrammatically, the Ψ functional corresponds to the sum of skeleton diagrams composed of bold Green's function lines and screened interaction lines, and each diagram should remain connected if one cut two Green's function lines or two interaction lines. The GW approximation corresponds to the first order diagram of the Ψ functional.

Although GW is a Φ-derivable approximation which conserves particle number, energy, momentum, and angular momentum, it does not respect the crossing symmetry due to the

fact that only part of the diagrams are included at order two and above. As discussed in Appendix A.4.d of Ref. [44], different treatments of the interaction tensor, e.g. whether anti-symmetrization (2.23) is performed, result in deviations of GW results.

5.3.4 Bold diagrammatic Monte Carlo

Instead of calculating low-order approximations to Φ deterministically, The bold diagrammatic Monte Carlo (BDMC) algorithm [20, 21] uses Monte Carlo integration described in Chapter 4 to compute all skeleton diagrams from (5.16) up to a certain order. The same self-consistency loop is applied to reach a fixed point for G. In studies of Coulomb systems, BDMC often measures diagrams composed of both G and W [64, 65], which corresponds to the expansion of the $\Psi[G, W]$ functional.

Chapter 6

Direct Comparison of Many-Body Methods for Realistic Systems

*This chapter is based on the following publications: (1) M. Motta et al.: Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods, Physical Review X 7, 031059 (2017). I contributed to the benchmark data and method descriptions of the self-consistent GW method. (2) K. T. Williams et al.: Direct Comparison of Many-Body Methods for Realistic Electronic Hamiltonians, Physical Review X **10**, 011041 (2020). I contributed to the visualization and analysis of all benchmark data, as well as the benchmark data and method descriptions of the self-consistent GW method.*

Because the dimension of Hilbert space of the many-body electron Hamiltonian (2.1) increases dramatically as system size increases, there is no general, numerically exact method that can treat many-electron systems with low computational cost. In recent years, a vast number of theories, numerical methodologies, computational software and algorithms have been proposed, and more are being developed. These methods have different strengths and

weaknesses and different domains of applicability. Therefore, careful comparisons of different methods on the same realistic system are valuable for facilitating further development and accelerating progress.

Recently, there is significant progress in direct comparisons of many-body methods, many of which are diagrammatic methods, for both model systems [66, 67] and realistic Hamiltonians [44, 68, 69]. These benchmarks apply complementary methods to the same problem, which enables cross-check and validation, leading to a powerful new paradigm of attack on difficult problems. Cases where results from different methods agree provide valuable benchmarks against which new methods can be tested, thereby facilitating further development and accelerating progress.

This chapter briefly summarizes two benchmark projects of many-body method in realistic systems: one for Hydrogen chains [44], and the other for transition metal oxides [69].

6.1 Hydrogen chain

Hydrogen is the first element in the periodic table and the most abundant in nature. Studies of the H atom, H_2^+ cation, and H_2 molecule have served as landmarks in quantum physics and chemistry. Despite their deceptive simplicity, bulk H systems are rich and complex. The ground-state properties of the hydrogen chain can differ significantly from those of simpler systems such as the Hubbard model, and are, in fact, not completely understood. The linear H chain captures key features that are essential for generalizing model-system methods to real materials, in particular: the presence of strong electron correlations of diverse nature as the H–H distance is varied, and the need to treat the full physical Coulomb interaction and to work in the continuous space and thermodynamic limits.

The linear hydrogen chain is an ideal first benchmark system for testing the ability of

many-body theoretical methods to handle the challenges posed by real materials. It has multiple (in principle, infinite) orbitals per site, as well as long-range interactions. The use of basis sets, as introduced in Chapter 2, defines models of the hydrogen chain of increasing complexity. In a minimal basis, there is only one band, and the problem resembles a one-dimensional Hubbard model with long-range interactions. Larger, more realistic, basis sets bring back characteristics of real materials. Thus, one can neatly and systematically connect from a fundamental model of strong electron correlation to a real material system. On the other hand, the H chain eliminates complexities of other materials systems such as the need to separately treat core electrons or incorporate relativistic effects, and is thus accessible to many theoretical methods at their current state of development.

Ref. [44] studies finite H chains of increasing length and crosscheck the results against calculations performed using periodic boundary conditions, as well as results from calculations formulated in the thermodynamic limit. A systematic quantum chemistry sequence of basis sets of increasing size are used to investigate convergence towards CBS limit, and compared to methods formulated directly in real space. A vast amount of data are produced from more than a dozen many-body computational methods, providing detailed information in finite length chains and with finite basis sets, which are available in online repositories [44].

Most calculations were performed using standard Gaussian basis sets introduced in Chapter 2 except density matrix renormalization group (DMRG) which uses a specialized basis. The correlation-consistent cc-pVxZ basis set [43] is used, with x =D, T, Q, and 5, which correspond to m =5, 14, 30, and 55 orbitals per atom, respectively. H-chains with nearest-neighbor proton separation (bond length) R of 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.4, 2.8, 3.2, and 3.6 a_0 are studied. The target quantity of this work is the ground-state energy $\mathscr{E}(N,R)$ for different chain sizes and lengths, and the *energy per atom*, $E(N,R) = \mathscr{E}(N,R)/N$, at the

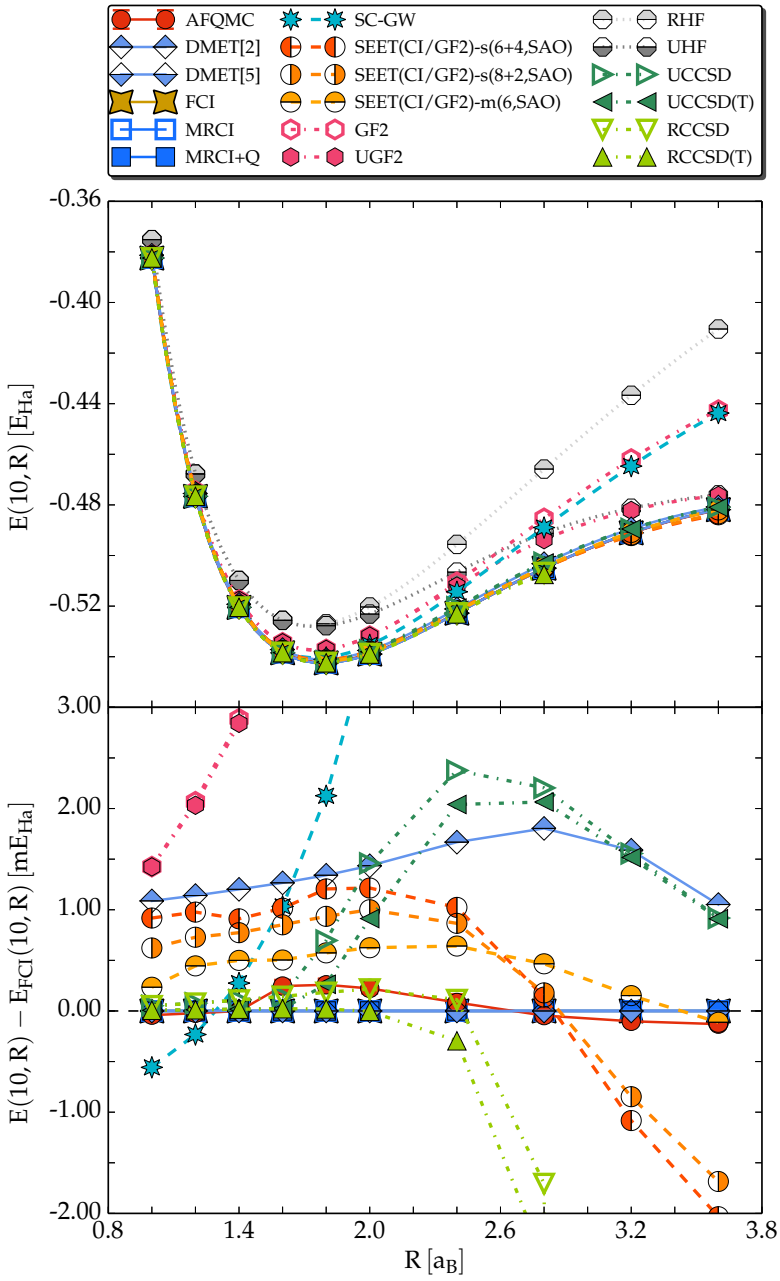

Figure 6.1: Potential energy curve of H_{10} (top) and deviations from FCI (bottom), in the minimal STO-6G basis. See Ref. [44] for a comprehensive description of the methods.

thermodynamic limit (TDL)

$$E_{\mathrm{TDL}}(R) = \lim_{N\to\infty} E(N,R)\,. \tag{6.1}$$

In the chemistry convention, $E(N,R)$ is referred to as the potential energy curve (PEC), and $E_{\mathrm{TDL}}(R)$ in TDL as the equation of state (EOS).

Figure 6.1 shows a detailed comparison of the PEC, $E(N=10,R)$ vs. R, obtained by a variety of methods, in the minimal STO-6G basis. For all methods, the PEC features a familiar short-range repulsion, due to the combined effect of Coulomb repulsion and Pauli exclusion, followed by a decrease to a minimum value E_0, attained at the equilibrium bondlength R_e. Beyond R_e, the PEC monotonically increases to an asymptotic value E_∞, the ground-state energy of a single H atom. The well depth gives a dissociation energy $D_e = (E_\infty - E_0)$. Owing to the small size of this chain and the STO-6G basis, the PEC can be calculated using the FCI method, giving the exact values $R_e = 1.786$ $\mathrm{a_0}$, $E_0 = -0.542\,457$ $\mathrm{E_h}$, and $E_\infty = -0.471\,039$ $\mathrm{E_h}$.

As weak coupling methods, the diagrammatic GF2 and self-consistent GW (SC-GW) methods have difficulties in the strong coupling regime at large bond lengths. Allowing methods to break spin symmetry may lead to an improvement of the energetics. As illustrated with GF2, using an unrestricted reference state provides a better estimate of the ground state energy in that regime but generates a spurious magnetization. Deviations at small distances (corresponding to the weak coupling regime) show that terms beyond the bare second-order or screened first-order approximation are needed to reach the accuracy of other methods.

Figure 6.2 shows results for the STO-6g basis extrapolated to the TDL. DMRG calculations can be carried out for large system sizes in this basis, and serve as a near-FCI quality benchmark. DMRG results for finite chains, after extrapolation to $N \to \infty$, yield an equilibrium geometry $R_e = 1.831(3)$ $\mathrm{a_0}$ and ground-state energy per atom of $E_0 = -0.5407(2)$ $\mathrm{E_h}$ at R_e. BDMC with maximum order 3 yields converged results up to $R = 2.4$. For $R = 2.8$ conver-

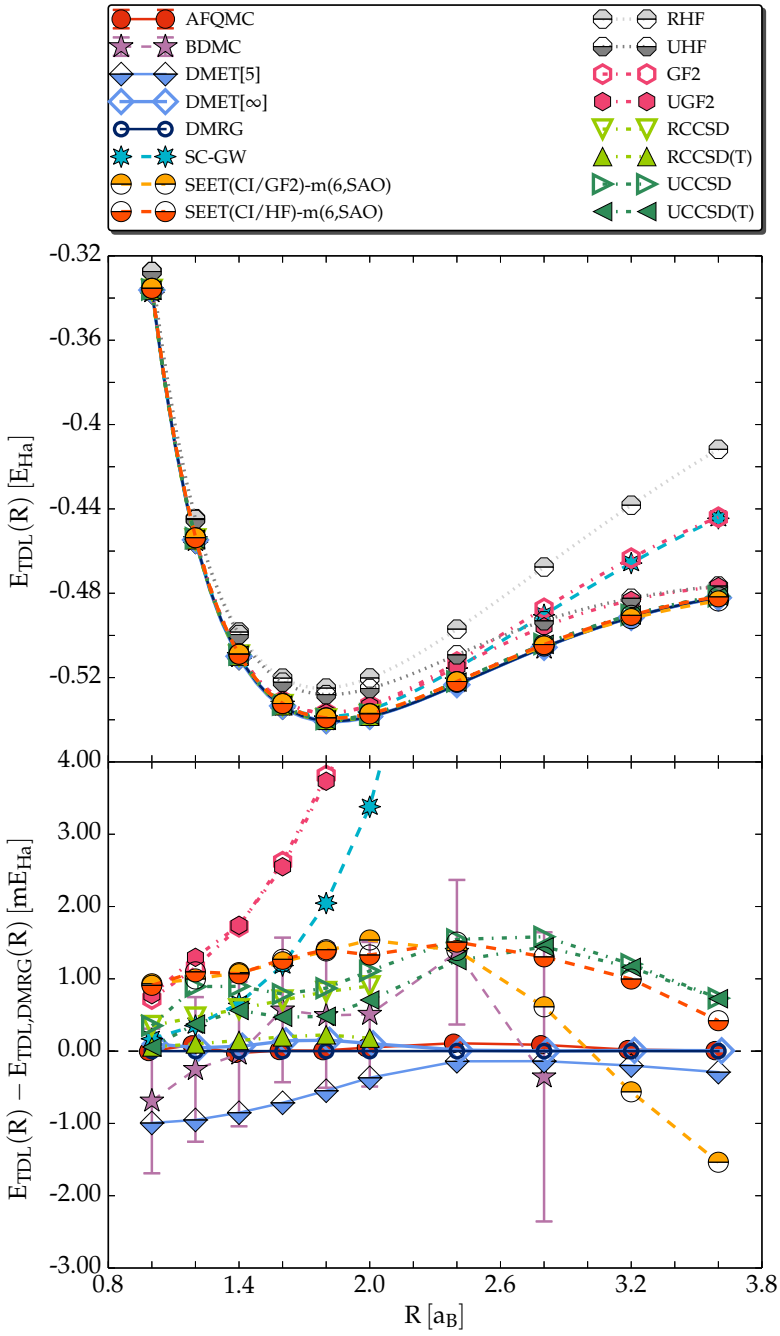

Figure 6.2: Top: Computed EOS in the minimal basis at the thermodynamic limit. Bottom: Detailed comparison using DMRG results as reference.

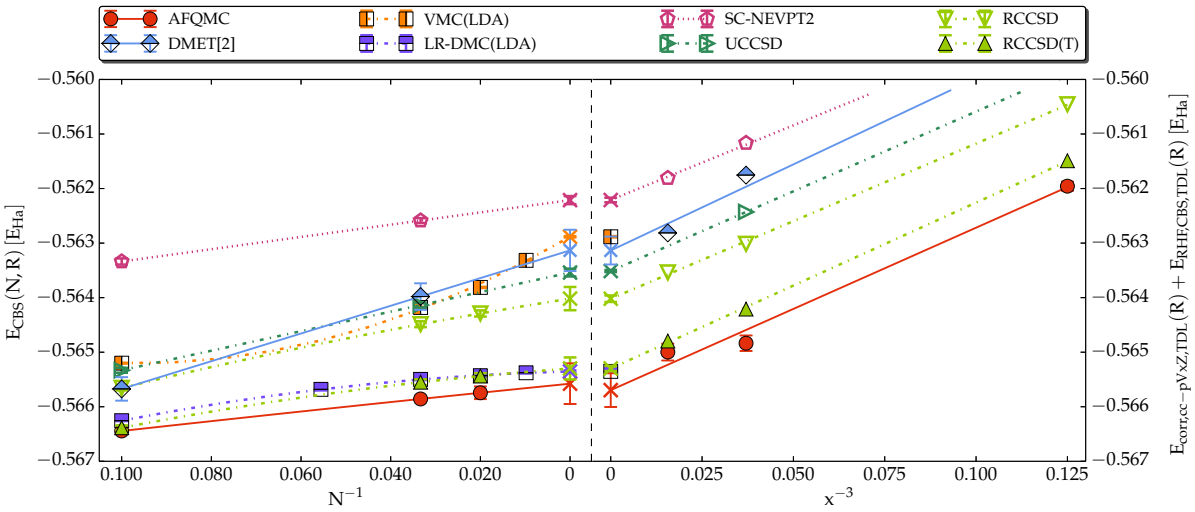

Figure 6.3: Illustration of the extrapolations to the CBS and TDL limits. Results are shown for $R = 1.8\ a_0$. The left panel shows extrapolation of $E_{\mathrm{CBS}}(N)$ vs. $1/N$, while the right panel shows extrapolation of $E_{\mathrm{cc-pV}x\mathrm{Z}}(N \to \infty)$ vs. $1/x^3$. (The correlation energy is shown on the right, shifted by the CBS RHF energy.) Final results are consistent within statistical errors and independent of the order with which the limits are taken.

gence is reached only with maximum expansion order of 5; reaching convergence for larger values of R requires even higher orders. The calculated EOS is in good agreement with the exact results, and its final error bar of 1 $\mathrm{mE_h}$ for $R \leq 2.4$ is dominated by the resolution of the grid of 512 Matsubara frequencies used. For $R = 2.8$ the error bar of 2 $\mathrm{mE_h}$ is dominated by statistical noise in high diagrammatic orders.

The finite basis set results are extrapolated to the CBS limit by standard procedures [70, 71], taking care to reach large basis sets. We first fit the HF energies $E_{\mathrm{HF},x}(N,R)$ computed at the cc-pVxZ basis set level, to an exponential function

$$E_{\mathrm{HF},x}(N,R) = A(N,R) + B(N,R)e^{-xC(N,R)}\,. \tag{6.2}$$

The correlation energy

$$E_{\mathrm{corr},x}(N,R) \equiv E(N,R) - E_{\mathrm{HF},x}(N,R) \tag{6.3}$$

is then fitted to a power law:

$$E_{\mathrm{corr},x}(N,R) = \alpha(N,R) + \frac{\beta(N,R)}{x^3}\,. \tag{6.4}$$

The CBS result is taken as $\alpha(N,R) + A(N,R)$, with a combined uncertainty estimated from the fitting procedures. To extrapolate the finite-N results to the TDL, we assume that the PEC has the following size dependence:

$$E(N,R) = \sum_{i=0}^{k} \frac{A_i(R)}{N^i}\,, \tag{6.5}$$

where k is a small integer. In this work we typically used $k = 2$, employing $N = 10, 30, 50$ and, when necessary, $N = 18$, 22, 70 and 102.

There are multiple strategies for finite basis-set methods to approach the combined limits of CBS and TDL. One could extrapolate to the CBS limit for each finite chain of fixed N and then extrapolate the results in N to the TDL. Alternatively, one could extrapolate each basis set to the TDL, and then extrapolate to the CBS limit, or use a joint Ansatz and extrapolate both simultaneously. As illustrated in Fig. 6.3, exchanging the order of the extrapolation leads to consistent and robust results.

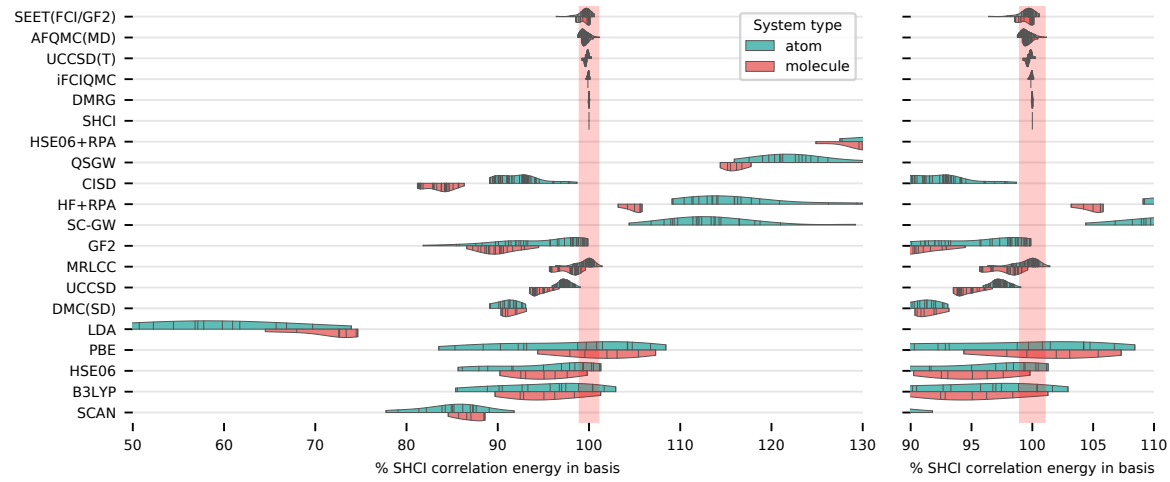

Figure 6.4: The percent of the reference (SHCI) correlation energy within each basis obtained by each of the methods in the benchmark set. All basis sets available are plotted; individual data points are indicated by small lines.

6.2 Transition metal oxides

In Ref. [69], the same methodology of directly comparing many-body methods is applied to more complex systems: transition metal atoms and their oxide molecules. Core electrons in the benchmark systems are removed using effective core potentials [72–74], which accurately represent the core [75] in many-body simulations and allow all the methods considered in this work to use the same Hamiltonian. In addition, they provide an easy way to include scalar relativistic effects, needed for meaningful comparison to experiment. These potentials are available for O, Sc, Ti, V, Cr, Mn, Fe, and Cu, which defines the test set. We consider these atoms, their ions, and the corresponding transition metal monoxide molecules. To simplify the comparison, the molecules were computed at their equilibrium geometry. For method which works in a finite basis set, CBS limit is reached in the same way as the hydrogen benchmark by choosing an ascending basis set from dZ to 5Z, i.e. dZ, tZ, qZ, 5Z, and extrapolating the correlation energy following a $1/x^3$ scaling, where $x = 2, 3, 4, 5$ respectively.

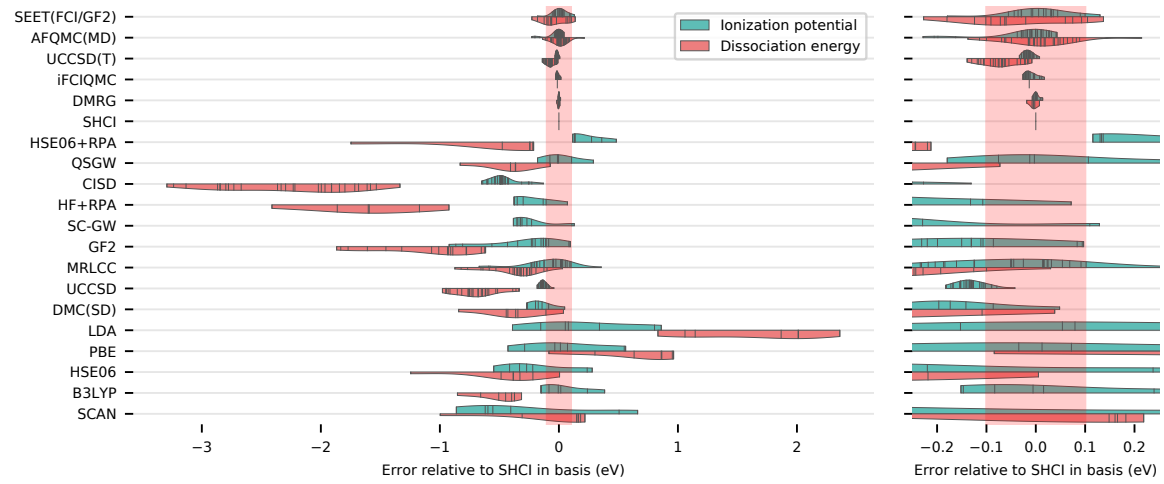

Figure 6.5: Dissociation energy and ionization potential of molecules and atoms compared to SHCI reference calculations.

Figure 6.4 shows the correlation energy computed by each method, plotted as a percentage of the semistochastic heat-bath configuration interaction (SHCI) [76, 77] reference data for the same system and basis set. Figure 6.5 compares the performance of methods in computing the ionization potential of the atoms and dissociation energy of the molecules. See Ref. [69] for detailed definitions and analysis for all methods listed in these plots.

The low-order self-consistent diagrammatic methods, GF2 and SC-*GW*, either overestimate or underestimate the total energy, as shown in Fig. 6.4. These errors tend to cancel significantly when calculating the energy differences, such as the ionization potential and the dissociation energy, resulting in more accurate results.

We survey 20 advanced many-electron techniques on precisely defined realistic Hamiltonians for transition metal systems. For a given basis set, we achieve approximately 1 mhartree agreement on the total energy between highaccuracy methods, which provides a total energy benchmark for many-body methods. To our knowledge, such an agreement is unprecedented for first-principles calculations of transition metal systems. Our accurate ref-

erence energies should enable the development of approximate, but more computationally efficient, many-body techniques as well as better density functionals, without the necessity of experimental reference values. These systems are also a useful test for future quantum computing algorithms.

This benchmark project assesses the state of the art in achieving high accuracy in realistic systems by surveying 20 advanced many-electron techniques on precisely defined realistic Hamiltonians for transition metal systems. The benchmark set includes systems with large Hilbert spaces with dimensions around 10^{44}. The systematically converged techniques used in this work are able to achieve excellent agreement but can be applied only to relatively small systems due to their computational cost. It is thus important to understand the errors in lower-scaling techniques that can be applied to larger systems and whether performance on small systems is transferable to larger systems. This work takes a step in that direction, since we are able to achieve converged results for both correlated atoms and molecules, and indeed we observe that the accuracy of some techniques degrades with system size.

Chapter 7

Diagrammatic Monte Carlo for Realistic Impurities

This chapter is based on JL, Markus Wallerberger, and Emanuel Gull: "Diagrammatic Monte Carlo method for impurity models with general interactions and hybridizations." Phys. Rev. Research 2, 033211 (2020).

In this chapter, we present a diagrammatic Monte Carlo method for quantum impurity problems with general interactions and general hybridization functions. Our method uses a recursive determinant scheme to sample diagrams for the scattering amplitude. Unlike in other methods for general impurity problems, an approximation of the continuous hybridization function by a finite number of bath states is not needed, and accessing low temperature does not incur an exponential cost. We test the method for the example of molecular systems, where we systematically vary temperature, interatomic distance, and basis set size. We further apply the method to an impurity problem generated by a self-energy embedding calculation of correlated antiferromagnetic NiO. We find that the method is ideal for quantum impurity problems with a large number of orbitals but only moderate correlations.

7.1 Introduction

Quantum impurity models, originally introduced to describe magnetic impurities such as iron or copper atoms with partially filled d-shells in a non-magnetic host material [78], have since found applications in nanoscience as representations of quantum dots and molecular conductors [79], and in surface science to understand the adsorption of atoms on surfaces [80, 81]. In addition, they form the central part of embedding theories such as the DMFT [15, 16] and its variants [82–94], as well as the SEET [35–37], where they describe the behavior of a few "strongly correlated" orbitals embedded into a weakly correlated or non-interacting background of other orbitals. These methods promise a systematic route for the simulation of strongly correlated quantum many-body problems [95].

While the original formulation of a quantum impurity model [78] only describes a single correlated orbital coupled to a non-interacting environment, in general the impurities occurring in the context of surface science and embedding theories contain many orbitals with general four-fermion interactions and few symmetries [96]. The time-dependent hybridization function describing the hopping between the impurity and its environment is typically such that it cannot be diagonalized for all frequencies at once.

Solving quantum impurity problems, i.e. obtaining the impurity Green's function given an impurity Hamiltonian and a hybridization function, requires the use of numerical methods. A wide range of such methods exist. Hamiltonian-based methods, such as exact diagonalization [97–101] and its variants [102], configuration-interactions [103], or coupled cluster theory [104, 105], solve the impurity problem by mapping the impurity problem onto a system with a local Hamiltonian and a finite number of auxiliary "bath" states chosen to fit the time-dependent hybridization function. The methods are limited to a relatively small set of strongly interacting sites or break down at moderate correlation strength. The bath fitting,

which typically approximates a continuous bath dispersion by a non-linear fit to a small number of delta-function peaks, introduces additional approximations [99, 101]. Numerical renormalization techniques [106, 107] overcome this issue by providing an almost continuous bath density of states but are in turn limited to a few orbitals in highly symmetrical situations.

A complementary approach is given by Monte Carlo techniques such as the CT-QMC methods [14]. These methods are based on a stochastic sampling of the terms in a diagrammatic expansion of the partition function. For particle-hole symmetric systems with on-site density-density interactions, interaction expansion methods [9, 12, 14] can solve systems with hundreds of strongly correlated orbitals [66]. Away from particle hole symmetry and at low temperature, they are typically limited to around eight orbitals, and their naive adaptation to general four-fermion operator terms suffers from a severe sign problem [108]. In contrast, a partition function expansion in the hybridization [10, 11, 109] is able to work with general local Hamiltonians of up to five orbitals, but is similarly restricted to diagonal hybridization functions. A reformulation [33] in terms of "inchworm" diagrams [27] overcomes the restriction of diagonal hybridizations, but so far remains limited to impurities with up to three orbitals.

There is therefore a need for impurity solver methods that can treat the problems of embedding theory and surface science, where several orbitals with general interactions and hybridizations occur. DiagMC methods [18–21, 110], which expand physical observables rather than partition functions, along with efficient ways of evaluating the resulting diagrammatic series via the CDet approach [22–24, 26, 31, 111], are promising. While these methods suffer from other limitations, including divergences of the series in the strong correlation regime, they do not require to approximate the hybridization function by a fit, and are not based on a diagonalization of the local Hamiltonian.

In this chapter, we show a formulation of the DiagMC method for impurity problems with general interactions and hybridizations based on the CDet framework. We test the method on the example of molecular systems, for which a broad range of very mature Hamiltonian methods exist. From the point of view of the algorithmic formulation, the molecular systems exhibit the full complexity of general impurity problems. The only difference between molecules and quantum impurities is that the latter are formulated with a time-dependent hybridization, rather than an instantaneous hopping. This hybridization function modifies the bare propagator but otherwise leaves the system and our algorithmic approach invariant. Applications to molecular systems therefore form an ideal testbed for impurity solver methods of this type. We complete our benchmark by applying the impurity solver to an impurity generated by a self-energy embedding calculation of antiferromagnetic solid NiO [112].

We carefully analyze the convergence behavior of the diagrammatic expansion and the computational cost of the method as a function of varying temperature, basis sets, intermolecular distance, and system size. We emphasize that we do not intend to present our method as a viable method for quantum chemistry systems without retardation effects. Rather, we exploit the rigorous and controlled framework of molecular simulations to generate a series of test cases that illustrate various parameter regimes in quantum impurities.

This chapter is organized as follows. In Sec. 7.2 we introduce the computational problem, the diagrammatic formulation, and the algorithmic description. In Sec. 7.3 we present applications to molecular systems and benchmark results for quantum impurities. Finally, Sec. 7.4 presents conclusions.

7.2 Method

7.2.1 Partition function expansion

We describe molecular electrons using the Hamiltonian from Eq. (2.24):

$$\hat{H} = \underbrace{\sum_{ab} h_{ab} \hat{c}_a^\dagger \hat{c}_b}_{\hat{H}_0} + \underbrace{\frac{1}{4} \sum_{abcd} U_{abcd} \hat{c}_a^\dagger \hat{c}_c^\dagger \hat{c}_d \hat{c}_b}_{\hat{H}_V}, \tag{7.1}$$

where a, b, c, d denote spin-orbitals, $1, \ldots, N$. We employ second quantization: $\hat{c}_a$ and $\hat{c}_a^\dagger$ annihilates and creates, respectively, an electron in the spin-orbital a. The interacting term $\hat{H}_V$ is parametrized by the antisymmetrized two-electron integrals U_{abcd} defined in (2.23), which avoids ambiguities in the diagrammatic expansions [44]. We orthonormalize the basis following Section 2.2 such that $\{\hat{c}_a^\dagger, \hat{c}_b\} = \delta_{ab}$, as we empirically found this to improve the error bars in the subsequent Monte Carlo procedure.

We perform perturbation expansions introduced in Chapter 3 by introducing the expansion parameter ξ:

$$\hat{H}_\xi = \hat{H}_0 + \xi \hat{H}_V. \tag{7.2}$$

The non-interacting case is given by $\hat{H}_{\xi=0}$, whereas $\hat{H}_{\xi=1}$ recovers the full Hamiltonian (7.1). There perturbation series of the partition function follows (3.10):

$$\begin{aligned} \frac{Z_\xi}{Z_0} = \sum_{k=0}^{\infty} \frac{(-\xi)^k}{k!} \sum_{a_1 b_1 c_1 d_1} \cdots \sum_{a_k b_k c_k d_k} \int_0^\beta \mathrm{d}\tau_1 \cdots \int_0^\beta \mathrm{d}\tau_k \\ \times \left(\frac{U_{a_1 b_1 c_1 d_1}}{4}\right) \cdots \left(\frac{U_{a_k b_k c_k d_k}}{4}\right) \langle \hat{c}_{a_1}^\dagger(\tau_1) \hat{c}_{c_1}^\dagger(\tau_1) \\ \times \hat{c}_{d_1}(\tau_1) \hat{c}_{b_1}(\tau_1) \ \cdots \ \hat{c}_{a_k}^\dagger(\tau_k) \hat{c}_{c_k}^\dagger(\tau_k) \hat{c}_{d_k}(\tau_k) \hat{c}_{b_k}(\tau_k) \rangle_0. \end{aligned} \tag{7.3}$$

In order to simplify our notation we combine four spin-orbitals a, b, c, d and an imaginary time τ into a single "vertex" $v = (a_v, b_v, c_v, d_v, \tau_v)$. We also introduce the following shorthands:

$$\int \mathrm{d}^k \mathscr{V} := \frac{1}{k!} \prod_{v \in \mathscr{V}} \sum_{a_v b_v c_v d_v} \int_0^\beta \mathrm{d}\tau_v, \tag{7.4a}$$

$$D(\mathscr{V}) := \prod_{v \in \mathscr{V}} \left(-\frac{U_{a_v b_v c_v d_v}}{4} \right) \times \left\langle \prod_{v \in \mathscr{V}} \hat{c}^\dagger_{a_v}(\tau_v) \hat{c}^\dagger_{c_v}(\tau_v) \hat{c}_{d_v}(\tau_v) \hat{c}_{b_v}(\tau_v) \right\rangle_0 . \tag{7.4b}$$

Eq. (7.4b) emphasizes the fact that expectation value in Eq. (7.3) corresponds to the sum over all disconnected and connected Feynman diagrams with vertices $\mathscr{V} = (v_1, \ldots, v_k)$, while Eq. (7.4a) just corresponds to the sum over all internal degrees of freedom of the diagrams. With these substitutions, Eq. (7.3) simplifies to:

$$\frac{Z_\xi}{Z_0} = \sum_{k=0}^{\infty} \xi^k \int \mathrm{d}^k \mathscr{V} \; D(\mathscr{V}). \tag{7.5}$$

To evaluate Eq. (7.5), we first introduce the non-interacting Green's function:

$$g_{ba}(\tau) = -\langle \hat{c}_b(\tau) \hat{c}^\dagger_a(0) \rangle_0 = [(-\partial_\tau + \mu) \boldsymbol{I} - \boldsymbol{h}]^{-1}_{ba}. \tag{7.6}$$

Given a diagram $\mathscr{V} = (v_1, \ldots, v_k)$ with $v_i = (a_i, b_i, c_i, d_i, \tau_i)$, we can use Wick's theorem to write Eq. (7.4b) as:

$$D(\mathscr{V}) = \prod_{i=1}^{k} \left(-\frac{U_{a_i b_i c_i d_i}}{4} \right) \det \boldsymbol{G}(\mathscr{V}), \tag{7.7}$$

where $\boldsymbol{G}$ is the $2k \times 2k$ matrix defined in (3.22) in which the rows (columns) correspond to the $2k$ annihilation (creation) operators. Introducing the column and row indices $\alpha, \beta, \ldots$ such

that

$$\begin{aligned}
\{a_\alpha\} &:= \{a_1, c_1, a_2, c_2, \dots, a_k, c_k\}, \\
\{b_\beta\} &:= \{b_1, d_1, b_2, d_2, \dots, b_k, d_k\}, \\
\{\tau_\alpha\} = \{\tau_\beta\} &:= \{\tau_1, \tau_1, \tau_2, \tau_2, \dots, \tau_k, \tau_k\},
\end{aligned} \tag{7.8}$$

we define the matrix elements

$$\begin{aligned}
[G(\mathscr{V})]_{\beta\alpha} &:= -\langle \hat{c}_{b_\beta}(\tau_\beta)\hat{c}^\dagger_{a_\alpha}(\tau_\alpha)\rangle_0 \\
&= g_{b_\beta a_\alpha}(\tau_\beta - \tau_\alpha + 0^-).
\end{aligned} \tag{7.9}$$

The full matrix can be written in a block form as

$$G(\mathscr{V}) := \begin{bmatrix} \boldsymbol{g}_{11} & \boldsymbol{g}_{12} & \cdots & \boldsymbol{g}_{1k} \\ \boldsymbol{g}_{21} & \boldsymbol{g}_{22} & \cdots & \boldsymbol{g}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_{k1} & \boldsymbol{g}_{k2} & \cdots & \boldsymbol{g}_{kk} \end{bmatrix}, \tag{7.10}$$

where each 2×2 block is given by

$$\boldsymbol{g}_{ij} := \begin{bmatrix} g_{b_i a_j}(\tau_i - \tau_j + 0^-) & g_{d_i a_j}(\tau_i - \tau_j + 0^-) \\ g_{b_i c_j}(\tau_i - \tau_j + 0^-) & g_{d_i c_j}(\tau_i - \tau_j + 0^-) \end{bmatrix}. \tag{7.11}$$

Eqs. (7.5) and (7.7) serve as the basis of interaction-expansion continuous time quantum Monte Carlo (CT-INT): one generates random configurations $(v_1 \dots v_k)$ and evaluates the corresponding weight by computing the determinant [9, 14, 108].

7.2.2 Free energy expansion

While the partition function expansion can be efficiently computed as determinants (with scaling $\mathcal{O}(k^3)$) and the series is guaranteed to converge, it is also plagued by the negative sign problem, which is expected to worsen exponentially as the system size is increased or the temperature reduced. The sign problem is typically manageable in Hubbard model calculations up to moderate correlations and system size, where it only stems from negative determinant contributions. In contrast, the sign problem is particularly severe in molecules and surface science quantum impurity problems [108], where both Coulomb interaction terms and determinants generate negative coefficients.

In order to overcome these difficulties, we move to the grand potential Ω, defined as

$$Z_\xi = \exp(-\beta\Omega_\xi). \tag{7.12}$$

Ω_ξ serves as a CGF for correlations functions [39] and its power series in ξ is given by:

$$\Omega_\xi = \Omega_0 - \frac{1}{\beta}\sum_{k=1}^{\infty}\xi^k \int \mathrm{d}^k\boldsymbol{\mathscr{V}}\; D_c(\boldsymbol{\mathscr{V}}), \tag{7.13}$$

where Ω_0 is defined as $Z_0 = \exp(-\beta\Omega_0)$.

The symbol D_c indicates that unlike in Eq. (7.5), the sum is to be performed over connected Feynman diagrams only, as shown in Fig. 7.1. As a result of Wick's theorem (7.7), all vacuum diagrams $D(\boldsymbol{\mathscr{V}})$ for a fixed vertex configuration $\boldsymbol{\mathscr{V}}$ can be partitioned into a connected subdiagram and the remainder of the vacuum components. Since no external legs exist to serve as reference points for defining connectivity, we start by picking a specific vertex $v \in \boldsymbol{\mathscr{V}}$ as

the "reference" and consider connectivity with respect to v, i.e.

$$D(\mathscr{V}) = \sum_{\substack{\mathscr{S}\subseteq\mathscr{V} \\ \mathscr{S}\ni v}} D_c(\mathscr{S}) D(\mathscr{V}\setminus\mathscr{S}). \tag{7.14}$$

As the choice of v is arbitrary, it can be any of the $k = |\mathscr{V}|$ vertices in $\mathscr{V}$, therefore

$$D(\mathscr{V}) = \frac{1}{|\mathscr{V}|} \sum_{v\in\mathscr{V}} \sum_{\substack{\mathscr{S}\subseteq\mathscr{V} \\ \mathscr{S}\ni v}} D_c(\mathscr{S}) D(\mathscr{V}\setminus\mathscr{S}). \tag{7.15}$$

This is equivalent to iterating all possible subsets $\mathscr{S}$ of $\mathscr{V}$ where the reference v can be any vertex in $\mathscr{S}$:

$$\begin{aligned} D(\mathscr{V}) &= \frac{1}{|\mathscr{V}|} \sum_{\mathscr{S}\subseteq\mathscr{V}} \sum_{v\in\mathscr{S}} D_c(\mathscr{S}) D(\mathscr{V}\setminus\mathscr{S}) \\ &= \sum_{\mathscr{S}\subseteq\mathscr{V}} \frac{|\mathscr{S}|}{|\mathscr{V}|} D_c(\mathscr{S}) D(\mathscr{V}\setminus\mathscr{S}). \end{aligned} \tag{7.16}$$

We now extract the term where $\mathscr{S} = \mathscr{V}$ from the right-hand side and obtain the recursive formula for $D_c(\mathscr{V})$:

$$D_c(\mathscr{V}) = D(\mathscr{V}) - \sum_{\mathscr{S}\subsetneq\mathscr{V}} \frac{|\mathscr{S}|}{|\mathscr{V}|} D_c(\mathscr{S}) D(\mathscr{V}\setminus\mathscr{S}). \tag{7.17}$$

The initial condition is the zeroth order contribution $D_c(\emptyset) = 0$. Eqs. (7.17) and (7.7) allow the computation of connected diagrams as a hierarchy of determinants at a cost of $\mathcal{O}(3^k)$.

A more general framework of deriving the recursion relations using idempotent polynomials is described in Ref. [111]. This framework does not resort to topological arguments.

We note that even in simple cases, the convergence radius R of the series (7.13) is not

infinite, with the value of R depending on h, U, and β. Whenever $R < 1$, an order-by-order summation of the series will fail. We will discuss strategies to extend the convergence radius in Sec. 7.2.5.

For convergent series ($R > 1$), one can employ the DiagMC algorithm to sample the series (7.13) by generating random vertices and computing the weight using the recursion (7.17). One observes that the relative statistical error diverges exponentially with diagrammatic order k [113], which requires truncation of the series to a finite order $k_{\max}$.

7.2.3 Scattering amplitude expansion

Other than free energy, we are primarily interested in thermal correlation function of some operators $(\hat{X}_1, \ldots, \hat{X}_m)$:

$$\langle \hat{X}_1 \ldots \hat{X}_m \rangle := \frac{1}{Z} \mathrm{Tr}[\mathrm{e}^{-\beta(\hat{H}-\mu\hat{N})} \mathscr{T}(\hat{X}_1 \ldots \hat{X}_m)], \tag{7.18}$$

in particular the single-particle Green's function:

$$G_{ba}(\tau) = -\langle \hat{c}_b(\tau) \hat{c}_a^\dagger(0) \rangle. \tag{7.19}$$

One can write down a diagrammatic expansion for the Green's function similar to Eq. (7.13) and a corresponding recursion relation [22]. We instead choose to perform the expansion for a vertex-like object.

In the case of the expansion of the free energy, the corresponding one-particle vertex is the scattering amplitude M [9, 12], defined as:

$$G(\tau) = g(\tau) + \int_0^\beta \mathrm{d}\tau_1 \mathrm{d}\tau_2 \; g(\tau - \tau_1) M(\tau_1 - \tau_2) g(\tau_2), \tag{7.20}$$

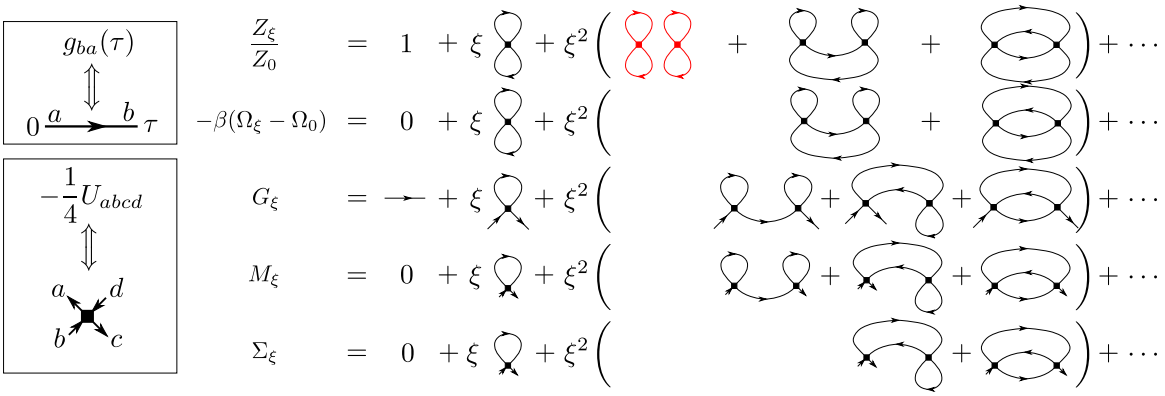

Figure 7.1: Schematic example of diagrams up to order 2. Diagrams shown here should be understood as "labeled" diagrams as described in Ref. [39]. Duplicate diagrams with the same topology are not shown. In the expansion of Z_ξ, the red diagram is an example of disconnected diagram, which is absent in the expansion of Ω due to linked cluster theorem.

where multiplication is to be understood as matrix-matrix multiplication in spin-orbitals. Sampling a one-particle vertex is advantageous because it is independent of the choice of "external legs" and thus allows measurements of both imaginary time-dependent quantities (G and Σ) and fixed-time quantities (density, kinetic energy, etc.) in the same simulation.

M arises naturally as a functional derivative of the grand potential:

$$M_{ab}(\tau) = \frac{\delta(\Omega - \Omega_0)}{\delta g_{ba}(-\tau)}. \tag{7.21}$$

We show this relation in Appendix A. Eq. (7.21) expresses the fact that by removing one line from a (closed) free-energy diagram, we get an interaction correction to the Green's function, which is exactly what the scattering amplitude encodes.

Combining Eq. (7.21) with Eq. (7.13) yields a series expansion for M:

$$M_{\xi,ab}(\tau) = -\frac{1}{\beta}\sum_{k=1}^{\infty}\xi^k \int \mathrm{d}^k \mathscr{V} \frac{\delta D_c(\mathscr{V})}{\delta g_{ba}(-\tau)}. \tag{7.22}$$

We thus need to evaluate the functional derivative of the recursion relation (7.17).

We start with the derivative of the sum of all diagrams $D(\mathscr{V})$, where we rely on the following identity:

$$\frac{\delta \det A}{\delta A_{\alpha\beta}} = (\mathrm{adj}\, A)_{\beta\alpha} := (-1)^{\alpha+\beta} \det A_{\bar{\alpha}\bar{\beta}}, \tag{7.23}$$

where A is an $n \times n$ matrix, $\mathrm{adj}(A)$ denotes the $n \times n$ adjugate matrix of A, and $A_{\bar{\alpha}\bar{\beta}}$ is the $(n-1)\times(n-1)$ submatrix of A with the α-th row and β-th column removed. The adjugate matrix $\mathrm{adj}\, A$ can be computed in $\mathcal{O}(n^3)$ time. The adjugate (or cofactor) matrix arises naturally in determinantal methods as a result of the Wick's theorem [9, 12, 114], and is often absorbed into the inverse matrix if the matrix A is not singular. In the context of CDet, however, care must be taken because A may be singular while $\mathrm{adj}\, A$ is still meaningful [115]. We elaborate on the numerical calculation in Appendix C.

Combining Eq. (7.7) with Eq. (7.23), we have

$$\begin{aligned}
\frac{\delta D(\mathscr{V})}{\delta g_{ba}(-\tau)} &= \prod_{i=1}^{k} \left(-\frac{U_{a_i b_i c_i d_i}}{4} \right) \sum_{\alpha,\beta}^{2k} [\mathrm{adj}\, G(\mathscr{V})]_{\alpha\beta} \delta_{a_\alpha a} \delta_{b_\alpha b} \\
&\quad \times [\delta(\tau_\alpha - \tau_\beta - \tau) - \delta(\tau_\alpha - \tau_\beta + \beta - \tau)] \\
&= -\sum_{\alpha,\beta}^{2k} [\mathbf{A}(\mathscr{V})]_{\alpha\beta} \delta_{a_\alpha a} \delta_{b_\alpha b} [\delta(\tau_\alpha - \tau_\beta - \tau) - \delta(\tau_\alpha - \tau_\beta + \beta - \tau)]
\end{aligned} \tag{7.24}$$

for $0 < \tau \leq \beta$, where a_α, b_β, and $\tau_{\alpha(\beta)}$ takes the same meaning as in Eq. (7.9), and we have defined the $2k \times 2k$ matrix

$$\mathbf{A}(\mathscr{V}) := -\prod_{i=1}^{k} \left(-\frac{U_{a_i b_i c_i d_i}}{4} \right) \mathrm{adj}\, \boldsymbol{G}(\mathscr{V}), \tag{7.25}$$

which includes all connected and disconnected amputated diagrams in which internal legs corresponding to $\hat{c}^\dagger_{a_\alpha}(\tau_\alpha)$ and $c_{b_\beta}(\tau_\beta)$ are removed.

For the functional derivative of a *connected* free-energy diagram (7.26), the sum over all

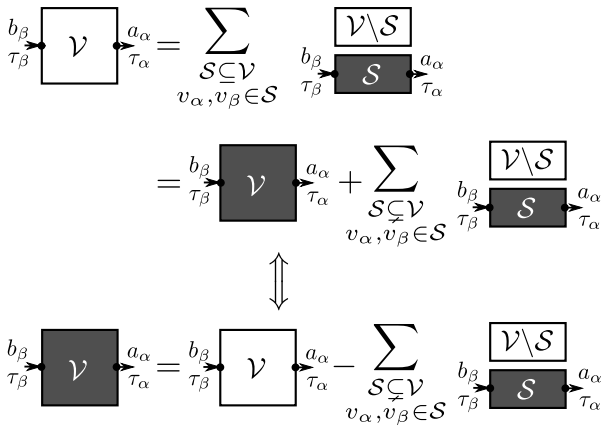

Figure 7.2: Schematic illustration of the recursive removal of disconnected amputated diagrams. Empty boxes stand for the contribution of all diagrams (D or $\mathbf{A}$), and filled ones for that connected diagrams only ($\mathbf{A}_c$). Symbols inside boxes denote the set of vertices included in each component. The top relation shows all partitions of $[\mathbf{A}(\mathscr{V})]_{\alpha\beta}$ into a subset fully connected to the amputated legs and a disconnected complement set. It is reorganized as the bottom relation which recursively defines $\mathbf{A}_c(\mathscr{V})$ by removing all disconnected components.

diagrams in Eq. (7.24) with amputated legs $\mathbf{A}(\mathscr{V})$ needs to be replaced with the sum over connected diagrams with amputated legs $\mathbf{A}_c(\mathscr{V})$:

$$\begin{aligned}\frac{\delta D_c(\mathscr{V})}{\delta g_{ba}(-\tau)} = -\sum_{\alpha,\beta}^{2k}\big\{&[\mathbf{A}_c(\mathscr{V})]_{\alpha\beta}\delta_{a_\alpha a}\delta_{b_\alpha b}\\ &\times[\delta(\tau_\alpha-\tau_\beta-\tau)-\delta(\tau_\alpha-\tau_\beta+\beta-\tau)]\big\}\end{aligned} \tag{7.26}$$

for $0<\tau\leq\beta$. The expansion of M (7.22) can now be expressed in terms of $\mathbf{A}_c$ as

$$\begin{aligned}[M_\xi(\tau)]_{\alpha\beta} = \frac{1}{\beta}\sum_{k=1}^{\infty}\xi^k\int \mathrm{d}^k\mathscr{V}\sum_{\alpha,\beta}^{2k}\big\{&[\mathbf{A}_c(\mathscr{V})]_{\alpha\beta}\delta_{a_\alpha a}\delta_{b_\alpha b}\\ &\times[\delta(\tau_\alpha-\tau_\beta-\tau)-\delta(\tau_\alpha-\tau_\beta+\beta-\tau)]\big\}.\end{aligned} \tag{7.27}$$

The sum over connected amputated diagrams $\mathbf{A}_c(\mathscr{V})$ can be built up from an recursion technique similar to Eq. (7.17). Defining v_α and v_β as vertices where the α-th and β-th opera-

tors are located, respectively, diagrams in $[\mathbf{A}(\mathscr{V})]_{\alpha\beta}$ can always be partitioned to a connected part which contains v_α and v_β, and the disconnected vacuum diagrams, i.e.

$$[\mathbf{A}(\mathscr{V})]_{\alpha\beta} = \sum_{\substack{\mathscr{S}\subseteq\mathscr{V} \\ v_\alpha, v_\beta \in S}} [\mathbf{A}_c(\mathscr{S})]_{\alpha'\beta'} D(\mathscr{V}\backslash\mathscr{S}), \tag{7.28}$$

where α', β' are row and column indices within $\mathscr{S}$ that correspond to the row and column indices α, β in $\mathscr{V}$. Extracting the term with $\mathscr{S} = \mathscr{V}$, we have the recursion relation for $\mathbf{A}_c$:

$$[\mathbf{A}_c(\mathscr{V})]_{\alpha\beta} = [\mathbf{A}(\mathscr{V})]_{\alpha\beta} - \sum_{\substack{\mathscr{S}\subsetneq\mathscr{V} \\ v_\alpha, v_\beta \in S}} [\mathbf{A}_c(\mathscr{S})]_{\alpha'\beta'} D(\mathscr{V}\backslash\mathscr{S}). \tag{7.29}$$

This partitioning process is illustrated in Fig. 7.2. Since M captures the interaction correction to the Green's function which starts at the first order in interaction, the zeroth order contribution $\mathbf{A}_c(\emptyset) = 0$. For each fixed $\mathscr{V}$, we apply Eq. (7.29) to recursively to compute $\mathbf{A}_c(\mathscr{V})$, which in turn yields M following Eq. (7.27). Algorithmically, Eq. (7.29) can be evaluated by following Algorithm 1. Algorithm 1 runs in $\mathcal{O}(3^k k^2)$ time.

Algorithm 1 Recursive evaluation of $\mathbf{A}_c(\mathscr{V})$

Require: Vertices $\mathscr{V}$, $G(\mathscr{V})$ defined in Eq. (7.10).

function RECURSION($\mathscr{V}$, G)
  if $\mathscr{V} = \emptyset$ **then**
    return $\mathbf{A}_c(\emptyset) = 0$.
  else
    Compute $\mathbf{A}(\mathscr{V})$ from G following Eq. (7.25).
    Initialize $\mathbf{A}_c(\mathscr{V}) \leftarrow \mathbf{A}(\mathscr{V})$.
    for $\mathscr{S} \subsetneq \mathscr{V}$ **do**
      Compute $D(\mathscr{V}\backslash\mathscr{S})$ following Eq. (7.7).
      $\mathbf{A}_c(\mathscr{S}) \leftarrow$ RECURSION($\mathscr{S}$, $G_{[\mathscr{S},\mathscr{S}]}$). ▷ $G_{[\mathscr{S},\mathscr{S}]}$ is the submatrix of G whose rows and columns correspond to the subset $\mathscr{S}$. Same definition applies to $[\mathbf{A}_c(\mathscr{V})]_{[\mathscr{S},\mathscr{S}]}$.
      Subtract $\mathbf{A}_c(\mathscr{S})D(\mathscr{V}\backslash\mathscr{S})$ from $[\mathbf{A}_c(\mathscr{V})]_{[\mathscr{S},\mathscr{S}]}$.
    end for
    return $\mathbf{A}_c(\mathscr{V})$.
  end if
end function

7.2.4 Observables from scattering amplitude

The electron self-energy Σ relates the Green's function G to the non-interacting propagator g via the Dyson's equation (3.34)

$$G(\tau) = g(\tau) + \int_0^\beta \mathrm{d}\tau_1 \mathrm{d}\tau_2 g(\tau - \tau_1)\Sigma(\tau_1 - \tau_2)G(\tau_2). \tag{7.30}$$

The expansion of the self-energy Σ can be interpreted as 1PI amputated diagrams, which stay connected even when any single propagator line is removed (cf. Fig. 7.1). The self-energy is thus not directly sampled, M and Σ are related to each other by [12]

$$\int_0^\beta \mathrm{d}\tau' \Sigma(\tau - \tau') G(\tau') = \int_0^\beta \mathrm{d}\tau' M(\tau - \tau') g(\tau'). \tag{7.31}$$

Replacing G with Eq. (7.20), we have

$$\Sigma^{-1}(\mathrm{i}\omega_n) = M(\mathrm{i}\omega_n)^{-1} + g(\mathrm{i}\omega_n), \tag{7.32}$$

where $X(\mathrm{i}\omega_n)$ denotes the Fourier transform of $X(\tau)$ ($X = \Sigma, M, \ldots$) and $\mathrm{i}\omega_n$ is a fermionic Matsubara frequency.

The one- and two-body contribution to the electronic energy follow from Eqs. (7.30) and (7.20):

$$E = E_0 + E_V, \tag{7.33a}$$

$$E_0 = \langle \hat{H}_0 \rangle = \frac{1}{\beta} \mathrm{Tr}[hG] = \sum_{ab} h_{ab} \rho_{ab}, \tag{7.33b}$$

$$\begin{aligned} E_V &= \langle \hat{H}_V \rangle = \frac{1}{2\beta} \mathrm{Tr}[\Sigma G] \\ &= \frac{1}{2} \int_0^\beta \mathrm{d}\tau \sum_{ab} M_{ab}(\tau) g_{ba}(-\tau). \end{aligned} \tag{7.33c}$$

Here $\rho_{ij} \equiv \langle \hat{c}_i^\dagger \hat{c}_j \rangle$ is the electron density matrix. Note that $\hat{H}_0$ does not include the Hartree and Fock terms of the interaction. See also Appendix B.

7.2.5 Hartree-Fock shifted Hamiltonian

In systems with significant electron-electron correlations where E_V has significant contribution to the full energy E, the perturbation expansions in Eqs. (7.13) and (7.22) may not converge at $\xi = 1$.

In order to achieve better convergence by starting from a "better" non-interacting solution such that $\hat{H}_0$ is closer to $\hat{H}$, we change the partition of the Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}_V$ by adding physically-motivated counter terms to $\hat{H}_0$ and subtracting the same terms from $\hat{H}_V$, following the shifted action formalism described in Section 3.5.

We start by adding the simplest counter term in the quadratic form

$$\Delta\hat{H}_\alpha = \sum_{ab} \alpha_{ab} \hat{c}_a^\dagger \hat{c}_b \tag{7.34}$$

to $\hat{H}_0$ and subtract it from $\hat{H}_V$, such that

$$\hat{H}_{0,\alpha} = \sum_{ab} (h_{ab} + \alpha_{ab}) \hat{c}_a^\dagger \hat{c}_b \tag{7.35}$$

$$\hat{H}_{V,\alpha} = \frac{1}{4} \sum_{abcd} U_{abcd} \hat{c}_a^\dagger \hat{c}_c^\dagger \hat{c}_d \hat{c}_b - \alpha_{ab} \hat{c}_a^\dagger \hat{c}_b. \tag{7.36}$$

The total Hamiltonian $\hat{H} = \hat{H}_{0,\alpha} + \hat{H}_{V,\alpha}$ is unchanged, whereas the perturbation expansion of $\hat{H}_\xi = \hat{H}_{0,\alpha} + \xi \hat{H}_{V,\alpha}$ can be controlled by choosing different α. The counter term need not be quadratic in general. Though quadratic choices are convenient in the determinantal setup, recursion schemes have been developed for general counter terms [26].

The shifted non-interacting propagator

$$g^\alpha(\tau) = [(-\partial_\tau + \mu)\boldsymbol{I} - h - \alpha]^{-1}_{aa'} \tag{7.37}$$

can be seen as a Green's function with an *a priori* self-energy α.

In the molecular context, a significant contribution to electron correlations can be obtained by the HF approximation, which is briefly introduced in Section 5.3.1. We therefore choose α to be the HF self-energy, i.e. Σ_{HF}. Σ_{HF} is given by the self-consistent equations at finite temperature

$$[\Sigma_{\mathrm{HF}}]_{ab} = \sum_{cd} U_{abcd}\rho_{cd}, \tag{7.38a}$$

$$\rho = f(h + \Sigma_{\mathrm{HF}} - \mu\boldsymbol{I}). \tag{7.38b}$$

Here $f(A) = [\boldsymbol{I} + \exp(\beta A)]^{-1}$ is the matrix-valued Fermi distribution function, and μ is the chemical potential which may be adjusted so that the total number of electrons in the system is adjusted to charge neutrality.

Diagrammatically, the Hartree-Fock shift renormalizes the propagators lines to g_α, and an additional effective two-point vertex α has to be included in diagrams. The effective vertex α cancels any diagram which has at least one vertex connecting to itself with *exactly* one propagator line. This removes all "tadpole" diagrams in expansions of G and M, as well as that of Ω except for the first order diagram whose vertex connects to itself with two propagator lines. Fig. 7.3 illustrates the cancellation of such diagrams.

Given a specific set of vertices $\mathscr{V}$, the removal of all tadpole diagrams is achieved by

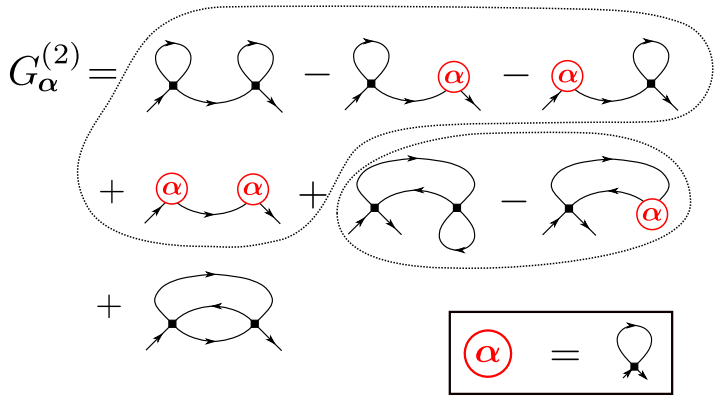

Figure 7.3: Schematic example of diagram cancellations due to the Hartree-Fock counter term. Here we show all second-order Green's function diagrams generated by the counter term, where the red circle indicates the counter term α, each introduces a factor of -1. Terms in each dashed curve cancel each other, leaving only the last term.

replacing the $\boldsymbol{G}$ matrix (7.10) defined on internal vertices $\mathscr{V}$ with:

$$
\begin{aligned}
G(\mathscr{V}) &:= \begin{bmatrix} 0 & \boldsymbol{g}^{\alpha}_{12} & \cdots & \boldsymbol{g}^{\alpha}_{1k} \\ \boldsymbol{g}^{\alpha}_{21} & 0 & \cdots & \boldsymbol{g}^{\alpha}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}^{\alpha}_{k1} & \boldsymbol{g}^{\alpha}_{k2} & \cdots & 0 \end{bmatrix} \\
\boldsymbol{g}^{\alpha}_{ij} &:= \begin{bmatrix} g^{\alpha}_{b_i a_j}(\tau_i - \tau_j + 0^-) & g^{\alpha}_{d_i a_j}(\tau_i - \tau_j + 0^-) \\ g^{\alpha}_{b_i c_j}(\tau_i - \tau_j + 0^-) & g^{\alpha}_{d_i c_j}(\tau_i - \tau_j + 0^-) \end{bmatrix}.
\end{aligned}
\tag{7.39}
$$

i.e. by setting all 2×2 diagonal blocks (corresponding to self-connections of vertices) to zero, and replacing bare propagators with g^{α}. Using the modified definition of $\boldsymbol{G}$ in Eqs. (7.7) and (7.25), one can carry out the same recursive calculations in Eq. (7.26) to obtain corresponding connected quantities.

Note that this introduces a bias in the free-energy evaluation by setting the first order

contribution (the "dumbbell" diagram) to zero, which needs to be corrected:

$$\begin{aligned}\Omega_\alpha^{(1)} &= \frac{1}{2}\sum_{abcd} U_{abcd} g_{ba}^\alpha(0^-) g_{dc}^\alpha(0^+) - \sum_{ab} \alpha_{ab} g_{ba}^\alpha(0^-) \\ &= -\frac{1}{2}\sum_{ab} [\Sigma_{\mathrm{HF}}]_{ab} [\rho_{\mathrm{HF}}]_{ba}.\end{aligned} \tag{7.40}$$

In the remainder of this chapter, we will always use a HF counter term and omit the α subscripts.

7.2.6 Monte Carlo integration of diagrammatic series

Evaluations of diagrammatic series, such as Eqs. (7.13) and (7.27), can be formally summarized as

$$X = \sum_{k=0}^{\infty} \int \mathrm{d}^k \boldsymbol{\mathcal{V}} \mathscr{C}(\boldsymbol{\mathcal{V}}), \tag{7.41}$$

where X is the physical variable (G, M, ...), $\boldsymbol{\mathcal{V}} = (v_1, \ldots, v_k)$ denotes space time indices of internal vertices, and $\mathscr{C}$ the contribution of each fixed configuration of $\boldsymbol{\mathcal{V}}$ to X. Here we take the "physical" value of the coupling constant $\xi = 1$. To perform a Monte Carlo integral, we introduce a cutoff $k_{\max}$ of the expansion order, and an *a priori* probability distribution of vertex space-time indices $p(\boldsymbol{\mathcal{V}})$ such that

$$p(\boldsymbol{\mathcal{V}}) \geq 0, \quad \sum_{k=0}^{k_{\max}} \int \mathrm{d}^k \boldsymbol{\mathcal{V}}\, p(\boldsymbol{\mathcal{V}}) \equiv 1. \tag{7.42}$$

Table 7.1: Measurements for physical observables. The imaginary time convolution is defined as $[f * g](\tau) = \int_0^\beta \mathrm{d}\tau' f(\tau - \tau')g(\tau')$. E_0^{HF} and E_V^{HF} are kinetic and potential energies from the HF solution.

X	$\mathscr{C}(\boldsymbol{\mathscr{V}})$
$M_{ab}(\tau)$	$\frac{1}{\beta}\sum_{\alpha,\beta=1}^{2\lvert\boldsymbol{\mathscr{V}}\rvert}[\mathbf{A}_c(\boldsymbol{\mathscr{V}})]_{\alpha\beta}\delta_{a_\alpha a}\delta_{b_\beta b}[\delta(\tau_\alpha - \tau_\beta - \tau) - \delta(\tau_\alpha - \tau_\beta + \beta - \tau)]$
$M_{ab}(\mathrm{i}\omega_n) = \int_0^\beta \mathrm{d}\tau M_{ab}(\tau)e^{\mathrm{i}\omega_n\tau}$	$\frac{1}{\beta}\sum_{\alpha,\beta=1}^{2\lvert\boldsymbol{\mathscr{V}}\rvert}[\mathbf{A}_c(\boldsymbol{\mathscr{V}})]_{\alpha\beta}\delta_{a_\alpha a}\delta_{b_\beta b}e^{\mathrm{i}\omega_n(\tau_\alpha-\tau_\beta)}$
$G_{ba}(\mathrm{i}\omega_n) - g_{ba}(\mathrm{i}\omega_n) = [gMg]_{ba}$	$\frac{1}{\beta}\sum_{\alpha,\beta=1}^{2\lvert\boldsymbol{\mathscr{V}}\rvert}[\mathbf{A}_c(\boldsymbol{\mathscr{V}})]_{\alpha\beta}g_{ba_\alpha}(\mathrm{i}\omega_n)g_{b_\beta a}(\mathrm{i}\omega_n)e^{\mathrm{i}\omega_n(\tau_\alpha-\tau_\beta)}$
$E_0 - E_0^{\mathrm{HF}}$	$\frac{1}{\beta}\sum_{\alpha,\beta=1}^{2\lvert\boldsymbol{\mathscr{V}}\rvert}[\mathbf{A}_c(\boldsymbol{\mathscr{V}})]_{\alpha\beta}\sum_{ab}h_{ab}[g_{ba_\alpha} * g_{b_\beta a}](\tau_\beta - \tau_\alpha)$
$E_V - E_V^{\mathrm{HF}}$	$\frac{1}{2\beta}\sum_{\alpha,\beta=1}^{2\lvert\boldsymbol{\mathscr{V}}\rvert}[\mathbf{A}_c(\boldsymbol{\mathscr{V}})]_{\alpha\beta}\Big\{g_{b_\beta a_\alpha}(\tau_\beta - \tau_\alpha) + \sum_{ab}[\Sigma_{\mathrm{HF}}]_{ab}[g_{ba_\alpha} * g_{b_\beta a}](\tau_\beta - \tau_\alpha)\Big\}$

In addition, we require that $p(\boldsymbol{\mathscr{V}}) > 0$ whenever $\mathscr{C}(\boldsymbol{\mathscr{V}}) \neq 0$. The order-$k_{\max}$ approximation to X can be estimated stochastically as

$$\begin{aligned} X_{k_{\max}} &= \sum_{k=0}^{k_{\max}} \int \mathrm{d}^k\boldsymbol{\mathscr{V}} \frac{\mathscr{C}(\boldsymbol{\mathscr{V}})}{p(\boldsymbol{\mathscr{V}})} p(\boldsymbol{\mathscr{V}}) = \left\langle \frac{\mathscr{C}}{p} \right\rangle_p \\ &\approx \frac{1}{\mathscr{N}} \sum_{i=1}^{N} \frac{\mathscr{C}(\boldsymbol{\mathscr{V}}_i)}{p(\boldsymbol{\mathscr{V}}_i)}, \quad \boldsymbol{\mathscr{V}}_1, \ldots, \boldsymbol{\mathscr{V}}_N \sim p \end{aligned} \tag{7.43}$$

with a large number $\mathscr{N}$ of Monte Carlo samples $\{\boldsymbol{\mathscr{V}}_i\}$ generated following distribution $p(\boldsymbol{\mathscr{V}})$.

Since the Green's function G, self-energy Σ, as well as the total electronic energy $E = E_0 + E_V$ can all be derived from the scattering matrix M using Eqs. (7.20), (7.32), (7.33), it is sufficient to only keep track of the amputated diagrams $\mathbf{A}_c(\boldsymbol{\mathscr{V}})$ and obtain all other observables as derived quantities. Table 7.1 summarizes some of these measurements. In our implementation, we only measure the energy and $M(\mathrm{i}\omega_n)$ with fermionic Matsubara frequen-

cies $\mathrm{i}\omega_n$ on the fly, and construct $G(\mathrm{i}\omega_n)$ and $\Sigma(\mathrm{i}\omega_n)$ from $M(\mathrm{i}\omega_n)$ following

$$G(\mathrm{i}\omega_n) = g(\mathrm{i}\omega_n) + g(\mathrm{i}\omega_n)M(\mathrm{i}\omega_n)g(\mathrm{i}\omega_n), \tag{7.44}$$

$$\Sigma(\mathrm{i}\omega_n) - \Sigma_{\mathrm{HF}} = [M(\mathrm{i}\omega_n)^{-1} + g(\mathrm{i}\omega_n)]^{-1} \tag{7.45}$$

for each frequency. Resampling techniques such as the jackknife or the bootstrap are applied to avoid biased error estimations.

For efficient Monte Carlo simulations, it is important to choose the *a priori* distribution $p(\mathscr{V})$ to achieve importance sampling, such that the simulation samples more frequently when $|\mathscr{C}(\mathscr{V})|$ is large and less frequently otherwise. Since we measure multiple observables in one simulation, we need to define such a distribution that works for all measurements. We find in practice that the following choices provides efficient samplings for most measurements:

$$p_{\mathbf{A}}(\mathscr{V}) = \frac{\|\mathbf{A}_c(\mathscr{V})\|}{W_{\mathbf{A}}}, \quad W_{\mathbf{A}} = \sum_{k=0}^{k_{\max}} \int \mathrm{d}^k\mathscr{V} \|\mathbf{A}_c(\mathscr{V})\|, \tag{7.46}$$

$$p_E(\mathscr{V}) = \frac{|\epsilon(\mathscr{V})|}{W_E}, \quad W_E = \sum_{k=0}^{k_{\max}} \int \mathrm{d}^k\mathscr{V} |\epsilon(\mathscr{V})|, \tag{7.47}$$

where $\|\cdot\|$ denotes the Frobenius norm of a matrix, and $\epsilon(\mathscr{V})$ is the energy measurement defined in Table 7.1

$$\begin{aligned}\epsilon(\mathscr{V}) =& \frac{1}{2\beta} \sum_{\alpha\beta=1}^{2|\mathscr{V}|} [\mathbf{A}_c(\mathscr{V})]_{\alpha\beta} \Big\{ g_{b_\beta a_\alpha}(\tau_\beta - \tau_\alpha) + \\ &+ \sum_{ab} [2h + \Sigma_{\mathrm{HF}}]_{ab} [g_{ba_\alpha} * g_{b_\beta a}](\tau_\beta - \tau_\alpha) \Big\},\end{aligned} \tag{7.48}$$

where $[f * g](\tau) = \int_0^\beta \mathrm{d}\tau' f(\tau - \tau')g(\tau')$ denotes a convolution in τ. p_E performs well for the

energy measurements, whereas $p_{\mathbf{A}}$ is more robust when measurement of M is needed.

At high expansion order $k_{\max}$, the normalization factors $W_{\mathbf{A}}$ and W_E are difficult to calculate analytically. Instead, we measure an auxillary quantity whose exact value can be calculated analytically, and normalize all other measurements against it. For example, we can normalize against the second-order contribution to the total energy

$$E^{(2)} = \left\langle \frac{\epsilon(\boldsymbol{\mathscr{V}})\delta_{|\boldsymbol{\mathscr{V}}|,2}}{p_E(\boldsymbol{\mathscr{V}})} \right\rangle_{p_E} = W_E \langle \mathrm{sgn}[\epsilon(\boldsymbol{\mathscr{V}})]\delta_{|\boldsymbol{\mathscr{V}}|,2} \rangle_{p_E}. \tag{7.49}$$

Here we have chosen p_E as the *a priori* distribution. Any other measurements can now be estimated as

$$\begin{aligned} X &= \left\langle \frac{\mathscr{C}(\boldsymbol{\mathscr{V}})}{p_E(\boldsymbol{\mathscr{V}})} \right\rangle_{p_E} = W_E \left\langle \frac{\mathscr{C}(\boldsymbol{\mathscr{V}})}{\epsilon(\boldsymbol{\mathscr{V}})} \right\rangle_{p_E} \\ &= E^{(2)} \frac{\langle \mathscr{C}(\boldsymbol{\mathscr{V}})/\epsilon(\boldsymbol{\mathscr{V}}) \rangle_{p_E}}{\langle \mathrm{sgn}[\epsilon(\boldsymbol{\mathscr{V}})]\delta_{|\boldsymbol{\mathscr{V}}|,2} \rangle_{p_E}}. \end{aligned} \tag{7.50}$$

Similar relations apply when we use other choices of *a priori* distributions or normalization measurements.

Once $p(\boldsymbol{\mathscr{V}})$ is defined, we generate Monte Carlo samples as a Markov chain using the Metropolis–Hastings algorithm detailed in Chapter 4. We adopt the definition II which treats $\boldsymbol{\mathscr{V}}$ as a set of vertices, and the integration measure includes the permutation sign $1/k!$.

In molecular systems, due to the complexity in the multi-orbital Coulomb interaction tensor, as well as the energy differences in non-interacting energy levels, the configuration space of the Monte Carlo can be uneven and may lead to ergodicity problems in the random walk. We design the following set of updates which lead to an ergodic random walk in the configuration space for all systems we investigate in Sec. 7.3.

1. Vertex splitting: Split a random vertex $v = (a, b, c, d; \tau)$ to two new vertices $v_1 =$

$(a, b, c', d'; \tau)$ and $v_2 = (a', b', c, d; \tau')$. The new indices a', b', c', d', and τ' can be proposed by some *a priori* probability p^{ins}. The proposal probability distribution for this update from order k to $k+1$ is

$$w^{\text{prop}}(v_1, v_2; k+1|v; k) = \frac{p^{\text{ins}}(a', b', c', d', \tau')}{k} \tag{7.51}$$

2. Vertex merging: Pick two random vertices $v_1 = (a, b, c', d'; \tau)$ and $v_2 = (a', b', c, d; \tau')$ and merge them into $v = (a, b, c, d; \tau)$. The proposal probability distribution from order $k+1$ to k is

$$w^{\text{prop}}(v; k|v_1, v_2; k+1) = \frac{1}{k(k+1)}. \tag{7.52}$$

3. Vertex shift in time: Update the time label τ of a vertex v to a new value τ'.

4. Vertex shift in orbitals: Update one of the orbital labels a, b, c, d of a vertex v to a random new value.

Vertex shift in time or orbitals are self-balancing moves, hence the acceptance ratios shares the same form

$$R(\mathscr{V}_2|\mathscr{V}_1) = \frac{p(\mathscr{V}_2)}{p(\mathscr{V}_1)}. \tag{7.53}$$

Vertex splitting and merging are mutually inverse updates. The acceptance ratios are therefore

$$\begin{aligned} &R(v_1, v_2; k+1|v; k) = R(v; k|v_1, v_2; k+1)^{-1} \\ =&\frac{w^{\text{prop}}(v; k|v_1, v_2; k+1)p(v_1, v_2; k+1)}{w^{\text{prop}}(v_1, v_2; k+1|v; k)p(v; k)} \\ =&\frac{k+1}{p^{\text{ins}}(a', b', c', d', \tau')}\frac{p(v_1, v_2; k+1)}{p(v; k)}. \end{aligned} \tag{7.54}$$

There is considerable freedom in choosing p^{ins}. For all systems we study in this work, we choose p^{ins} such that

$$p^{\text{ins}}(a', b', c', d', \tau') = p^{\text{orb}}(a', b', c', d')p^{\text{time}}(\tau'), \tag{7.55}$$

where $p^{\text{orb}}(a', b', c', d')$ is uniformly distributed if the inserted indices can form non-zero propagator connections and zero otherwise, and

$$p^{\text{time}}(\tau') = \varphi(|\tau' - \bar{\tau}[\mathscr{V}]|). \tag{7.56}$$

where $\bar{\tau}[\mathscr{V}] = \frac{1}{k}\sum_{i=1}^{k} \tau_k$ is the average time coordinate of the existing vertices, and we choose $\varphi(\tau)$ as a function in $[0, \beta]$ which has more weight near $\tau = 0$ and β but still non-negligible weight in between. Since the HF propagators decay exponentially away from 0 and β, this makes sure that the new vertex are more likely to stay close to existing vertices so that the resulting configuration has sizable contribution. In our implementation, we define

$$f(\tau) = \frac{\lambda}{\arctan(\beta\lambda)}\left\{\frac{1}{1 + (\tau\lambda)^2} + \frac{1}{1 + [(\beta - \tau)\lambda]^2}\right\}, \tag{7.57}$$

as a Lorentzian distribution where λ is an estimation of the overall energy scale of the system proportional to e.g. the standard deviation of the HF energy levels.

7.3 Results

7.3.1 Series convergence

We first present a test of our method on a minimal molecular system: H_2 in the STO-6g basis set [42]. Two hydrogen atoms are placed at distance r and finite temperature $T = 1/\beta$. The basis set only contains the 1s orbital in each atom. This setup allows us to easily perform exact diagonalization (ED) calculations of the full molecular Hamiltonian at any temperature, such that exact benchmark results for our CDet results are available.

In Fig. 7.4, we compare the total energy E_{tot} from CDet with order truncation $k_{\max}$ up to 6 to the ED energy at $T = 50^{-1}$ E_h, both as a function of r. Around equilibrium distance $r \approx 1.4$ a_0, the CDet energy converges well to the ED solution. The system moves to the strongly correlated regime (i.e. a regime far from the HF solution), as we "stretch" the molecule by increasing r. At $r > 2.0$ a_0 we start to observe significant systematic deviation at $k_{\max} = 6$. Since the kinetic energy of electrons moving between two atoms is significantly reduced as we increase r but the long-range Coulomb repulsion between electrons changes slowly, the electron-electron interaction becomes more important at larger r, and hence it is expected that the perturbation expansion becomes more difficult to converge. This setup is standard in quantum chemistry [38] and is similar in spirit to lattice model setups in which a metal-to-insulator transition is induced by gradually increasing an on-site interaction.

Analytically, the convergence behavior is determined by the properties of the expanded quantity (e.g. $E[\xi]$) as a function of the coupling constant ξ on the complex plane, similar to the convergence analysis for many-body perturbation theory (MBPT) calculations at $T = 0$ [116–120]. We evaluate the electron energy $E[\xi]$ for complex values of ξ near $\xi = 0$ using

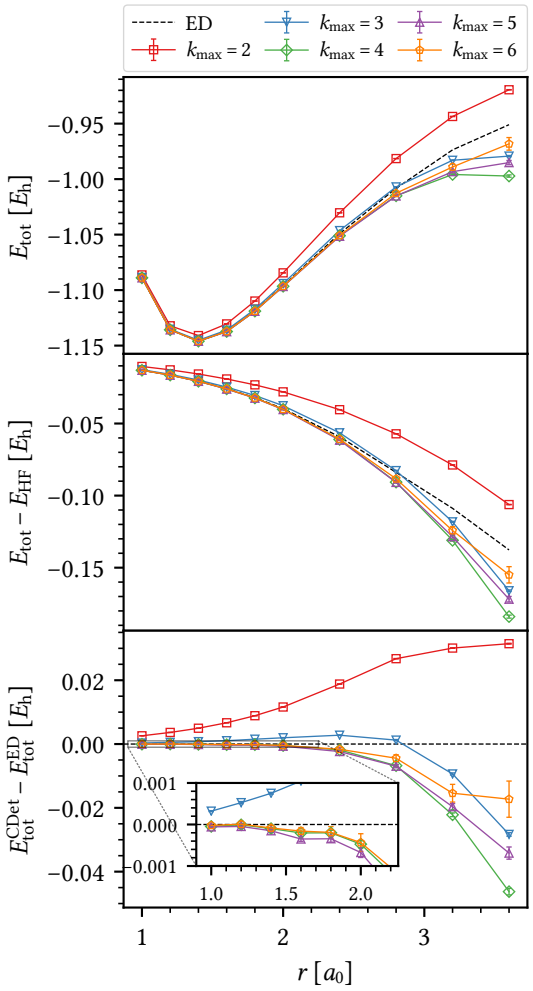

Figure 7.4: Total energy E_{tot} with Monte Carlo errors for H_2, STO-6g, $T = 50^{-1}$ E_h. Top panel: comparison of ED and CDet at different k_{max}. Middle panel: total energy with HF contribution removed. Bottom panel: difference between ED and CDet.

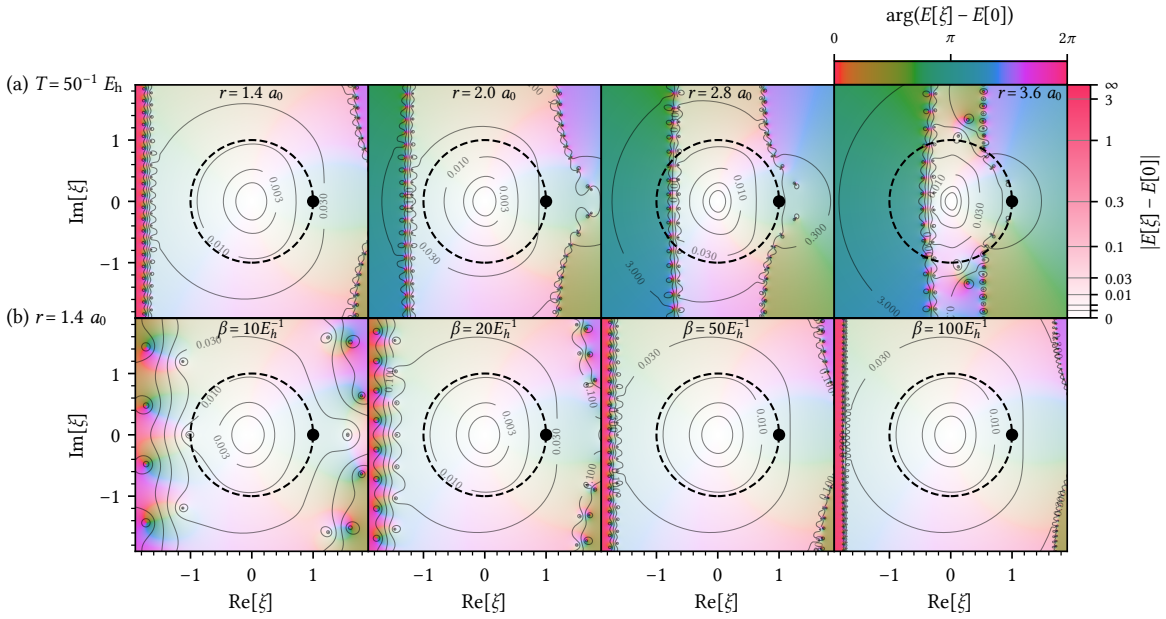

Figure 7.5: Analytic structure of electron total energy evaluated with ED as a function of the complex coupling constant ξ. H_2, STO-6g. Colors represent complex phases and brightness indicates the magnitude (see color bars). The black dot at $\xi = 1$ represents the "physical" result. The dashed black circle indicates minimum convergence radius necessary for the perturbation series to converge at $\xi = 1$. (a) Effect of changing r at fixed temperature $T = 50^{-1}$ E_h. At $r = 1.4$ a_0 and $r = 2.0$ a_0, no singularity is visible in the unit circle and the series is convergent at $\xi = 1$. At $r = 2.8$ a_0 and 3.6 a_0, poles appear in the unit circle, resulting in a divergent series at $\xi = 1$. (b) Effect of changing T at fixed $r = 1.4$ a_0. The real-axis locations of the vertical "walls" of poles does not change significantly as temperature decreases, while the imaginary-axis spacing of the poles decreases proportionally with T.

ED at $r = 1.4$, 2.0, 2.8, and 3.6 a_0, following

$$Z[\xi] = \mathrm{Tr}\left\{e^{-\beta[(\hat{H}_0+\hat{H}_\alpha)+\xi(\hat{H}_V-\hat{H}_\alpha)-\mu\hat{N}]}\right\}, \tag{7.58}$$

$$E[\xi] = \frac{1}{Z[\xi]}\,\mathrm{Tr}\left\{[(\hat{H}_0 + \hat{H}_\alpha/2) + \xi(\hat{H}_V - \hat{H}_\alpha/2)] \times e^{-\beta[(\hat{H}_0+\hat{H}_\alpha)+\xi(\hat{H}_V-\hat{H}_\alpha)-\mu\hat{N}]}\right\}, \tag{7.59}$$

where $\hat{H}_\alpha$ is the HF counter term introduced in Sec. 7.2.5. One can show via a straight forward substitution that $E[1]$ gives the "physical" electron energy and $E[0]$ recovers the HF energy. Figure 7.5.a shows the interaction correction $E[\xi] - E[0]$ to the total energy, where the black dot represents the physical value at $\xi = 1$. Since the convergence radius of the power series around $\xi = 0$ is determined by the singularity (pole or branch cut) closest to the origin, the series is convergent at the "physical" point $\xi = 1$ if and only if there are no singularities in the unit circle (dashed circles in Fig. 7.5). At $r = 1.4$ a_0, all poles are far outside the unit circle, indicating a rapidly convergent series. As we increase r, poles move closer to the unit circle at $r = 2.0$ a_0, implying a slower convergence of the series, and finally enter the unit circle at $r = 2.8$ a_0 and 3.6 a_0, resulting in divergent series at $\xi = 1$.

The analytic properties are reflected directly in the convergence behavior of the CDet results. For a direct comparison, we calculate the contribution of each order k to the total energy $E_{\mathrm{tot}}^{(k)}$ up to $k_{\max} = 8$ for the same values of r, as shown in Fig. 7.6. At $r = 1.4$ a_0, $E_{\mathrm{tot}}^{(k)}$ quickly converges to zero at $k > 4$. At $r = 2.0$ a_0, we observe tendency to converge at $k = 8$ but non-zero systematic deviations remain. For $r = 2.8$ a_0 and 3.6 a_0, no signs of convergence are observed up to $k = 8$.

The CDet approach can be applied to different temperatures without adding significant computational cost, as we will show in Sec. 7.3.2. This is fundamentally different from methods such as CT-QMC, where reaching lower T is only possible at an exponential cost away

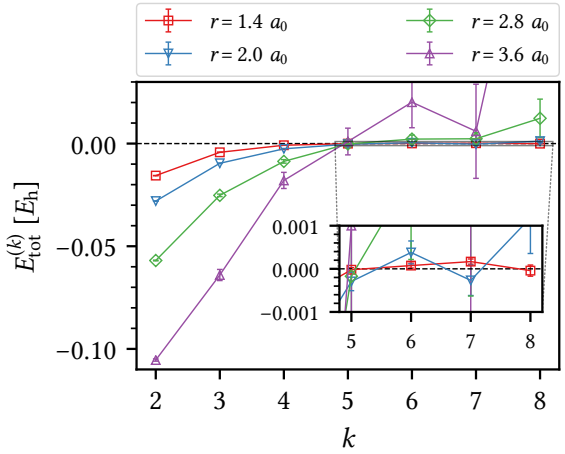

Figure 7.6: Contribution $E^{(k)}_{\mathrm{tot}}$ of each order k to total energy. H_2, STO-6g, $T = 50^{-1}$ E_h. Convergence is observed at bond lengths $r \leq 2.0$ a_0 but not at $r \geq 2.8$ a_0.

from half filling [14]. In Fig. 7.7, we show the temperature dependence of the CDet total energy for H_2, STO-6g at $k_{\mathrm{max}} = 6$, from $T = 10.0^{-1}$ E_h down to $T = 500.0^{-1}$ E_h, in comparison to the ED solution at $T = 0$. All calculations use the same algorithmic setup and the same number of Monte Carlo steps. Convergence to the zero-temperature solution is observed as T decreases, while the stochastic error estimation does not change significantly. Systematic deviations can be observed at similar locations ($r > 2.0$ a_0) for different temperatures, indicating similar convergence behavior for the same system at different temperature. This can be shown by the temperature dependence of the analytic structure of $E[\xi]$, as plotted in Fig. 7.5.b. As temperature is reduced, the spacing of the poles along the imaginary direction decreases proportionally, but the real-axis locations of the vertical "walls" of poles stay almost unchanged, which leads to similar convergence radii at different temperature.

The HF shifted action plays an important role in achieving better series convergence in CDet. Figure 7.8 compares the ED analytic structure of the total energy $E[\xi]$ with and without the HF shift. Without the shift, even for the equilibrium distance $r = 1.4$ a_0 (usually con-

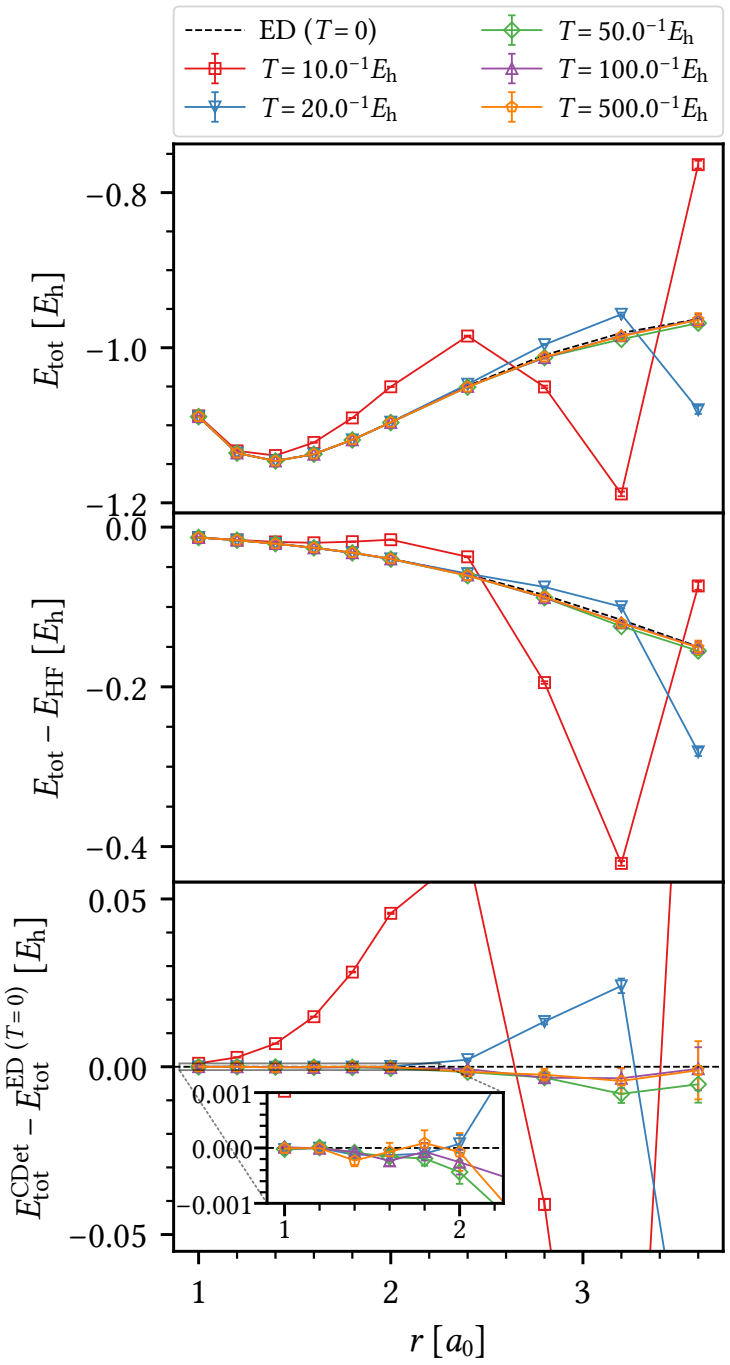

Figure 7.7: Temperature dependence of CDet total energy E_{tot} for H_2, STO-6g, $k_{max} = 6$. Top panel: comparison of ED at $T = 0$ and CDet at different T. Middle panel: total energy with HF contribution removed. Bottom panel: difference between ED and CDet.

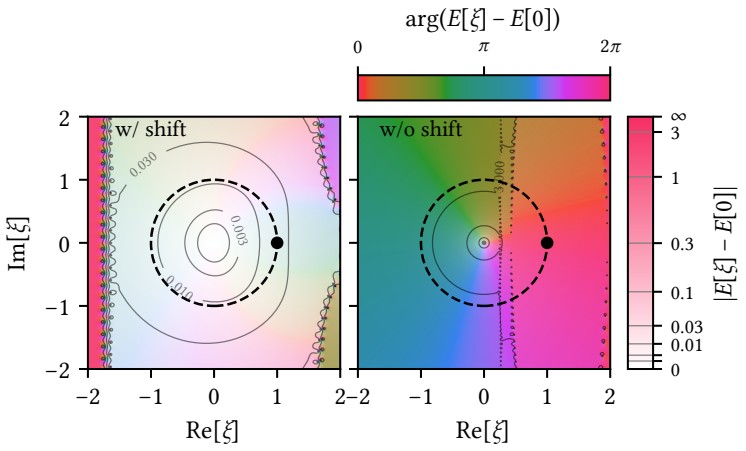

Figure 7.8: Analytic structure of electron total energy as a function of a complex coupling constant ξ with and without HF shift. Left (right) panel presents values of $E[\xi] - E[0]$ evaluated on the complex plane for H_2, STO-6g, $T = 50^{-1}$ E_h and $r = 1.4$ a_0 with (without) HF shift.

sidered "weakly correlated"), there are poles deep inside the unit circle, implying a highly divergent series at $\xi = 1$. In contrast, the HF shift pushes the poles away from the origin, which leads to a convergent series as seen in Fig. 7.4 and Fig. 7.6.

The CDet approach gives access to dynamic quantities, such as the Green's function G and the self-energy Σ, through the scattering amplitude M. The left column of Fig. 7.9 shows the CDet measurement of $\hat{M}(i\omega_n)$ in Matsubara frequency space up to $k_{max} = 6$ for H_2, STO-6g at $r = 1.4$ a_0 and $T = 50^{-1}$ E_h. As we increase the expansion order, CDet results gradually converge to the ED solution (black lines), and at order 6 we observe only a small systematic error due to order truncation. The CDet self-energy Σ is calculated from M following Eq. (7.45). Both quantities exhibit similar behavior, as shown in the right column of Fig. 7.9. At order 3 and higher, the real part of $\hat{\Sigma}(i\omega_n)$ takes non-zero value at high-frequency limit, corresponding to the correction to the frequency-independent HF self-energy Σ_{HF}. The CDet Green's function, derived from M following Eq. (7.44), is shown in Fig. 7.10. Good agreement with ED is observed at $k_{max} = 6$ on the top panel, where both the Monte Carlo error estimation

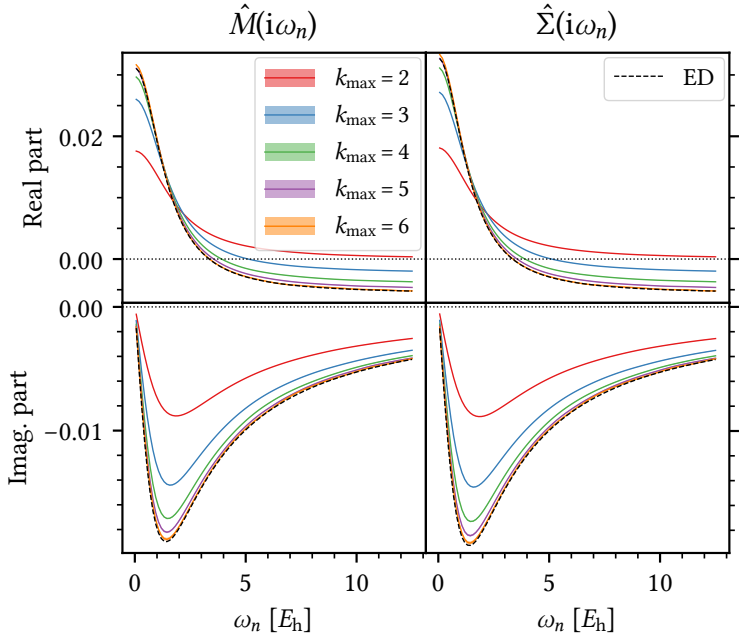

Figure 7.9: CDet dynamic quantities for H_2, STO-6g, $T = 50^{-1}\,\mathrm{E_h}$, $r = 1.4\,\mathrm{a_0}$. Shadings indicate Monte Carlo error estimates. Left column: Measured CDet $\mathrm{Re}M(\mathrm{i}\omega_n)$ (top panel) and $\mathrm{Im}M(\mathrm{i}\omega_n)$ (bottom panel) compared to ED. Right column: CDet self-energy in comparison to ED (excluding HF contribution Σ_{HF}), top (bottom) panel showing real (imaginary) part of $\Sigma(\mathrm{i}\omega_n)$. Here we show the diagonal matrix element at orbital 1 for both M and Σ.

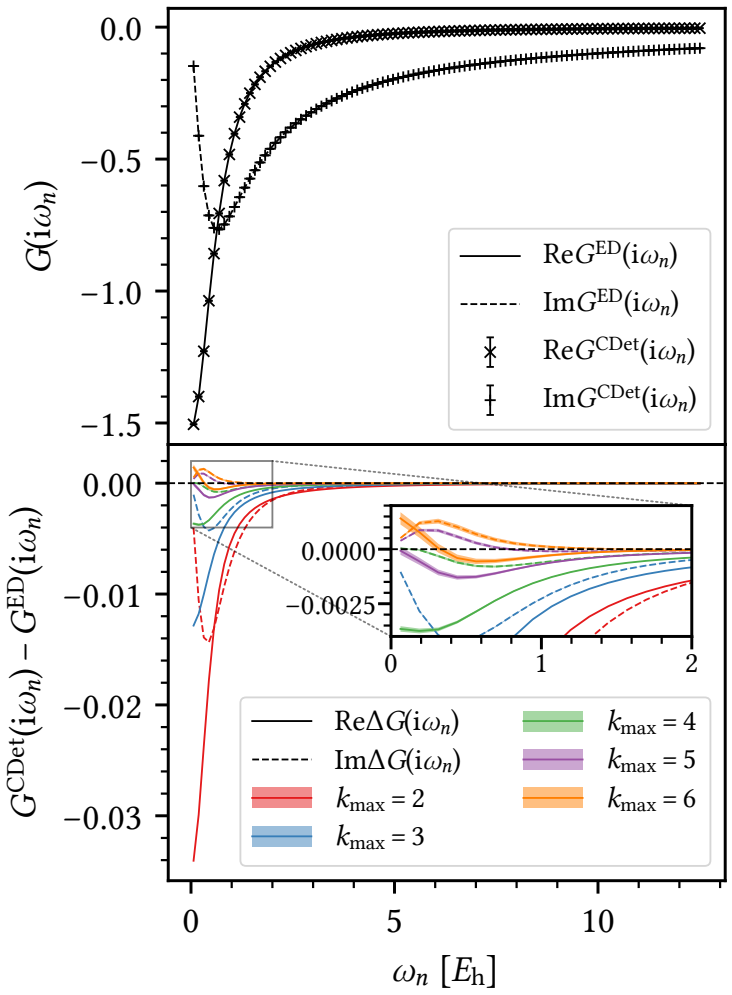

Figure 7.10: CDet Green's function in comparison to ED. H_2, STO-6g, $T = 50^{-1}$ E_h, $r = 1.4$ a_0 Top panel: values of $\hat{G}(\mathrm{i}\omega_n)$ at orbital 1. CDet results at $k_{\mathrm{max}} = 6$ are plotted as symbols and ED values as lines. Error bars are indicated but much smaller than symbol size. Bottom panel: deviations of CDet results from ED at different k_{max}. Solid (dashed) lines represent real (imaginary) part of $\hat{G}_{11}(i\omega_n)$. Shadings indicate stochastic uncertainties of CDet.

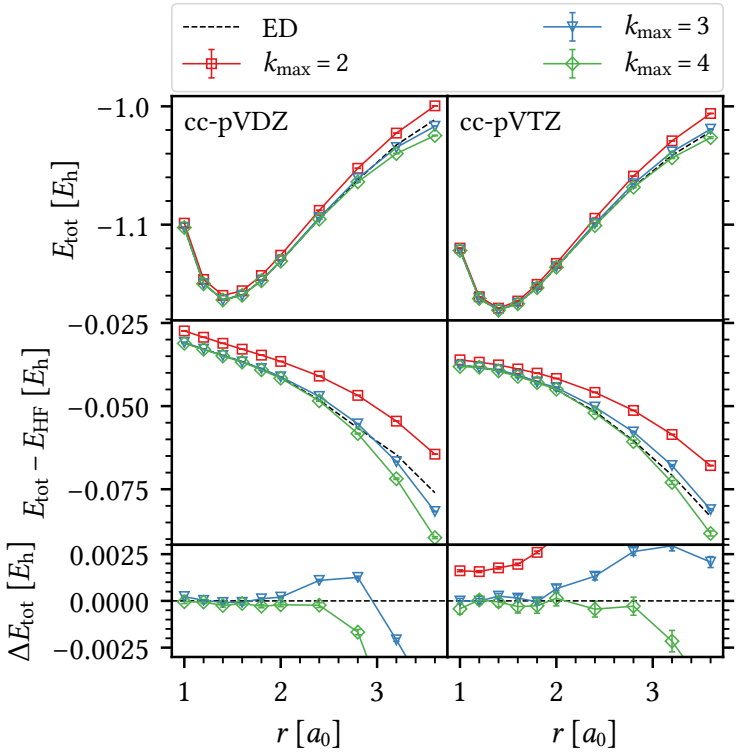

Figure 7.11: Total energy E_{tot} with Monte Carlo errors for H_2 with cc-pVDZ (left column) and cc-pVTZ(right column) basis sets, $T = 50^{-1}$ E_h. Top panels: comparison of ED and CDet at different k_{max}. Middle panel: total energy with HF contribution removed. Bottom panels: difference between ED and CDet.

and the systematic error due to order truncation is much smaller than the symbol size. The bottom panel shows convergence of CDet Green's function to ED by increasing k_{max}, with a small but visible systematic deviation at low frequency when $k_{max} = 6$.

The generality of our CDet implementation allows a straightforward extension to much larger basis sets. Going beyond the minimal basis, we compute the CDet total energy of H_2 using cc-pVDZ and cc-pVTZ basis sets with 10 and 28 orbitals in total, respectively, and compare to the ED solution as shown in Fig. 7.11. For $r < 2.0$ a_0, CDet gives decent convergence to ED at $k_{max} = 4$, with both stochastic and systematic error below 1 mE_h. The 2 s and 2 p orbitals added by cc-pVDZ basis and 3 s, 3 p and 3 d orbitals by cc-pVTZ basis are mostly un-

occupied, and the most electron excitation occur near the lowest 1 s orbitals. Consequently, the convergence behavior and computational cost of CDet do not change significantly from the minimal basis STO-6g.

Finally, we extend our method to bigger molecules by adding more hydrogen atoms to the system. We consider a chain of 10 hydrogen atoms on a straight line with equal spacing r, the same benchmark system used in Ref. [44]. At minimal basis STO-6g, all ten 1 s orbitals contribute equally to the active space of 10 electrons. Compared to H_2 with cc-pVDZ, which has the same number of orbitals, H_{10} with STO-6g has more orbitals relevant to electron correlations, and the cost of CDet is higher (for a detailed analysis see Sec. 7.3.2). The left column of Fig. 7.12 plots the CDet total energy up to $k_{\max} = 4$ in comparison to ED solution at $T = 50^{-1}$ E_h. Convergence within 5 mE_h is achieved at $k_{\max} = 4$ for $r < 2.4$ a_0, and systematic deviations are evident for $r > 2.4$ a_0. Similar behavior can be found in the the zero-temperature coupled cluster (CCSD) result (dotted lines), as both methods rely on the perturbative expansions of electron-electron interactions in different forms. The computational cost becomes much higher as we go to a bigger basis for H_{10}. With cc-pVDZ, there are 50 atomic orbitals in total, with potential excitations to the empty orbitals from all 10 electrons. As shown in the right column Fig. 7.12, CDet still agrees with the reference method (MRCI+Q data from Ref. [44] at $T = 0$) for small values of r, but the Monte Carlo errors are significantly larger. Although our generic implementation has achieved decent extensibility without fine-tuning for each specific system, more efficient Monte Carlo estimators and sampling schemes as well as analytical resummation techniques should advance the limit of CDet to more complex systems.

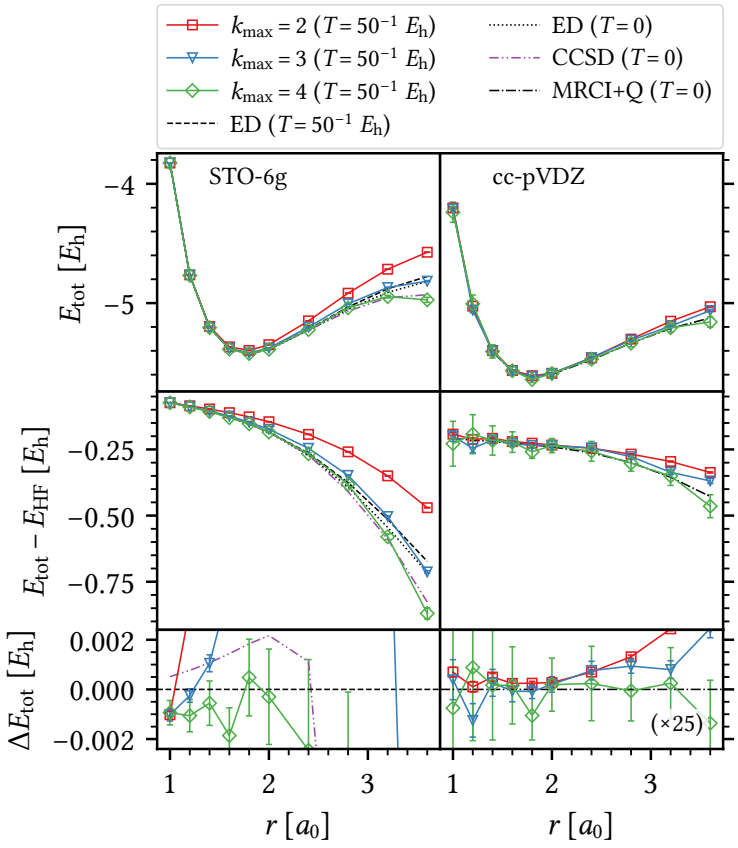

Figure 7.12: Total energy E_{tot} with Monte Carlo errors for H_{10} with STO-6g (left column) and cc-pVDZ (right column) basis. ED results are used as reference for STO-6g and MRCI+Q ($T = 0$) from Ref. [44] for cc-pVDZ. Top panels: comparison of reference data and CDet at different k_{max} at finite temperature $T = 50^{-1}$ E_h, along with ED and CCSD results at $T = 0$ for STO-6g basis. Middle panel: total energy with HF contribution removed. Bottom panels: difference between CDet and reference data at finite temperature, (for STO-6g) in comparison to difference between CCSD and ED at zero temperature.

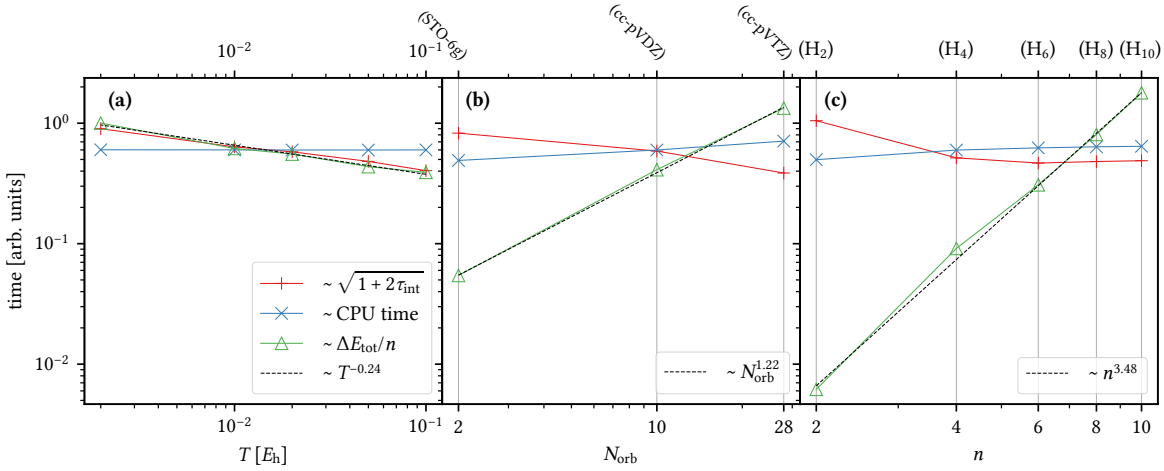

Figure 7.13: Empirical cost analysis of CDet simulations of hydrogen chain H_n at $r = 1.4\ a_0$. In each panel, all simulations are carried out using the same setup of Monte Carlo updates and number of iterations. We estimate the contribution of integrated autocorrelation time τ_{int} to the stochastic error (blue), the computational cost (orange), and the total stochastic uncertainty of energy ΔE_{tot} (green) for each simulation, and scale them to the same range on double-logarithmic plots. (a) Temperature dependence, H_2, STO-6g, $k_{max} = 6$. (b) Basis set dependence, H_2, $T = 50^{-1}\ E_h$, $k_{max} = 4$. (c) System size dependence, H_n, cc-pVDZ, $T = 50^{-1}\ E_h$, $k_{max} = 4$.

7.3.2 Analysis of computational cost

The computational cost of a Markov chain Monte Carlo simulation, measured as the computational time needed for reaching a result for observable X within a desired accuracy ΔX, is determined by three factors. First, the cost of each individual update, which is $O(k^3 2^k)$ for a configuration at expansion order k according to Algorithm 1. Second, the number of configuration updates needed to reach an independent sample by transversing a Markov chain of potentially correlated configurations, described by the integrated autocorrelation time τ_{int}. Finally, the variance $\text{Var}(X)$ of the estimator of the quantity of interest (Table 7.1), such that

$$\Delta X = \sqrt{\frac{\text{Var}(X)}{N}(2\tau_{\text{int}} + 1)}. \tag{7.60}$$

To assess the computational cost of our CDet implementation for reaching a certain uncertainty level, as well as how the effort changes with respect to temperature, choice of basis set, and system size, we perform a series of simulations of convergent series for the hydrogen chain H_n with the same Monte Carlo updates and measurements for a fixed number of Markov chain iterations. In Fig. 7.13, we show estimates of autocorrelation effects, actual computational costs, and stochastic uncertainties in total energy, as functions of temperature T, the number of orbitals N_{orb}, or the number of hydrogen atoms n in log-log plots. We rescale the y-values by an arbitrary factor to emphasize the respective scaling of these quantities in the same plot.

Figure 7.13.a shows the temperature dependence of CDet simulations of a fixed system (H_2, STO-6g, $r = 1.4\ a_0$) at $k_{\text{max}} = 6$. We observe that the simulation time does not change significantly as we decrease temperature, indicating similar distributions of the expansion order (usually tilted to the highest order). The error estimate in total energy follows almost the same tendency as the factor of the autocorrelation effect $\sqrt{2\tau_{\text{int}} + 1}$, indicating the under-

lying energy estimator does not have strong temperature dependence. The autocorrelation effect shows a slow power-law increase as temperature is lowered, implying that our Monte Carlo updates remain efficient at low temperature.

A similar analysis is shown in Fig. 7.13.b for the basis set dependence of the same system (H_2, $r = 1.4\ a_0$) at fixed temperature. We perform CDet simulations with $k_{\max} = 4$ for basis sets STO-6g, cc-pVDZ, and cc-pVTZ, with 2, 10, and 28 atomic orbitals, respectively. As we add more "virtual" orbitals to the system, the computational time increases slowly, and the autocorrelation time even decreases as the additional orbitals improve the connectivity of Monte Carlo configurations. However, the stochastic error shows a different trend from the autocorrelation effect and increases (a fit with a power law results in $\sim N_{\text{orb}}^{1.22}$), meaning that the additional orbitals introduce more diagrammatic configurations with alternating signs that lead to stronger Monte Carlo fluctuations.

As we increase the systems size in Fig. 7.13.c by adding more hydrogen atoms, the stochastic error (normalized by the system size n) at fixed computational time increases with a much larger power law than Fig. 7.13.b (fitted $\sim n^{3.48}$), while the autocorrelation time barely changes. This implies that adding electrons that contribute to excitations near the Fermi level rapidly increases the complexity of the diagrammatics. The result is very different from the situation where additional basis states for the same number of electrons are added (Fig. 7.13.b).

The behavior illustrated in Fig. 7.13.c also differs from DiagMC with short-range or on-site interactions, which are formulated directly in the thermodynamic limit [18–20] and usually do not show strong scaling dependencies on system size. We suspect the difference is caused by the long range nature of the bare Coulomb interaction, which introduces significant non-local electronic correlations as the system size increases. In this case, the use of "bold" (or "screened") interactions instead of the bare Coulomb interactions, as performed in Ref. [44],

may alleviate the problem. However, "bold" methods must deal with intrinsic issues of misconvergence to unphysical solutions [58]. Moreover, adapting such a method to the CDet framework requires further algorithmic development. This topic is under active development [26, 48].

Thus, through the empirical analysis above, we have shown that for convergent series, the computational cost of our CDet implementation is not very sensitive to changes in temperature or basis sets, but depends strongly on the size of the system, or more specifically, on the number of valence electrons directly participating in electron excitations.

7.3.3 Realistic impurity: SEET for NiO

Finally, we test our CDet implementation in a general quantum impurity problem setup that includes the coupling to a non-interacting bath. We employ the SEET framework [95] for the antiferromagnetic compound NiO, which was studied by Mott [121] as one of the original correlated insulators. Following the computational setup in Ref. [112], we choose fcc NiO with lattice constant $a = 4.1705$ Å at temperature $T \sim 451$ K ($\beta = 700\ \mathrm{E_h}^{-1}$). The unit cell is doubled along the [111] direction to capture the antiferromagnetic ordering, which contains two nickel atoms and two oxygen atoms. We use a $4 \times 4 \times 4$ momentum discretization and the gth-dzvp-molopt-sr basis set [122] with gth-pbe pseudopotential [123]. The Coulomb integral is decomposed using density fitting with the def2-svp-ri auxiliary basis [124]. For benchmark purposes, we select the e_g orbitals of both Ni atoms in the unit cell as the strongly correlated 'impurities', which is the minimal choice of impurities to capture correlation effects. This yields two independent impurities each with two orbitals. 'Non-interacting' impurity propagators are generated from a converged *GW* simulation of the complete unit cell following the SEET framework (for details of the computational setup see

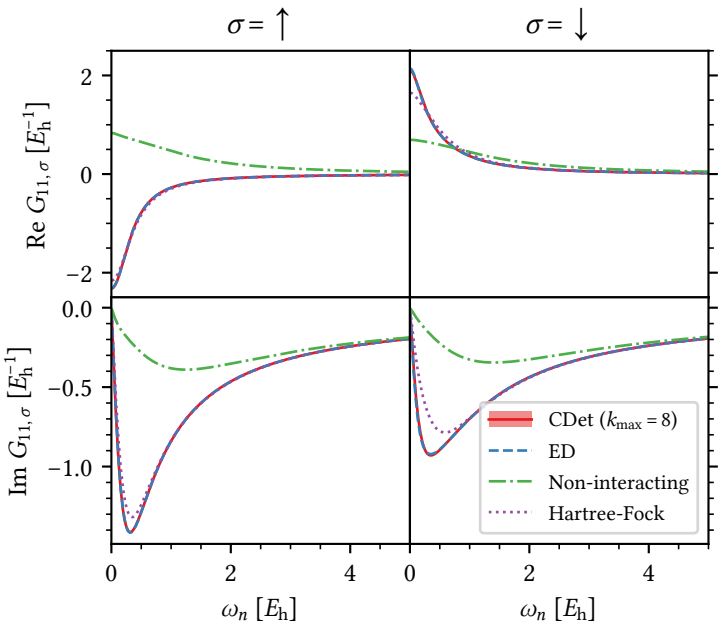

Figure 7.14: Matsubara Green's function for a NiO e_g impurity. Top (bottom) row shows the real (imaginary) part of the Green's functions, and left (right) column shows values for spin up (down). Red solid lines: Impurity Green's function from CDet with $k_{max} = 8$. Monte Carlo error estimations are plotted as color shadings but smaller than line width. Blue dashed lines: Impurity Green's function from ED, mostly overlapping with the CDet lines within line width. Green dash-dotted lines: "Non-interacting" impurity propagator with discretized hybridization from ED. Purple dotted lines: Impurity Green's function with HF counter term as the starting point of CDet.

Ref. [112]).

As a benchmark, we compare our CDet impurity solver to the ED [125] results used in Ref. [112]. ED requires the discretization of the continuous bath spectrum and its approximation by a few states. In order to separate ED bath fitting errors from the performance of the CDet method, we run our method for the "non-interacting" impurity Green's function g corresponding to the discretized non-interacting problem solved by ED. Figure 7.14 shows the impurity Green's functions for one of the two e_g impurities. At $k_{max} = 8$, the impurity Green's function from CDet agrees with the ED solution within line width, and

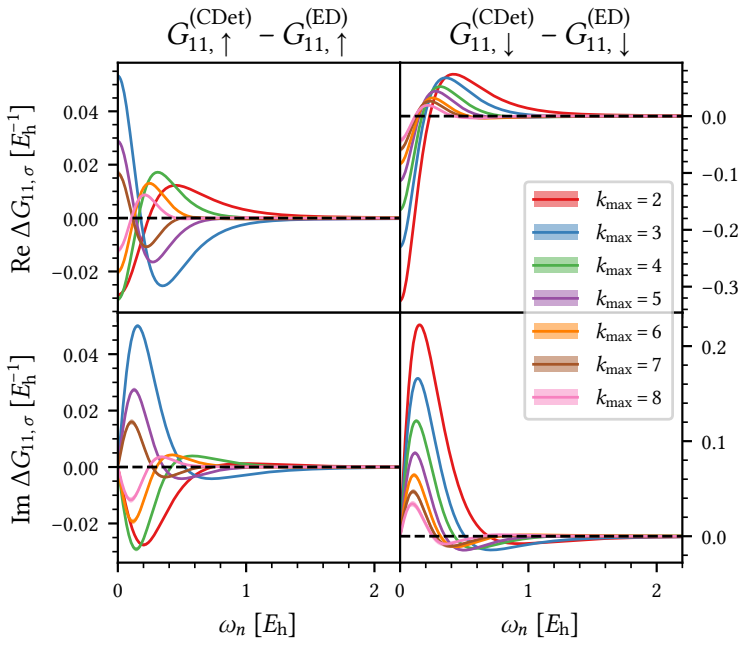

Figure 7.15: Difference of CDet impurity Green's functions to ED at different truncation order k_max. Top (bottom) row shows the real (imaginary) part, and left (right) column shows values for spin up (down). Shadings indicate stochastic uncertainties of CDet.

the stochastic uncertainty is almost invisible. The spin polarization due to the antiferromagnetic ordering is greatly enhanced in both the ED and the CDet solutions, indicating that dynamical correlations plays an important role and are well captured by the selected impurity.

In Fig. 7.15, we take a closer look at the convergence of CDet series in comparison to ED by plotting the differences of CDet impurity Green's functions to ED at different order truncations up to $k_\mathrm{max} = 8$. We observe that for both spins, the CDet result consistently converges to the ED result, giving agreement to within a percent for $k_\mathrm{max} = 8$. The convergence of the spin down component is slower than spin up, which is consistent to what can be observed in Fig. 7.14, i.e. the Hartree-Fock contribution already accounts for a greater part of the overall interaction contribution for the spin up component than for spin down.

The comparison to ED illustrates that our CDet solver can be reliably applied to general impurity problems as part of a quantum embedding theory using the exact same framework as developed for molecules. We emphasize that, at the same complexity, systems with continuous bath states can be solved. Our method is therefore a controlled method for quantum impurities with general multi-orbital interactions and hybridizations, not limited by the systematic error introduced by the bath discretization procedure. The application of the solver to more complex impurities, where ED calculations are impractical, is a topic of subsequent publications.

7.4 Conclusion

In conclusion, we have presented a diagrammatic Monte Carlo method for quantum impurity models with general interactions and hybridizations using the connected determinant formalism [22]. We have tested the method at the example of molecular systems, which presents a systematic way of changing correlation strength, system size, basis size, and temperature. We have also tested our method for impurity problems occurring in realistic quantum impurity calculations.

Our method is formulated in the language of Green's functions and self-energies. As a grand-canonical finite-temperature method, it is able to describe systems with particle number fluctuations and excited states. However, similar to other perturbative methods, the diagrammatic series breaks down in the strong correlation regime. This breakdown is clearly evident in the order-by-order convergence of the series and, as we have shown in detail, can be traced back to the pole structure of the diagram series.

Our method fills a crucial need of impurity solvers able to treat general four-fermion interaction and general off-diagonal hybridizations in large multi-orbital problem. It should

therefore find applications in moderately correlated real-material simulations such as those occurring in DMFT [15, 16] and SEET [35–37, 95].

Further methodological progress, such as the use of higher order counter terms [26], better integration methods [126], complex conformal mapping techniques [34, 49, 127], and other types of Monte Carlo updates will expand the accessible parameter regime of the method and may make simulations in the strongly correlated regime possible.

Chapter 8

Inchworm Algorithm for Interaction Expansion

This chapter is based on JL, Yang Yu, Emanuel Gull, and Guy Cohen: "Interaction expansion inchworm Monte Carlo for lattice and impurity models." (in preparation.)

In a perturbative method described in Chapter 3, it is typically assumed that the non-interacting Hamiltonian $\hat{H}_0$ is close to the physical one $\hat{H}$ and captures the majority of the physics in the system. The perturbation series is expected to converge within expansion orders that is numerically accessible. However, as the interaction strength of the system becomes stronger, the perturbation expansion of the system becomes increasingly difficult to converge. Any finite order cutoff in this scenario leads to large truncation errors.

In this chapter, we introduce a Monte Carlo algorithm based on incremental perturbation expansions of the electron interaction, where each expansion is based on the result of the previous one. Such a gradual increase of complexity is analogous to the "inchworm" Monte Carlo algorithm for hybridization expansion methods [27, 29, 31, 33].

8.1 Imaginary time perturbation theory

We solve electronic systems given by the general Hamiltonian (2.19)

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{V}, \\ \hat{H}_0 &= \sum_{ij,\sigma} h_{ij} \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma}, \\ \hat{V} &= \frac{1}{2} \sum_{ijkl} \sum_{\sigma\sigma'} U_{ijkl} \hat{c}^\dagger_{i\sigma} \hat{c}^\dagger_{k\sigma'} \hat{c}_{l\sigma'} \hat{c}_{j\sigma}\end{aligned} \tag{8.1}$$

where $\hat{c}_i^\dagger, \hat{c}_i$ are electron creation and annihilation operators in orbital i, h is the single-particle Hamiltonian, and U the electronic interaction tensor.

We take a perturbative approach following the interaction-expansion formalism, by treating the non-interacting Hamiltonian $\hat{H}_0$ as the unperturbed system and the interaction $\hat{V}$ as the perturbation. The partition function of the system at inverse temperature β can be expanded as a Dyson series in the interaction picture [46, 128],

$$\begin{aligned}Z &= \operatorname{Tr} e^{-\beta \hat{H}} = \operatorname{Tr}[e^{-\beta \hat{H}_0} \hat{U}_I(\beta)] = Z_0 \langle \hat{U}_I(\beta) \rangle_0, \\ \hat{U}_I(\beta) &:= e^{\beta \hat{H}_0} e^{-\beta \hat{H}} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^\beta \mathrm{d}\tau_1 \int_0^\beta \mathrm{d}\tau_2 \cdots \\ &\quad \cdots \int_0^\beta \mathrm{d}\tau_k \mathscr{T}_\tau \{\hat{V}_I(\tau_1) \hat{V}_I(\tau_2) \cdots \hat{V}_I(\tau_k)\}\end{aligned} \tag{8.2}$$

where the subscript I denotes operators in the interaction picture, $\hat{U}_I(\tau) = e^{\tau \hat{H}_0} e^{-\tau \hat{H}}$ the time evolution operator, $Z_0 = \operatorname{Tr} e^{-\beta \hat{H}_0}$ is the non-interacting partition function, $\langle \cdot \rangle_0 = Z_0^{-1} \operatorname{Tr}[e^{-\beta \hat{H}_0}(\cdot)]$ is the non-interacting thermal expectation value, and $\mathscr{T}_\tau$ the time order-

ing operator. Similarly, the electronic Green's function in imaginary time, defined as

$$G_{ij}(\tau,\tau') = G_{ij}(\tau-\tau') = -\langle \mathscr{T}_\tau \hat{c}_i(\tau)\hat{c}_j^\dagger(\tau'+0^+)\rangle, \tag{8.3}$$

where $\langle\cdot\rangle = Z^{-1}\,\mathrm{Tr}[e^{-\beta\hat{H}}(\cdot)]$, can be expanded as [9, 14, 46]

$$\begin{aligned} G(\tau,\tau') = -\frac{Z_0}{Z}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_0^\beta \mathrm{d}\tau_1\int_0^\beta \mathrm{d}\tau_2\cdots\int_0^\beta \mathrm{d}\tau_k\times \\ \times\langle \mathscr{T}_\tau \hat{c}_I(\tau)\hat{c}_I^\dagger(\tau')\hat{V}_I(\tau_1)\hat{V}_I(\tau_2)\cdots\hat{V}_I(\tau_k)\rangle_0. \end{aligned} \tag{8.4}$$

We introduce a control parameter $\theta\in[0,\beta]$, and define an "auxiliary" partition function

$$Z_\theta := Z_0\langle\hat{U}_I(\theta)\rangle_0 = \mathrm{Tr}[e^{-(\beta-\theta)\hat{H}_0}e^{-\theta\hat{H}}]. \tag{8.5}$$

Since $\hat{U}_I(0)$ is the identity operator, Z_θ connects $Z_0 = Z_0\langle\hat{U}_I(0)\rangle_0$ and $Z = Z_0\langle\hat{U}_I(\beta)\rangle_0$ continuously via the parameter θ, such that $Z_{\theta=0}=Z_0$, $Z_{\theta=\beta}=Z$. With $\hat{U}_I$ from Eq. (8.2), Z_θ can be expanded as

$$\begin{aligned} Z_\theta = Z_0\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_0^\theta \mathrm{d}\tau_1\int_0^\theta \mathrm{d}\tau_2\cdots\int_0^\theta \mathrm{d}\tau_k\times \\ \times\langle \mathscr{T}_\tau \hat{V}_I(\tau_1)\hat{V}_I(\tau_2)\cdots\hat{V}_I(\tau_k)\rangle_0, \end{aligned} \tag{8.6}$$

which corresponds to replacing all upper integration bounds from β to θ in the expansion of the physical partition function Z. $Z_\theta = \mathrm{Tr}[e^{-(\beta-\theta)\hat{H}_0}e^{-\theta\hat{H}}]$ can be understood as a trace of a "partially dressed" time evolution: from 0 to θ the system is propagated with the full Hamiltonian $\hat{H}$, and then from θ to β with the non-interacting Hamiltonian $\hat{H}_0$.

We can similarly apply this parametrization to the Green's function, by changing the bounds of the imaginary time integrals in Eq. (8.4) to θ. This defines an "auxiliary Green's

function"

$$G_\theta(\tau,\tau') = -\frac{Z_0}{Z_\theta}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_0^\theta \mathrm{d}\tau_1\int_0^\theta \mathrm{d}\tau_2\cdots\int_0^\theta \mathrm{d}\tau_k\times \\ \times\langle\mathscr{T}_\tau\hat{c}_I(\tau)\hat{c}_I^\dagger(\tau')\hat{V}_I(\tau_1)\hat{V}_I(\tau_2)\cdots\hat{V}_I(\tau_k)\rangle_0. \tag{8.7}$$

G_θ recovers the non-interacting Green's function at $\theta = 0$ and the full Green's function G at $\theta = \beta$, and thus continuously connects the two. In Appendix D we show an explicit non-perturbative definition of G_θ. Since θ breaks the time-translational invariance, G_θ cannot be formulated as a function of a single time parameter as in Eq. (8.3).

8.2 Diagrammatic evaluation of auxiliary quantities

The expansions of physical quantities Z and G, when applied to the electronic Hamiltonian (8.1), can be represented graphically as a sum over Feynman diagrams [128]. A diagram at order k is composed of k interaction vertices representing U_{ijkl}, each assigned to an imaginary time index $\tau_i \in [0,\beta]$, $i = 1,\ldots,k$. Propagator lines representing the non-interacting Green's function G_0 connect these vertices. For the partition function Z, the non-interacting expectation values in Eq. (8.2) can be evaluated using Wick's theorem, which generates closed "vacuum" diagrams which can be either connected or disconnected. The Green's function expansion in Eq. (8.4) involves two "external" operators $\hat{c}_I(\tau)$ and $\hat{c}_I^\dagger(\tau')$ which become external "legs" in Feynman diagrams, and the disconnected components are canceled by the partition function diagrams of Z in the denominator, leaving diagrams in which all internal vertices and external legs are fully connected [39, 128]. Figure 8.1 shows examples of such "bare" Feynman diagrams.

Since expansions of the auxiliary quantities, Eqs. (8.6) and (8.7), only differ from Eqs. (8.2)

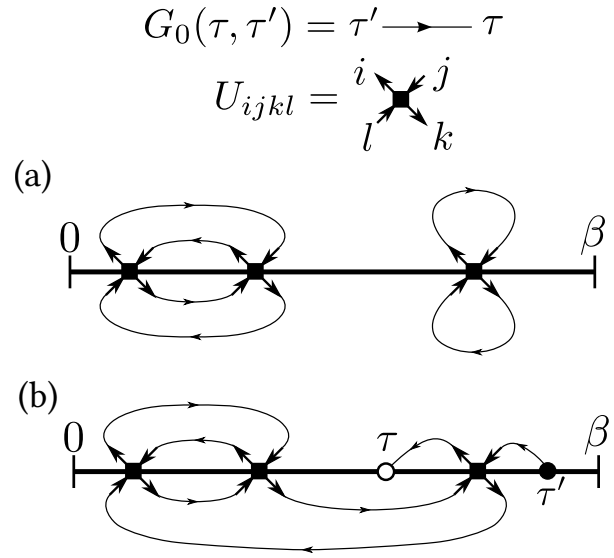

Figure 8.1: Examples of bare Feynman diagrams. Diagram (a) is a third order diagram for Z, which can be either connected or disconnected. Diagram (b) is a third order diagram for $G(\tau,\tau')$, which can only be connected. In both diagrams, filled squares are internal vertices representing U_{ijkl}, lines with arrows are bare propagators representing G_0, and open/closed circles represent the external operators.

and (8.4) in the integration bounds of internal time indices, the same diagram rules can be applied to compute Z_θ and G_θ, as long as the vertices U are confined to the imaginary time interval $[0,\theta]$, as illustrated in Fig. 8.2. The expansions of G and G_θ can be formally written as

$$\begin{aligned} G(\tau,\tau') &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^\beta \mathrm{d}\tau_1 \int_0^\beta \mathrm{d}\tau_2 \cdots \int_0^\beta \mathrm{d}\tau_k \times \\ &\quad \times D^{\mathrm{bare}}(\tau,\tau';\tau_1,\tau_2,\ldots,\tau_k), \\ G_\theta(\tau,\tau') &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^\theta \mathrm{d}\tau_1 \int_0^\theta \mathrm{d}\tau_2 \cdots \int_0^\theta \mathrm{d}\tau_k \times \\ &\quad \times D^{\mathrm{bare}}(\tau,\tau';\tau_1,\tau_2,\ldots,\tau_k), \end{aligned} \tag{8.8}$$

where D^{bare} denotes the sum of all connected bare diagrams [39, 128].

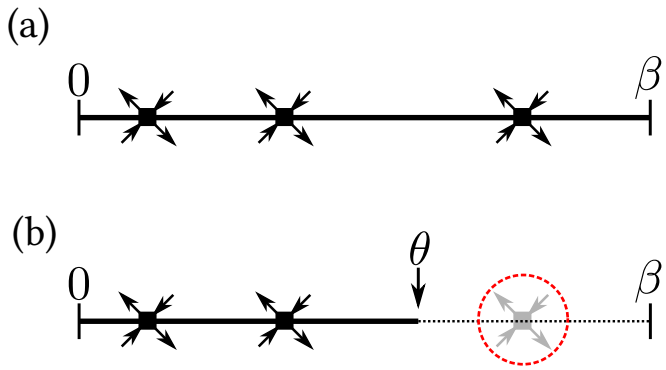

Figure 8.2: Examples of valid vertex locations of bare expansions. (a) shows vertices contributing to expansions of Z or G, which can be inserted at any imaginary times in $[0, \beta]$. In (b) we show valid vertices contributing to expansions of Z_θ or G_θ with solid color, which are only allowed in $[0, \theta]$ shown as the solid segment. The shaded vertex in the red dashed circle is excluded.

Assuming that we know G_θ for some θ, we aim to express $G_{\theta'}$ for $\theta' > \theta$ as a diagrammatic series in terms of G_θ instead of G_0, i.e. using G_θ as partially "dressed" propagator lines. As each G_θ contains infinitely many bare diagram components, each of which is a valid bare diagram for G_θ where internal vertices reside in the interval $[0, \theta]$, diagram topologies for $G_{\theta'}$ which can be obtained from replacing the G_θ propagator with these bare components would be overcounted if the unmodified diagram rules were applied. Therefore, diagram rules need to be modified to exclude the overcounted diagram topologies. Panels (d) and (e) in Fig. 8.3 illustrate an example of such overcounting. We summarize the updated diagram rules for computing $G_{\theta'}$ from G_θ as follows:

1. For a given set of vertices U at $\tau_1, \ldots, \tau_k \in [0, \theta']$ and external operators $\hat{c}$, $\hat{c}^\dagger$ at τ, τ':

2. Generate all possible graphs by connecting vertices and operators with propagator lines;

3. Eliminate all disconnected graphs;

4. Sort the vertices into two categories:

- "Type 1" if $0 < \tau_i < \theta$,
- "Type 2" if $\theta < \tau_i < \theta'$;

5. Eliminate all graphs that only contain Type-1 vertices;
6. Eliminate all graphs that contain subgraphs of Type-1 vertices connected with exactly two propagators to the remainder of the graph.

Figure 8.3 illustrates these rules. The first three rules are the same as in a bare diagram [39, 128], and the additional rules exclude overcounted diagrams. Note that rule 6 is analogous to the "skeleton" diagram rules of the self-energy for bold-line perturbation theory [52].

The diagrammatic series can be formally written as

$$G_{\theta'}(\tau,\tau') = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^{\theta'} \mathrm{d}\tau_1 \int_0^{\theta'} \mathrm{d}\tau_2 \cdots \int_0^{\theta'} \mathrm{d}\tau_k \times \\ \times D_\theta(\tau,\tau';\tau_1,\tau_2,\ldots,\tau_k) \tag{8.9}$$

where D_θ denotes the sum of all diagrams following the updated diagram rules in which G_θ is used as the propagator. We emphasize here that all the internal time indices $\tau_1,\ldots,\tau_k$ are integrated from 0 to θ', whereas the external indices τ and τ' are unconstrained and take values from 0 to β.

For $\theta' \to \theta$, $G_{\theta'}$ continuously approaches G_θ, and the expansion Eq. (8.9) includes substantially fewer diagrams than the bare expansion Eq. (8.8). As we will show in Sec. 7.3, G_θ is typically a much better starting point than G_0 for a perturbation expansion of $G_{\theta'}$.

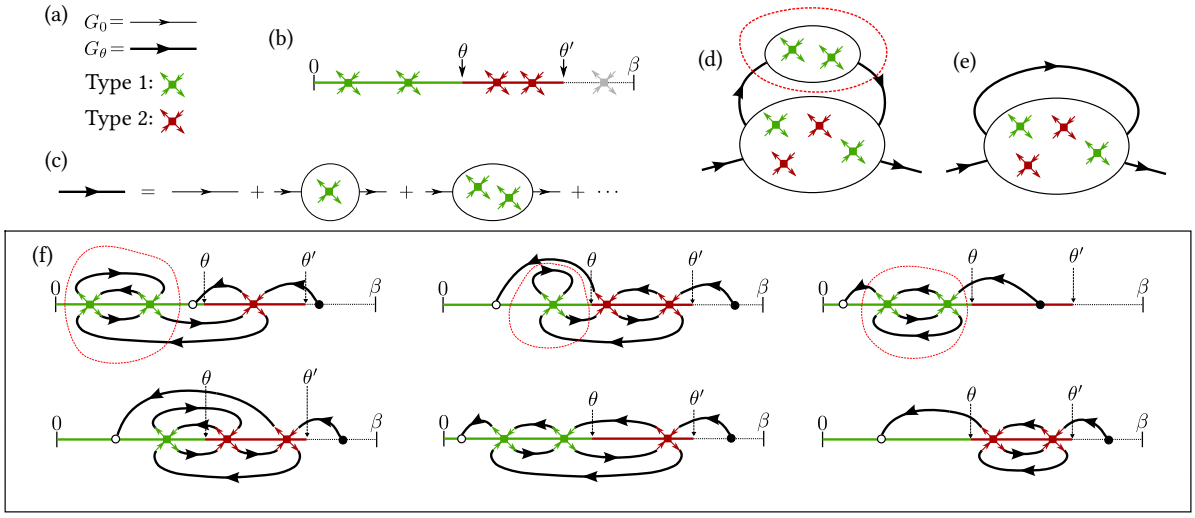

Figure 8.3: Diagram rules for the inchworm expansion from G_θ to $G_{\theta'}$ with $\theta < \theta'$. (a) Thin lines stand for the bare propagator G_0, and the "dressed" lines for G_θ. Green (red) crosses represent Type 1 (2) vertices. (b) Type 1 vertices can only be inserted in $[0, \theta]$ (green segment), and Type 2 vertices in $(\theta, \theta']$ (red segment). Neither type of vertices are allowed in the dashed segment. (c) Each "dressed" line can be expanded in to a bare series following Eq. (8.7), in terms of connected diagrams with only Type 1 vertices. Diagram (d) is an example of a connected diagram that needs to be excluded from the inchworm expansion, since it is already included in Diagram (e). In Box (f), the top row of diagrams are excluded by the diagram rules, where the overcounted components are circled out by dashed red curves; the bottom row shows valid diagrams for the expansion Eq. (8.9).

8.3 Inchworm Monte Carlo algorithm

The ability to efficiently obtain $G_{\theta'}$ from G_θ with $\theta < \theta'$ suggests an iterative algorithm where a sequence of N simulations for $\theta_1, \ldots, \theta_N$ is performed to obtain G_{θ_n} from $G_{\theta_{n-1}}$, with $\theta_0 = 0$ and $\theta_N = \beta$. The proximity of G_{θ_n} and $G_{\theta_{n-1}}$ reduces the number of diagrams to be evaluated and thereby accelerates the simulation. Due to its similarity to the inchworm algorithm for the hybridization expansion [27, 33, 129, 130], which utilizes the same strategy for gradually increasing the the propagator intervals, we term the parameter θ the "inchworm time", and the expansion from Eq. (8.9) as the "inchworm expansion". The final solution is guaranteed to be exact if (1) the perturbation series converges in each inchworm expansion calculation, and (2) the series is computed to infinite orders.

By making the difference in inchworm time $\Delta\theta = \theta' - \theta$ sufficiently small, such that $G_{\theta'}$ is well approximated by G_θ, we observe that in practic the first assumption is satisfied for all systems we study in Sec. 7.3. In Appendix E, we connect the convergence of inchworm series to the skeleton expansion [52], which is typically obtained self-consistently and may converge to an unphysical fixed point [58].

Satisfying the second assumption is harder as the summation of all diagram is not possible in practice. However, one may hope that contributions to the observable decay with in the accessible orders. This behavior is dependent on the systems studied. Section 7.3 shows examples where this assumption is valid, and systems where contributions do not decay within the accessible orders.

A complete inchworm simulation proceeds as follows. We first construct two imaginary time grids: one "inchworm grid" $\{\theta_n | n = 0, \ldots, N, \theta_{n+1} > \theta_n\}$ for the sequence of inchworm times θ, and one "interpolation grid" $\{\tau_i | i = 0, \ldots, N_\tau\}$ for measuring and interpolating the auxiliary Green's function. The final Green's function is then computed via N "inchworm

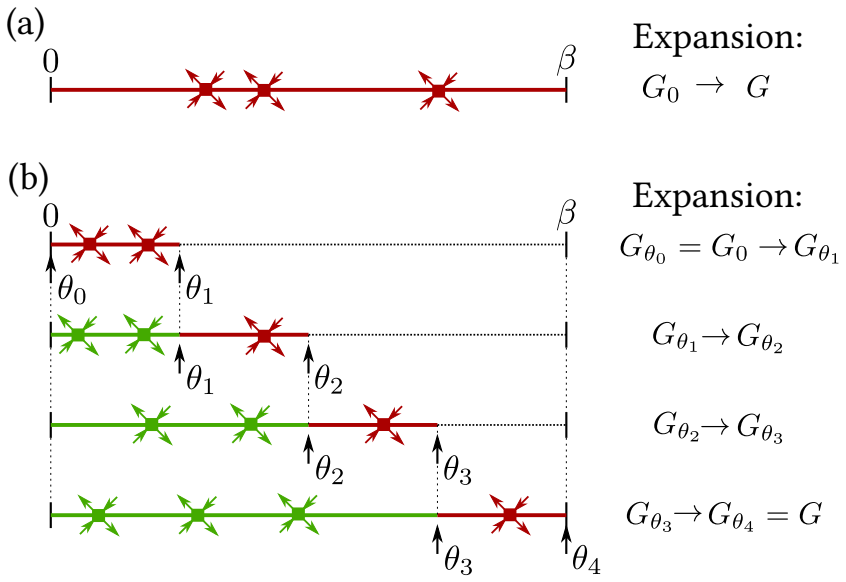

Figure 8.4: Comparison of bare and inchworm Monte Carlo for Green's function. (a) shows a vertex configuration for the bare expansion in Eq. 8.4 which is equivalent to an inchworm simulation with $N = 1$. (b) shows configurations for an inchworm Monte Carlo simulation with $N = 4$ at each inchworm step. In Monte Carlo samplings of each expansion, Type-1 (Type-2) vertices are sampled in green (red) segments on the imaginary time axis.

steps": In the n-th step, we perform an inchworm expansion of G_{θ_n} with respect to $G_{\theta_{n-1}}$, and calculate $G_{\theta_n}(\tau_i, \tau_j)$ for each pair of $i, j = 0, \ldots, N_\tau$ using Monte Carlo as detailed in Sec. 8.4, with the non-interacting initial condition $G_{\theta_0} = G_0$. Figure 8.4 illustrates the "inching" process, in comparison with the bare expansion which is equivalent to doing only one inchworm step. G_{θ_n} is interpolated on the interpolation grid for continuous-time evaluations of the next inchworm step. Since $G_\theta(\tau, \tau')$ is generally not smooth when $\tau = \theta$ or $\tau' = \theta$, the interpolation grid $\{\tau_i\}$ should be chosen to include all points on the inchworm grid $\{\theta_n\}$ for good interpolation results.

In this work, we choose equidistant time points for both grids, and perform linear interpolation for measured auxiliary Green's functions, which provide decent accuracy at high temperatures. More advanced time grids or interpolation schemes may be employed to optimize the method for lower temperatures.

8.4 Continuous-time Monte Carlo evaluation of inchworm expansions

We evaluate each inchworm expansion (8.9) using the standard continuous-time quantum Monte Carlo approach [9, 12, 14, 17, 18, 22, 131]. We employ a finite cutoff of the expansion order $k_{\max}$, and perform a Monte Carlo importance sampling of the internal spacetime coordinates following the *a priori* distribution

$$p(\mathscr{C}) \propto |D_\theta(\tau, \tau'; \mathscr{C})|, \tag{8.10}$$

where $\mathscr{C} = \{\tau_1, \ldots, \tau_k\}$ is a Monte Carlo configuration. Since D_θ has varying signs due to its fermionic nature, the absolute value is necessary to ensure $p(\mathscr{C})$ is positive, whereas the fermionic sign $\mathrm{sgn}(D_\theta)$ enters measurements of all physical observables. For a given vertex configuration, D_θ is computed explicitly by summing over all proper inchworm diagrams according to the diagram rules. In our implementation, we use a graph theory routine to precompute and save all valid diagram topologies for each expansion order.

We generate Monte Carlo samples as Markov chain via the Metropolis-Hastings algorithm. From each configuration $\mathscr{C}$, a new configuration $\mathscr{C}'$ is proposed following some proposal probability distribution $w^{\mathrm{prop}}(\mathscr{C}'|\mathscr{C})$. To ensure detailed balance, an acceptance ratio R is calculated after each proposal as

$$R(\mathscr{C}'|\mathscr{C}) = \frac{w^{\mathrm{prop}}(\mathscr{C}|\mathscr{C}')p(\mathscr{C}')}{w^{\mathrm{prop}}(\mathscr{C}'|\mathscr{C})p(\mathscr{C})}. \tag{8.11}$$

The proposal $\mathscr{C} \to \mathscr{C}'$ is accepted with probability

$$w^{\mathrm{acc}}(\mathscr{C}'|\mathscr{C}) = \min(1, R(\mathscr{C}'|\mathscr{C})). \tag{8.12}$$

This ensures the detailed balance of the Markov process, i.e.

$$w(\mathscr{C}'|\mathscr{C})p(\mathscr{C}) = w(\mathscr{C}|\mathscr{C}')p(\mathscr{C}'), \tag{8.13}$$

where

$$w(\mathscr{C}'|\mathscr{C}) = w^{\text{acc}}(\mathscr{C}'|\mathscr{C})w^{\text{prop}}(\mathscr{C}'|\mathscr{C}). \tag{8.14}$$

which guarantees that samples obtain the equilibrium distribution $p(\mathscr{C})$ after thermalization. We employ the same Monte Carlo updates as in CT-INT [9] which guarantees ergodicity for all systems we study in this work, including random insertions and removals of a single vertex or a pair of vertices. Auxiliary Green's function G_θ is measured from the Monte Carlo procedure and normalized against quantities that are analytically available, such as low-order diagrams.

8.5 Generalization of the inchworm construction

The inchworm parametrization scheme we have introduced so far performs a partial "dressing" in the interval $[0, \theta]$ in the imaginary time evolution. This can be generalized to any open set of the imaginary time $\Theta \subset [0, \beta]$, such that imaginary time evolves with $\hat{H}$ in all disjoint intervals in Θ, and with $\hat{H}_0$ in the rest of $[0, \beta]$. This gives auxiliary quantities Z_Θ and G_Θ that are "dressed" in Θ. The bare diagrammatic expansions of these quantities are the same as in Eqs. (8.6) and (8.7), except that the interaction vertices can only be inserted in Θ.

Incremental inchworm expansions of $G_{\Theta'}$ with respect to G_Θ can be formulated in analogy to Eq. (8.9) for $\Theta \subsetneq \Theta'$, by simply changing the classification of vertices. A Type 1 vertex now has imaginary time $\tau \in \Theta$, and a Type 2 vertex has $\tau \in \Theta' \backslash \Theta$. The inchworm grid of

the imaginary time is thus replaced by a sequence of open sets $\Theta_0, \Theta_1, \ldots, \Theta_N$, where $\Theta_0 = \emptyset$ and $\Theta_N = [0, \beta]$, and $\Theta_n \subsetneq \Theta_{n+1}$ for $n = 0, \ldots, N-1$. This brings flexibility to the inchworm Monte Carlo setup, which can now choose a different sequence of intervals or sets when left-to-right "inching" is unsuitable.

8.6 Results

This section presents preliminary results obtained from our prototype implementation of the inchworm algorithm.

We first apply the algorithm to two Hubbard atoms with on-site interaction $U = 2$, coupled with a hopping amplitude of $t = 1$. In Fig. 8.5, we show inchworm calculation results of the Green's function (blue lines) in comparison to bare diagrammatic series (orange lines) and ED results (dashed black lines), computed at different temperatures. Series expansions in both the inchworm simulations and the bare ones are truncated at $k_{\max} = 6$. Hartree–Fock shift (see, e.g., Sec. 3.5 or Sec. 7.2.5) is applied to remove the tadpole diagrams. We observe that inchworm agrees with ED at all temperature points we compute here. The fact that bare simulations also give decent agreement indicates that this problem is weakly-interacting and the bare series converges fast.

Once we remove the Hartree–Fock shift and simulate directly from the non-interacting Green's function g, the bare series starts to break down, as shown in Fig. 8.6 where the same system is calculated without the shift. The bare series results do not converge at $k_{\max} = 6$, whereas the inchworm show decent agreement with ED. A closer look at the convergence behavior is presented in Fig. 8.7, in which the top panel shows the same results as the bottom panel of Fig. 8.6, the middle panel shows the order-by-order contributions in the final inchworm step, and the bottom panel shows order-by-order contributions to the Green's function

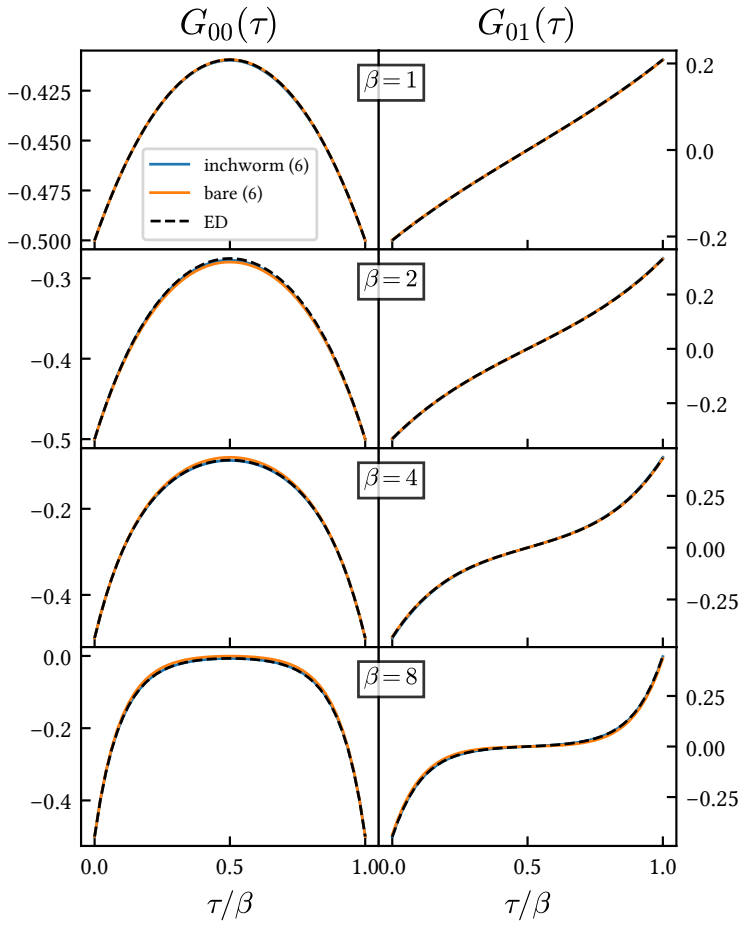

Figure 8.5: Results at different β. Hubbard dimer with $U = 2$ and Hartree-shift, half-filling.

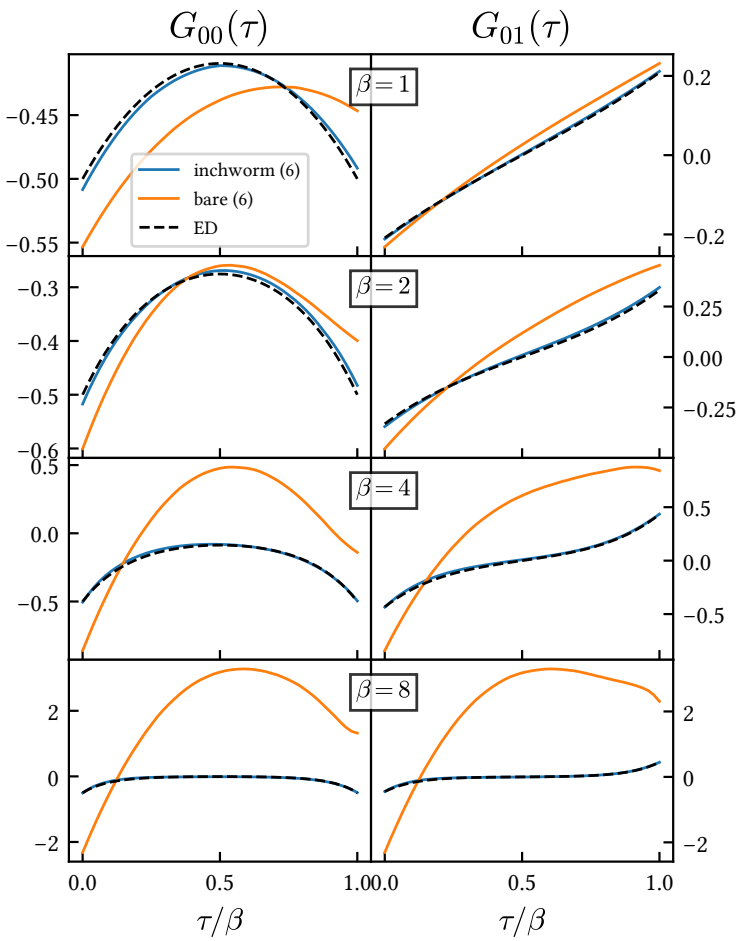

Figure 8.6: Results at different β. Hubbard dimer with $U = 2$ without Hartree-shift, half-filling, compared to bare.

in the bare calculation. The inchworm calculation converges rapidly and only the first three expansion orders show significant contribution, whereas in the bare case no convergence can be observed within $k_{\max} = 6$. This shows that inchworm avoids the slow convergence or divergence problems of the bare series.

We now move on to a slightly more complex system: a triangular cluster of three Hubbard atoms at $U = 2$. As shown in Fig. 8.8, even if the system is away from particle–hole symmetry and a Hartree–Fock shift becomes difficult, inchworm calculations at $k_{\max} = 6$ agrees with

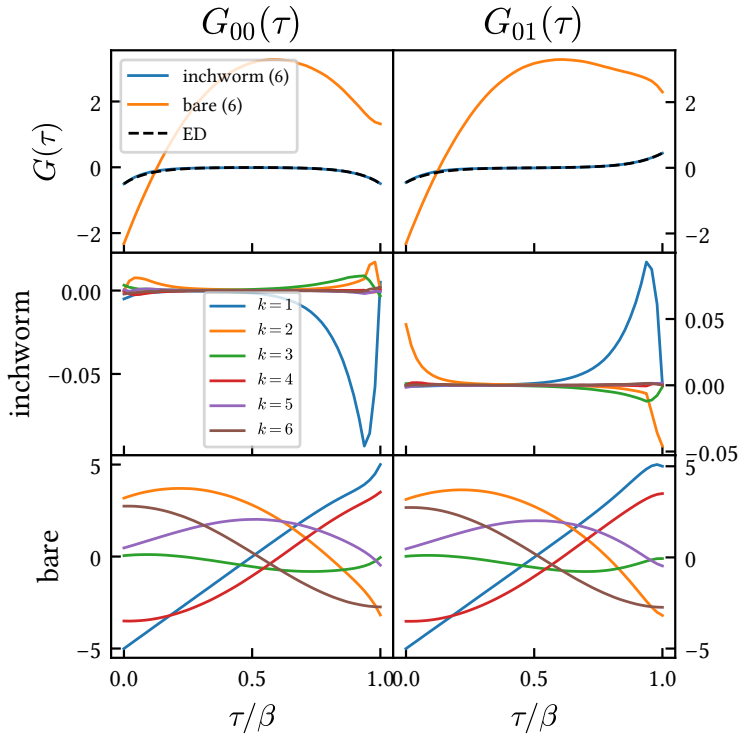

Figure 8.7: Convergence comparison between bare and inchworm. Hubbard dimer with $U = 2$, $\beta = 8$ without Hartree-shift, half-filling.

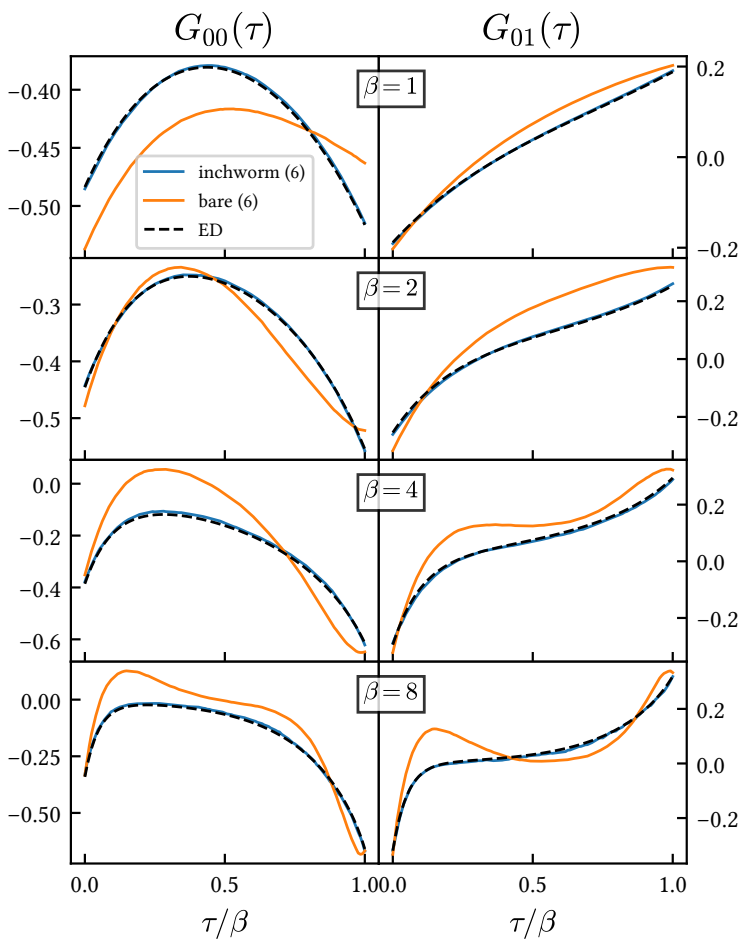

Figure 8.8: Results at different β. Hubbard trimer with $U = 2$ and $\mu = 1$, no Hartree-shift.

ED within line width, whereas the bare calculation with the same order truncation fails.

Figure 8.9 illustrates individual inchworm steps and the behavior of the auxiliary Green's function G_θ. At each θ_i, $G_\theta(\tau, \tau')$ shows discontinuity of the first derivative that $\tau = \theta$ or $\tau' = \theta$. For each inchworm step $\theta_i \to \theta_{i+1}$, the inchworm expansion corrects the "kink" from G_{θ_i} and introduces a new "kink" to $G_{\theta_{i+1}}$, as shown by the "strips" in the bottom panel.

In Fig. 8.10, we show the deviations of inchworm results from ED when using the generalized inchworm scheme with different choices of "dressed" intervals. Results labeled as

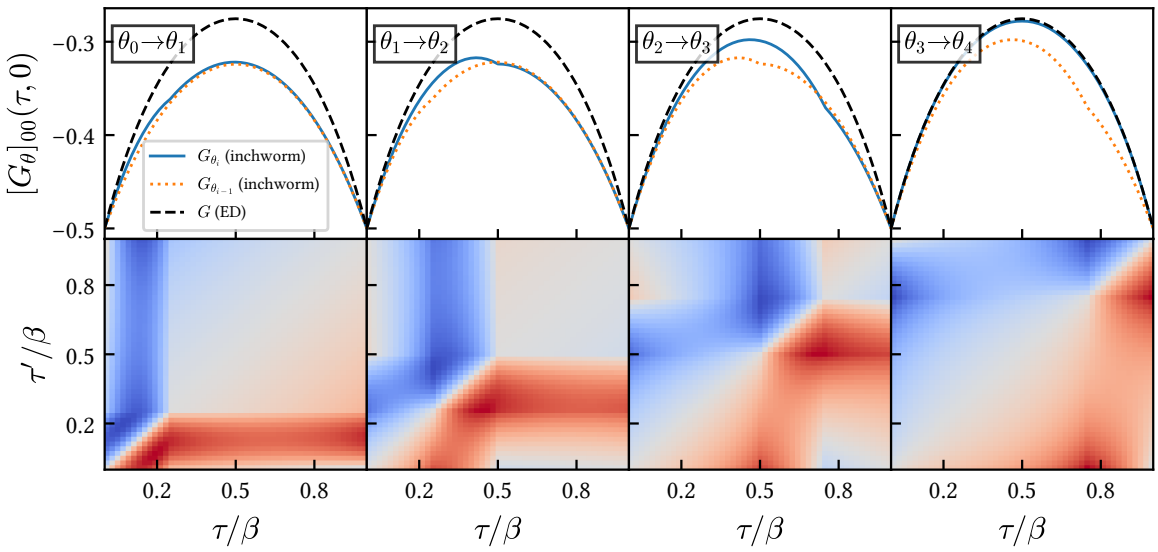

Figure 8.9: Auxiliary Green's functions at each inchworm steps for Hubbard dimer, $U = 2$, $\beta = 2$. Top panel: $G_{\theta_i}(\tau, 0)$ at each step, compared to $G_{\theta_{i-1}}(\tau, 0)$ and ED. Bottom panel: $G_{\theta_i}(\tau, \tau') - G_{\theta_i}(\tau, \tau')$. Colors not to scale. Red means positive, blue means negative. Hubbard dimer with $U = 2$, $\beta = 2$ and Hartree-shift, half-filling.

"left" are obtained from the original formulation where $\Theta_n = [0, n\Delta\theta]$. The meanings of the other labels are: "right" corresponds to $\Theta_n = [\beta - n\Delta\theta, \beta]$, "both" corresponds $\Theta_n = [0, n\Delta\theta/1] \cup [\beta - n\Delta\theta/2, \beta]$, and "middle" corresponds to $\Theta_n = [\beta/2 - n\Delta\theta/2, beta/2 + n\Delta\theta/2]$. Since the system is particle–hole symmetric, the bare Green's function g already captures the physics near $\tau = 0$ and $\tau = \beta$ well. The "both" mode therefore show the largest deviation, whereas the "middle" result shows a much better agreement. This shows that one can achieve better inchworm results by choosing different propagation schemes according to the physical properties of the systems. Note that this only matters if the finite order truncation is applied, not if the series converges.

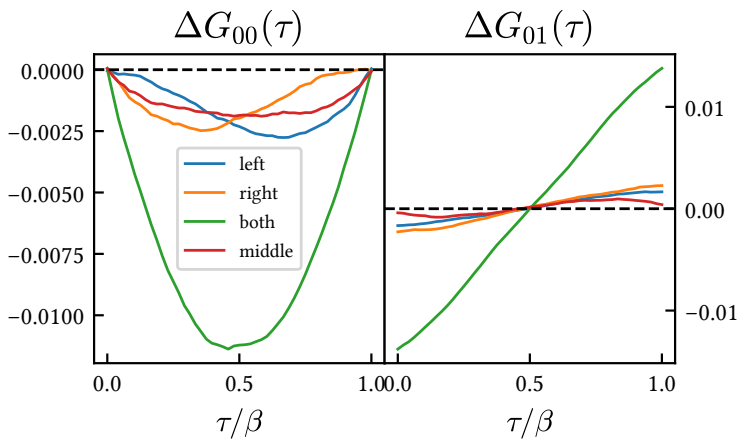

Figure 8.10: Difference of final inchworm GF vs ED with different propagation modes. Hubbard dimer with $U = 2$, $\beta = 2$ and Hartree-shift, half-filling.

8.7 Future directions

The inchworm algorithm introduced in this chapter is the first attempt in applying incremental diagrammatic expansions to the interaction expansion scheme. In this section, we list a few directions for future developments along this route.

Algorithmic improvements and extensions can be applied to the current theoretical setup:

- The on-site Hubbard interaction can be extended to general four-fermion interaction tensor U_{ijkl}. Monte Carlo updates may need to be adjusted to a similar scheme to ones used in Chapter 7. This will enable applications to realistic impurities with general interactions.

- The maximum expansion order is the main limitation of this method. Potential improvements include: in-memory caching of diagram topologies, e.g., using the least-frequently-used (LFU) cache; direct sampling of diagrammatic topologies (which requires carefully designed Monte Carlo updates); CDet-like recursion relations (may require two-particle objects).

On the theory aspect, one could explore one of the following directions:

- Different integration scheme. In the current integration scheme with finite order cutoff $k_{\max}$, in early inchworm steps where the dressed interval is small, the effective "density" of vertices is higher than in later steps, which lead to the asymmetry observed in the results. If the series can be reformulated to separate integrals over the "old" interval from those over the "new" interval, the situation may be improved. Note that this implies a different perturbation series, not simply reweighing the Monte Carlo integration, which only affects the stochastic error, not the truncation error.

- If we have dressed an interval $\Theta = [0, \theta]$, by the circular property of the trace, we also obtain auxiliary quantities with dressed interval $\Theta' = [\Delta\theta, \theta + \Delta\theta]$ for some $\Delta\theta$ automatically. It would be interesting to see if one can "merge" two dressed propagators, such that $[0, \theta](+)[\theta, 2\theta] \to [0, 2\theta]$. This implies that in the diagrammatic series, we now have three types of vertices: vertices belonging to either interval or neither of them, as well as (at least) two types of propagators.

- The diagram rules are designed assuming the perturbation expansion is summed up to infinite order, whereas in practice we only perform a finite order simulation. One might gain additional insight by formulating the incremental expansion in a "partially-bold" scheme [48].

Chapter 9

Numerical Representations of the Green's Function

This chapter is partially based on JL, Markus Wallerberger, Naoya Chikano, Chia-Nan Yeh, Emanuel Gull, and Hiroshi Shinaoka: "Sparse sampling approach to efficient ab initio calculations at finite temperature." Phys. Rev. B 101, 035144 (2020).

The key quantity in diagrammatic methods for correlated electrons is the electron Green's function $G(\tau)$, both as a component (the "propagator") in diagrammatic evaluations and as the target quantity of these methods. In *ab initio* calculations, the difference between the scales of the bare Hamiltonian (which sets a frequency range), temperature (which dictates the frequency resolution), and the energy scales for competing quantum phenomena spans many orders of magnitude. As a consequence, a naive representation of the Green's function requires an imaginary-time grid too large to store in memory, and solving equations such as the Dyson equation to the required accuracy becomes prohibitively expensive. This issue becomes even more pronounced for two-particle response functions, which are generically a function of multiple time and orbital indices.

Compact representations of Green's functions are crucial to address this problem. Representations based on power meshes [132, 133], Legendre polynomials [134, 135], Chebyshev polynomials [136, 137], intermediate numerical representations (IR) [138–140], quadrature rules [141, 142], and spline interpolations [143] have been proposed, as well as high frequency tail expansions [5, 144, 145]. In this chapter, we briefly review some of the representations, and introduce the sparse sampling method [146, 147] which allows compact representations of the Green's function in both frequency and in time, and provides efficient transformations between the two.

9.1 High-frequency expansion

We start with the non-interacting Green's function of a single-band atom at energy level $\varepsilon = 0$ as an example. Following (3.13), we have

$$g(\mathrm{i}\omega_n) = \frac{1}{\mathrm{i}\omega_n - \varepsilon} = \frac{1}{\mathrm{i}\omega_n}. \tag{9.1}$$

The inverse Fourier transform writes

$$g(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} \frac{e^{-\mathrm{i}\omega_n \tau}}{\mathrm{i}\omega_n}. \tag{9.2}$$

If we take the limit $\tau \to 0^-$, which corresponds to the density matrix, the Fourier transform become a harmonic series, which does not converge absolutely and requires to be evaluated as principle values. As a result, any truncations of the frequency, which is necessary in numerical methods, results in incorrect results. A common solution to this problem [144, 145] is to remove the high-frequency "tails", defined as polynomials of the inverse frequency $1/\mathrm{i}\omega_n$, when performing Fourier transforms, and add back the corresponding functions in τ

representation afterwards:

$$
\begin{aligned}
G^{\text{tail}}(\mathrm{i}\omega_n) &:= \frac{G^{\text{I}}}{\mathrm{i}\omega_n} + \frac{G^{\text{II}}}{(\mathrm{i}\omega_n)^2} + \cdots, \\
\tilde{G}(\mathrm{i}\omega_n) &:= G(\mathrm{i}\omega_n) - G^{\text{tail}}(\mathrm{i}\omega_n), \\
\tilde{G}(\tau) &\xleftarrow{\text{Inverse Fourier}} \tilde{G}(\mathrm{i}\omega_n), \\
G(\tau) &= \tilde{G}(\tau) + G^{\text{tail}}(\tau),
\end{aligned}
\tag{9.3}
$$

where $G^{\text{tail}}(\tau)$ is obtained from inverse Fourier transforms of $1/(\mathrm{i}\omega_n)^k$. Here we list examples of Fermion "tails" for the first three orders:

$$
\begin{aligned}
\frac{1}{\mathrm{i}\omega_n} &\sim -\frac{1}{2}, \\
\frac{1}{(\mathrm{i}\omega_n)^2} &\sim \frac{2\tau - \beta}{4}, \\
\frac{1}{(\mathrm{i}\omega_n)^3} &\sim \frac{\tau(\beta - \tau)}{4}, \\
&\cdots
\end{aligned}
\tag{9.4}
$$

Suppose we have a time-ordered correlation function defined as $F(\tau) = \langle \mathscr{T}_\tau \hat{A}(\tau)\hat{B}(0)\rangle$, where $\hat{A}$ and $\hat{B}$ are Heisenberg picture operators of the same (Bosonic/Fermionic) statistics. The Matsubara representation of F is

$$
F(\mathrm{i}\omega_n) = \int_0^\beta F(\tau) e^{\mathrm{i}\omega_n \tau} \mathrm{d}\tau, \tag{9.5}
$$

where $\mathrm{i}\omega_n$ are the corresponding Bosonic/Fermionic Matsubara frequencies.

The high frequency moments of $F(\mathrm{i}\omega_n)$ can be derived from integration by parts, i.e.

$$\begin{aligned} F(\mathrm{i}\omega_n) &= \int_0^\beta F(\tau) e^{\mathrm{i}\omega_n \tau} \mathrm{d}\tau \\ &= \left. \frac{F(\tau) e^{\mathrm{i}\omega_n \tau}}{\mathrm{i}\omega_n} \right|_0^\beta - \frac{1}{\mathrm{i}\omega_n} \int_0^\beta F'(\tau) e^{\mathrm{i}\omega_n \tau} \mathrm{d}\tau \\ &= \cdots \\ &= \frac{\zeta F(\beta) - F(0)}{\mathrm{i}\omega_n} - \frac{\zeta F'(\beta) - F'(0)}{(\mathrm{i}\omega_n)^2} + \frac{\zeta F''(\beta) - F''(0)}{(\mathrm{i}\omega_n)^3} - \cdots \end{aligned} \tag{9.6}$$

where $\zeta = \pm 1$ for Bosonic/Fermionic statistics. We can induce from above that the high frequency expansion of $F(\mathrm{i}\omega_n)$ is

$$F(\mathrm{i}\omega_n) = \sum_{k=1}^{\infty} \frac{m_k}{(\mathrm{i}\omega_n)^k}, \tag{9.7}$$

where

$$\begin{aligned} m_k &= (-1)^{k-1} \left[\zeta F^{(k-1)}(\beta) - F^{(k-1)}(0) \right] \\ &= (-1)^{k-1} \left[F^{(k-1)}(0^-) - F^{(k-1)}(0^+) \right] \\ &= (-1)^{k} \left[F^{(k-1)}(0^+) - F^{(k-1)}(0^-) \right] \end{aligned} \tag{9.8}$$

is the k-th moment. Using the equation of motion of the Heisenberg picture operators, we have

$$F^{(k-1)}(\tau) = \langle \mathscr{T}_\tau \partial_\tau^k \hat{A}(\tau) \hat{B}(0) \rangle = \langle \mathscr{T}_\tau [\hat{H}, \hat{A}]_{(k)}(\tau) \hat{B}(0) \rangle, \tag{9.9}$$

where $[\cdot, \cdot]_{(k)}$ is the sequential commutator defined as

$$[\hat{H}, \hat{A}]_{(k)} = [\hat{H}, [\hat{H}, \hat{A}]_{(k-1)}], \quad [\hat{H}, \hat{A}]_{(0)} = \hat{A}. \tag{9.10}$$

With this, the high-frequency moments can be expressed as

$$\begin{aligned} m_k &= (-1)^k \left[F^{(k-1)}(0^+) - F^{(k-1)}(0^-) \right] \\ &= (-1)^k \left[\langle [\hat{H}, \hat{A}]_{(k-1)} \hat{B} \rangle - \zeta \langle \hat{B} [\hat{H}, \hat{A}]_{(k-1)} \rangle \right] \\ &= (-1)^k \langle [[\hat{H}, \hat{A}]_{(k-1)}, \hat{B}]_\zeta \rangle, \end{aligned} \tag{9.11}$$

where $[\cdot, \cdot]_\zeta$ is commutator for Bosons, anticommutator for Fermions.

In Appendix F, the first three high-frequency tails of the Green's function are derived for the general Hamiltonian (2.24).

9.2 Chebyshev polynomials

The Green's function $G(\tau)$ is a continuous and bounded function in the interval $[0, \beta]$. When mapped into the interval $[-1, 1]$, it can be exactly represented by orthogonal polynomials, such as Legendre or Chebyshev polynomials. These polynomials can serve as efficient basis representations of the Green's function (see, e.g., Refs. [134, 135] for Legendre, Refs. [136, 137] for Chebyshev.) In addition, these orthogonal systems are accompanied by Gauss quadrature rules for numerical integration, which provides a natural replacement to the uniform imaginary time grid which is inefficient in realistic calculations. In this section, we briefly summarize properties of the Chebyshev representation.

The Chebyshev polynomials of the first kind $T_l(x)$ form an orthogonal system in the interval $[-1, 1]$, which can be mapped in to the interval $[0, \beta]$ via

$$x(\tau) = \frac{2\tau}{\beta} - 1, \quad \tau(x) = \frac{\beta(x+1)}{2} \tag{9.12}$$

such that $T_l^\alpha(\tau) = T_l[x(\tau)]$. We use the notation $T_l(\tau)$ to represent the order l Chebyshev

polynomial mapped onto the interval $[0, \beta]$.

Approximating an analytical function with the first N Chebyshev polynomials is convenient due to the discrete orthogonality on the roots of the $(N+1)$-th Chebyshev polynomial

$$\frac{1}{N}\sum_{k=0}^{N-1} T_i(x_k)T_j(x_k) = \frac{1+\delta_{i,0}}{2}\delta_{ij} \tag{9.13}$$

where x_k are the roots of $T_N(x)$. The Chebyshev coefficients are therefore well approximated by Clenshaw–Curtis quadrature:

$$G_l^\alpha = \frac{2}{N(1+\delta_{0,l})}\sum_{k=1}^{N-1} G^\alpha(\tau_k)T_l(\tau_k) + \mathcal{O}(2^{-N}) \tag{9.14}$$

where $\tau_k = \tau(x_k)$ defined in (9.12).

With the Chebyshev coefficients G_l^α, one can perform fast interpolation of $G^\alpha(\tau)$ at any $\tau \in [0, \beta]$ using recursion relations. Fourier transforms of the basis function $\hat{T}_l^\alpha(\mathrm{i}\omega_n^\alpha)$ can also be computed, see Ref. [136].

9.3 Intermediate representation

The intermediate representation (IR) basis introduced in Refs. [138, 139] is designed to better capture properties of Green's functions in physical systems rather than arbitrary analytic functions. The IR basis has been applied to numerical analytic continuation [148] and DMFT calculations [149]. This section provides a brief description of the IR basis following the notation used in Ref. [140]. The IR basis originates from the Lehmann representation of the

single-particle Green's function

$$G^{\alpha}(\tau) = -\int_{-\omega_{\max}}^{\omega_{\max}} \mathrm{d}\omega K^{\alpha}(\tau,\omega)\rho^{\alpha}(\omega), \tag{9.15}$$

where α denotes the statistics (F for fermions and B for bosons), and the spectrum $\rho^{\alpha}(\omega)$ is bounded in the interval $[-\omega_{\max}, \omega_{\max}]$ ($\omega_{\max}$ is a cutoff frequency). The kernel $K^{\alpha}(\tau,\omega)$ reads

$$K^{\alpha}(\tau,\omega) \equiv \omega^{\delta_{\alpha,\mathrm{B}}} \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}} \tag{9.16}$$

for $\tau \in [0,\beta]$, where the + and − signs are used for fermions and bosons, respectively. The extra ω factor for bosons in Eq. (9.16) is introduced in order to avoid a singularity of the kernel at $\omega = 0$.

For a fixed value of β and $\omega_{\max}$, the IR basis functions are defined through the singular value decomposition (SVD)

$$K^{\alpha}(\tau,\omega) = \sum_{l=0}^{\infty} S_l^{\alpha} U_l^{\alpha}(\tau) V_l^{\alpha}(\omega) \tag{9.17}$$

where one observes an exponential decay of the singular values S_l^{α} (> 0) with increasing l. $U_l(\tau)$ and $V_l(\omega)$ form an orthonormal system for $\tau \in [0,\beta]$ and $y \in [-\omega_{\max}, \omega_{\max}]$, respectively.

A Green's function can be expanded as

$$G^{\alpha}(\tau) = \sum_{l=0}^{\infty} G_l^{\alpha} U_l^{\alpha}(\tau), \tag{9.18}$$

$$G_l^{\alpha} = -S_l^{\alpha} \rho_l^{\alpha}, \tag{9.19}$$

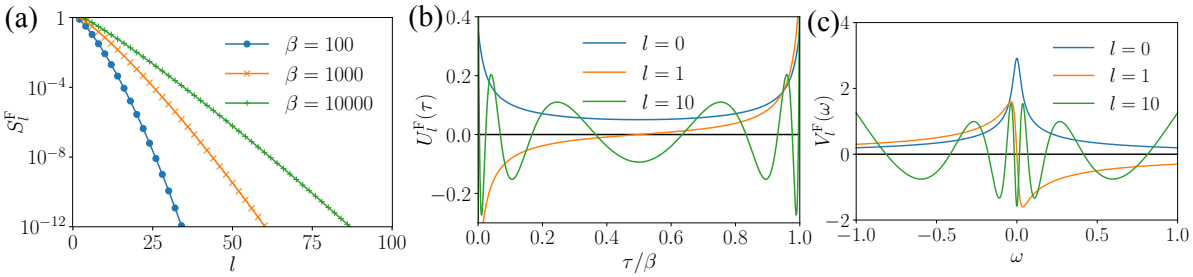

Figure 9.1: (From Ref. [147].) (a) Singular value S_l^α computed for various values of β, (b), (c) IR basis functions $U_l^\alpha(\tau)$ and $V_l^\alpha(\omega)$ computed for $\beta = 100$. Here, we present the results for fermions and $\omega_{\max} = 1$. The data were calculated using `irbasis` [140].

where

$$\rho_l^\alpha \equiv \int_{-\omega_{\max}}^{\omega_{\max}} \mathrm{d}\omega \rho^\alpha(\omega) V_l^\alpha(\omega). \tag{9.20}$$

If $|\rho_l^\alpha|$ does not grow, the exponential decay of S_l^α ensures exponential convergence of G_l^α. The accuracy of the expansion can be controlled by applying a cut-off on the singular values.

$U_l^\alpha(\tau)$ can be Fourier transformed to Matsubara frequencies:

$$\hat{U}_l^\alpha(\mathrm{i}\omega_n^\alpha) = \int_0^\beta \mathrm{d}\tau U_l^\alpha(\tau) e^{\mathrm{i}\omega_n^\alpha \tau}, \tag{9.21}$$

where the "hat" indicates quantities in Matsubara frequency representation. Figure 9.1 shows examples of the IR basis functions.

Note that $\hat{U}_l^\alpha(\mathrm{i}\omega_n^\alpha)$ does not compactly describe a constant shift in Matsubara frequency which corresponds to an unbounded spectrum. Thus, any constant term must be subtracted beforehand particularly when expanding the self-energy. Also, $U_l^{\mathrm{B}}(\tau)$ does not describe a constant shift in imaginary time, corresponding to a zero-energy mode. Such terms must be treated separately as well [139].

In calculations of realistic systems, one can set $\omega_{\max}$ large enough to capture the expected

spectral width. The basis functions change their shapes through the change of the dimensionless quantity $\Lambda = \beta\omega_{\max}$ as temperature is lowered. This leads to a logarithmic growth of the basis size with respect to β.

The dimensionless form of the IR basis is defined as

$$U_l^\alpha(\tau) = \sqrt{\frac{2}{\beta}} u_l^\alpha(x(\tau)), \tag{9.22}$$

$$\begin{aligned}\hat{U}_l^\alpha(\mathrm{i}\omega_n^\alpha) &= \sqrt{\beta} u_{ln}^\alpha \\ &= \sqrt{\frac{\beta}{2}} \int_{-1}^{1} \mathrm{d}x e^{\mathrm{i}\pi\{n+(1/2)\delta_{\alpha,\mathrm{F}}\}(x+1)} u_l^\alpha(x),\end{aligned} \tag{9.23}$$

where $u_l^\alpha(x)$ form an orthonormalized basis for $x \in [-1, 1]$.

9.4 Sparse sampling

Green's function representations, in addition to being compact, also need to enable efficient calculations. The two main stages in most methods introduced in Chapter 5 are the evaluation of self-energy diagrams (usually best done in imaginary time, as the interaction is instantaneous) and the solution of the Dyson equation (usually best done in Matsubara space, where the equation is diagonal in frequency). Some representations, such as the cubic splines [132, 133] or the high frequency tails, are only compact in either time or frequency, and transforming between those domains is expensive or involves a loss of accuracy. Others, such as the orthogonal polynomial bases, can efficiently and accurately be transformed between coefficients and imaginary time sampling points, but frequency transforms result in a loss of compactness.

It is therefore natural to ask if there is a set of "sparse" sampling points in both frequency and time such that, if the Green's function is evaluated on these points, one may reconstruct

the continuous Green's functions in both time and frequency to high precision. This will then allow to perform diagram calculations in time, Dyson equation solutions in frequency, and transformations in between with minimal loss of accuracy.

The sparse sampling method [146] is a framework to generate compact representations in both time and frequency by proposing a sparse sampling scheme for finite temperature Green's functions, following prior efforts using frequency interpolation [143, 150] and MP2 quadratures [142]. We illustrate our scheme at the example of Chebyshev [136] and IR basis functions [138]. It accurately resolves all the information contained within finite temperature Green's functions in a compact set of sampling points, and enables efficient and accurate transforms between imaginary time and Matsubara frequency. The sparsity of the sampling points directly corresponds to the compactness of the basis representation, which leads to system-independent time and frequency grids with few control parameters.

9.4.1 General description and notation

We expand the Green's function G^α into a compact representation in terms of N basis functions, such that in imaginary time and Matsubara frequencies

$$G^\alpha(\tau) = \sum_{l=0}^{N-1} G_l^\alpha F_l^\alpha(\tau), \tag{9.24}$$

$$\hat{G}^\alpha(\mathrm{i}\omega_n^\alpha) = \sum_{l=0}^{N-1} G_l^\alpha \hat{F}_l^\alpha(\mathrm{i}\omega_n^\alpha) \tag{9.25}$$

$$\hat{F}_l^\alpha(\mathrm{i}\omega_n^\alpha) = \int_0^\beta \mathrm{d}\tau F_l^\alpha(\tau) e^{\mathrm{i}\omega_n^\alpha \tau}, \tag{9.26}$$

where G_l^α are expansion coefficients, $F_l^\alpha(\tau)$ imaginary time basis functions with Fourier transform $\hat{F}_l^\alpha(\mathrm{i}\omega_n^\alpha)$ (the "hat" denoting frequency representations), $\omega_n^\alpha = \pi(2n + \delta_{\alpha,F})/\beta$ Mat-

subara frequencies, and α denotes the statistics (F for fermions and B for bosons).

In the Chebyshev representation,

$$F_l^\alpha(\tau) \equiv T_l[x(\tau)], \tag{9.27}$$

where $T_l(x)$ are Chebyshev polynomials of the first kind and $x(\tau) = 2\tau/\beta - 1$. In the IR basis,

$$F_l^\alpha(\tau) \equiv U_l^\alpha(\tau) \tag{9.28}$$

where $U_l^\alpha(\tau)$ depend on the statistics and a dimensionless parameter $\Lambda = \beta\omega_{\max}$ with a cutoff frequency $\omega_{\max}$.

To determine G_l^α from $G^\alpha(\tau)$, we choose a finite set of M sampling points $\bar{\tau}_0^\alpha, \ldots, \bar{\tau}_{M-1}^\alpha \in [0, \beta]$ ($M \geq N$). If these points are chosen such that the discretized basis vectors $\{F_0^\alpha(\bar{\tau}_k^\alpha)\}, \ldots, \{F_{N-1}^\alpha(\bar{\tau}_k^\alpha)\}$ are linearly independent, the exact values of G_l^α can be computed (transformed) from sampled values of $G^\alpha(\tau)$. Similarly, if a subset of Matsubara frequencies $\{\mathrm{i}\bar{\omega}_k^\alpha\}$ is chosen such that the basis functions are linearly independent, G_l^α can be obtained from $\hat{G}^\alpha(\mathrm{i}\bar{\omega}_k^\alpha)$.

If as many sampling points M are chosen as imaginary time points N, these transformations are

$$G_l^\alpha = \sum_{k=0}^{N-1} [\mathbf{F}_\alpha^{-1}]_{lk} G^\alpha(\bar{\tau}_k^\alpha) \tag{9.29}$$

$$= \sum_{k=0}^{N-1} [\hat{\mathbf{F}}_\alpha^{-1}]_{lk} \hat{G}^\alpha(\mathrm{i}\bar{\omega}_k^\alpha), \tag{9.30}$$

where $\mathbf{F}_\alpha$ and $\hat{\mathbf{F}}_\alpha$ are $N \times N$ matrices:

$$[\mathbf{F}_\alpha]_{kl} = F_l^\alpha(\tilde{\tau}_k^\alpha) \tag{9.31}$$

$$[\hat{\mathbf{F}}_\alpha]_{kl} = \hat{F}_l^\alpha(\mathrm{i}\tilde{\omega}_k^\alpha). \tag{9.32}$$

This procedure only requires evaluating the Green's function on N sampling points, and linear transforms between the time or frequency domain and the basis representation G_l^α become invertible. G_l^α can thus serve as a proxy to transform between imaginary time and frequency sampling points, as well as evaluation at arbitrary τ and $\mathrm{i}\omega_n$ values, as illustrated in Fig. 9.2.

For $M > N$ (more sampling points than basis coefficients), the inverses in Eqs. (9.29) and (9.30) are replaced by the corresponding pseudoinverses $\mathbf{F}^+ \equiv (\mathbf{F}^\dagger\mathbf{F})^{-1}\mathbf{F}^\dagger$, and the exact transform is replaced by a least squares fitting procedure.

In practical calculations, different choices of basis functions and sampling points lead to differently conditioned equation systems. A naive choice of sampling points, such as uniformly distributed time or frequency grids, results in almost linearly dependent basis vectors and ill-conditioned transforms, which improve very slowly when additional grid points are added. For an efficient method, a minimal set of sampling points that generates well-conditioned transformation matrices is desired in order to minimize the number of function evaluations and the loss of accuracy during transforms.

In the remainder of this section, we show that such a set with $M = N$ can be generated according to the distribution of the roots of the basis functions.

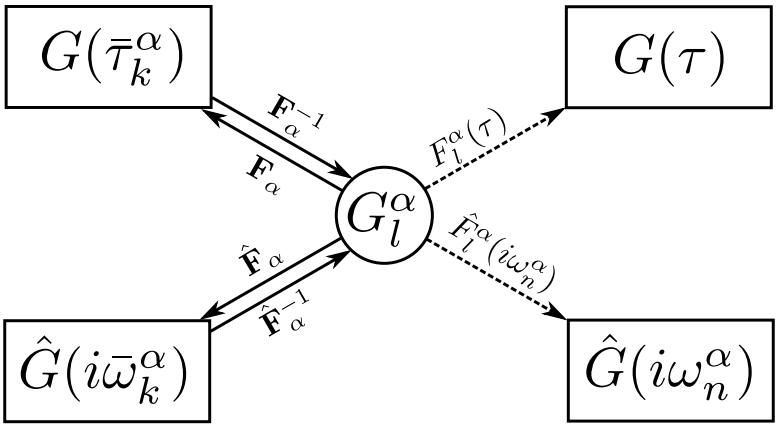

Figure 9.2: Schematic illustration of relations between different representations. Solid lines denote transformations between the basis representation coefficients G_l^α (center) and Green's functions evaluated at imaginary time or frequency sampling points via transformation matrices. Dashed lines represent basis expansions of G_l^α to arbitrary imaginary time or frequency points.

9.4.2 Imaginary time sampling

Chebyshev.– For a truncated Chebyshev basis of size N, the N sampling points in τ are naturally given by the roots of the $(N+1)$-th basis function $T_N(\tau)$ as

$$\bar{\tau}_k^\alpha \equiv \tau\left(\cos\left(\pi\frac{2k+1}{2N}\right)\right), \tag{9.33}$$

for $k = 0, \ldots, N-1$ with the mapping $\tau(x) = \beta(x+1)/2$. These sampling points lead to very well conditioned transformation matrices due to the discrete orthogonality of Chebyshev polynomials, and the condition number of $\mathbf{F}_\alpha$ (defined as $\|\mathbf{F}_\alpha\|_2\|\mathbf{F}_\alpha^{-1}\|_2$) takes the constant value of $\sqrt{2}$.

IR basis.– For the IR basis with N basis functions and given β and $\omega_{\max}$, we choose the sampling points $\{\bar{\tau}_k^\alpha\}$ to be the midpoints of the grid composed of the $N-1$ roots of the highest order basis function $U_{N-1}^\alpha(\tau)$ and the boundary points 0 and β. We choose not to use the roots of the next basis function $U_N^\alpha(\tau)$ like in the Chebyshev case due to the fact that the

IR basis is a numerical basis, and thus it is more convenient to determine sampling points from the available basis functions.

9.4.3 Matsubara frequency sampling

We generate sampling points in Matsubara frequencies following an algorithm similar to the imaginary time cases, with two additional considerations. First, function values should only be evaluated on the discrete Matsubara frequencies. Second, fermionic and bosonic Matsubara frequencies have to be treated separately, and the zero bosonic frequency (which represents static physics) has to be considered explicitly.

Chebyshev.–In the Chebyshev representation, we follow the same heuristics as in the τ sampling by finding or approximating zeros of the next basis function $\hat{T}_N^\alpha(\mathrm{i}\omega_n^\alpha)$ defined in Matsubara frequency.

For fermions and even N, when continued to continuous Matsubara frequency space, $\hat{T}_N^\mathrm{F}(\mathrm{i}\omega_n^\alpha)$ has N roots on the imaginary axis $(-\mathrm{i}\infty, \mathrm{i}\infty)$. We take N Matsubara frequencies closest to these roots as sampling points.

For bosons and odd N, $\hat{T}_N^\mathrm{B}(\mathrm{i}\omega_n^\alpha)$ has $N-1$ roots. We define $N-1$ sampling points as the Matsubara frequencies closest to the roots. We take the zero bosonic frequency $\mathrm{i}\omega_n^\mathrm{B} = 0$ as the last sampling point. The zero bosonic frequency, which corresponds to a constant offset in τ and often has to be treated separately, serves as a natural complement.

The requirement that even N should be used for fermions and odd N for bosons is necessary because the other cases (odd N for fermions or even N for bosons) will not yield adequate number of sampling points due to the analytic structure of $\hat{T}_N^\alpha(\mathrm{i}\omega_n^\alpha)$.

IR basis.–For the IR basis, the procedure for getting frequency sampling points is more empirical due to the numerical nature of the basis function. We partition all Matsubara

frequencies into contiguous groups such that the highest order basis function $\hat{U}^{\alpha}_{l_{\max}}(\mathrm{i}\omega_n^{\alpha})$ has the same sign within each group ($l_{\max} \equiv N-1$). $\hat{U}^{\alpha}_{l_{\max}}(\mathrm{i}\omega_n^{\alpha})$ is either purely real (for even $l_{\max}$) or purely imaginary (for odd $l_{\max}$). We therefore use the sign of the corresponding real or imaginary part as the sign of $\hat{U}^{\alpha}_{l_{\max}}(\mathrm{i}\omega_n^{\alpha})$. The sampling points $\mathrm{i}\bar{\omega}_k^{\alpha}$ are chosen to be those that maximize $|\hat{U}^{\alpha}_{l_{\max}}(\mathrm{i}\omega_n^{\alpha})|$ in each group.

By checking the resulting sampling points numerically, we conclude that by requiring N to be even ($l_{\max}$ odd) for fermionic basis and odd ($l_{\max}$ even) for bosonic basis, the number of sampling points is exactly N, and the bosonic sampling points naturally include zero.

9.4.4 Numerical demonstration

The transformation defined in Eqs. (9.29) and (9.30) is exact if the Green's function is a linear combination of a finite set of basis functions (9.24). With physical Green's functions, this is seldom the case, and any finite expansion incurs a truncation error. Fortunately, in both the IR and the Chebyshev expansion, the truncation error is controlled: the analyticity of the finite-temperature Green's function in $(0, \beta)$ guarantees exponential convergence of the Chebyshev expansion, and the construction of the IR basis from analytic continuation guarantees the same thing for the IR expansion [138].

To demonstrate the behavior of the sparse sampling scheme when applied to physical Green's functions, we consider a model with semicircular density of states (full bandwidth is 2) for the IR basis in Fig. 9.3:

$$\rho(\omega) = \frac{2}{\pi}\sqrt{1-\omega^2}. \tag{9.34}$$

As an example, the top left panel shows U^{F}_{34} as a function of imaginary time (blue lines) and illustrates the $N = 34$ sampling points for the fermionic basis of $\beta = 100$ and $\omega_{\max} = 1$

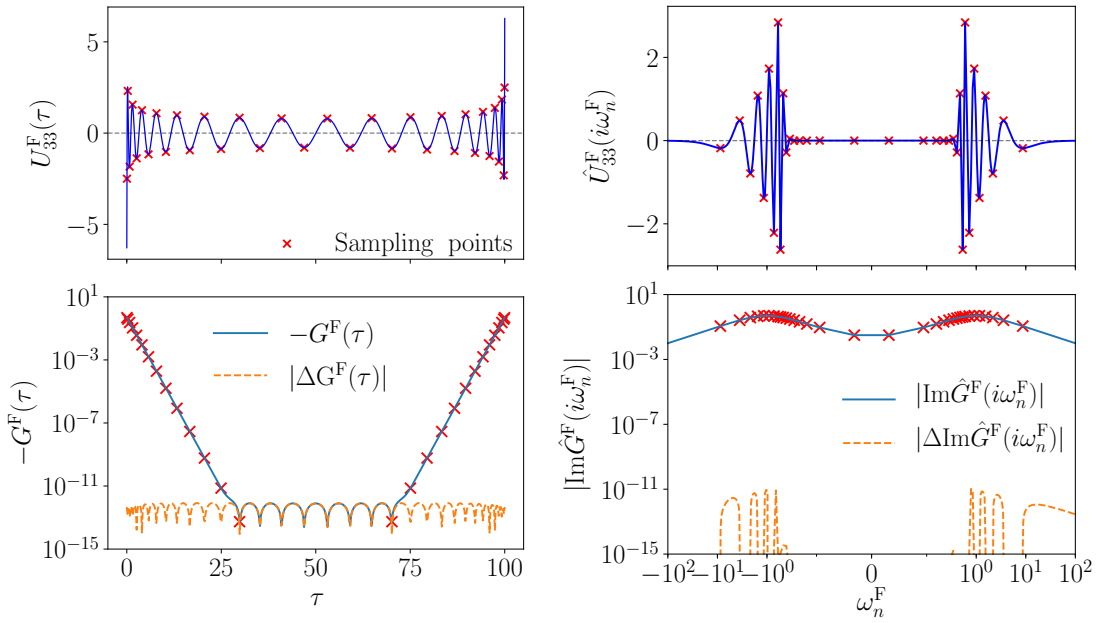

Figure 9.3: Distribution of sampling points and results transformed from imaginary time (left panels) and Matsubara frequency (right panels) for the IR basis by sparse sampling. We consider a model of a semicircular density of states of half width 1 at $\beta = 100$ defined in Eq. (9.34). We take $\omega_{\mathrm{max}} = 1$ for the IR basis. The sampling points are denoted by crosses. Top row: The basis functions used to generate sampling points ($l = 33$). Bottom row: Comparison of the reconstructed Green's function to exact results.

(red crosses). The sampling points cluster near $\tau = 0$ and β, where this basis function is rapidly oscillating.

The top right panel of Fig. 9.3 shows the distribution of the Matsubara frequency sampling points generated for the same basis for $N = 34$. The Fourier transformed basis function $\hat{U}_{33}^{\mathrm{F}}(\mathrm{i}\omega_n)$ (blue lines) exhibits $N - 1 = 33$ sign changes, which define $N = 34$ sampling points (red crosses). The sampling points are distributed almost logarithmically, which allows us to capture all the features of $\hat{U}_{l_{\max}}^{\mathrm{F}}(\mathrm{i}\omega_n)$ from low to high frequency.

In the bottom row of Fig. 9.3, the left and right panels illustrate the sampling of $G(\tau)$ and $\hat{G}(\mathrm{i}\omega_n)$, respectively. The sampling points capture relevant features of the Green's function in both cases. We compare the interpolated and extrapolated results with the numerically exact values. For imaginary time, one can see agreement at the level of $\sim 10^{-12}$ in the whole interval of $[0, \beta]$, which matches the singular value cutoff we used (10^{-12}).For Matsubara frequency, the coefficients obtained by the sparse sampling not only interpolate $\hat{G}(\mathrm{i}\omega_n)$ but also extrapolate it precisely beyond the highest sampling frequency.

9.4.5 Technical details

In practical applications, it is advisable to precompute the sampling points and transformation matrices for the basis functions employed, to avoid unnecessary evaluations of $\hat{T}_l^{\alpha}(\mathrm{i}\omega_n^{\alpha})$ and $\hat{U}_l^{\alpha}(\mathrm{i}\omega_n^{\alpha})$.

The `irbasis` library [140] provides numerical data of the IR basis functions in the dimensionless form for selected values of Λ from $\Lambda = 10$ up to $\Lambda = 10^7$. The numerical evaluation of the basis functions in Matsubara frequency is also implemented.

The procedures we have presented are not unique, and we do not claim that they are optimal definitions of sampling points. One may design other choices with similar or, po-

tentially, better numerical performance. The number of sampling points may also exceed the basis size N, as long as inversions in Eqs. (9.29) and (9.30) are replaced with a pseudoinverse, and the resulting transformations are well-conditioned. Nevertheless, the algorithms introduced in this section provide a systematic and unambiguous way to obtain the minimum sets of sampling points which yield well-conditioned numerical transforms and high accuracy.

9.4.6 Application to GF2 and *GW*

We first briefly recap definitions of GF2 and *GW* introduced in Chapter 5. The self-consistency in GF2 and *GW* is defined by the Dyson equation (3.34)

$$\hat{G}(\mathrm{i}\omega_n^{\mathrm{F}}) = [(\mathrm{i}\omega_n^{\mathrm{F}} + \mu)I - F - \hat{\tilde{\Sigma}}(\mathrm{i}\omega_n^{\mathrm{F}})]^{-1}. \tag{9.35}$$

The Fock matrix $F = h + \Sigma^{\mathrm{HF}}$ includes the frequency independent Hartree-Fock contribution

$$\Sigma_{ij}^{\mathrm{HF}} = (2V_{ijkl} - V_{ilkj})\rho_{kl}, \tag{9.36}$$

where i, j, k, l are orbital indices. GF2 approximates the frequency-dependent self-energy $\tilde{\Sigma}$ as

$$\begin{aligned}\tilde{\Sigma}_{ij}(\tau) = -G_{kl}(\tau)G_{qm}(\tau)G_{np}(-\tau)\times \\ \times V_{ikpq}(2V_{ljmn} - V_{mjln}),\end{aligned} \tag{9.37}$$

while *GW* approximates $\tilde{\Sigma}$ as

$$\tilde{\Sigma}_{ij}(\tau) = -G_{lk}(\tau)\tilde{W}_{ilkj}(\tau), \tag{9.38}$$

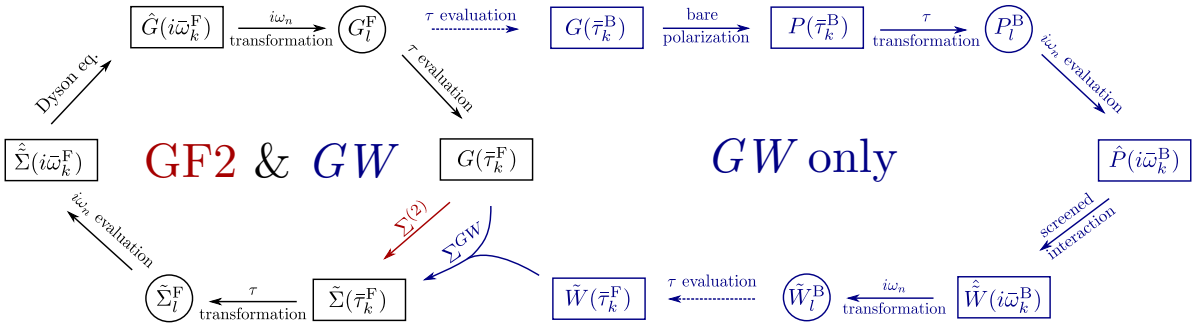

Figure 9.4: Illustration of GF2 and GW procedures using the sparse sampling scheme. The red and blue lines denote GF2-only and GW-only steps, respectively. Dashed arrows indicate evaluations that change the statistics of the representation.

where $\tilde{W} = W - V$ is the frequency-dependent part of the screened interaction W. W is calculated by the random phase approximation (RPA) [151] as

$$\hat{W}_{ijkl}(\mathrm{i}\omega_n^{\mathrm{B}}) = V_{ijkl} + V_{ijpq}\hat{P}_{qpsr}(\mathrm{i}\omega_n^{\mathrm{B}})\hat{W}_{rskl}(\mathrm{i}\omega_n^{\mathrm{B}}), \tag{9.39}$$

where the bare polarization is given by

$$P_{ijkl}(\tau) = -G_{il}(\tau)G_{jk}(-\tau). \tag{9.40}$$

At self-consistency, physical properties are evaluated from G and Σ. For example, density matrix ρ is given by $\rho_{ij} = G_{ji}(0^-) = -G_{ji}(\beta)$, and the total electronic energy is

$$\begin{aligned} E = &\mathrm{Tr}\left[\rho H_0\right] + \frac{1}{2}\mathrm{Tr}\left[\rho\Sigma_{\mathrm{HF}}\right] \\ &+ \frac{1}{2\beta}\sum_n \mathrm{Tr}\left[\hat{\tilde{\Sigma}}(\mathrm{i}\omega_n^{\mathrm{F}})\hat{G}(\mathrm{i}\omega_n^{\mathrm{F}})\right]. \end{aligned} \tag{9.41}$$

Fig. 9.4 illustrates how the self-consistent calculations can be performed by sparse sampling. In a GF2 calculation, we first evaluate Green's function at sampling points $\bar{\tau}_k^{\mathrm{F}}$, and construct the self-energy $\tilde{\Sigma}(\bar{\tau}_k^{\mathrm{F}})$ following second-order approximation (9.37). The self-energy is

then transformed to the basis representation $\tilde{\Sigma}_l^{\mathrm{F}}$ following Eq. (9.29), which is then evaluated on the frequency sampling points to get $\hat{\tilde{\Sigma}}(\mathrm{i}\bar{\omega}_k^{\mathrm{F}})$. The Dyson equation (3.34) is solved for each $\mathrm{i}\bar{\omega}_k^{\mathrm{F}}$ to obtain $\hat{G}(\mathrm{i}\bar{\omega}_k^{\mathrm{F}})$. We then transform the Green's function to its basis representation G_l^{F} following Eq. (9.30). The updated Green's function in τ is recovered by evaluating G_l^{F} on sampling points $\bar{\tau}_k^{\mathrm{F}}$. The procedure is repeated using the updated Green's function until self-consistency, which corresponds to the inner loop in Fig. 9.4. We can see that the compact basis representations G_l^{F} and $\tilde{\Sigma}_l^{\mathrm{F}}$ serve as proxies to convert back and forth between τ and frequency domains, all evaluated on corresponding sampling points.

While all quantities involved in GF2 are fermionic, in GW we have to switch between fermionic quantities (G and Σ) and bosonic quantities (P and W). This is achieved again by using compact basis representations as proxies: when calculating the polarization P, we evaluate G_l^{F} on the *bosonic* sampling points $\bar{\tau}_k^{\mathrm{B}}$ and assemble $P(\bar{\tau}_k^{\mathrm{B}})$ following Eq. (9.40). We then carry out the calculation of $\tilde{W}$ on the frequency sampling points $\mathrm{i}\bar{\omega}_k^{\mathrm{B}}$, and obtain the compact basis representation $\tilde{W}_l^{\mathrm{B}}$. Finally we evaluate $\tilde{W}_l^{\mathrm{B}}$ back on the *fermionic* sampling points $\bar{\tau}_k^{\mathrm{F}}$, and compute self-energy using the GW approximation (9.38). The rest of the procedure is identical to GF2. Two additional matrices may be precomputed to allow fast switching between fermionic and bosonic representations in GW:

$$[\mathbf{F}^{\mathrm{F}\to\mathrm{B}}]_{kl} = F_l^{\mathrm{F}}(\bar{\tau}_k^{\mathrm{B}}) \tag{9.42}$$

$$[\mathbf{F}^{\mathrm{B}\to\mathrm{F}}]_{kl} = F_l^{\mathrm{B}}(\bar{\tau}_k^{\mathrm{F}}). \tag{9.43}$$

Note that the inverse transform of those matrices are not well defined in general, since sampling points generated for one type of statistics usually do not serve as good sampling points for the other.

In the total energy evaluation (9.41), the frequency summation term can be rewritten using

an auxiliary scalar quantity S such that

$$\frac{1}{2\beta}\sum_n \mathrm{Tr}\,[\hat{\tilde{\Sigma}}(\mathrm{i}\omega_n^{\mathrm{F}})\hat{G}(\mathrm{i}\omega_n^{\mathrm{F}})] = \frac{1}{2\beta}\sum_n \hat{S}(\mathrm{i}\omega_n^{\mathrm{F}}) \\ = \frac{1}{2}S(0^-) = -\frac{1}{2}S(\beta) \tag{9.44}$$

where

$$\hat{S}(\mathrm{i}\omega_n^{\mathrm{F}}) = \mathrm{Tr}\,[\hat{\tilde{\Sigma}}(\mathrm{i}\omega_n^{\mathrm{F}})\hat{G}(\mathrm{i}\omega_n^{\mathrm{F}})]. \tag{9.45}$$

We first evaluate $\hat{S}(\mathrm{i}\bar{\omega}_k^{\mathrm{F}})$ on the frequency sampling points $\mathrm{i}\bar{\omega}_k^{\mathrm{F}}$, which is then transformed to the basis representation S_l^{F}. The value of $S(\beta)$ is now a straightforward basis expansion at $\tau = \beta$

$$S(\beta) = \sum_{l=0}^{N-1} S_l^{\mathrm{F}} F_l^{\mathrm{F}}(\beta). \tag{9.46}$$

For the Chebyshev basis we simply have $T_l(\beta) = 1$. In the case of IR basis, it is desirable to also tabulate the values $U_l^{\mathrm{F}}(\beta)$ along with the transformation matrices for efficient evaluations of relevant quantities.

Similarly, the calculation of the density matrix ρ is a straightforward evaluation at $\tau = \beta$ from G_l^{F}. In calculations where the number of electrons is fixed, the chemical potential μ needs to be adjusted in each self-consistent iteration through a root finding procedure to conserve particle number [4]. This step involves repeated density evaluations and solutions of the Dyson equation using the frequency-dependent self-energy, and sometimes becomes the bottleneck of the calculation. The sparse sampling scheme massively reduces the number of frequency points needed in this process, which leads to a significant speedup over traditional approaches.

Note that since most basis functions, including Chebyshev and IR, cannot capture constant

shifts in frequency (which correspond to a delta function at $\tau = 0$), it is important that one only expands the frequency-dependent components such as $\tilde{\Sigma}$ and $\tilde{W}$ using compact basis representations. One also needs to be careful when zero-energy mode exists in a bosonic quantity, in which case the IR basis function cannot describe the constant shift in imaginary time [139].

9.4.7 Results

Hydrogen chain

We first apply our sparse sampling scheme to GF2 and *GW* caculations of a system composed of 10 hydrogen atoms placed on a straight line with equal spacing r. The hydrogen chain, due to its simplicity, serves as a benchmark platform for testing numerical methods of correlated electrons. Reference data for the hydrogen chain were carefully compared and analyzed in Ref. [44] with many methods including GF2 and *GW*. It is therefore convenient to use this system to analyze the sparse sampling scheme.

We perform GF2 and *GW* calculations for H_{10} with $r = 1$ a_0 and $\beta = 1000$ ${E_h}^{-1}$ ($T \sim 315.8$ K). We use the minimal basis set STO-6g with only one $1s$ orbital per atom. Hartree-Fock calculations show that the difference between the highest and the lowest Hartree-Fock energy levels is about $\Delta E \approx 5.76$ E_h (~ 156 eV). The dimensionless parameter for the IR basis can thus be estimated by taking Λ to bound the value $\beta\Delta E \approx 5.76 \times 10^3$. We take $\Lambda = 10^5$ in all our calculations.

Both Chebyshev and IR basis functions are used together with the sparse sampling scheme. To demonstrate convergence, we examine a series of calculations with different sizes of basis functions. Typically we choose N to be an even number, which is then used as the size of the fermionic basis. For corresponding bosonic basis functions in *GW*, we used the closest odd

number $N-1$ as the basis size. The Python library `irbasis` version 2.0.0b1 [140] is used for calculating IR basis functions.

With each basis of fixed size N, we perform self-consistent GF2 and GW calculations following the procedures described in the previous section. The initial guess of the Green's function is constructed from Hartree-Fock calculations using the PySCF library [152]. Self-consistent calculations, as illustrated in Fig. 9.4, are then executed repeatedly until the energy difference is below the convergence tolerance $E_{\mathrm{tol}} = 10^{-8}\ \mathrm{E_h}$ between two consecutive iterations. Energy values converged with respect to basis size N are cross-checked with reference data to ensure correctness. All results are then compared to converged values to assess errors in energy and density matrix as a function of basis size N.

In the left panel of Fig. 9.5, we show the convergence of errors in total energy with respect to N. GF2 and GW share similar convergence behavior, in that the error decreases almost exponentially with respect to N in either basis representation. In the Chebyshev representation, we obtain convergence of the total energy below the tolerance E_{tol} with around 350 Chebyshev polynomials. With IR basis, convergence is reached with less than 100 basis functions. Similarly, the right panel of Fig. 9.5 illustrates the convergence of the maximum error in density matrix, exhibiting an exponential decay of errors. This indicates that with a reasonable number of sampling points, we can reach very high precision in both global observables such as the energy and local properties like the density matrix. The observation that GW shows a convergence behavior similar to GF2 indicates that no substantial additional errors are introduced during the frequent switching between fermionic and bosonic representations.

The sparse sampling scheme is stable thanks to the well-conditioned transformation matrices generated from the sampling points. We demonstrate this in Fig. 9.6, which shows the relative magnitude of basis expansion coefficients for the converged solution. N is chosen

large enough to ensure that all quantities are well approximated by the basis representations, with $N = 600$ for Chebyshev and $N = 130$ for IR with $\Lambda = 10^5$. Even after several iterations with multiple transforms forward and backward between different types of sampling points, we see that for all quantities, expansion coefficients decay at least exponentially as we expect from the properties of the basis, down to a relative size below 10^{-12}. The truncation error due to the finite basis expansion is therefore controlled, and no amplification of error is observed during the self-consistent iterations.

The sparse sampling scheme ensures that the number of τ grid points and the number of Matsubara frequencies is the same as the basis size N. We reach a precision of 8 digits in total energy with only hundreds of τ points and Matsubara frequencies, a significant improvement from the conventional approach used in Ref. [44], where $\sim 10^4$ Matsubara frequencies were used for higher temperature ($\beta \sim 100\,{\mathrm{E_h}}^{-1}$) and bigger convergence threshold ($E_{\mathrm{tol}} \sim 10^{-6}\,\mathrm{E_h}$ or $10^{-7}\,\mathrm{E_h}$). This greatly reduces the computational cost and memory requirement in all parts of the calculations while still being accurate. The sparse sampling scheme thus allows to tackle problems that were too costly to calculate, especially low temperature calculations of systems with large energy scales.

Noble gas atoms

Noble gas atoms such as Kr have deep core states, if no pseudopotentials are employed. Due to the large energy scale caused by the core states, it is computationally very demanding for conventional methods to resolve sharp features close to $\tau = 0$ or β in $G(\tau)$ or the slow decay of $\hat{G}(i\omega_n)$ at high frequency. Even with a compact polynomial basis such as Chebyshev, thousands of basis functions are required to represent the Green's functions [136], and effective core potentials (ECP), which absorb electron in inner orbitals to the ionic potential, have to be employed in most practical calculations. We choose this problem to demonstrate the

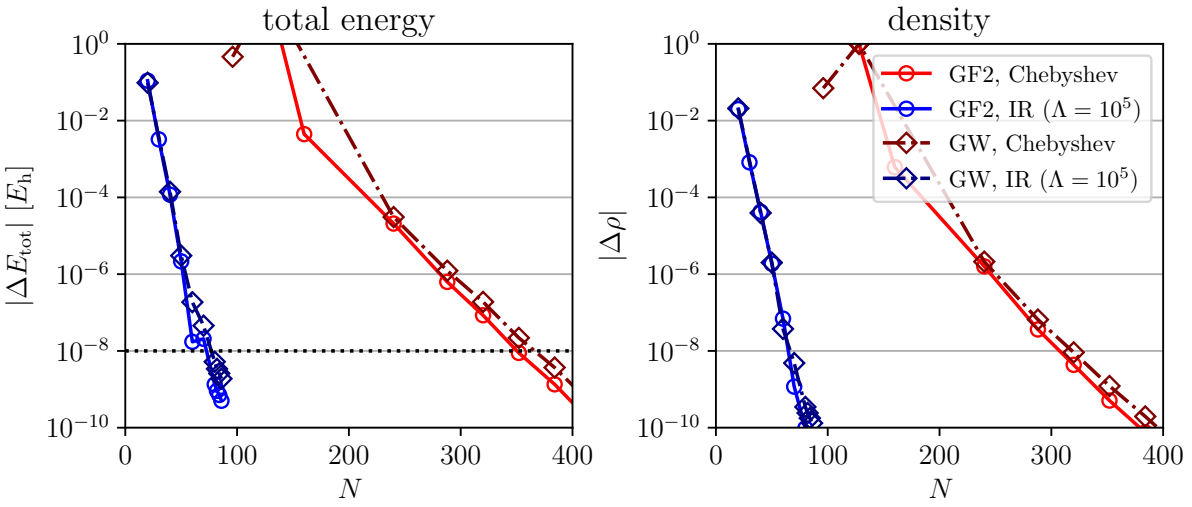

Figure 9.5: Error in total energy and density from converged GF2 and *GW* calculations for H_{10} in minimal basis at $\beta = 10^3\ {E_h}^{-1}$. The left panel shows the convergence of the total energy with the dashed horizontal line representing the convergence threshold of $10^{-8}\ E_h$. The right panel shows the convergence of the density.

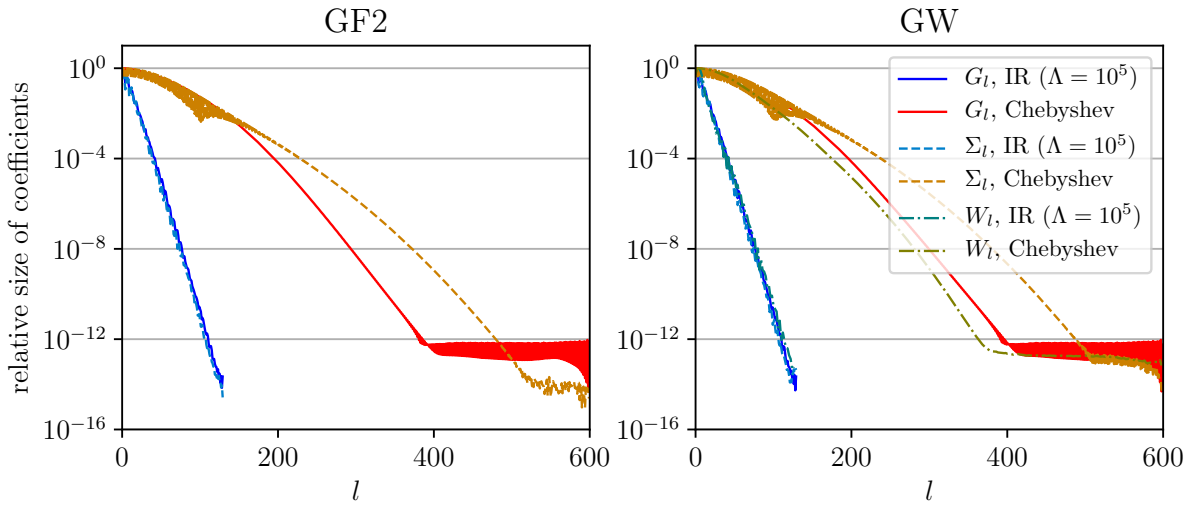

Figure 9.6: Relative size of basis expansion coefficients with converged GF2 (left panel) and *GW* (right panel) calculations for H_{10} in STO-6g at $r = 1.0$ a.u., $\beta = 1000\ {E_h}^{-1}$.

power of the sparse sampling method when dealing with large energy scales, while avoiding additional physical or technical difficulties.

The IR basis is a natural choice for systems with large energy scales. As long as the spectral cutoff $\omega_{\max}$ is chosen to include all energy scales in the system, exponential convergence of the coefficients is guaranteed by construction, usually with no more than a couple hundred basis functions [138]. The full potential of IR basis is realized when combined with the sparse sampling scheme developed in this paper. Numerical difficulties in either τ or frequency domain are reduced to a single issue: whether the basis functions can capture all relevant quantities well. Therefore, by using a sparse sampling method with the IR basis, we are able to treat noble gas atoms efficiently even in all-electron calculations.

We choose the all-electron correlation consistent basis set cc-pVDZ [43], and perform GF2 and GW calculations of four noble gas atoms: He, Ne, Ar, and Kr, at $\beta = 1000\ {\mathrm{E_h}}^{-1}$. Similar to the case for H_{10}, we estimate the dimensionless parameter Λ according to the Hartree-Fock energy spectrum for each individual atom: $\Lambda = 10^4$ for He, $\Lambda = 10^5$ for Ne, and $\Lambda = 10^6$ for Ar and Kr. The sparse sampling scheme is used in all calculations, and we vary the basis size N to explore the convergence behavior. The energy convergence threshold is set to $E_{\mathrm{tol}} = 10^{-8}\ \mathrm{E_h}$, which is much smaller than the energy scale of Kr ($\sim 10^3\ \mathrm{E_h}$).

Fig. 9.7 shows the energy convergence of all four atoms with respect to basis size N with GF2 (left column) and GW (right column). The upper panels indicate that the basis converges for all atoms, with absolute difference in energy dropping in an exponential trend below the convergence tolerance E_{tol}. The lower panels put the convergence in a relative scale, where all atoms in both GF2 and GW can reach $\sim 10^{-10}$ relative convergence.

Our results show that we can obtain fast basis convergence with around 100 basis functions for all systems studied. This is consistent with the property of IR basis that the basis size N scales only logarithmically with Λ [139].

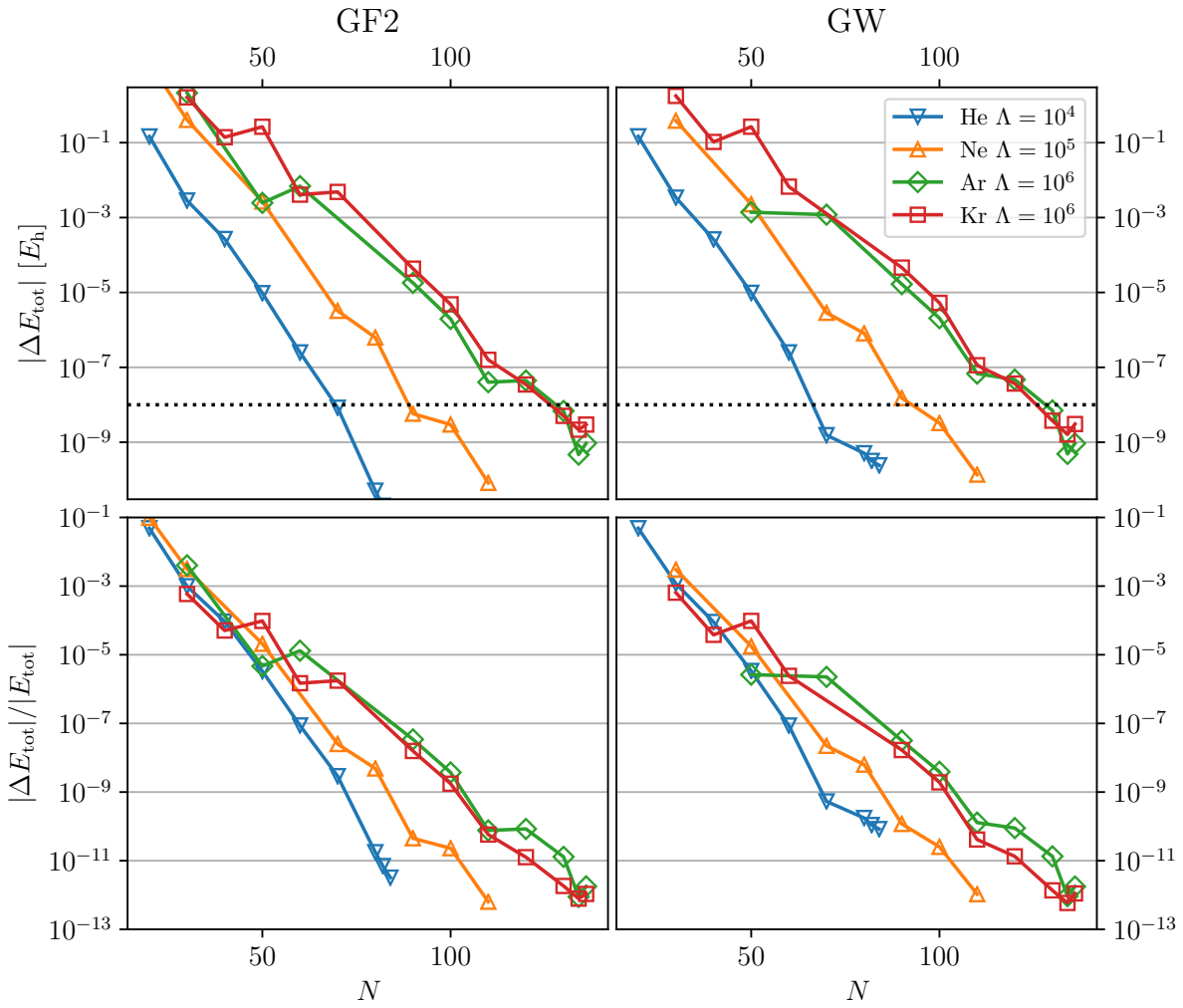

Figure 9.7: Total energy convergence in GF2 and *GW* calculations of noble gas atoms with respect to the number of IR basis functions N. Left and right columns show GF2 and *GW* results respectively. Top row: absolute differences in total energy to the converged value. Dashed black horizontal line illustrates the energy convergence threshold of 10^{-8} $\mathrm{E_h}$. Bottom row: relative differences in total energy.

9.4.8 Numerical stability of sparse sampling

When the expansion coefficients are numerically evaluated using the sparse sampling procedures, numerical errors, such as round-off errors from floating point operations or truncation errors from a finite basis cutoff, may be amplified due to the (pseudo-)inversion process. This error amplification can be quantified by the condition number of the transformation matrices $\mathbf{F}_\alpha$ and $\hat{\mathbf{F}}_\alpha$, defined as the product of the 2-norms of the matrix and its inverse. In Fig. 9.8 we show the behaviors of such condition numbers for the IR basis as a function of the basis size $N = L$ (left panel, compared to the Chebyshev representation), and as a function of Λ (right panel). We can see that up to a significant number of basis functions, the condition numbers are $< 10^4$, which indicates well-conditioned inversion problems. In addition, the condition numbers show an approximate scaling of $O(L^{1/2})$, which is slower that of the Chebyshev representation $O(L^{3/2})$. Since the values of L and Λ shown in Fig. 9.8 cover most values used in the calculations reviewed in this paper, the sparse sampling scheme guarantees stable numerical routines to get accurate results.

In practical applications of the sparse sampling method, one should take care not to introduce large systematic errors in the basis representation, such as large truncation errors due to insufficient basis size L or control parameter Λ. For example, a systematic error at the level of 10^{-3} in the basis representation, amplified by a condition number of 10^3 of the "fitting" procedure, may lead to a numerical error greater than the actual result. As shown in Fig. 9.9, such a situation could make the simulation unstable. Therefore, to ensure stable numerical calculations with the sparse sampling method, one should choose the appropriate basis parameters L and Λ such that the basis representation stays accurate.

Figure 9.9 shows examples of stable and unstable GW calculation of the Krypton atom in cc-pVDZ basis with $\beta = 10\,000\ {\mathrm{E_h}}^{-1}$ using sparse sampling. We carried out five GW

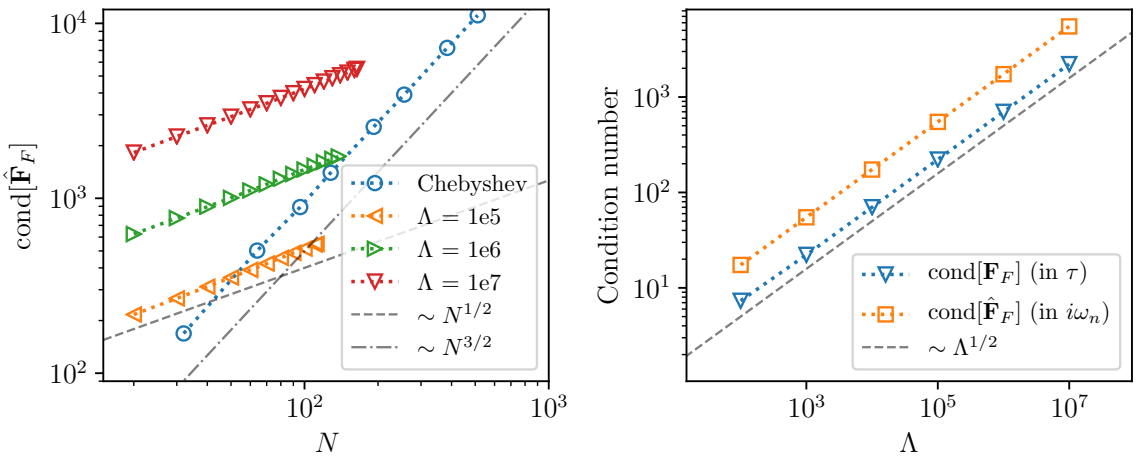

Figure 9.8: Condition number of the IR transformation matrices $\hat{\mathbf{F}}_\mathrm{F}$ and $\mathbf{F}_\mathrm{F}$ [146]. Left panel shows the condition number of frequency transformation matrices $\hat{\mathbf{F}}_\mathrm{F}$ as a function of basis size $N = L$, in comparison with the Chebyshev representation. Right panel shows the condition number of both τ and $i\omega_n$ transformation matrices with respect to Λ, where N is chosen to be the maximum number of coefficients with the same cutoff in singular values S_l^α, provided in the `irbasis` library [139].

iterations with different choices of Λ, and plotted the norms of G_l from each iteration. The left panel of Fig. 9.9 shows results for $\Lambda = 10^7$. At each iteration, the Green's function is well approximated by the IR basis and the basis truncation error is small, resulting in a stable simulation. In the right panel, a smaller $\Lambda = 10^6$ is used, which is insufficient for this system and introduces a large systematic error (around 10^{-3}), while the condition number of $\mathbf{F}$ is $\sim 10^3$. The systematic error is amplified by fitting procedure, rendering the *GW* simulation unstable.

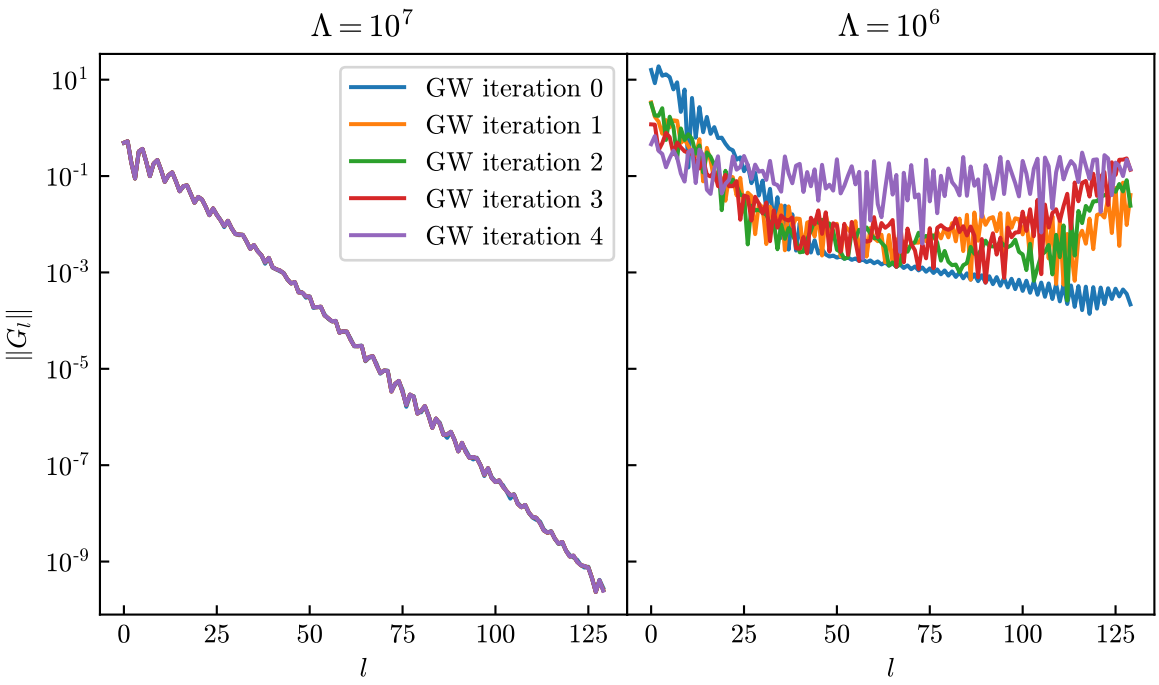

Figure 9.9: Examples of stable and unstable *GW* calculations of the Krypton atom.

Chapter 10

Conclusion and Outlook

The ability of computing the electronic properties of strongly-correlated materials from the first principles would profoundly deepen our quantitative understanding of electron correlation effects, and potentially enable predictive calculations for discovering new correlated materials with desired functionalities. Numerous theoretical and numerical methods are being proposed to achieve that goal. Among these developments, diagrammatic methods, based on a straightforward perturbative expansion in terms of the electron–electron interaction, have seen rapid progress and show significant potential, due to their theoretical simplicity, flexibility in formulation, and improving efficiency in numerical calculations. This thesis provides an overview of the fundamentals for developing diagrammatic methods for realistic systems, and presents our progress towards a truly general diagrammatic solver for realistic systems.

The first half of the thesis briefly summarizes basic notations and concepts common in diagrammatic methods, especially DiagMC. In Chapter 2, we define the general mathematical problem we aim to solve — the electronic Hamiltonian, as well as its atomic basis sets representation. In Chapter 3, we introduce the general formalism for performing pertur-

bation expansion of the electron Hamiltonian, and basic rules for Feynman diagrammatic representations. Chapter 4 reviews the Monte Carlo integration technique with importance sampling and its application to diagrammatic series. In Chapter 5, we review the definition of the Luttinger–Ward functional, the general self-consistent diagrammatic framework based on the functional, and examples of such self-consistent methods.

Chapter 6 summarizes two benchmark projects for state-of-the-art many-body methods in realistic systems, including many diagrammatic methods. Strengths and limitations of these methods are investigated systematically, and the results are carefully cross-checked with other methods. These benchmarks demonstrate a systematic approach for reaching a theoretical "consensus" of methods with different areas of applicability, and provide valuable reference for future developments

Chapter 7 proposes a general DiagMC solver for systems with realistic interactions following the CDet approach, and systematically tests the solver on realistic molecules. The method handles multiple orbitals, low temperatures, and sparse interaction matrices well at moderate electron correlation. In the strongly-correlated regime, the method is limited by the potential divergence of the perturbation series.

Chapter 8 shows our attempt to mitigate the convergence problem from a direct diagrammatic expansion as used in Chapter 7. The interaction-expansion inchworm Monte Carlo method introduced in this Chapter obtains the physical observables via a sequence of incremental diagrammatic expansions. Preliminary results from inchworm show convergent solutions in systems where bare DiagMC fails.

Chapter 9 summarizes the technical methods for representing the electron correlation functions efficiently. We reviewed multiple compact representations, and introduced the sparse sampling method which enables efficient storage and fast computation in both imaginary time and Matsubara frequencies.

Going forward, two fundamental (and often intertwined) problems remain that limit the applicability of DiagMC methods in realistic systems. The first one is the efficient computation of high-order Feynman diagrams. CDet has been a recent breakthrough, which significantly simplifies the Monte Carlo sampling scheme and speeds up calculations of high-order diagrams. However, because CDet eliminates individual diagrammatic topologies from the procedure to gain efficiency, it looses some flexibilities from the original DiagMC framework. For example, it cannot be applied to self-consistent skeleton series (so far) due to its reliance to Wick's theorem. Formulating CDet in frequency and momentum representations may be difficult. Arbitrary counter terms for better perturbation starting points often lead to a redesign of the recursive procedure. Any progress along these directions would further widen the applicability of CDet. Alternatively, a smart partial summation based on diagrammatic topologies may retain the advantages of explicit diagram sampling in DiagMC, while avoiding a severe sign problem at high expansion orders.

The second problem is the convergence (or the lack thereof) of the diagrammatic series. Bare diagrammatic expansions are destined to diverge in the strong-correlation limit. The inchworm Monte Carlo approach introduced in Chapter 8 improves the convergence behavior, but still does not guarantee a fast convergence. Analytical resummation techniques may be useful if they can be fully automatized and can work with multi-orbital systems. More complex counter terms based on the shifted action formalism, introduced in Section 3.5, may enable systematic control of the convergence by changing the starting points of diagrammatic expansions if the diagrammatic series can be computed efficiently. Reformulating the diagrammatic language in terms of two-particle quantities may also provide useful insights, e.g., by normalizing the interaction vertices. A faster convergence would also reduce the number of expansion orders required for a certain numerical precision, which in turn alleviates the first problem.

In essence, diagrammatic methods all try to solve the unknown problem starting from the known in a perturbative manner. It would be interesting to know the true limit of these methods, through careful theoretical thinking and numerical optimizations, which may hint at fundamental challenges of either the interacting electron problem, or the classical computing paradigm.

Appendix A

Definitions and Properties of the Scattering Amplitude

The intuition in defining the M object is similar to the relation between the self-energy Σ and the Luttinger-Ward functional $\Phi[G]$ introduced in Chapter 5

$$\Sigma(x', x) = \frac{\delta \Phi[G]}{\delta G(x, x')}, \tag{A.1}$$

which gives the 1PI amputated diagrams with the "bold" propagator G [52]. Here we have employed the compound space-time indices $x = (a, \tau)$. To get the connected amputated diagrams with the "bare" propagator instead, we define a similar relation

$$M(x', x) = \beta \frac{\delta(\Omega - \Omega_0)}{\delta g(x, x')} = -\frac{\delta \log Z/Z_0}{\delta g(x, x')}. \tag{A.2}$$

We show that by carrying out this functional derivative, we will recover the definition of M as in Eq. (7.20).

Switching to the action formalism using coherent state path-integrals of Grassmann vari-

ables [39], we rewrite the partition function as

$$Z = \int \mathscr{D}[\bar{c}, c] e^{-S[\bar{c},c]}, \tag{A.3}$$

where the action is given as

$$\begin{aligned} S &= S_0 + S_V, \\ S_0 &= -\int \mathrm{d}y\mathrm{d}y' \bar{c}(y') g^{-1}(y', y) c(y), \\ S_V &= \frac{1}{4} \int \mathrm{d}\tau \sum_{abcd} U_{abcd} \bar{c}_a(\tau) \bar{c}_c(\tau) c_d(\tau) c_b(\tau). \end{aligned} \tag{A.4}$$

Observe that

$$\begin{aligned} \frac{\delta \log Z}{\delta g(x, x')} &= \frac{1}{Z} \frac{\delta}{\delta g(x, x')} \int \mathscr{D}[\bar{c}, c] \\ &\quad \times \exp\left[\int \mathrm{d}y\mathrm{d}y' \bar{c}(y') g^{-1}(y', y) c(y) - S_V \right] \\ &= \int \mathrm{d}y\mathrm{d}y' \frac{\delta g^{-1}(y', y)}{\delta g(x, x')} \frac{1}{Z} \int \mathscr{D}[\bar{c}, c] \bar{c}(y') c(y) e^{-S} \\ &= \int \mathrm{d}y\mathrm{d}y' \frac{\delta g^{-1}(y', y)}{\delta g(x, x')} G(y, y'). \end{aligned} \tag{A.5}$$

Using the fact that for an invertible matrix $\boldsymbol{A}$,

$$(\boldsymbol{A} + \delta \boldsymbol{A})^{-1} - \boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1} \delta \boldsymbol{A} \boldsymbol{A}^{-1} \tag{A.6}$$

$$\frac{\delta [\boldsymbol{A}^{-1}]_{ij}}{\delta \boldsymbol{A}_{kl}} = -[\boldsymbol{A}^{-1}]_{ik} [\boldsymbol{A}^{-1}]_{lj}, \tag{A.7}$$

we have

$$\frac{\delta \log Z}{\delta g(x, x')} = -\int \mathrm{d}y\mathrm{d}y' g^{-1}(x', y) G(y, y') g^{-1}(y', x). \tag{A.8}$$

Similarly in the non-interacting case,

$$\begin{aligned}\frac{\delta \log Z_0}{\delta g(x,x')} &= -\int \mathrm{d}y\mathrm{d}y' g^{-1}(x',y)g(y,y')g^{-1}(y',x) \\ &= -g^{-1}(x',x).\end{aligned} \tag{A.9}$$

Putting it all together, we have

$$\begin{aligned}M(x',x) &= -\frac{\delta \log(Z/Z_0)}{\delta g(x,x')} \\ &= \int \mathrm{d}y\mathrm{d}y' g^{-1}(x',y)[G(y,y') - g(y,y')]g^{-1}(y',x).\end{aligned} \tag{A.10}$$

Therefore

$$G(y,y') = g(y,y') + \int \mathrm{d}x\mathrm{d}x' g(y,x')M(x',x)g(x,y') \tag{A.11}$$

which is exactly the same as Eq. (7.20).

Expanding compound indices $x, y, \ldots$, Eq. (A.2) can be rewritten as

$$M_{ab}(\tau_1,\tau_2) = \beta\frac{\delta(\Omega-\Omega_0)}{\delta g_{ba}(\tau_2,\tau_1)}. \tag{A.12}$$

In practice, we usually work with the time-translational invariant functions $M(\tau)$ and $g(\tau)$ instead of their two-variable form. To that effect, we consider for $0 < \tau \leq \beta$,

$$\begin{aligned}\beta\frac{\delta(\Omega-\Omega_0)}{\delta g_{ba}(-\tau)} &= \sum_{a'b'}\int_0^\beta \mathrm{d}\tau_1\mathrm{d}\tau_2 \frac{\beta\delta(\Omega-\Omega_0)}{\delta g_{b'a'}(\tau_2,\tau_1)}\frac{\delta g_{b'a'}(\tau_2,\tau_1)}{\delta g_{ba}(-\tau)} \\ &= \sum_{a'b'}\int_0^\beta \mathrm{d}\tau_1\mathrm{d}\tau_2 M_{a'b'}(\tau_1-\tau_2)\frac{\delta g_{b'a'}(\tau_2,\tau_1)}{\delta g_{ba}(-\tau)},\end{aligned} \tag{A.13}$$

in which

$$\begin{aligned}\frac{\delta g_{b'a'}(\tau_2,\tau_1)}{\delta g_{ba}(-\tau)} &= \frac{\delta g_{b'a'}(\tau_2-\tau_1)}{\delta g_{ba}(-\tau)}\\ &= \delta_{b'b}\delta_{a'a}[\delta(\tau_2-\tau_1+\tau)\Theta(\tau_1-\tau_2)\\ &\quad -\delta(\tau_2-\tau_1+\tau-\beta)\Theta(\tau_2-\tau_1)].\end{aligned} \tag{A.14}$$

Therefore

$$\begin{aligned}\beta\frac{\delta(\Omega-\Omega_0)}{\delta g_{ba}(-\tau)} &= \sum_{a'b'}\int_0^\beta \mathrm{d}\tau_1\mathrm{d}\tau_2 M_{a'b'}(\tau_1-\tau_2)\\ &\quad\times\delta_{b'b}\delta_{a'a}[\delta(\tau_2-\tau_1+\tau)\Theta(\tau_1-\tau_2)\\ &\quad -\delta(\tau_2-\tau_1+\tau-\beta)\Theta(\tau_2-\tau_1)]\\ &= \int_0^\beta \mathrm{d}\tau_1\mathrm{d}\tau_2[M_{ab}(\tau)\delta(\tau_2-\tau_1+\tau)\Theta(\tau_1-\tau_2)\\ &\quad -M_{ab}(\tau-\beta)\delta(\tau_2-\tau_1+\tau-\beta)\Theta(\tau_2-\tau_1)]\\ &= M_{ab}(\tau)\int_0^\beta \mathrm{d}\tau_1\mathrm{d}\tau_2[\delta(\tau_2-\tau_1+\tau)\Theta(\tau_1-\tau_2)\\ &\quad +\delta(\tau_2-\tau_1+\tau-\beta)\Theta(\tau_2-\tau_1)]\\ &= M_{ab}(\tau)\int_0^\beta \mathrm{d}\tau_1\int_{\tau_1}^{\tau_1+\beta}\mathrm{d}\tau_2\delta(\tau_2-\tau_1+\tau-\beta)\\ &= \beta M_{ab}(\tau),\end{aligned} \tag{A.15}$$

which gives Eq. (7.21).

Appendix B

Thermal Expectation Value of the Electron Energy

The one-body energy is straightforward:

$$E_0 = \langle \hat{H}_0 \rangle = \sum_{ab} h_{ab} \langle \hat{c}_a^\dagger \hat{c}_b \rangle = \sum_{ab} h_{ab} \rho_{ab}. \tag{B.1}$$

Expression for the two-body energy term can be derived in multiple ways such as using the equation of motion or the Schwinger-Dyson equation. Here we provide a simple derivation following Ref. [56]. We introduce a coupling constant ξ to the action defined in Eq. (A.4) such that $S_\xi = S_0 + \xi S_V$ and $\xi \to 1$ recovers the "physical" results. Now we have

$$\begin{aligned}
\left.\frac{\mathrm{d}Z_\xi}{\mathrm{d}\xi}\right|_{\xi=1} &= -\int \mathscr{D}[\bar{c}, c] S_V e^{-S_0 - \xi S_V} \Big|_{\xi=1} \\
&= -Z \left\langle \int_0^\beta \mathrm{d}\tau \frac{U_{abcd}}{4} \bar{c}_a(\tau) \bar{c}_c(\tau) c_d(\tau) c_b(\tau) \right\rangle \\
&= -Z\beta \langle \hat{H}_V \rangle.
\end{aligned} \tag{B.2}$$

Introducing a change of variables such that $c \to c/\xi^{1/4}$ and $\bar{c} \to \bar{c}/\xi^{1/4}$, then

$$\begin{aligned}
\mathscr{D}[\bar{c}, c] &= \lim_{\mathscr{N}\to\infty} \prod_{\alpha=1}^{\mathscr{N}} \mathrm{d}\bar{c}_\alpha \mathrm{d}c_\alpha \\
&\to \lim_{\mathscr{N}\to\infty} \xi^{+\mathscr{N}/2} \prod_{\alpha=1}^{\mathscr{N}} \mathrm{d}\bar{c}_\alpha \mathrm{d}c_\alpha \\
&= \lim_{\mathscr{N}\to\infty} \xi^{+\,\mathrm{Tr}[I]/2} \prod_{\alpha=1}^{\mathscr{N}} \mathrm{d}\bar{c}_\alpha \mathrm{d}c_\alpha \\
&= \xi^{+\,\mathrm{Tr}[I]/2} \mathscr{D}[\bar{c}, c]
\end{aligned} \tag{B.3}$$

where the indices α denote states at each discretized time point on the integration path, $\mathrm{Tr}[I] = \int \mathrm{d}x \delta(x, x)$ in which x is the compound spacetime index, and the plus sign on the exponent is due to the nature of Grassmann integrals. The partition function is unaffected by the change of variables, which now takes the form

$$Z_\xi = \xi^{\mathrm{Tr}[I]/2} \int \mathscr{D}[\bar{c}, c] e^{-\xi^{-1/2} S_0 - S_V}. \tag{B.4}$$

Therefore

$$\begin{aligned}
\left.\frac{\mathrm{d}Z_\xi}{\mathrm{d}\xi}\right|_{\xi=1} &= \frac{\mathrm{Tr}[I]}{2} Z + \int \mathscr{D}[\bar{c}, c] \frac{\xi^{-3/2}}{2} S_0 e^{-\xi^{-1/2} S_0 - S_V} \Big|_{\xi=1} \\
&= \frac{\mathrm{Tr}[I]}{2} Z - \frac{1}{2} \int \mathscr{D}[\bar{c}, c] g^{-1}(x', x) \bar{c}(x') c(x) e^{-S} \\
&= \frac{\mathrm{Tr}[I]}{2} Z - \frac{Z}{2} g^{-1}(x', x) G(x, x') \\
&= \frac{Z}{2} \mathrm{Tr}[I - g^{-1} G]
\end{aligned} \tag{B.5}$$

Comparing (B.2) and (B.5), we have

$$\begin{aligned}\langle \hat{H}_V \rangle &= \frac{1}{2\beta} \mathrm{Tr}[g^{-1}G - I] = \frac{1}{2\beta} \mathrm{Tr}[(g^{-1} - G^{-1})G] \\ &= \frac{1}{2\beta} \mathrm{Tr}[\Sigma G] \end{aligned} \tag{B.6}$$

Appendix C

Numerical Computation of the Adjugate Matrix

We calculate the adjugate $\text{adj}(A)$ of a matrix $A \in \mathbb{R}^{n \times n}$ numerically by first performing a rank-revealing factorization on the matrix [153], such as the pivoted QR via the Householder algorithm

$$A = QDRP \tag{C.1}$$

where Q is an orthogonal matrix of Householder reflections, D is a diagonal matrix, R is an upper triangular matrix in which all diagonal elements equal 1, and P is a permutation matrix of the columns. The rank of the matrix $r = \text{rank}(A)$ is determined by the number of nonzero diagonal elements of D.

If A is not singular, i.e. $r = n$, then the adjugate is given by

$$\begin{aligned}\operatorname{adj}(A) &= \det(A)A^{-1} \\ &= \det(P)\det(D)\det(Q)P^T R^{-1} D^{-1} Q^T \\ &= \left[(-1)^{n_P n_Q} \prod_{i=1}^{n} d_i\right] P^T R^{-1} D^{-1} Q^T, \end{aligned} \tag{C.2}$$

where n_P is the number of transpositions in the permutation P, n_Q is the number of Householder reflections in Q, and d_i are diagonal elements of D. The scaling as a function of n for the complexity of calculating the adjugate is the same as the one for calculating A^{-1}, and we obtain $\det(A)$ at the same time.

If A is singular, i.e. $r < n$, $\det(A)$ becomes zero, and Eq. (C.2) is replaced by

$$\operatorname{adj}(A) = \left[(-1)^{n_P n_Q}\right] P^T R^{-1} \operatorname{adj}(D) Q^T. \tag{C.3}$$

If $r = n - 1$, there is one zero in the diagonal of D. Assuming $d_n = 0$ and $d_i \neq 0$ for $i = 1, \ldots, n-1$, the adjugate of D follows directly from the definition (7.23)

$$\operatorname{adj}(D) = \operatorname{diag}\left(\left[0, \ldots, 0, \prod_{i=1}^{n-1} d_i\right]\right). \tag{C.4}$$

If $r < n - 1$, $\operatorname{adj}(D) = 0$, therefore $\operatorname{adj}(A) = 0$.

In the presence of off-diagonal propagators, it is possible that the amputated diagrams $\mathbf{A}(\mathscr{V})$ is nonzero while the vacuum diagrams $D(\mathscr{V}, \emptyset)$ vanish. Therefore it is crucial to implement the adjugate of singular matrices as discussed above.

Appendix D

Explicit Definition of the Auxiliary Green's Function

Introducing

$$\hat{S}_I(\tau,\tau') = \hat{U}_I(\tau)\hat{U}_I^{-1}(\tau'), \tag{D.1}$$

we can write the physical Green's function G in interaction picture as

$$G(\tau,\tau') = \begin{cases} -\frac{1}{Z}\,\mathrm{Tr}[e^{-\beta\hat{H}_0}\hat{S}_I(\beta,\tau)\hat{c}_I(\tau)\hat{S}_I(\tau,\tau')\hat{c}_I^\dagger(\tau')\hat{S}_I(\tau',0)] & \tau > \tau' \\ \frac{1}{Z}\,\mathrm{Tr}[e^{-\beta\hat{H}_0}\hat{S}_I(\beta,\tau')\hat{c}_I^\dagger(\tau')\hat{S}_I(\tau',\tau)\hat{c}_I(\tau)\hat{S}_I(\tau,0)] & \tau < \tau' \end{cases} \tag{D.2}$$

The auxiliary Green's function can be formulated in a similar manner:

$$G_\theta(\tau,\tau') = \begin{cases} -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{c}_I(\tau)\hat{c}_I^\dagger(\tau')\hat{S}_I(\theta,0)], & \tau>\tau'>\theta, \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{c}_I^\dagger(\tau')\hat{c}_I(\tau)\hat{S}_I(\theta,0)], & \tau'>\tau>\theta, \\ -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{c}_I(\tau)\hat{S}_I(\theta,\tau')\hat{c}_I^\dagger(\tau')U(\tau')], & \tau>\theta>\tau', \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{c}_I^\dagger(\tau')\hat{S}_I(\theta,\tau)\hat{c}_I(\tau)U(\tau)], & \tau'>\theta>\tau, \\ -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{S}_I(\theta,\tau)\hat{c}_I(\tau)\hat{S}_I(\tau,\tau')\hat{c}_I^\dagger(\tau')U(\tau')], & \theta>\tau>\tau', \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-\beta\hat{H}_0}\hat{S}_I(\theta,\tau')\hat{c}_I^\dagger(\tau')\hat{S}_I(\tau',\tau)\hat{c}_I(\tau)U(\tau)], & \theta>\tau'>\tau. \end{cases} \tag{D.3}$$

One can verify that this is equivalent to Eq. (8.7) by plugging in the Dyson series for $\hat{S}_I$ [46]:

$$\hat{S}_I(\tau,\tau') = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_{\tau'}^{\tau}\mathrm{d}\tau_1\int_{\tau'}^{\tau}\mathrm{d}\tau_2\cdots\int_{\tau'}^{\tau}\mathrm{d}\tau_k\mathscr{T}_\tau\{\hat{V}_I(\tau_1)\hat{V}_I(\tau_2)\cdots\hat{V}_I(\tau_k)\}. \tag{D.4}$$

Expanding all interaction picture operators explicitly, we have

$$G_\theta(\tau,\tau') = \begin{cases} -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\tau)\hat{H}_0}\hat{c}e^{-(\tau-\tau')\hat{H}_0}\hat{c}^\dagger e^{-(\tau'-\theta)\hat{H}_0}e^{-\theta\hat{H}}], & \tau>\tau'>\theta, \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\tau')\hat{H}_0}\hat{c}^\dagger e^{-(\tau'-\tau)\hat{H}_0}\hat{c}e^{-(\tau-\theta)\hat{H}_0}e^{-\theta\hat{H}}], & \tau'>\tau>\theta, \\ -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\tau)\hat{H}_0}\hat{c}e^{-(\tau-\theta)\hat{H}_0}e^{-(\theta-\tau')\hat{H}}\hat{c}^\dagger e^{-\tau'\hat{H}}], & \tau>\theta>\tau', \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\tau')\hat{H}_0}\hat{c}^\dagger e^{-(\tau'-\theta)\hat{H}_0}e^{-(\theta-\tau)\hat{H}}\hat{c}e^{-\tau\hat{H}}], & \tau'>\theta>\tau, \\ -\frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\theta)\hat{H}_0}e^{-(\theta-\tau)\hat{H}}\hat{c}e^{-(\tau-\tau')\hat{H}}\hat{c}^\dagger e^{-\tau'\hat{H}}], & \theta>\tau>\tau', \\ \frac{1}{Z_\theta}\operatorname{Tr}[e^{-(\beta-\theta)\hat{H}_0}e^{-(\theta-\tau')\hat{H}}\hat{c}^\dagger e^{-(\tau'-\tau)\hat{H}}\hat{c}e^{-\tau\hat{H}}], & \theta>\tau'>\tau, \end{cases} \tag{D.5}$$

which can be used to compute G_θ analytically.

Appendix E

Connection of the Inchworm Expansion to the Skeleton Series

The diagram rules for the inchworm expansion are reminiscent of the skeleton diagram rules [52] due to the exclusion of two-particle reducible Type 1 components. The connection between the inchworm expansion (8.9) and the skeleton series can be revealed in the limit where $\theta' = \beta$, $\theta = \beta - \Delta\theta$, and $\Delta\theta \to 0$. For convenience, rewrite Eq. (8.6) as a coherent state path integral [39]

$$Z_\theta = \int \mathscr{D}[\bar{c}, c] e^{-S_0} \exp\left(-\int_0^\theta \mathrm{d}\tau V(\tau)\right), \tag{E.1}$$

where $\bar{c}(\tau)$ and $c(\tau)$ are Grassmann fields, S_0 the non-interacting action, and $V(\tau)$ the Grassmann function obtained by replacing operators $\hat{c}^\dagger$ and $\hat{c}$ in $\hat{V}$ with the Grassmann fields. The generating function $\mathscr{W}_\theta$ of the auxiliary Green's function is the logarithm of Z_θ with a

bilinear source term J [39]:

$$
\begin{aligned}
Z_\theta[J] &= \int \mathscr{D}[\bar{c}, c] \exp\Big(- S_0 - \int_0^\theta \mathrm{d}\tau V(\tau) + \\
&\quad + \int_0^\beta \mathrm{d}\tau' \mathrm{d}\tau \bar{c}(\tau') J(\tau', \tau) c(\tau) \Big), \\
\mathscr{W}_\theta[J] &:= \log Z_\theta[J], \qquad \frac{\delta \mathscr{W}_\theta}{\delta J(\tau', \tau)}\bigg|_{J=0} = G_\theta(\tau, \tau').
\end{aligned}
\tag{E.2}
$$

When $\Delta\theta \to 0$, we have

$$
G_\theta - G_{\theta - \Delta\theta} \approx \frac{\partial G_\theta}{\partial \theta} \Delta\theta = \frac{\delta}{\delta J} \frac{\partial \mathscr{W}_\theta}{\partial \theta}\bigg|_{J=0} \Delta\theta. \tag{E.3}
$$

If the series expansion of $\partial_\theta \mathscr{W}_\theta$ uniformly converges near $J = 0$, its derivative is also expected to converge. The convergence of this infinitesimal inchworm expansion for G_θ is thus directly related to the convergence properties of $\partial_\theta \mathscr{W}_\theta|_{J=0} = \partial_\theta \log Z_\theta$.

From (E.1), we have

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \log Z_\theta = \frac{1}{Z_\theta} \frac{\partial Z_\theta}{\partial \theta} &= \frac{1}{Z_\theta} \int \mathscr{D}[\bar{c}, c](-V(\theta)) e^{-S_0} \times \\
&\quad \times \exp\Big(- \int_0^\theta \mathrm{d}\tau V(\tau) \Big).
\end{aligned}
\tag{E.4}
$$

When taking $\theta = \beta$, Z_θ becomes Z, and the integral of $V(\tau)$ recovers the interacting action S_V and we have

$$
\frac{\partial}{\partial \theta} \log Z_\theta\bigg|_{\theta=\beta} = \frac{1}{Z} \int \mathscr{D}[\bar{c}, c](-V(\beta)) e^{-S} = -\langle \hat{V} \rangle, \tag{E.5}
$$

where $S = S_0 + S_V$ is the full action of the system. For a standard four-fermion interaction $\hat{V}$, the expectation value can be formulated in terms of the Green's function G and the self-

energy Σ

$$\langle \hat{V} \rangle = \frac{1}{2\beta} \sum_n \mathrm{Tr}[\Sigma(\mathrm{i}\omega_n) G(\mathrm{i}\omega_n)]. \tag{E.6}$$

Σ can be obtained as a functional derivative of the Luttinger-Ward functional $\Phi[G]$ [52]:

$$\frac{\delta \Phi}{\delta G} = \Sigma[G], \tag{E.7}$$

and the skeleton series can be formally written as [52, 56]

$$\begin{aligned} \Sigma[G] &= \sum_{k=1}^{\infty} \Sigma^{(k)}[G], \quad \Phi[G] = \sum_{k=1}^{\infty} \Phi^{(k)}[G], \\ \Phi^{(k)} &= \frac{1}{2k} \mathrm{Tr}[\Sigma^{(k)} G] = \frac{1}{2k} \sum_n \mathrm{Tr}[\Sigma^{(k)}(\mathrm{i}\omega_n) G(\mathrm{i}\omega_n)], \end{aligned} \tag{E.8}$$

where $\Sigma^{(k)}$ is the sum of all k-th order skeleton diagrams. Combining Eqs. (E.5), (E.6), and (E.8), we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \log Z_\theta \Big|_{\theta=\beta} &= -\sum_{k=1}^{\infty} \frac{1}{2\beta} \sum_n \mathrm{Tr}[\Sigma^{(k)}(\mathrm{i}\omega_n) G(\mathrm{i}\omega_n)] \\ &= -\frac{1}{\beta} \sum_{k=1}^{\infty} k \Phi^{(k)}. \end{aligned} \tag{E.9}$$

This directly relates the inchworm expansion at $\theta = \beta$ to the skeleton expansion of the Luttinger-Ward functional. If the skeleton series (E.8) is absolutely convergent, so should Eq. (E.9), which implies a convergent inchworm expansion at $G_{\theta=\beta}$.

Appendix F

High-frequency Tails of the Green's Function

We start deriving the high-frequency tails for Green's functions with the most general type of interaction from (2.24):

$$\hat{H} = \sum_{ij} h_{ij} \hat{c}_i^\dagger \hat{c}_j + \frac{1}{4} \sum_{ijkl} U_{ijkl} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \tag{F.1}$$

where $i, j, \ldots$ are compound spin-orbital indices, h_{ij} is the non-interacting matrix elements (including the chemical potential contribution $-\mu\delta_{ij}$), and U_{ijkl} is the fully antisymmetrized interaction (2.23):

$$U_{ijkl} = V_{ijkl} - V_{ijlk} = -U_{ijkl} = -U_{jikl}. \tag{F.2}$$

The high frequency expansion of the Green's function $G_{ij}(\tau) = -\langle \mathscr{T}_\tau \hat{c}_i(\tau) \hat{c}_j^\dagger(0) \rangle$ is given by

$$G_{ij}(\mathrm{i}\omega_n) = \sum_{k=1}^{\infty} (-1)^k \frac{G_{ij}^{(k-1)}(\tau = 0^+) - G_{ij}^{(k-1)}(\tau = 0^-)}{(\mathrm{i}\omega_n)^k} = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\left\langle \{ [\hat{H}, \hat{c}_i]_{(k-1)}, \} \right\rangle}{(\mathrm{i}\omega_n)^k} \tag{F.3}$$

We focus on the first three moments of the Green's function.

Now we derive the first few moments of the Green's function. Here are some useful properties of the (anti-)commutators that will be used in the derivation.

$$\begin{aligned}
\{AB, C\} &= A[B, C] + \{A, C\}B \\
\{A, BC\} &= \{A, B\}C - B[A, C] \\
\{AB, C\} &= A\{B, C\} - [A, C]B \\
[AB, C] &= A\{B, C\} - \{A, C\}B
\end{aligned} \tag{F.4}$$

The derivations are given as follows. We use Einstein's summation rule to simplify notations.

- $G^{\mathrm{I}}_{ij} = \left\langle \left\{ c_i, \hat{c}_j^\dagger \right\} \right\rangle = \delta_{ij}$
- $G^{\mathrm{II}}_{ij} = -\left\langle \left\{ \left[\hat{H}, \hat{c}_i \right], \hat{c}_j^\dagger \right\} \right\rangle$, in which

$$\begin{aligned}
[\hat{H}, \hat{c}_i] =& h_{pq} \left[\hat{c}_p^\dagger \hat{c}_q, \hat{c}_i \right] + \frac{1}{4} U_{pqmn} \left[\hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_m, \hat{c}_i \right] \\
=& h_{pq} \left(\hat{c}_p^\dagger \cancelto{0}{\left\{ \hat{c}_q, \hat{c}_i \right\}} - \left\{ \hat{c}_p^\dagger, \hat{c}_i \right\} \hat{c}_q \right) + \frac{1}{4} U_{pqmn} \left(\hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_n \cancelto{0}{\left\{ \hat{c}_m, \hat{c}_i \right\}} - \left\{ \hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_n, \hat{c}_i \right\} \hat{c}_m \right) \\
=& - h_{pq} \delta_{ip} \hat{c}_q - \frac{1}{4} U_{pqmn} \left(\hat{c}_p^\dagger \hat{c}_q^\dagger \cancelto{0}{\left\{ \hat{c}_n, \hat{c}_i \right\}} - \left[\hat{c}_p^\dagger \hat{c}_q^\dagger, \hat{c}_i \right] \hat{c}_n \right) \hat{c}_m \\
=& - h_{iq} \hat{c}_q + \frac{1}{4} U_{pqmn} \left(\hat{c}_p^\dagger \left\{ \hat{c}_q^\dagger, \hat{c}_i \right\} - \left\{ \hat{c}_p^\dagger, \hat{c}_i \right\} \hat{c}_q^\dagger \right) \hat{c}_n \hat{c}_m \\
=& - h_{iq} \hat{c}_q + \frac{1}{4} U_{pqmn} \hat{c}_p^\dagger \hat{c}_n \hat{c}_m \delta_{qi} - \frac{1}{4} U_{pqmn} \hat{c}_q^\dagger \hat{c}_n \hat{c}_m \delta_{pi} \\
=& - h_{iq} \hat{c}_q + \frac{1}{4} U_{pimn} \hat{c}_p^\dagger \hat{c}_n \hat{c}_m - \frac{1}{4} U_{iqmn} \hat{c}_q^\dagger \hat{c}_n \hat{c}_m
\end{aligned} \tag{F.5}$$

Respect the symmetry of the interaction matrix (F.2) and renaming dummy indices, we have

$$[\hat{H}, \hat{c}_i] = -h_{iq} \hat{c}_q - \frac{1}{2} U_{iqmn} \hat{c}_q^\dagger \hat{c}_n \hat{c}_m \tag{F.6}$$

Therefore

$$
\begin{aligned}
G_{ij}^{\mathrm{II}} &= -\left\langle\left\{\left[\hat{H},\hat{c}_i\right],\hat{c}_j^\dagger\right\}\right\rangle = h_{iq}\left\langle\left\{\hat{c}_q,\hat{c}_j^\dagger\right\}\right\rangle + \frac{1}{2}U_{iqmn}\left\langle\left\{\hat{c}_q^\dagger\hat{c}_n\hat{c}_m,\hat{c}_j^\dagger\right\}\right\rangle \\
&= h_{iq}\delta_{qj} + \frac{1}{2}U_{iqmn}\left\langle\hat{c}_q^\dagger\left[\hat{c}_n\hat{c}_m,\hat{c}_j^\dagger\right] - \cancelto{0}{\left\{\hat{c}_q^\dagger,\hat{c}_j^\dagger\right\}}\hat{c}_m\hat{c}_n\right\rangle \\
&= h_{ij} + \frac{1}{2}U_{iqmn}\left\langle\hat{c}_q^\dagger\hat{c}_n\left\{\hat{c}_m,\hat{c}_j^\dagger\right\} - \hat{c}_q^\dagger\left\{\hat{c}_m,\hat{c}_j^\dagger\right\}\hat{c}_n\right\rangle \\
&= h_{ij} + \frac{1}{2}U_{iqmn}\left(\rho_{qn}\delta_{mj} - \rho_{qm}\delta_{nj}\right) \\
&= h_{ij} + \frac{1}{2}\left(U_{iqjn}\rho_{qn} - U_{iqmj}\rho_{qm}\right) \\
&= h_{ij} + U_{iqjn}\rho_{qn}
\end{aligned}
\tag{F.7}
$$

where $\rho_{ij} = \left\langle\hat{c}_i^\dagger\hat{c}_j\right\rangle$ is the density matrix.

- $G_{ij}^{\mathrm{III}} = \left\langle\left\{\left[\hat{H},\left[\hat{H},\hat{c}_i\right]\right],\hat{c}_j^\dagger\right\}\right\rangle$. From (F.6) we have

$$
\left[\hat{H},\left[\hat{H},\hat{c}_i\right]\right] = -h_{iq}\left[\hat{H},\hat{c}_q\right] - \frac{1}{2}U_{iqmn}\left[\hat{H},\hat{c}_q^\dagger\hat{c}_n\hat{c}_m\right] \tag{F.8}
$$

The first term goes as

$$
\begin{aligned}
-h_{iq}[\hat{H},\hat{c}_q] &= -h_{iq}\left(-h_{qr}\hat{c}_r - \frac{1}{2}U_{qrst}\hat{c}_r^\dagger\hat{c}_t\hat{c}_s\right) \\
&= h_{iq}h_{qr}\hat{c}_r + \frac{1}{2}h_{iq}U_{qrst}\hat{c}_r^\dagger\hat{c}_t\hat{c}_s.
\end{aligned}
\tag{F.9}
$$

From the previous calculation we have $\left\langle\left\{\hat{c}_q^\dagger\hat{c}_m\hat{c}_n,\hat{c}_j^\dagger\right\}\right\rangle = \rho_{qm}\delta_{nj} - \rho_{qn}\delta_{mj}$, therefore

$$
\left\langle\left\{-h_{iq}[\hat{H},\hat{c}_q],\hat{c}_j^\dagger\right\}\right\rangle = h_{iq}h_{qr}\delta_{rj} + \frac{1}{2}h_{iq}U_{qrst}\left(\rho_{rt}\delta_{sj} - \rho_{rs}\delta_{tj}\right) = \left(h^2\right)_{ij} + h_{iq}U_{qrjs}\rho_{rs}. \tag{F.10}
$$

To evaluate the second term, we first investigate the following commutation relation

$$
\begin{aligned}
\left[\hat{H}, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right] &= h_{ij}\left[\hat{c}_i^\dagger \hat{c}_j, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right] + \frac{1}{4} U_{ijkl}\left[\hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right] \\
&= h_{ij}\left(\left[\hat{c}_i^\dagger, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right] \hat{c}_j + \hat{c}_i^\dagger \left[\hat{c}_j, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right]\right) \\
&\quad + \frac{1}{4} U_{ijkl}\left(\left[\hat{c}_i^\dagger \hat{c}_j^\dagger, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right] \hat{c}_l \hat{c}_k + \hat{c}_i^\dagger \hat{c}_j^\dagger \left[\hat{c}_l \hat{c}_k, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right]\right) \\
&= h_{ij}\left(\cancelto{0}{\left\{\hat{c}_i^\dagger, \hat{c}_q^\dagger\right\}} \hat{c}_n \hat{c}_m \hat{c}_j - \hat{c}_q^\dagger \left\{\hat{c}_i^\dagger, \hat{c}_n \hat{c}_m\right\} \hat{c}_j + \hat{c}_i^\dagger \left\{\hat{c}_j, \hat{c}_q^\dagger\right\} \hat{c}_n \hat{c}_m - \hat{c}_i^\dagger \hat{c}_q^\dagger \left\{\hat{c}_j, \hat{c}_n \hat{c}_m\right\}\right) \\
&\quad + \frac{1}{4} U_{ijkl}\left(\hat{c}_i^\dagger \left\{\hat{c}_j^\dagger, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right\} \hat{c}_l \hat{c}_k - \left\{\hat{c}_i^\dagger, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right\} \hat{c}_j^\dagger \hat{c}_l \hat{c}_k + \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \left\{\hat{c}_k, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right\} - \hat{c}_i^\dagger \hat{c}_j^\dagger \left\{\hat{c}_l, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m\right\} \hat{c}_k\right) \\
&= h_{ij}\left(-\hat{c}_q^\dagger \left\{\hat{c}_i^\dagger, \hat{c}_n\right\} \hat{c}_m \hat{c}_j + \hat{c}_q^\dagger \hat{c}_n \left[\hat{c}_i^\dagger, \hat{c}_m\right] \hat{c}_j + \hat{c}_i^\dagger \hat{c}_n \hat{c}_m \delta_{jq} - \hat{c}_i^\dagger \hat{c}_q^\dagger \cancelto{0}{\left\{\hat{c}_j, \hat{c}_n\right\}} \hat{c}_m + \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \left[\hat{c}_j, \hat{c}_m\right]\right) \\
&\quad + \frac{1}{4} U_{ijkl}\Big(\hat{c}_i^\dagger \cancelto{0}{\left\{\hat{c}_j^\dagger, \hat{c}_q^\dagger\right\}} \hat{c}_n \hat{c}_m \hat{c}_l \hat{c}_k - \hat{c}_i^\dagger \hat{c}_q^\dagger \left[\hat{c}_j^\dagger, \hat{c}_n \hat{c}_m\right] \hat{c}_l \hat{c}_k - \cancelto{0}{\left\{\hat{c}_i^\dagger, \hat{c}_q^\dagger\right\}} \hat{c}_j^\dagger \hat{c}_n \hat{c}_m \hat{c}_l \hat{c}_k + \hat{c}_q^\dagger \left[\hat{c}_i^\dagger, \hat{c}_n \hat{c}_m\right] \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \\
&\quad + \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \left\{\hat{c}_k, \hat{c}_q^\dagger\right\} \hat{c}_n \hat{c}_m - \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \cancelto{0}{\left[\hat{c}_k, \hat{c}_n \hat{c}_m\right]} - \hat{c}_i^\dagger \hat{c}_j^\dagger \left\{\hat{c}_l, \hat{c}_q^\dagger\right\} \hat{c}_n \hat{c}_m \hat{c}_k + \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_q^\dagger \cancelto{0}{\left[\hat{c}_l, \hat{c}_n \hat{c}_m\right]} \hat{c}_k\Big) \\
&= h_{ij}\left(-\hat{c}_q^\dagger \hat{c}_m \hat{c}_j \delta_{in} + \hat{c}_q^\dagger \hat{c}_n \left(-\delta_{im} + 2\hat{c}_i^\dagger \hat{c}_m\right) \hat{c}_j + \hat{c}_i^\dagger \hat{c}_n \hat{c}_m \delta_{jq} + \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \left(2\hat{c}_j \hat{c}_m\right)\right) \\
&\quad + \frac{1}{4} U_{ijkl}\Big(-\hat{c}_i^\dagger \hat{c}_q^\dagger \left(-\hat{c}_n \left\{\hat{c}_j^\dagger, \hat{c}_m\right\} + \left\{\hat{c}_j^\dagger, \hat{c}_n\right\} \hat{c}_m\right) \hat{c}_l \hat{c}_k + \hat{c}_q^\dagger \left(-\hat{c}_n \left\{\hat{c}_i^\dagger, \hat{c}_m\right\} + \left\{\hat{c}_i^\dagger, \hat{c}_n\right\} \hat{c}_m\right) \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \\
&\quad + \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_n \hat{c}_m \delta_{kq} - \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_n \hat{c}_m \cancelto{\left(-\hat{c}_l \delta_{kq}\right)}{\hat{c}_k \delta_{lq}}\Big) \\
&= h_{ij}\Big(-\hat{c}_q^\dagger \hat{c}_m \hat{c}_j \delta_{in} - \hat{c}_q^\dagger \hat{c}_n \hat{c}_j \delta_{im} + \hat{c}_i^\dagger \hat{c}_n \hat{c}_m \delta_{jq} + 2\cancelto{\hat{c}_q^\dagger \hat{c}_m \hat{c}_j \delta_{in}}{\left(\hat{c}_q^\dagger \hat{c}_n \hat{c}_i^\dagger \hat{c}_m \hat{c}_j + \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_j \hat{c}_m\right)}\Big) \\
&\quad + \frac{1}{4} U_{ijkl}\left(\hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_l \hat{c}_k \delta_{jm} - \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_m \hat{c}_l \hat{c}_k \delta_{jn} - \hat{c}_q^\dagger \hat{c}_n \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \delta_{im} + \hat{c}_q^\dagger \hat{c}_m \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \delta_{in} + 2\hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_n \hat{c}_m \delta_{kq}\right) \\
&= h_{ij}\left(\hat{c}_q^\dagger \hat{c}_m \hat{c}_j \delta_{in} - \hat{c}_q^\dagger \hat{c}_n \hat{c}_j \delta_{im} + \hat{c}_i^\dagger \hat{c}_n \hat{c}_m \delta_{jq}\right) \\
&\quad + \frac{1}{4} U_{ijkl}\Big(\hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_l \hat{c}_k \delta_{jm} - \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_m \hat{c}_l \hat{c}_k \delta_{jn} + \cancelto{\hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_l \hat{c}_k \delta_{jm}}{\hat{c}_q^\dagger \hat{c}_j^\dagger \hat{c}_n \hat{c}_l \hat{c}_k \delta_{im}} \\
&\quad - \cancelto{\hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_m \hat{c}_l \hat{c}_k \delta_{jn}}{\hat{c}_q^\dagger \hat{c}_j^\dagger \hat{c}_m \hat{c}_l \hat{c}_k \delta_{in}} - \hat{c}_q^\dagger \hat{c}_l \hat{c}_k \delta_{nj} \delta_{im} + \cancelto{-\hat{c}_q^\dagger \hat{c}_l \hat{c}_k \delta_{im} \delta_{jn}}{\hat{c}_q^\dagger \hat{c}_l \hat{c}_k \delta_{jm} \delta_{in}} + 2\hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_n \hat{c}_m \delta_{kq}\Big) \\
&= h_{ij}\left(\hat{c}_q^\dagger \hat{c}_m \hat{c}_j \delta_{in} - \hat{c}_q^\dagger \hat{c}_n \hat{c}_j \delta_{im} + \hat{c}_i^\dagger \hat{c}_n \hat{c}_m \delta_{jq}\right) \\
&\quad + \frac{1}{2} U_{ijkl}\left(\hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_n \hat{c}_l \hat{c}_k \delta_{jm} - \hat{c}_i^\dagger \hat{c}_q^\dagger \hat{c}_m \hat{c}_l \hat{c}_k \delta_{jn} - \hat{c}_q^\dagger \hat{c}_l \hat{c}_k \delta_{nj} \delta_{im} + \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_n \hat{c}_m \delta_{kq}\right)
\end{aligned}
\tag{F.11}
$$

To avoid confusion, we change the dummy indices in the result above from $\{i,j,k,l\}$ to $\{r,s,t,v\}$, which gives

$$\begin{aligned}[\hat{H},\hat{c}_q^\dagger\hat{c}_m\hat{c}_n] =& h_{rs}(\hat{c}_q^\dagger\hat{c}_m\hat{c}_s\delta_{rn}-\hat{c}_q^\dagger\hat{c}_n\hat{c}_s\delta_{rm}+\hat{c}_r^\dagger\hat{c}_n\hat{c}_m\delta_{sq})\\ &+\frac{U_{rstv}}{2}(\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_n\hat{c}_v\hat{c}_t\delta_{sm}-\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_m\hat{c}_v\hat{c}_t\delta_{sn}+\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v\hat{c}_n\hat{c}_m\delta_{tq}-\hat{c}_q^\dagger\hat{c}_v\hat{c}_t\delta_{ns}\delta_{rm})\end{aligned} \tag{F.12}$$

Plugging this into the second term of (F.8), we have

$$\begin{aligned}-\frac{1}{2}U_{iqmn}\left[\hat{H},\hat{c}_q^\dagger\hat{c}_n\hat{c}_m\right] =& -\frac{1}{2}U_{iqmn}h_{rs}(\cancel{\hat{c}_q^\dagger\hat{c}_m\hat{c}_s\delta_{rn}}^{-\hat{c}_q^\dagger\hat{c}_n\hat{c}_s\delta_{rm}}-\hat{c}_q^\dagger\hat{c}_n\hat{c}_s\delta_{rm}+\hat{c}_r^\dagger\hat{c}_n\hat{c}_m\delta_{sq})\\ &-\frac{U_{iqmn}U_{rstv}}{4}(\cancel{\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_n\hat{c}_v\hat{c}_t\delta_{sm}}^{-\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_m\hat{c}_v\hat{c}_t\delta_{sn}}-\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_m\hat{c}_v\hat{c}_t\delta_{sn}+\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v\hat{c}_n\hat{c}_m\delta_{tq}-\hat{c}_q^\dagger\hat{c}_v\hat{c}_t\delta_{ns}\delta_{rm})\\ =& \frac{U_{iqmn}h_{rs}}{2}\left(2\hat{c}_q^\dagger\hat{c}_n\hat{c}_s\delta_{rm}-\hat{c}_r^\dagger\hat{c}_n\hat{c}_m\delta_{sq}\right)\\ &+\frac{U_{iqmn}U_{rstv}}{4}\left(2\hat{c}_r^\dagger\hat{c}_q^\dagger\hat{c}_m\hat{c}_v\hat{c}_t\delta_{sn}-\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v\hat{c}_n\hat{c}_m\delta_{tq}+\hat{c}_q^\dagger\hat{c}_v\hat{c}_t\delta_{ns}\delta_{rm}\right)\end{aligned} \tag{F.13}$$

We then analyze the following anti-commutation relation

$$\begin{aligned}\{\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v\hat{c}_n\hat{c}_m,\hat{c}_j^\dagger\} =& \hat{c}_r^\dagger\hat{c}_s^\dagger\{\hat{c}_v\hat{c}_n\hat{c}_m,\hat{c}_j^\dagger\}-\cancel{[\hat{c}_r^\dagger\hat{c}_s^\dagger,\hat{c}_j^\dagger]}^{0}\hat{c}_v\hat{c}_n\hat{c}_m\\ =& \hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v[\hat{c}_n\hat{c}_m,\hat{c}_j^\dagger]+\hat{c}_r^\dagger\hat{c}_s^\dagger\{\hat{c}_v,\hat{c}_j^\dagger\}\hat{c}_n\hat{c}_m\\ =& \hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v(\hat{c}_n\{\hat{c}_m,\hat{c}_j^\dagger\}-\{\hat{c}_n,\hat{c}_j^\dagger\}\hat{c}_m)+\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_n\hat{c}_m\delta_{jv}\\ =& \hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_v(\hat{c}_n\delta_{mj}-\hat{c}_m\delta_{nj})+\hat{c}_r^\dagger\hat{c}_s^\dagger\hat{c}_n\hat{c}_m\delta_{jv}\end{aligned} \tag{F.14}$$

Using this together with that $\left\langle\left\{\hat{c}_q^\dagger\hat{c}_m\hat{c}_n,\hat{c}_j^\dagger\right\}\right\rangle = \rho_{qm}\delta_{nj}-\rho_{qn}\delta_{mj}$, we have (define χ_{ijkl} =

$\langle \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k \rangle)$

$$
\begin{aligned}
&\left\langle \left\{ -\frac{1}{2} U_{iqmn} \left[\hat{H}, \hat{c}_q^\dagger \hat{c}_n \hat{c}_m \right], \hat{c}_j^\dagger \right\} \right\rangle \\
&= \frac{U_{iqmn} h_{rs}}{2} \Bigg(2 \left(\rho_{qn} \delta_{sj} - \rho_{qs} \delta_{nj} \right) \delta_{rm} - \cancelto{2\rho_{rm}\delta_{nj}}{\left(\rho_{rm} \delta_{nj} - \rho_{rn} \delta_{mj} \right) \delta_{sq}} \Bigg) \\
&\quad + \frac{U_{iqmn} U_{rstv}}{4} \Bigg(\Bigg\langle 2 \Bigg(\hat{c}_r^\dagger \hat{c}_q^\dagger \hat{c}_m \cancelto{2\hat{c}_v \delta_{tj}}{\left(\hat{c}_v \delta_{tj} - \hat{c}_t \delta_{vj} \right)} + \hat{c}_r^\dagger \hat{c}_q^\dagger \hat{c}_v \hat{c}_t \delta_{jm} \Bigg) \delta_{sn} \\
&\quad - \Bigg(\hat{c}_r^\dagger \hat{c}_s^\dagger \hat{c}_v \cancelto{-2\hat{c}_m \delta_{nj}}{\left(\hat{c}_n \delta_{mj} - \hat{c}_m \delta_{nj} \right)} + \hat{c}_r^\dagger \hat{c}_s^\dagger \hat{c}_n \hat{c}_m \delta_{jv} \Bigg) \delta_{tq} \Bigg\rangle \\
&\quad + \cancelto{2\rho_{qv}\delta_{tj}}{\left(\rho_{qv} \delta_{tj} - \rho_{qt} \delta_{vj} \right)} \delta_{ns} \delta_{rm} \Bigg) \\
&= U_{iqmn} h_{rs} (\rho_{qn} \delta_{sj} \delta_{rm} - \rho_{qs} \delta_{nj} \delta_{rm} - \rho_{rm} \delta_{nj} \delta_{sq}) \\
&\quad + \frac{U_{iqmn} U_{rstv}}{4} \Bigg(\Bigg\langle 4 \hat{c}_r^\dagger \hat{c}_q^\dagger \hat{c}_m \hat{c}_v \delta_{tj} \delta_{sn} + 2 \hat{c}_r^\dagger \hat{c}_q^\dagger \hat{c}_v \hat{c}_t \delta_{jm} \delta_{sn} \\
&\quad + 2 \hat{c}_r^\dagger \hat{c}_s^\dagger \hat{c}_v \hat{c}_m \delta_{nj} \delta_{tq} - \hat{c}_r^\dagger \hat{c}_s^\dagger \hat{c}_n \hat{c}_m \delta_{jv} \delta_{tq} \Bigg\rangle + 2 \rho_{qv} \delta_{tj} \delta_{ns} \delta_{rm} \Bigg) \\
&= U_{iqmn} h_{mj} \rho_{qn} - U_{iqmj} h_{ms} \rho_{qs} + U_{iqmj} h_{rq} \rho_{rm} + \frac{1}{2} U_{iqmn} U_{mnjv} \rho_{qv} \\
&\quad + U_{iqmn} U_{rnjv} \chi_{rqvm} + \frac{1}{2} U_{iqjn} U_{rntv} \chi_{rqtv} + \frac{1}{2} U_{iqmj} U_{rsqv} \chi_{rsmv} - \frac{1}{4} U_{iqmn} U_{rsqj} \chi_{rsmn}
\end{aligned}
\tag{F.15}
$$

To sum up, we have

$$
\begin{aligned}
G_{ij}^{\mathrm{III}} &= \left(h^2 \right)_{ij} + h_{iq} U_{qrjs} \rho_{rs} + U_{iqmn} h_{mj} \rho_{qn} - U_{iqmj} h_{ms} \rho_{qs} + U_{iqmj} h_{rq} \rho_{rm} + \frac{1}{2} U_{iqmn} U_{mnjv} \rho_{qv} \\
&\quad + U_{iqmn} U_{rnjv} \chi_{rqvm} + \frac{1}{2} U_{iqjn} U_{rntv} \chi_{rqtv} + \frac{1}{2} U_{iqmj} U_{rsqv} \chi_{rsmv} - \frac{1}{4} U_{iqmn} U_{rsqj} \chi_{rsmn}
\end{aligned}
\tag{F.16}
$$

Recall that $G^{\mathrm{II}}_{ij} = h_{ij} + U_{iqjn}\rho_{qn}$, which gives $(G^{\mathrm{II}})^2_{ij} = (h^2)_{ij} + h_{iq}U_{qmjn}\rho_{mn} + U_{iqmn}h_{mj}\rho_{qn} + U_{iqmn}U_{msjv}\rho_{qn}\rho_{sv}$. Collecting identical terms in G^{III}

$$
\begin{aligned}
G^{\mathrm{III}}_{ij} =& (G^{\mathrm{II}})^2_{ij} - \cancel{U_{iqmn}U_{msjv}\rho_{qn}\rho_{sv}} + U_{iqmn}U_{rnjv}\rho_{rv}\rho_{qm} - U_{iqmj}h_{ms}\rho_{qs} + U_{iqmj}h_{rq}\rho_{rm} + \frac{1}{2}U_{iqmn}U_{mnjv}\rho_{qv} \\
&+ U_{iqmn}U_{rnjv}\chi_{rqvm} + \frac{1}{2}U_{iqjn}U_{rntv}\chi_{rqtv} + \frac{1}{2}U_{iqmj}U_{rsqv}\chi_{rsmv} - \frac{1}{4}U_{iqmn}U_{rsqj}\chi_{rsmn} \\
=& (G^{\mathrm{II}})^2_{ij} \\
&+ U_{iqmn}U_{rnjv}(\chi_{rqvm} - \rho_{rv}\rho_{qm}) \\
&- U_{iqmj}h_{ms}\rho_{qs} + U_{iqmj}h_{rq}\rho_{rm} \\
&+ \frac{1}{2}U_{iqmn}U_{mnjv}\rho_{qv} + \frac{1}{2}U_{iqjn}U_{rntv}\chi_{rqtv} + \frac{1}{2}U_{iqmj}U_{rsqv}\chi_{rsmv} \\
&- \frac{1}{4}U_{iqmn}U_{rsqj}\chi_{rsmn}
\end{aligned}
\tag{F.17}
$$

If we assume that all elements in h_{ij} and U_{ijkl} are real, and so is ρ_{ij} and χ_{ijkl}, then since the Hamiltonian is Hermitian, we have the following additional symmetry:

$$
\begin{aligned}
h_{ij} =& h_{ji} \\
U_{ijkl} =& U_{klij}
\end{aligned}
\tag{F.18}
$$

ρ_{ij} (G_{ij}) and χ_{ijkl} follows the same symmetry as h_{ij} and U_{ijkl} respectively. We can make

use of these properties to simplify the result above.

$$
\begin{aligned}
G^{\mathrm{III}}_{ij} =& (G^{\mathrm{II}})^2_{ij} \\
&+ U_{iqnm}U_{rnvj}(\chi_{rqvm} - \rho_{rv}\rho_{qm}) \\
&+ \frac{1}{2}U_{iqmn}U_{mnjv}\rho_{qv} \\
&+ \underbrace{\frac{1}{2}(U_{iqjn} - U_{injq})U_{rntv}\chi_{rqtv}}_{\text{zero under symmetry } i\leftrightarrow j} \\
&- \frac{1}{4}U_{iqmn}U_{rsqj}\chi_{rsmn} \\
=& (G^{\mathrm{II}})^2_{ij} \\
&+ U_{iqnm}U_{rnvj}(\chi_{rqvm} - \rho_{rv}\rho_{qm}) \\
&- \frac{1}{4}U_{iqmn}U_{rsqj}\chi_{rsmn} \\
&+ \frac{1}{2}U_{iqmn}U_{mnjv}\rho_{qv}
\end{aligned}
\tag{F.19}
$$

Now we treat the orbital and spin indices separately. To simplify notations, we introduce the following convention: suppose we have a compound index $i = \{o_i, \sigma_i\}$ where o_i is the orbital index and σ_i is the spin index, we denote the corresponding vector quantity to be $A_i = A^i_i$, in which the subscript is always the orbital index, and the superscript is always the spin index. Einstein's summation rule is applied to all orbital indices, namely the subscript. Spin summations should always be written with the summation symbol explicitly, and the indices under the summation symbol only refer to the spin. We further assume that the Hamiltonian takes the following spin-preserving form:

$$
H = \sum_s h^s_{ij}\hat{c}^{s\dagger}_i\hat{c}^s_j + \frac{1}{2}\sum_{ij} V^{ij}_{ijkl}\hat{c}^{i\dagger}_i\hat{c}^{j\dagger}_j\hat{c}^j_l\hat{c}^i_k \tag{F.20}
$$

which relates to the previous spin-orbital notations as follows (Einstein's summation rule is NOT implied here):

$$\begin{aligned} h_{ij}^{ij} &= h_{ij}^{i}\delta^{ij} \\ V_{ijkl}^{ijkl} &= V_{ijkl}^{ij}\delta^{ik}\delta^{jl} \\ U_{ijkl}^{ijkl} &= V_{ijkl}^{ijkl} - V_{ijlk}^{ijlk} = V_{ijkl}^{ij}\delta^{ik}\delta^{jl} - V_{ijlk}^{ij}\delta^{il}\delta^{jk} \end{aligned} \tag{F.21}$$

Since the Hamiltonian preserves spin, the observables should also reflect such preservation (Einstein's summation rule is NOT implied here):

$$\begin{aligned} G_{ij}^{ij} &= G_{ij}^{i}\delta^{ij} \\ \rho_{ij}^{ij} &= \rho_{ij}^{i}\delta^{ij} \\ \chi_{ijkl}^{ijkl} &= \chi_{ijkl}^{ij}\delta^{ik}\delta^{jl} - \chi_{ijlk}^{ij}\delta^{il}\delta^{jk}(1-\delta^{ij}) \end{aligned} \tag{F.22}$$

where $\chi_{ijkl}^{ij} = \langle \hat{c}_i^{i\dagger}\hat{c}_j^{j\dagger}\hat{c}_l^{j}\hat{c}_k^{i}\rangle$. Now we evaluate the tails under this convention.

- The first tail is simply $(G^{\mathrm{I}})_{ij}^{s} = \delta_{ij}$
- The second tails turns into

$$\begin{aligned} (G^{\mathrm{II}})_{ij}^{s} &= h_{ij}^{s} + \sum_{iqjn} \delta^{is}\delta^{js}(V_{iqjn}^{iq}\delta^{ij}\delta^{qn} - V_{iqnj}^{in}\delta^{in}\delta^{qj})\rho_{qn}^{q}\delta^{qn} \\ &= h_{ij}^{s} + \left(\sum_{s'} V_{iqjn}^{ss'}\rho_{qn}^{s'} - V_{iqnj}^{ss}\rho_{qn}^{s}\right) \end{aligned} \tag{F.23}$$

  from which we recognize the familiar Hartree and Fock self energies.

- The third tail goes as (here $\bar{\sigma}$ denotes the opposite spin of σ)

$$
\begin{aligned}
(G^{\mathrm{III}})^i_{ij} =&(G^{\mathrm{II}})^i_{ik}(G^{\mathrm{II}})^i_{kj} \\
&+\sum_{qmnrv}(V^{iq}_{iqmn}\delta^{im}\delta^{qn}-V^{iq}_{iqnm}\delta^{in}\delta^{qm})(V^{rn}_{rnjv}\delta^{ri}\delta^{nv}-V^{rn}_{rnvj}\delta^{rv}\delta^{ni}) \\
&((\chi^{rq}_{rqvm}\delta^{rv}\delta^{qm}-\chi^{rq}_{rqmv}\delta^{rm}\delta^{qv}(1-\delta^{rq}))-\rho^{r}_{rv}\delta^{rv}\rho^{q}_{qm}\delta^{qm}) \\
&+\frac{1}{2}\sum_{qmnv}(V^{iq}_{iqmn}\delta^{im}\delta^{qn}-V^{iq}_{iqnm}\delta^{in}\delta^{qm})(V^{mn}_{mnjv}\delta^{mi}\delta^{nv}-V^{mn}_{mnvj}\delta^{mv}\delta^{ni})\rho^{q}_{qv}\delta^{qv} \\
&-\frac{1}{4}\sum_{qmnrs}(V^{iq}_{iqmn}\delta^{im}\delta^{qn}-V^{iq}_{iqnm}\delta^{in}\delta^{qm})(V^{rs}_{rsqj}\delta^{rq}\delta^{si}-V^{rs}_{rsjq}\delta^{ri}\delta^{sq}) \\
&(\chi^{rs}_{rsmn}\delta^{rm}\delta^{sn}-\chi^{rs}_{rsnm}\delta^{rn}\delta^{sm}(1-\delta^{rs})) \\
\\
&=(G^{\mathrm{II}})^i_{ik}(G^{\mathrm{II}})^i_{kj} \\
&+\big(V^{ii}_{iqmn}V^{ii}_{rnjv}(\chi^{ii}_{rqvm}-\rho^{i}_{rv}\rho^{i}_{qm})-V^{i\bar{i}}_{iqmn}V^{i\bar{i}}_{rnjv}\chi^{i\bar{i}}_{rqmv} \\
&\quad-\sum_{r}V^{ii}_{iqmn}V^{ri}_{rnvj}(\chi^{ri}_{rqvm}-\rho^{r}_{rv}\rho^{i}_{qm})-\sum_{q}V^{iq}_{iqnm}V^{ii}_{rnjv}(\chi^{iq}_{rqvm}-\rho^{i}_{rv}\rho^{q}_{qm}) \\
&\quad+\sum_{qr}V^{iq}_{iqnm}V^{ri}_{rnvj}(\chi^{rq}_{rqvm}-\rho^{r}_{rv}\rho^{q}_{qm})\big) \\
&+\frac{1}{2}\Big(\sum_{q}V^{iq}_{iqmn}V^{iq}_{mnjv}\rho^{q}_{qv}-V^{ii}_{iqmn}V^{ii}_{mnvj}\rho^{i}_{qv}-V^{ii}_{iqnm}V^{ii}_{mnjv}\rho^{i}_{qv}+\sum_{q}V^{iq}_{iqnm}V^{qi}_{mnvj}\rho^{q}_{qv}\Big) \\
&-\frac{1}{4}\big(V^{ii}_{iqmn}V^{ii}_{rsqj}\chi^{ii}_{rsmn}-V^{i\bar{i}}_{iqmn}V^{\bar{i}i}_{rsqj}\chi^{\bar{i}i}_{rsnm}-\sum_{q}V^{iq}_{iqmn}V^{iq}_{rsjq}\chi^{iq}_{rsmn} \\
&\quad-\sum_{q}V^{iq}_{iqnm}V^{qi}_{rsqj}\chi^{qi}_{rsmn}+V^{ii}_{iqnm}V^{ii}_{rsjq}\chi^{ii}_{rsmn}-V^{i\bar{i}}_{iqnm}V^{i\bar{i}}_{rsjq}\chi^{i\bar{i}}_{rsnm}\big)
\end{aligned}
\tag{F.24}
$$

In the case of Coulomb interactions, the interaction is spin-independent, i.e.

$$V_{ijkl}^{ij} = V_{ijkl} = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\phi_l(\mathbf{r}_2)\phi_k(\mathbf{r}_1) \tag{F.25}$$

We further make the following assumptions, which is usually the case in typical quantum chemistry calculations

1. Non-interacting Hamiltonian is real and spin-independent: $h_{ij}^s = h_{ij} = h_{ji}$;
2. Interaction is real: $V_{ijkl} = V_{kjil} = V_{ilkj} = V_{jilk}$.

In this case, the tails become

- $(G^{\mathrm{I}})_{ij}^s = \delta_{ij}$
- $(G^{\mathrm{II}})_{ij}^s = h_{ij}^s + \sum_{s'} V_{iqjn}\rho_{qn}^{s'} - V_{iqnj}\rho_{qn}^s = h_{ij} + V_{iqjn}\sum_{s'}\rho_{qn}^{s'} - V_{iqnj}\rho_{qn}^s$
- It can be shown that $G^{\mathrm{III}} = (G^{\mathrm{II}})^2 + \Sigma^{\mathrm{I}}$, where Σ^{I} is the $1/\mathrm{i}\omega_n$ moment of the self-energy,

which we can get from the previous derivations as

$$
\begin{aligned}
(\Sigma^{\mathrm{I}})^{\sigma}_{ij} =& \big(V_{iqmn}V_{rnjv}(\chi^{\sigma\sigma}_{rqvm} - \rho^{\sigma}_{rv}\rho^{\sigma}_{qm}) - V_{iqmn}V_{rnjv}\chi^{\sigma\bar{\sigma}}_{rqmv} - V_{iqmn}V_{rnvj}\sum_{\sigma'}(\chi^{\sigma'\sigma}_{rqvm} - \rho^{\sigma'}_{rv}\rho^{\sigma}_{qm}) \\
& - V_{iqnm}V_{rnjv}\sum_{\sigma'}(\chi^{\sigma\sigma'}_{rqvm} - \rho^{\sigma}_{rv}\rho^{\sigma'}_{qm}) + V_{iqnm}V_{rnvj}\sum_{\sigma'\tau'}(\chi^{\sigma'\tau'}_{rqvm} - \rho^{\sigma'}_{rv}\rho^{\tau'}_{qm})\big) \\
& + \frac{1}{2}\big(V_{iqmn}V_{mnjv}\sum_{\sigma'}\rho^{\sigma'}_{qv} - V_{iqmn}V_{mnvj}\rho^{\sigma}_{qv} - V_{iqnm}V_{mnjv}\rho^{\sigma}_{qv} + V_{iqnm}V_{mnvj}\sum_{\sigma'}\rho^{\sigma'}_{qv}\big) \\
& - \frac{1}{4}\big(V_{iqmn}V_{rsqj}\chi^{\sigma\sigma}_{rsmn} - V_{iqmn}V_{rsqj}\chi^{\bar{\sigma}\sigma}_{rsnm} - V_{iqmn}V_{rsjq}\sum_{\sigma'}\chi^{\sigma\sigma'}_{rsmn} \\
& - V_{iqnm}V_{rsqj}\sum_{\sigma'}\chi^{\sigma'\sigma}_{rsmn} + V_{iqnm}V_{rsjq}\chi^{\sigma\sigma}_{rsmn} - V_{iqnm}V_{rsjq}\chi^{\sigma\bar{\sigma}}_{rsnm}\big) \\
=& \big(V_{iqmn}V_{rnjv}(\chi^{\sigma\sigma}_{rqvm} - \rho^{\sigma}_{rv}\rho^{\sigma}_{qm}) - V_{iqmn}V_{rnjv}\chi^{\sigma\bar{\sigma}}_{rqmv} - V_{iqmn}V_{rnvj}\sum_{\sigma'}(\chi^{\sigma'\sigma}_{rqvm} - \rho^{\sigma'}_{rv}\rho^{\sigma}_{qm}) \\
& - V_{iqnm}V_{rnjv}\sum_{\sigma'}(\chi^{\sigma\sigma'}_{rqvm} - \rho^{\sigma}_{rv}\rho^{\sigma'}_{qm}) + V_{iqnm}V_{rnvj}\sum_{\sigma'\tau'}(\chi^{\sigma'\tau'}_{rqvm} - \rho^{\sigma'}_{rv}\rho^{\tau'}_{qm})\big) \\
& + V_{iqmn}V_{mnjv}\sum_{\sigma'}\rho^{\sigma'}_{qv} - V_{iqmn}V_{mnvj}\rho^{\sigma}_{qv} \\
& - \frac{1}{2}(V_{iqmn}V_{rsqj}\chi^{\sigma\sigma}_{rsmn} - V_{iqmn}V_{rsqj}\chi^{\bar{\sigma}\sigma}_{rsnm} - V_{iqnm}V_{rsqj}\sum_{\sigma'}\chi^{\sigma'\sigma}_{rsmn})
\end{aligned}
\tag{F.26}
$$

Bibliography

[1] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", Phys. Rev. **140**, A1133–A1138 (1965).

[2] C. C. J. Roothaan, "Self-consistent field theory for open shells of electronic systems", Rev. Mod. Phys. **32**, 179–185 (1960).

[3] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics", Rev. Mod. Phys. **20**, 367–387 (1948).

[4] J. J. Phillips and D. Zgid, "Communication: The description of strong correlation within self-consistent Green's function second-order perturbation theory", Journal of Chemical Physics **140**, 241101 (2014).

[5] A. A. Rusakov, J. J. Phillips, and D. Zgid, "Local Hamiltonians for quantitative Green's function embedding methods", Journal of Chemical Physics **141**, 194105 (2014).

[6] J. J. Phillips, A. A. Kananenka, and D. Zgid, "Fractional charge and spin errors in self-consistent Green's function theory", Journal of Chemical Physics **142**, 194108 (2015).

[7] A. R. Welden, A. A. Rusakov, and D. Zgid, "Exploring connections between statistical mechanics and Green's functions for realistic systems: Temperature dependent electronic entropy and internal energy from a self-consistent second-order Green's function", The Journal of Chemical Physics **145**, 204106 (2016).

[8] L. Hedin, "New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem", Physical Review **139**, A796–A823 (1965).

[9] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, "Continuous-time quantum Monte Carlo method for fermions", Physical Review B **72**, 035122 (2005).

[10] P. Werner, A. Comanac, L. de'Medici, M. Troyer, and A. J. Millis, "Continuous-time solver for quantum impurity models", Phys. Rev. Lett. **97**, 076405 (2006).

[11] P. Werner and A. J. Millis, "Hybridization expansion impurity solver: general formulation and application to kondo lattice and two-orbital models", Phys. Rev. B **74**, 155107 (2006).

[12] E. Gull, P. Werner, O. Parcollet, and M. Troyer, "Continuous-time auxiliary-field Monte Carlo for quantum impurity models", EPL (Europhysics Letters) **82**, 57003 (2008).

[13] E. Gull, P. Staar, S. Fuchs, P. Nukala, M. S. Summers, T. Pruschke, T. C. Schulthess, and T. Maier, "Submatrix updates for the continuous-time auxiliary-field algorithm", Phys. Rev. B **83**, 075122 (2011).

[14] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, "Continuous-time Monte Carlo methods for quantum impurity models", Reviews of Modern Physics **83**, 349–404 (2011).

[15] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, "Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions", Reviews of Modern Physics **68**, 13–125 (1996).

[16] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, "Electronic structure calculations with dynamical mean-field theory", Reviews of Modern Physics **78**, 865–951 (2006).

[17] N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn, "Exact, complete, and universal continuous-time worldline Monte Carlo approach to the statistics of discrete quantum systems", Journal of Experimental and Theoretical Physics **87**, 310–321 (1998).

[18] N. V. Prokof'ev and B. V. Svistunov, "Polaron problem by diagrammatic quantum Monte Carlo", Physical Review Letters **81**, 2514–2517 (1998).

[19] N. Prokof'ev and B. Svistunov, "Fermi-polaron problem: diagrammatic Monte Carlo method for divergent sign-alternating series", Physical Review B **77**, 020408(R) (2008).

[20] N. V. Prokof'ev and B. V. Svistunov, "Bold diagrammatic Monte Carlo: a generic sign-problem tolerant technique for polaron models and possibly interacting many-body problems", Physical Review B **77**, 125101 (2008).

[21] K. van Houcke, E. Kozik, N. Prokof'ev, and B. Svistunov, "Diagrammatic monte carlo", Physics Procedia **6**, Computer Simulations Studies in Condensed Matter Physics XXI, 95–105 (2010).

[22] R. Rossi, "Determinant diagrammatic Monte Carlo algorithm in the thermodynamic limit", Physical Review Letters **119**, 045701 (2017).

[23] A. Moutenet, W. Wu, and M. Ferrero, "Determinant Monte Carlo algorithms for dynamical quantities in fermionic systems", Physical Review B **97**, 085117 (2018).

[24] F. Šimkovic and E. Kozik, "Determinant Monte Carlo for irreducible Feynman diagrams in the strongly correlated regime", Physical Review B **100**, 121102 (2019).

[25] F. Šimkovic, R. Rossi, and M. Ferrero, "Efficient one-loop-renormalized vertex expansions with connected determinant diagrammatic monte carlo", Phys. Rev. B **102**, 195122 (2020).

[26] R. Rossi, F. Šimkovic, and M. Ferrero, "Renormalized perturbation theory at large expansion orders", EPL (Europhysics Letters) **132**, 11001 (2020).

[27] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, "Taming the dynamical sign problem in real-time evolution of quantum many-body problems", Phys. Rev. Lett. **115**, 266802 (2015).

[28] H.-T. Chen, G. Cohen, and D. R. Reichman, "Inchworm monte carlo for exact non-adiabatic dynamics. i. theory and algorithms", J. Chem. Phys. **146**, 054105 (2017).

[29] A. E. Antipov, Q. Dong, J. Kleinhenz, G. Cohen, and E. Gull, "Currents and green's functions of impurities out of equilibrium: results from inchworm quantum monte carlo", Phys. Rev. B **95**, 085144 (2017).

[30] Q. Dong, I. Krivenko, J. Kleinhenz, A. E. Antipov, G. Cohen, and E. Gull, "Quantum monte carlo solution of the dynamical mean field equations in real time", Phys. Rev. B **96**, 155126 (2017).

[31] A. Boag, E. Gull, and G. Cohen, "Inclusion-exclusion principle for many-body diagrammatics", Phys. Rev. B **98**, 115152 (2018).

[32] I. Krivenko, J. Kleinhenz, G. Cohen, and E. Gull, "Dynamics of kondo voltage splitting after a quantum quench", Phys. Rev. B **100**, 201104 (2019).

[33] E. Eidelstein, E. Gull, and G. Cohen, "Multiorbital quantum impurity solver for general interactions and hybridizations", Phys. Rev. Lett. **124**, 206405 (2020).

[34] R. Rossi, T. Ohgoe, K. van Houcke, and F. Werner, "Resummation of diagrammatic series with zero convergence radius for strongly correlated fermions", Phys. Rev. Lett. **121**, 130405 (2018).

[35] A. A. Kananenka, E. Gull, and D. Zgid, "Systematically improvable multiscale solver for correlated electron systems", Physical Review B **91**, 121111(R) (2015).

[36] T. N. Lan, A. A. Kananenka, and D. Zgid, "Communication: towards *ab initio* self-energy embedding theory in quantum chemistry", The Journal of Chemical Physics **143**, 241102 (2015).

[37] T. N. Lan, A. A. Kananenka, and D. Zgid, "Rigorous ab initio quantum embedding for quantum chemistry using green's function theory: screened interaction, nonlocal self-energy relaxation, orbital basis, and chemical accuracy", Journal of Chemical Theory and Computation **12**, 4856–4870 (2016).

[38] A. Szabo and N. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Dover Books on Chemistry (Dover Publications, 1996).

[39] J. W. Negele and H. Orland, *Quantum many-particle systems* (Addison-Wesley, 1988).

[40] J. C. Slater, "Atomic shielding constants", Phys. Rev. **36**, 57–64 (1930).

[41] S. F. Boys and A. C. Egerton, "Electronic wave functions - i. a general method of calculation for the stationary states of any molecular system", Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **200**, 542–554 (1950).

[42] W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-consistent molecular-orbital methods. i. use of gaussian expansions of slater-type atomic orbitals", The Journal of Chemical Physics **51**, 2657–2664 (1969).

[43] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen", The Journal of Chemical Physics **90**, 1007–1023 (1989).

[44] M. Motta, D. M. Ceperley, G. K.-l. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang, "Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods", Physical Review X **7**, 031059 (2017).

[45] S. Lehtola, "A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules", International Journal of Quantum Chemistry **119**, e25968 (2019).

[46] G. D. Mahan, *Many particle physics, third edition* (Plenum, New York, 2000).

[47] J. Li, M. Wallerberger, and E. Gull, "Diagrammatic monte carlo method for impurity models with general interactions and hybridizations", Phys. Rev. Research **2**, 033211 (2020).

[48] R. Rossi, F. Werner, N. V. Prokof'ev, and B. Svistunov, "Shifted-action expansion and applicability of dressed diagrammatic schemes", Physical Review B **93**, 161102(R) (2016).

[49] W. Wu, M. Ferrero, A. Georges, and E. Kozik, "Controlling Feynman diagrammatic expansions: physical nature of the pseudogap in the two-dimensional Hubbard model", Phys. Rev. B **96**, 041105 (2017).

[50] M. Srednicki, *Quantum field theory* (Cambridge University Press, 2007).

[51] D. Landau and K. Binder, *A guide to monte carlo simulations in statistical physics* (Cambridge university press, 2021).

[52] J. M. Luttinger and J. C. Ward, "Ground-state energy of a many-fermion system. ii", Phys. Rev. **118**, 1417–1427 (1960).

[53] G. Baym and L. P. Kadanoff, "Conservation laws and correlation functions", Phys. Rev. **124**, 287–299 (1961).

[54] G. Baym, "Self-consistent approximations in many-body systems", Phys. Rev. **127**, 1391–1401 (1962).

[55] M. Potthoff, "Self-energy-functional approach to systems of correlated electrons", The European Physical Journal B - Condensed Matter and Complex Systems **32**, 429–436 (2003).

[56] L. Lin and M. Lindsey, "Bold Feynman diagrams and the Luttinger-Ward formalism via Gibbs measures. Part I: perturbative approach", arXiv preprint arXiv:1809.02900 (2018).

[57] L. Lin and M. Lindsey, "Variational structure of luttinger–ward formalism and bold diagrammatic expansion for euclidean lattice field theory", Proceedings of the National Academy of Sciences **115**, 2282–2286 (2018).

[58] E. Kozik, M. Ferrero, and A. Georges, "Nonexistence of the luttinger-ward functional and misleading convergence of skeleton diagrammatic series for hubbard-like models", Phys. Rev. Lett. **114**, 156402 (2015).

[59] O. Gunnarsson, G. Rohringer, T. Schäfer, G. Sangiovanni, and A. Toschi, "Breakdown of traditional many-body theories for correlated electrons", Phys. Rev. Lett. **119**, 056402 (2017).

[60] K. V. Houcke, E. Kozik, R. Rossi, Y. Deng, and F. Werner, "Physical and unphysical regimes of self-consistent many-body perturbation theory", arXiv preprint arXiv:2102.04508 (2021).

[61] A. Stan, N. E. Dahlen, and R. van Leeuwen, "Fully self-consistent GW calculations for atoms and molecules", Europhysics Letters (EPL) **76**, 298–304 (2006).

[62] A. Stan, N. E. Dahlen, and R. van Leeuwen, "Levels of self-consistency in the GW approximation", The Journal of Chemical Physics **130**, 114105 (2009).

[63] C.-O. ALMBLADH, U. V. BARTH, and R. V. LEEUWEN, "Variational total energies from Φ- and Ψ- derivable theories", International Journal of Modern Physics B **13**, 535–541 (1999).

[64] S. A. Kulagin, N. Prokof'ev, O. A. Starykh, B. Svistunov, and C. N. Varney, "Bold diagrammatic monte carlo method applied to fermionized frustrated spins", Phys. Rev. Lett. **110**, 070601 (2013).

[65] I. S. Tupitsyn, A. S. Mishchenko, N. Nagaosa, and N. Prokof'ev, "Coulomb and electron-phonon interactions in metals", Phys. Rev. B **94**, 155145 (2016).

[66] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull (Simons Collaboration on the Many-Electron Problem), "Solutions of the two-dimensional hubbard model: benchmarks and results from a wide range of numerical algorithms", Phys. Rev. X **5**, 041041 (2015).

[67] T. Schäfer, N. Wentzell, F. Šimkovic, Y.-Y. He, C. Hille, M. Klett, C. J. Eckhardt, B. Arzhang, V. Harkov, F. ç.-M. Le Régent, A. Kirsch, Y. Wang, A. J. Kim, E. Kozik, E. A. Stepanov, A. Kauch, S. Andergassen, P. Hansmann, D. Rohe, Y. M. Vilk, J. P. F. LeBlanc, S. Zhang, A.-M. S. Tremblay, M. Ferrero, O. Parcollet, and A. Georges, "Tracking the footprints of spin fluctuations: a multimethod, multimessenger study of the two-dimensional hubbard model", Phys. Rev. X **11**, 011058 (2021).

[68] M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis, U. Ray, E. Ronca, H. Shi, S. Sorella, E. M. Stoudenmire, S. R. White, and S. Zhang (Simons Collaboration on the Many-Electron Problem), "Ground-state properties of the hydrogen chain: dimerization, insulator-to-metal transition, and magnetic phases", Phys. Rev. X **10**, 031058 (2020).

[69] K. T. Williams, Y. Yao, J. Li, L. Chen, H. Shi, M. Motta, C. Niu, U. Ray, S. Guo, R. J. Anderson, J. Li, L. N. Tran, C.-N. Yeh, B. Mussard, S. Sharma, F. Bruneval, M. van Schilfgaarde, G. H. Booth, G. K.-L. Chan, S. Zhang, E. Gull, D. Zgid, A. Millis, C. J. Umrigar, and L. K. Wagner (Simons Collaboration on the Many-Electron Problem), "Direct comparison of many-body methods for realistic electronic hamiltonians", Phys. Rev. X **10**, 011041 (2020).

[70] D. Feller, "Application of systematic sequences of wave functions to the water dimer", The Journal of Chemical Physics **96**, 6104–6114 (1992).

[71] T. Helgaker, W. Klopper, H. Koch, and J. Noga, "Basis-set convergence of correlated calculations on water", The Journal of Chemical Physics **106**, 9639–9646 (1997).

[72] J. R. Trail and R. J. Needs, "Pseudopotentials for correlated electron systems", The Journal of Chemical Physics **139**, 014101 (2013).

[73] J. R. Trail and R. J. Needs, "Correlated electron pseudopotentials for 3d-transition metals", The Journal of Chemical Physics **142**, 064110 (2015).

[74] J. R. Trail and R. J. Needs, "Shape and energy consistent pseudopotentials for correlated electron systems", The Journal of Chemical Physics **146**, 204107 (2017).

[75] M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials for correlated calculations", The Journal of Chemical Physics **147**, 224106 (2017).

[76] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, "Heat-bath configuration interaction: an efficient selected ci algorithm inspired by heat-bath sampling", J. Chem. Theory Comput. **12**, 3674–3680 (2016).

[77] J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, "Fast semistochastic heat-bath configuration interaction", J. Chem. Phys. **148**, 214110 (2018).

[78] P. W. Anderson, "Localized magnetic states in metals", Phys. Rev. **124**, 41–53 (1961).

[79] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen, "Spins in few-electron quantum dots", Reviews of Modern Physics **79**, 1217, 1217 (2007).

[80] R. Brako and D. M. Newns, "Slowly varying time-dependent local perturbations in metals: a new approach", Journal of Physics C: Solid State Physics **14**, 3065–3078 (1981).

[81] D. C. Langreth and P. Nordlander, "Derivation of a master equation for charge-transfer processes in atom-surface collisions", Phys. Rev. B **43**, 2541–2557 (1991).

[82] A. I. Lichtenstein and M. I. Katsnelson, "Antiferromagnetism and d-wave superconductivity in cuprates: a cluster dynamical mean-field theory", Phys. Rev. B **62**, R9283(R) (2000).

[83] G. Kotliar, S. Y. Savrasov, G. Pálsson, and G. Biroli, "Cellular dynamical mean field approach to strongly correlated systems", Phys. Rev. Lett. **87**, 186401 (2001).

[84] M. H. Hettler, M. Mukherjee, M. Jarrell, and H. R. Krishnamurthy, "Dynamical cluster approximation: nonlocal dynamics of correlated electron systems", Phys. Rev. B **61**, 12739–12756 (2000).

[85] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, "Quantum cluster theories", Rev. Mod. Phys. **77**, 1027–1080 (2005).

[86] V. I. Anisimov, A. I. Poteryaev, M. A. Korotin, A. O. Anokhin, and G. Kotliar, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: dynamical mean-field theory", Journal of Physics: Condensed Matter **9**, 7359–7367 (1997).

[87] A. I. Lichtenstein and M. I. Katsnelson, "Ab initiocalculations of quasiparticle band structure in correlated systems: LDA++ approach", Physical Review B **57**, 6884–6895 (1998).

[88] K. Held, I. A. Nekrasov, G. Keller, V. Eyert, N. Blümer, A. K. McMahan, R. T. Scalettar, T. Pruschke, V. I. Anisimov, and D. Vollhardt, "Realistic investigations of correlated electron systems with LDA+DMFT", physica status solidi (b) **243**, 2599–2631 (2006).

[89] P. Sun and G. Kotliar, "Extended dynamical mean-field theory and *GW* method", Physical Review B **66**, 085120 (2002).

[90] S. Biermann, F. Aryasetiawan, and A. Georges, "First-principles approach to the electronic structure of strongly correlated systems: combining the *GW* approximation and dynamical mean-field theory", Physical Review Letters **90**, 086402 (2003).

[91] S. Biermann, F. Aryasetiawan, and A. Georges, "Electronic structure of strongly correlated materials: towards a first principles scheme", in Physics of spin in solids: materials, methods and applications, edited by S. Halilov (2005), pp. 43–65.

[92] L. Boehnke, F. Nilsson, F. Aryasetiawan, and P. Werner, "When strong correlations become weak: consistent merging of *GW* and DMFT", Physical Review B **94**, 201106(R) (2016).

[93] S. Choi, A. Kutepov, K. Haule, M. van Schilfgaarde, and G. Kotliar, "First-principles treatment of mott insulators: linearized QSGW+DMFT approach", npj Quantum Materials **1**, 16001 (2016).

[94] J. Lee and K. Haule, "Diatomic molecule as a testbed for combining DMFT with electronic structure methods such as *GW* and DFT", Phys. Rev. B **95**, 155104 (2017).

[95] D. Zgid and E. Gull, "Finite temperature quantum embedding theories for correlated systems", New Journal of Physics **19**, 023047 (2017).

[96] V. V. Mazurenko, S. N. Iskakov, A. N. Rudenko, V. I. Anisimov, and A. I. Lichtenstein, "Renormalized spectral function for co adatom on the pt(111) surface", Phys. Rev. B **82**, 193403 (2010).

[97] M. Caffarel and W. Krauth, "Exact diagonalization approach to correlated fermions in infinite dimensions: mott transition and superconductivity", Phys. Rev. Lett. **72**, 1545–1548 (1994).

[98] M. Capone, M. Civelli, S. S. Kancharla, C. Castellani, and G. Kotliar, "Cluster-dynamical mean-field theory of the density-driven mott transition in the one-dimensional Hubbard model", Phys. Rev. B **69**, 195105 (2004).

[99] E. Koch, G. Sangiovanni, and O. Gunnarsson, "Sum rules and bath parametrization for quantum cluster theories", Phys. Rev. B **78**, 115102 (2008).

[100] A. Liebsch and N.-H. Tong, "Finite-temperature exact diagonalization cluster dynamical mean-field study of the two-dimensional Hubbard model: pseudogap, non-fermi-liquid behavior, and particle-hole asymmetry", Phys. Rev. B **80**, 165126 (2009).

[101] D. Sénéchal, "Bath optimization in the cellular dynamical mean-field theory", Phys. Rev. B **81**, 235125 (2010).

[102] Y. Lu, M. Höppner, O. Gunnarsson, and M. W. Haverkort, "Efficient real-frequency solver for dynamical mean-field theory", Phys. Rev. B **90**, 085102 (2014).

[103] D. Zgid, E. Gull, and G. K.-L. Chan, "Truncated configuration interaction expansions as solvers for correlated quantum impurity models and dynamical mean-field theory", Phys. Rev. B **86**, 165128 (2012).

[104] A. Shee and D. Zgid, "Coupled cluster as an impurity solver for green's function embedding methods", Journal of Chemical Theory and Computation **15**, PMID: 31518129, 6010–6024 (2019).

[105] T. Zhu, C. A. Jiménez-Hoyos, J. McClain, T. C. Berkelbach, and G. K.-L. Chan, "Coupled-cluster impurity solvers for dynamical mean-field theory", Phys. Rev. B **100**, 115154 (2019).

[106] A. Weichselbaum and J. von Delft, "Sum-rule conserving spectral functions from the numerical renormalization group", Phys. Rev. Lett. **99**, 076402 (2007).

[107] R. Bulla, T. A. Costi, and T. Pruschke, "Numerical renormalization group method for quantum impurity systems", Rev. Mod. Phys. **80**, 395–450 (2008).

[108] E. Gorelov, T. O. Wehling, A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, "Relevance of the complete coulomb interaction matrix for the kondo problem: co impurities in cu hosts", Phys. Rev. B **80**, 155132 (2009).

[109] K. Haule, "Quantum Monte Carlo impurity solver for cluster dynamical mean-field theory and electronic structure calculations with adjustable cluster base", Phys. Rev. B **75**, 155113 (2007).

[110] K. Chen and K. Haule, "A combined variational and diagrammatic quantum Monte Carlo approach to the many-electron problem", Nature Communications **10**, 3725 (2019).

[111] R. Rossi, "Direct sampling of the self-energy with Connected Determinant Monte Carlo", arXiv preprint arXiv:1802.04743 (2018).

[112] S. Iskakov, C.-N. Yeh, E. Gull, and D. Zgid, "Ab initio self-energy embedding for the photoemission spectra of nio and mno", Phys. Rev. B **102**, 085105 (2020).

[113] Rossi, R., Prokof'ev, N., Svistunov, B., Van Houcke, K., and Werner, F., "Polynomial complexity despite the fermionic sign", EPL **118**, 10004 (2017).

[114] C. Bertrand, O. Parcollet, A. Maillard, and X. Waintal, "Quantum Monte Carlo algorithm for out-of-equilibrium green's functions at long times", Phys. Rev. B **100**, 125129 (2019).

[115] P. Gunacker, M. Wallerberger, E. Gull, A. Hausoel, G. Sangiovanni, and K. Held, "Continuous-time quantum monte carlo using worm sampling", Phys. Rev. B **92**, 155102 (2015).

[116] P. Knowles, K. Somasundram, N. Handy, and K. Hirao, "The calculation of higher-order energies in the many-body perturbation theory series", Chemical Physics Letters **113**, 8–12 (1985).

[117] J. Olsen, O. Christiansen, H. Koch, and P. Jørgensen, "Surprising cases of divergent behavior in møller-plesset perturbation theory", Journal of Chemical Physics **105**, 5082–5090 (1996).

[118] S. Hirata, M. R. Hermes, J. Simons, and J. V. Ortiz, "General-order many-body green's function method", Journal of Chemical Theory and Computation **11**, 1595–1606 (2015).

[119] S. Hirata, A. E. Doran, P. J. Knowles, and J. V. Ortiz, "One-particle many-body green's function theory: algebraic recursive definitions, linked-diagram theorem, irreducible-diagram theorem, and general-order algorithms", The Journal of Chemical Physics **147**, 044108 (2017).

[120] Z. Li, "Stochastic many-body perturbation theory for electron correlation energies", The Journal of Chemical Physics **151**, 244114 (2019).

[121] N. F. Mott, "The basis of the electron theory of metals, with special reference to the transition metals", Proceedings of the Physical Society. Section A **62**, 416–422 (1949).

[122] J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases", The Journal of Chemical Physics **127**, 114105 (2007).

[123] S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space gaussian pseudopotentials", Phys. Rev. B **54**, 1703–1710 (1996).

[124] C. Hättig, "Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: core–valence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr", Phys. Chem. Chem. Phys. **7**, 59–66 (2005).

[125] S. Iskakov and M. Danilov, "Exact diagonalization library for quantum electron models", Computer Physics Communications **225**, 128–139 (2018).

[126] M. Ma ček, P. T. Dumitrescu, C. Bertrand, B. Triggs, O. Parcollet, and X. Waintal, "Quantum quasi-monte carlo technique for many-body perturbative expansions", Phys. Rev. Lett. **125**, 047702 (2020).

[127] C. Bertrand, S. Florens, O. Parcollet, and X. Waintal, "Reconstructing nonequilibrium regimes of quantum many-body systems from the analytical structure of perturbative expansions", Phys. Rev. X **9**, 041008 (2019).

[128] A. A. Abrikosov, L. P. Gorkov, and I. Y. Dzyaloshinskii, *Methods of quantum field theory in statistical physics* (Pergamon, 1965).

[129] G. Cohen, D. R. Reichman, A. J. Millis, and E. Gull, "Green's functions from real-time bold-line Monte Carlo", Phys. Rev. B **89**, 115139 (2014).

[130] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, "Green's functions from real-time bold-line Monte Carlo calculations: spectral properties of the nonequilibrium anderson impurity model", Phys. Rev. Lett. **112**, 146802 (2014).

[131] A. N. Rubtsov and A. I. Lichtenstein, "Continuous-time quantum Monte Carlo method for fermions: beyond auxiliary field framework", Journal of Experimental and Theoretical Physics Letters **80**, 61–65 (2004).

[132] W. Ku, "Electronic Excitations in Metals and Semiconductors: Ab Initio Studies of Realistic Many-Particle Systems", PhD thesis (University of Tennessee, 2000).

[133] W. Ku and A. G. Eguiluz, "Band-Gap Problem in Semiconductors Revisited: Effects of Core States and Many-Body Self-Consistency", Physical Review Letters **89**, 126401 (2002).

[134] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, "Orthogonal polynomial representation of imaginary-time Green's functions", Physical Review B **84**, 075145 (2011).

[135] A. A. Kananenka, J. J. Phillips, and D. Zgid, "Efficient Temperature-Dependent Green's Functions Methods for Realistic Systems: Compact Grids for Orthogonal Polynomial Transforms", Journal of Chemical Theory and Computation **12**, 564–571 (2016).

[136] E. Gull, S. Iskakov, I. Krivenko, A. A. Rusakov, and D. Zgid, "Chebyshev polynomial representation of imaginary-time response functions", Physical Review B **98**, 75127 (2018).

[137] A. L. Kutepov, "Self-consistent solution of Hedin's equations: Semiconductors and insulators", Physical Review B **95**, 195120 (2017).

[138] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, "Compressing Green's function using intermediate representation between imaginary-time and real-frequency domains", Physical Review B **96**, 35147 (2017).

[139] N. Chikano, J. Otsuki, and H. Shinaoka, "Performance analysis of a physically constructed orthogonal representation of imaginary-time Green's function", Physical Review B **98**, 35104 (2018).

[140] N. Chikano, K. Yoshimi, J. Otsuki, and H. Shinaoka, "irbasis: Open-source database and software for intermediate-representation basis functions of imaginary-time Green's function", Computer Physics Communications **240**, 181–188 (2019).

[141] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. White, and R. Godby, "Enhancements to the GW space-time method", Computer Physics Communications **125**, 105–118 (2000).

[142] M. Kaltak, J. Klimeš, and G. Kresse, "Low scaling algorithms for the random phase approximation: Imaginary time and laplace transformations", Journal of Chemical Theory and Computation **10**, 2498–2507 (2014).

[143] A. A. Kananenka, A. R. Welden, T. N. Lan, E. Gull, and D. Zgid, "Efficient Temperature-Dependent Green's Function Methods for Realistic Systems: Using Cubic Spline Interpolation to Approximate Matsubara Green's Functions", Journal of Chemical Theory and Computation **12**, 2250–2259 (2016).

[144] A.-B. Comanac, "Dynamical Mean Field Theroy of Correlated Electron Systems: New Algorithms and Applications to Local Observables", PhD thesis (Columbia University, 2007).

[145] N. Blümer, "Mott-Hubbard Metal-Insulator Transition and Optical Conductivity in High Dimensions", PhD thesis (Universität Augsburg, 2002), p. 346.

[146] J. Li, M. Wallerberger, N. Chikano, C.-N. Yeh, E. Gull, and H. Shinaoka, "Sparse sampling approach to efficient ab initio calculations at finite temperature", Phys. Rev. B **101**, 035144 (2020).

[147] H. Shinaoka, N. Chikano, E. Gull, J. Li, T. Nomoto, J. Otsuki, M. Wallerberger, T. Wang, and K. Yoshimi, "Efficient ab initio many-body calculations based on sparse modeling of matsubara green's function", arXiv preprint arXiv:2106.12685 (2021).

[148] J. Otsuki, M. Ohzeki, H. Shinaoka, and K. Yoshimi, "Sparse modeling approach to analytical continuation of imaginary-time quantum Monte Carlo data", Physical Review E **95**, 061302(R)–6 (2017).

[149] Y. Nagai and H. Shinaoka, "Smooth Self-energy in the Exact-diagonalization-based Dynamical Mean-field Theory: Intermediate-representation Filtering Approach", Journal of the Physical Society of Japan **88**, 064004 (2019).

[150] A. Kutepov, K. Haule, S. Y. Savrasov, and G. Kotliar, "Electronic structure of Pu and Am metals by self-consistent relativistic *GW* method", Physical Review B **85**, 155129 (2012).

[151] D. Bohm and D. Pines, "A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas", Physical Review **92**, 609–625 (1953).

[152] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, "Pyscf: the python-based simulations of chemistry framework", Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).

[153] G. W. Stewart, "On the adjugate matrix", Linear Algebra and its Applications **283**, 151–164 (1998).