9 Hirsch-Fye Quantum Monte Carlo Method for Dynamical Mean-Field Theory

Nils Blümer
Institut für Physik
Johannes Gutenberg-Universität Mainz

Contents

E. Pavarini, E. Koch, Dieter Vollhardt, and Alexander Lichtenstein
The LDA+DMFT approach to strongly correlated materials
Modeling and Simulation Vol. 1
Forschungszentrum Jülich, 2011, ISBN 978-3-89336-734-4
http://www.cond-mat.de/events/correl11

1 Introduction

A conventional starting point for the study of strongly correlated electron systems is the Hubbard model, which in its single-band version reads

$$\hat{H} = -t \sum_{\langle ij \rangle, \sigma} \left(\hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} + \text{h.c.} \right) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \tag{1}$$

where $\hat{c}_{i\sigma}$ is the fermion annihilation operator at site i with spin $\sigma \in \{\uparrow, \downarrow\}$, $\hat{n}_{i\sigma} = \hat{c}^\dagger_{i\sigma} \hat{c}_{i\sigma}$ is the density of fermions with spin σ, t parameterizes the hopping amplitude, and U is the on-site interaction. Unfortunately, numerical methods for its direct solution are either restricted to one-dimensional cases or suffer in general from severe finite-size errors and/or sign problems. Insight into the physics of higher-dimensional systems, thus, requires the use of additional approximations. The dynamical mean-field theory (DMFT) neglects inter-site correlations by assuming a momentum-independent self-energy; it becomes exact in the limit of infinite coordination number. The DMFT maps the lattice problem onto a single-impurity Anderson model (SIAM), supplemented by a self-consistency condition [1]. Its enormous success within the last 20 years would not have been possible without the availability of controlled numerical solvers for (multi-orbital) SIAMs, in particular of the auxiliary-field Hirsch-Fye quantum Monte Carlo (HF-QMC) algorithm [2].

These lecture notes are aimed at a pedagogical introduction to the HF-QMC based method of solving the DMFT self-consistency problem, i.e, of computing electronic properties of Hubbard-type models (possibly derived *ab initio* for a specific material from density functional theory or using the GW method, see subsequent lecture by K. Held) at the DMFT level. As will become clear in the following, the usage of the HF-QMC impurity solver within the DMFT self-consistency cycle has several important implications that would not arise in HF-QMC solutions of some fixed SIAM. For example, statistical Monte Carlo errors might prevent the detection of metastable DMFT fixed points and can, in general, lead to systematic biases in observables (in all stochastic methods including continuous-time QMC); the formulation in imaginary time (in all QMC methods and some IPT variants) implies that (i) analytic continuation is necessary for obtaining spectral information and (ii) that Fourier transforms have to be used within the DMFT cycle which are especially problematic in the case of a uniform time grid (as in HF-QMC); finally, the discretization error of estimated observables in (conventional) HF-QMC is impacted in a quite irregular way due to the self-consistency (cf. Fig. 17) and makes the location of phase boundaries difficult. Thus, a large part of these lectures will go beyond the HF-QMC algorithm itself; we will try to discuss all issues that a researcher should understand (or at least be aware of) when evaluating or producing DMFT data on the basis of HF-QMC. On the other hand, we wish to avoid unnecessary complexity in the presentation, in particular in the formalism. Therefore, we will mainly restrict the treatment to the single-band case (1) which is representative also of multi-band models with (spin) density-density type interactions for which HF-QMC is most competitive and mainly used, but will also point out generalizations where appropriate.

In the remainder of this introduction, we will briefly recapitulate the DMFT and establish the

notation for the following. We will then discuss the HF-QMC algorithm for computing (an estimate of) the Green function of a given impurity problem (defined by the bath Green function $\mathcal{G}$), as established by Hirsch and Fye, in Sec. 2. In Sec. 3 we will outline all the steps necessary for achieving a self-consistent DMFT solution on the basis of HF-QMC; thus sections 2 and 3 together contain the essential steps for obtaining Green functions and self-energies for Hubbard-type models at the DMFT level (but with a systematic Trotter error). The computation of observables (such as the energy) and of spectra will follow in Sec. 4; here, we will also fully quantify errors and introduce extrapolation techniques for eliminating the Trotter error. Finally, we conclude and give an outlook in Sec. 5.

1.1 Dynamical mean-field theory

As discussed in the preceding lectures by Vollhardt and Kollar, the absence of momentum dependence in the self-energy greatly simplifies the treatment of the Hubbard model [3, 4]. It allows, in fact, to single out one of the lattice sites and replace the influence of its neighbors by the interaction with a single, frequency-dependent bath, i.e., map the Hubbard model onto a single impurity Anderson model (SIAM) in the limit of large coordination number. In order to restore the periodicity of the original lattice, this medium has to be determined self-consistently [1, 5–7]. Written in terms of fermionic Matsubara frequencies[1] $\omega_n = (2n+1)\pi T$, self-energy $\Sigma_{\sigma n} \equiv \Sigma_\sigma(i\omega_n)$, and Green function $G_{\sigma n} \equiv G_\sigma(i\omega_n)$ as well as its Fourier transform $G_\sigma(\tau)$ (cf. subsection 3.1) the resulting coupled equations read

$$G_{\sigma n} = \int_{-\infty}^{\infty} \mathrm{d}\varepsilon \, \frac{\rho(\varepsilon)}{i\omega_n + \mu - \Sigma_{\sigma n} - \varepsilon} \tag{2}$$

$$G_\sigma(\tau) = -\langle T_\tau \, \psi_\sigma(\tau) \, \psi_\sigma^*(0) \rangle_{\mathcal{A}}, \tag{3}$$

in the homogeneous phase. Here, T_τ is the time ordering operator; properties of the lattice only enter via the density of states (DOS) (for dispersion $\epsilon_{\boldsymbol{k}}$ and volume V_B of the Brillouin zone)

$$\rho(\epsilon) = \frac{1}{V_B} \int \mathrm{d}\boldsymbol{k} \; \delta(\epsilon - \epsilon_{\boldsymbol{k}}) \tag{4}$$

of the noninteracting electrons. The thermal average $\langle \hat{C} \rangle_{\mathcal{A}}$ of some observable $\hat{C}$ is defined as a functional integral over Grassmann variables[2] ψ, ψ^* (with differentials denoted by $\mathcal{D}$)

$$\langle \hat{C} \rangle_{\mathcal{A}} = \frac{1}{\mathcal{Z}} \int \mathcal{D}[\psi] \, \mathcal{D}[\psi^*] \, C[\psi, \psi^*] \, \mathrm{e}^{\mathcal{A}[\psi, \psi^*, \mathcal{G}]}, \tag{5}$$

using the partition function

$$\mathcal{Z} = \int \mathcal{D}[\psi] \, \mathcal{D}[\psi^*] \, \mathrm{e}^{\mathcal{A}[\psi, \psi^*, \mathcal{G}]}, \tag{6}$$

[1] Here, we choose an imaginary-time formulation in anticipation of its use in the context of the imaginary-time quantum Monte Carlo algorithm to be discussed in Sec. 2.

[2] Grassmann variables are anticommuting numbers for which integration and differentiation rules are defined [8, 9]. Strictly speaking, ψ and ψ^* are (independent) Grassmann fields, i.e., infinite-dimensional vectors with the discrete index σ and the continuous index/argument τ (or t in a real-time formulation). Thus, $\psi_\uparrow(\tau_i)$, $\psi_\downarrow(\tau_i)$, $\psi_\uparrow^*(\tau_i)$, and $\psi_\uparrow(\tau_j)$ are all linearly independent Grassmann variables for $\tau_i \neq \tau_j$.

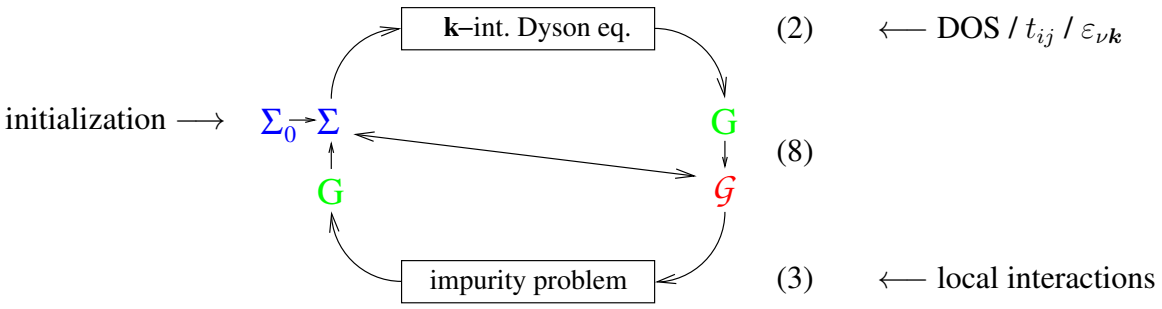


Fig. 1: *DMFT self-consistency cycle in conventional form: starting, e.g., with an initial guess for self-energy Σ, the $\boldsymbol{k}$ integrated Dyson equation (2) yields the lattice Green function G. Both Σ and G are used to compute the bath Green function $\mathcal{G}$ via (8) which defines the impurity problem (3). Its solution using QMC (or IPT, NCA, ED, NRG, etc.) provides a new estimate for G. The cycle is then closed by application of (8) to the new G and the old Σ.*

and the single-site action

$$\mathcal{A}[\psi,\psi^*,\mathcal{G}] = \int_0^\beta\int_0^\beta d\tau d\tau' \sum_\sigma \psi_\sigma^*(\tau)\mathcal{G}_\sigma^{-1}(\tau,\tau')\psi_\sigma(\tau') \;-\; U\int_0^\beta \mathrm{d}\tau\, \psi_\uparrow^*(\tau)\psi_\uparrow(\tau)\psi_\downarrow^*(\tau)\psi_\downarrow(\tau). \tag{7}$$

Here, $\beta = 1/k_\mathrm{B}T$ is the inverse temperature (in the following, we set $k_\mathrm{B} \equiv 1$) and

$$\mathcal{G}_{\sigma n}^{-1} = G_{\sigma n}^{-1} + \Sigma_{\sigma n} \tag{8}$$

is the effective local propagator.[3] Functions related by a Fourier transformation (here from imaginary time τ to fermionic Matsubara frequencies ω_n or vice versa) are denoted by the same symbol, but can be distinguished by their indices.

The solution of the DMFT problem by iteration is illustrated in Fig. 1. Here, the solution of the $\boldsymbol{k}$–integrated Dyson equation (2) is straightforward; it can be performed analytically for the semi-elliptic Bethe DOS commonly used for model studies in the literature. In contrast, the solution of the impurity problem (3) is highly nontrivial.[4] Most numerical methods developed for the treatment of SIAMs with fixed bath could be adapted to the DMFT problem, e.g., solutions based on exact diagonalization (ED) [1, 10], the non-crossing approximation (NCA) [11–14], the fluctuation-exchange approximation (FLEX) [15–17], the numerical renormalization group (NRG) [18–21], density-matrix renormalization group (DMRG) [22] and quantum Monte Carlo (QMC) algorithms. We will focus on the Hirsch-Fye QMC method in this lecture; recently developed diagrammatic continuous-time QMC algorithms will be covered by Philipp Werner.

[3] $\mathcal{G}_{\sigma n}$ may be regarded as a Weiss field in a (frequency-dependent) generalization of the usual static mean field. In contrast with spin models, for which the Weiss field replaces all (NN) interactions, the bath propagator replaces the hybridization of one site with the rest of the lattice.

[4] An exception is the application of (2) and (3) to a Lorentzian DOS $\rho(\epsilon) = t/(\pi(\epsilon^2 + t^2))$ which can be realized on lattices with long range hopping [6]. For this DOS (which is clearly pathological due to its infinite variance), the Weiss function is independent of U; furthermore (3) is solvable by Bethe ansatz in this case so that many properties can be obtained analytically.

2 Hirsch-Fye QMC solution of the single-impurity Anderson model

In this section, we will discuss the (original)[5] auxiliary-field quantum Monte Carlo (QMC) algorithm for solving the quantum impurity problem (3). It was originally formulated by Hirsch and Fye for treating a small number of magnetic impurities in metals [2] and later applied to arbitrary hybridization functions, i.e., in the form required for the solution of the DMFT problem [5, 24–26]. Here, we will concentrate on the solution of the impurity problem and postpone aspects specific to the DMFT context to Sec. 3. For simplicity, we also specialize on the single-band homogeneous case (1); generalizations for the multi-band case (and implications for the minus-sign problem) will be pointed out where appropriate, see also App. A. A very general formulation of the HF-QMC method in the multi-band case, including the cases of interorbital hybridization and complex interactions, can be found in [27].
The functional integral equation for the Green function has the structure

$$G_\sigma(\tau_1-\tau_2) = -\frac{1}{\mathcal{Z}}\int \mathcal{D}[\psi]\,\mathcal{D}[\psi^*]\;\; \psi_\sigma(\tau_1)\psi^*_\sigma(\tau_2)\;\; \exp\Big[\mathcal{A}_0 - U\int_0^\beta d\tau\, \psi^*_\uparrow\psi_\uparrow\psi^*_\downarrow\psi_\downarrow\Big], \quad (9)$$

where $\mathcal{A}_0$ denotes the hybridization part, i.e., the first term in (7). For $U = 0$, only quadratic terms in operators or Grassmann variables occur; then, this expression can be solved using Wick's theorem. The idea of the Hirsch-Fye QMC method is to transform the interaction term (quartic in operators or Grassmann variables) to a quadratic term and to solve the resulting problem, again, by Wicks theorem. Such a Hubbard Stratonovich transformation requires, however, a decoupling of interacting and noninteracting terms which only commute in the limit of infinite temperatures. Therefore, HF-QMC has to include some kind of numerical high-temperature expansion by Trotter decomposition. A high-level overview of the full scheme is shown in Fig. 2, which illustrates (in step iii) that the interaction between electrons with different spin (and/or different orbitals in the multi-band case) is replaced by the interaction of each electron with a binary auxiliary field. For each of the configurations the path integral in (9) evaluates to a term of the structure $M^{-1}\det\{M\}$ where the Matrix M depends onto the field configuration $\{s\}$. This analytic part will be discussed in the following subsection 2.1. We will then introduce the concept of Monte Carlo importance sampling in a general way in subsection 2.2 before applying it in the HF-QMC context in subsection 2.3.

2.1 Wick's theorem for the discretized impurity problem

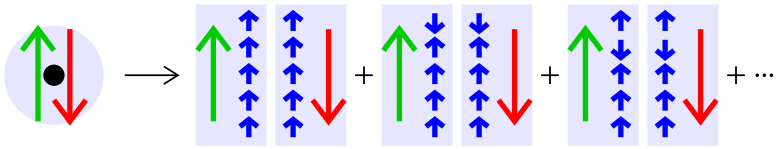
The difficulty in solving the functional integral equation (3) arises from the noncommutativity of the kinetic term and the interaction term in the single-site action (7). These terms can be

[5] Here and in the following, we discuss the original QMC version using a discretization of the imaginary time and based on Trotter decomposition and Hubbard-Stratonovich transformation. It should not be confused with the recent auxiliary-field formulation of the weak-coupling continuous-time QMC method often referred to as CT-AUX [23].

(i) Imaginary-time discretization ($\beta = \Lambda\,\Delta\tau$)

(ii) Trotter decoupling ($\rightsquigarrow$ systematic error $\mathcal{O}(\Delta\tau^2)$)

(iii) Hubbard-Stratonovich transformation (exact)



Wick theorem:

$$G = \frac{\sum_{\{s\}} M^{-1} \det\{M\}}{\sum_{\{s\}} \det\{M\}}$$

(iv) MC importance sampling over auxiliary Ising field $\{s\}$ ($\rightsquigarrow$ statistical error)

Fig. 2: *High-level overview of the HF-QMC approach.*

separated by use of the Trotter-Suzuki formula [28, 29] for operators $\hat{A}$ and $\hat{B}$:

$$\mathrm{e}^{-\beta(\hat{A}+\hat{B})} = \left(\mathrm{e}^{-\Delta\tau\hat{A}}\,\mathrm{e}^{-\Delta\tau\hat{B}}\right)^{\Lambda} + \mathcal{O}(\Delta\tau)\,, \tag{10}$$

where $\Delta\tau = \beta/\Lambda$ and Λ is the number of (imaginary) time slices. [6] Rewriting the action (7) in discretized form

$$\begin{aligned}\mathcal{A}_\Lambda[\psi,\psi^*,\mathcal{G},U] \;&=\; (\Delta\tau)^2 \sum_\sigma \sum_{l,l'=0}^{\Lambda-1} \psi^*_{\sigma l}(\boldsymbol{\mathcal{G}}^{-1}_\sigma)_{ll'}\,\psi_{\sigma l'} \\ &\quad -\Delta\tau U \sum_{l=0}^{\Lambda-1} \psi^*_{\uparrow l}\,\psi_{\uparrow l}\,\psi^*_{\downarrow l}\,\psi_{\downarrow l},\end{aligned} \tag{11}$$

where the matrix $\boldsymbol{\mathcal{G}}_\sigma$ consists of elements $\mathcal{G}_{\sigma ll'} \equiv \mathcal{G}_\sigma(l\Delta\tau - l'\Delta\tau)$ and $\psi_{\sigma l} \equiv \psi_\sigma(l\Delta\tau)$, we apply (10) and obtain to lowest order

$$\begin{aligned}\exp\left(\mathcal{A}_\Lambda[\psi,\psi^*,\mathcal{G},U]\right) \;&=\; \prod_{l=0}^{\Lambda-1}\Bigg[\exp\left((\Delta\tau)^2\sum_\sigma\sum_{l'=0}^{\Lambda-1}\psi^*_{\sigma l}(\boldsymbol{\mathcal{G}}^{-1}_\sigma)_{ll'}\,\psi_{\sigma l'}\right) \\ &\quad\times\;\exp\left(-\Delta\tau\,U\,\psi^*_{\uparrow l}\,\psi_{\uparrow l}\,\psi^*_{\downarrow l}\,\psi_{\downarrow l}\right)\Bigg].\end{aligned} \tag{12}$$

Shifting the chemical potential by $U/2$, the four-fermion term can be rewritten as a square of the magnetization (in terms of operators: $(\hat{n}_\uparrow - \hat{n}_\downarrow)^2 = \hat{n}^2_\uparrow + \hat{n}^2_\downarrow - 2\hat{n}_\uparrow\hat{n}_\downarrow = \hat{n}_\uparrow + \hat{n}_\downarrow - 2\hat{n}_\uparrow\hat{n}_\downarrow$) which makes it suitable for the following discrete Hubbard-Stratonovich transformation [30] (the correctness of which is checked easily by inserting the four possible combinations $[(0,0),(0,1),(1,0),(1,1)]$ of eigen values of the associated density operators):

$$\exp\left(\frac{\Delta\tau U}{2}(\psi^*_{\uparrow l}\,\psi_{\uparrow l} - \psi^*_{\downarrow l}\,\psi_{\downarrow l})^2\right) = \frac{1}{2}\sum_{s_l=\pm1}\exp\left(\lambda s_l(\psi^*_{\uparrow l}\,\psi_{\uparrow l} - \psi^*_{\downarrow l}\,\psi_{\downarrow l})\right), \tag{13}$$

[6] Since $\beta = 1/k_B T$ is the inverse temperature, small $\Delta\tau$ on each "time slice" corresponds to a higher temperature, for which the operators effectively decouple. Thus, we may view the Trotter approach as a numerical extension of a high-temperature expansion to lower temperatures.

with $\cosh\lambda = \exp(\Delta\tau U/2)$. Here, the interaction between electrons is replaced by the interaction with an auxiliary binary field $\{s\}$ with components $s_l = \pm 1$ for $0 \le l \le \Lambda$. Acting like a local, but time-dependent magnetic field, $\{s\}$ can be regarded as an ensemble of Ising spins (as depicted in Fig. 2).
Applying the Trotter formula again, these transformations yield an expression for the functional integral

$$G_{\sigma l_1 l_2} = \frac{1}{\mathcal{Z}} \sum_{\{s\}} \int \mathcal{D}[\psi]\, \mathcal{D}[\psi^*]\; \psi^*_{\sigma l_1} \psi_{\sigma l_2} \exp\Big(\sum_{\sigma,l,l'} \psi^*_{\sigma l} M^{s_l}_{\sigma l l'} \psi_{\sigma l'} \Big), \tag{14}$$

with[7]

$$M^{s_l}_{\sigma l l'} = (\Delta\tau)^2 (\boldsymbol{\mathcal{G}}^{-1}_\sigma)_{ll'} - \lambda\sigma\delta_{ll'} s_l, \tag{15}$$

where in (14) the sum is taken over all configurations of the Ising spin field, and each term of the sum involves independent fermions only. Now Wick's theorem [32] (or, equivalently, the rules for Gaussian Grassmann integrals) can be applied to get the solution[8]

$$G_{\sigma l l'} = \frac{1}{\mathcal{Z}} \sum_{\{s\}} \big(\boldsymbol{M}^{\{s\}}_\sigma\big)^{-1}_{ll'} \det \boldsymbol{M}^{\{s\}}_\uparrow \det \boldsymbol{M}^{\{s\}}_\downarrow, \tag{16}$$

where $\boldsymbol{M}^{\{s\}}_\sigma$ is the matrix with elements $M^{s_l}_{\sigma l l'}$, and the partition function has the value

$$\mathcal{Z} = \sum_{\{s\}} \det \boldsymbol{M}^{\{s\}}_\uparrow \det \boldsymbol{M}^{\{s\}}_\downarrow. \tag{17}$$

Equations (16) and (17) already constitute the solution of the impurity problem, as the physical problem has been reduced to algebra which can, in principle, be evaluated exactly for an arbitrary number Λ of time slices. In fact, such so-called full summation can be useful in some contexts (even though the number of terms is exponential). However, the independent computation of the matrix inverses and determinants for each auxiliary field (operations scaling with $\mathcal{O}(\Lambda^3)$ and worse than $\mathcal{O}(\Lambda^4)$, respectively) would be highly inefficient.

Fast update scheme

At least for the case of full summation (i.e., when all the terms are explicitly calculated) one may arrange the sum over the auxiliary spins in (16) so that only one auxiliary spin s_m is flipped between subsequent configuration (Gray code). For the matrices $\mathbf{M}^{\{s\}}_\sigma$ (with components l_1, l_2) this means [33]:

$$\begin{aligned} \mathbf{M}_\sigma \quad \overset{s_m \to -s_m}{\longrightarrow} \quad \mathbf{M}_\sigma{}' &= \mathbf{M}_\sigma + \boldsymbol{\Delta}^{\sigma\, m} \\ &= \big(1 + \boldsymbol{\Delta}^{\sigma\, m} (\mathbf{M}_\sigma)^{-1}\big)\, \mathbf{M}_\sigma \qquad (18) \\ &\quad \text{with} \qquad \Delta^{\sigma\, m}_{ll'} = \delta_{ll'}\delta_{lm}\, 2\Delta\tau\; \lambda\; \sigma s_l. \qquad (19) \end{aligned}$$

[7] A more precise form including subleading corrections is $M^{s_l}_{\sigma l l'} = (\Delta\tau)^2\, (\boldsymbol{\mathcal{G}}^{-1}_\sigma)_{ll'}\, \mathrm{e}^{\lambda\sigma s_{l'}} + \delta_{ll'} \left(1 - \mathrm{e}^{\lambda\sigma s_l}\right)$ [1, 31, 27].

[8] The form shown in Fig. 2 is obtained when $\boldsymbol{G}_\uparrow$ and $\boldsymbol{G}_\downarrow$ and, correspondingly, $\boldsymbol{M}_\uparrow$ and $\boldsymbol{M}_\downarrow$ are each arranged as the block-diagonal super-matrices $\boldsymbol{G}$ and $\boldsymbol{M}$, respectively.

The ratio of the determinants of new and old matrix can be easily determined using the inverse of the old matrix:[9]

$$\begin{aligned} R^{\sigma\, m} &:= \frac{\det(\mathbf{M}_\sigma{}')}{\det(\mathbf{M}_\sigma)} = \det\left(\mathbf{1} + \mathbf{\Delta}^{\sigma\, m}(\mathbf{M}_\sigma)^{-1}\right) \\ &= 1 + 2\Delta\tau\,\lambda\,\sigma s_m\,(M_\sigma)^{-1}_{mm}. \end{aligned} \tag{20}$$

The inversion of M is also elementary, one obtains:

$$(\mathbf{M}_\sigma{}')^{-1} = (\mathbf{M}_\sigma)^{-1} + \frac{1}{R^{\sigma\, m}}\,(\mathbf{M}_\sigma)^{-1}\mathbf{\Delta}^{\sigma\, m}(\mathbf{M}_\sigma)^{-1}. \tag{21}$$

This reduces the effort for the recalculation of a term of (16) after a spin flip to $\mathcal{O}(\Lambda^2)$.
Only for $\Lambda \lesssim 30$ can all terms be summed up exactly. Computations at larger Λ are made possible by Monte Carlo importance sampling which reduces the number of terms that have to be calculated explicitly from 2^Λ to order $\mathcal{O}(\Lambda)$.

2.2 Monte Carlo importance sampling

Monte Carlo (MC) procedures in general are stochastic methods for estimating large sums (or high-dimensional integrals) by picking out a comparatively small number of terms (or evaluating the integrand only for a relatively small number of points). Let us assume we want to compute the average $X := \frac{1}{M}\sum_{l=1}^{M} x_l$, where l is an index (e.g., an Ising configuration $l \equiv \{s\}$) and x some observable with the (true) variance $v_x = \frac{1}{M}\sum_{l=1}^{M}(x_l - X)^2$. In a simple MC approach, one may select a subset of $N \ll M$ indices independently with a uniform random distribution $P(l_j) = \text{const.}$ (for $1 \le j \le N$),

$$X_{\text{MC}} = \frac{1}{N}\sum_{j=1}^{N} x_{l_j}, \tag{22}$$

$$\Delta X_{\text{MC}} := \langle (X_{\text{MC}} - X)^2\rangle = \frac{v_x}{N} \approx \frac{1}{N(N-1)}\sum_{j=1}^{N}(x_{l_j} - X_{\text{MC}})^2\,. \tag{23}$$

Here, the averages are taken over all realizations of the random experiment (each consisting of a selection of N indices). In the limit of $N \to \infty$, the distribution of X_{MC} becomes Gaussian according to the central limit theorem. Only in this limit is the estimate of v_x from the QMC data reliable. An application for a continuous set is illustrated in Fig. 3.
Smaller errors and faster convergence to a Gaussian distribution for the estimate may be obtained by importance sampling. Here, the function x_l is split up,

$$x_l = p_l\, o_l; \quad p_l \ge 0; \quad \sum_{l=1}^{M} p_l = c\,, \tag{24}$$

[9] Since $\mathbf{\Delta}^{\sigma\, m}$ has only one non-zero element, it is clear that only the row m of $\mathbf{\Delta}^{\sigma\, m}(\mathbf{M}_\sigma)^{-1}$ will contain non-zero elements. The determinant of $1 + \mathbf{\Delta}^{\sigma\, m}(\mathbf{M}_\sigma)^{-1}$ is then equal to the product of its diagonal elements.

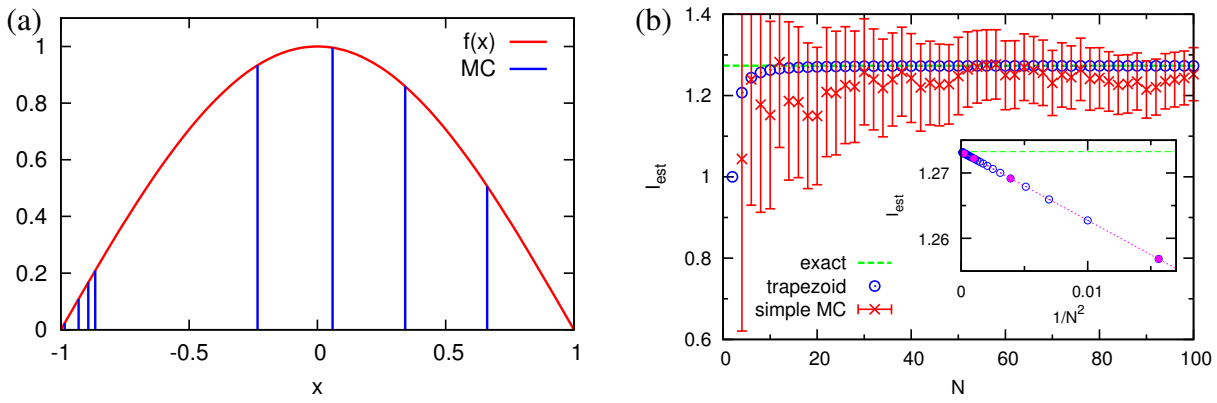


Fig. 3: *Illustration of simple MC: (a) instead of using deterministic quadrature schemes such as the trapezoid rule, integrals over some function $f(x)$ can be evaluated by averaging over function values $f(x_i)$ of stochastically generated support points x_i (and multiplying by the integration volume). (b) The resulting statistical error decreases as $1/\sqrt{N}$ with the number of function evaluations (crosses and error bars). The deterministic trapezoid discretization scheme (circles) converges with an error proportional to $N^{-2/d}$, i.e., is superior in dimensions $d < 4$; furthermore, an extrapolation of the systematic error is straightforward (inset).*

where we may regard p_l as a (unnormalized) probability distribution for the indices and o_l as a remaining observable. If both the normalization c is known (i.e., the sum over the weights p_l can be performed exactly) and the corresponding probability distribution can be realized (by drawing indices l with probability $P(l) = p_l/c$), we obtain

$$X^{\mathrm{MC}}_{\mathrm{imp}} = \frac{c}{N}\sum_{j=1}^{N} o_{l_j} \quad \text{and} \quad \Delta X^{\mathrm{MC}}_{\mathrm{imp}} = c\sqrt{\frac{v_o}{N}}\,. \tag{25}$$

Thus, the error can be reduced ($v_o < c^2 v_x$), when the problem is partially solvable, i.e., the sum over p_l with $p_l \approx x_l$ can be computed.[10] Since this is not possible in general, one usually has to treat the normalization c as an unknown and realize the probability distribution $P(l_j) = p_{l_j}/c$ in a stochastic Markov process: Starting with some initial configuration l_1, a chain of configurations is built up where in each step only a small subset of configurations l' is accessible in a "transition" from configuration l. Provided that the transition rules satisfy the detailed balance principle,

$$p_l\,\mathcal{P}(l \to l') = p_{l'}\,\mathcal{P}(l' \to l)\,, \tag{26}$$

and the process is ergodic (i.e., all configurations can be reached from some starting configuration), the distribution of configurations of the chain approaches the target distribution in the limit of infinite chain length, as illustrated in Fig. 4. Since the normalization remains unknown, importance sampling by a Markov process can only yield ratios of different observables evaluated on the same chain of configurations. Another consequence of using a Markov process is that initial configurations have to be excluded from averages since the true associated probabilities might be vanishingly small. They would otherwise be overrepresented in any run of

[10] The possible reduction of the variance is limited when x_l is of varying sign. This "minus-sign problem" seriously restricts the applicability of Monte Carlo methods for finite-dimensional fermion problems.

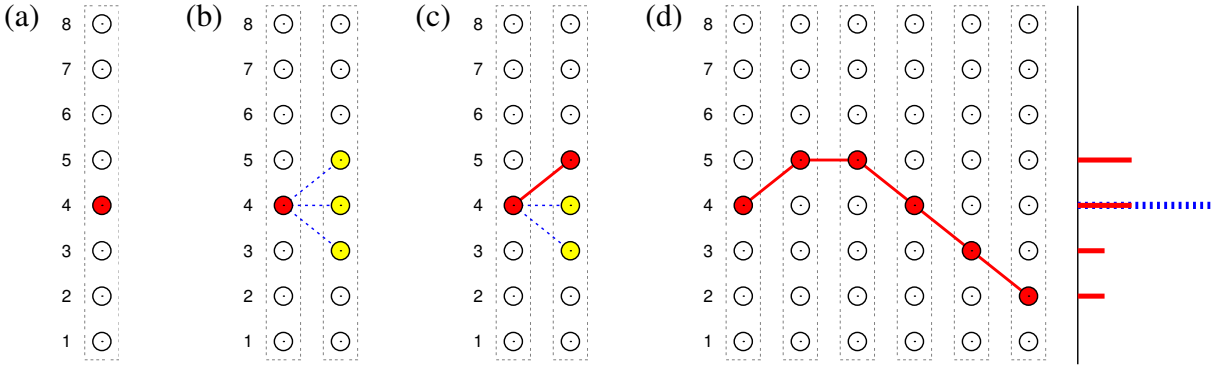


Fig. 4: *Illustration of importance sampling Monte Carlo, here for 8 states: (a) the simulation is initialized by randomly choosing a starting configuration, here state 4. (b) Transitions to the neighborhood of the old state are proposed, here for states* $s_{new} = s_{old} \pm 1$. *(c) A move (here* $+1$*) is accepted. (d) The history of the simulation (here with 5 attempted updates) consists both of accepted and declined updates. The normalized histogram of the visited configurations (red lines) approaches the target probability distribution (only) in the limit of infinite simulation length. For finite runs, states in the vicinity of the starting configuration may be overrepresented; this "warm-up" part should be discarded.*

finite length. Consequently, we will later distinguish "warm-up sweeps" from "measurement sweeps".

For the computation of errors, one has to take into account the finite autocorrelation induced by the Markov process, i.e., correlation between subsequent measurements. This correlation may be characterized by the autocorrelation time[11] $\kappa_o \geq 1$ which effectively reduces the number of independent samples, so that $\Delta X = c\sqrt{v_o \kappa_o / N}$. The numerical effort necessary to reach some target statistical accuracy ΔX nevertheless increases only as $(1/\Delta X)^2$.

2.3 Single-spin flip implementation of the HF-QMC method

Returning to the evaluation of the Green function using (16) and (17), the obvious choice is to sample configurations $\{s\}$ according to the (unnormalized) probability

$$P(\{s\}) = \left| \det \boldsymbol{M}_{\uparrow}^{\{s\}} \det \boldsymbol{M}_{\downarrow}^{\{s\}} \right| . \tag{27}$$

The Green function can then be calculated as an average $\langle \dots \rangle_s$ over these configurations:

$$G_{\sigma ll'} = \frac{1}{\tilde{\mathcal{Z}}} \left\langle \left(\boldsymbol{M}_{\sigma}^{\{s\}} \right)^{-1}_{ll'} \operatorname{sign} \left(\det \boldsymbol{M}_{\uparrow}^{\{s\}} \det \boldsymbol{M}_{\downarrow}^{\{s\}} \right) \right\rangle_s , \tag{28}$$

$$\tilde{\mathcal{Z}} = \left\langle \operatorname{sign} \left(\det \boldsymbol{M}_{\uparrow}^{\{s\}} \det \boldsymbol{M}_{\downarrow}^{\{s\}} \right) \right\rangle_s . \tag{29}$$

Here, $\tilde{\mathcal{Z}}$ deviates from the full partition function by an unknown prefactor which cancels in (28). The same is true for other expectation values of the form (5). The inability to compute

[11] For a set $\{o_1, o_2, \dots, o_N\}$ of measurements, the autocorrelation function (for the observable o) is $c_l^o = \langle (o_k - \langle o \rangle)(o_{k+l} - \langle o \rangle) \rangle_k$. An associated autocorrelation time may then be defined as $\kappa_o = c_0^o + 2 \sum_{l=1}^{N_0} c_l^o$, where the cutoff N_0 is determined by $c_l^o > 0$ for $l \leq N_0$ and $c_{N_0+1} < 0$.

the partition function itself (and, consequently, the free energy and entropy) is a consequence of the importance sampling and is thus a general characteristic of QMC methods which has severe consequences for the study of phase transitions.
If the sign in (28) does not depend on $\{s\}$, i.e. if the so-called sign problem is absent (as in HF-QMC for density-type interactions), the expressions can be simplified:

$$G_{\sigma ll'} = \frac{1}{\tilde{\mathcal{Z}}} \left\langle \left(\boldsymbol{M}_\sigma^{\{s\}}\right)^{-1}_{ll'} \right\rangle_s , \qquad \tilde{\mathcal{Z}} = \langle 1 \rangle_s . \tag{30}$$

A practical HF-QMC calculation proceeds as follows: (i) A starting configuration $\{s\}$ is chosen, e.g., taken from a uniform random distribution or from a previous run with similar parameters. (ii) The initial Green function matrix $\boldsymbol{M}^{-1}$ is computed for this configuration from scratch. Note that the associated determinant is not needed. (iii) A number N_{warm} of warm-up sweeps (see below) is performed in order to generate a Markov chain; intermediate Green functions $\boldsymbol{M}^{-1}$ are discarded. (iv) A (larger) number N_{meas} of measurement sweeps is performed, accumulating the sum of all occurring matrices $\boldsymbol{M}^{-1}$ (and possibly of other observables). (v) The estimate of the Green function and of observables is computed by dividing each of the sums by the number of terms, i.e., by the number of attempted configuration updates.
Each of the sweeps mentioned above consists of a loop in which Λ spins are selected for a possible update (either in sequence or randomly). Within the loop, the ratio of determinants for each candidate spin flip is evaluated using (20) and accepted or rejected according to a detailed balance principle. Often, the acceptance probability is chosen according to the Metropolis transition rule [34]

$$\mathcal{P}(\{s\} \to \{s'\}) = \min\{1, P(\{s'\})/P(\{s\})\}; \tag{31}$$

an alternative is the symmetric heat-bath algorithm.
Roughly, the number of sweeps has to be held constant (rather than the number of attempted single-spin flips) for constant statistical error when the number of spins Λ is varied. Since the systematic error depends on $\Delta\tau$, the necessary number of time slices is proportional to the inverse temperature, $\Lambda \propto \beta$, so that the total numerical cost for given accuracy is proportional to β^3.

Memory issues and parallelization

The numerically costly part of HF-QMC consists of the update of the Green function matrix (and the associated determinant), specifically of the addition of a dyadic (outer) product of two vectors (dimension Λ) onto a $\Lambda \times \Lambda$ matrix (for each spin and orbital). This BLAS-1 operation (DGER) breaks down to elementary multiply-add pairs where, in general, all operands of subsequent operations are different and have to be loaded from memory. Thus, an efficient execution depends on high memory bandwidth; in practice, this requires that the matrices can be held in the processor cache (typically L2). Fortunately, usual cache sizes (of a few megabytes) suffice for many applications; our code can use fine-grained OpenMP parallelization in order to extend these limits (by spreading the matrices across multiple processor cores and caches). Another strategy for optimizing memory issues is the use of delayed updates [35, 36].

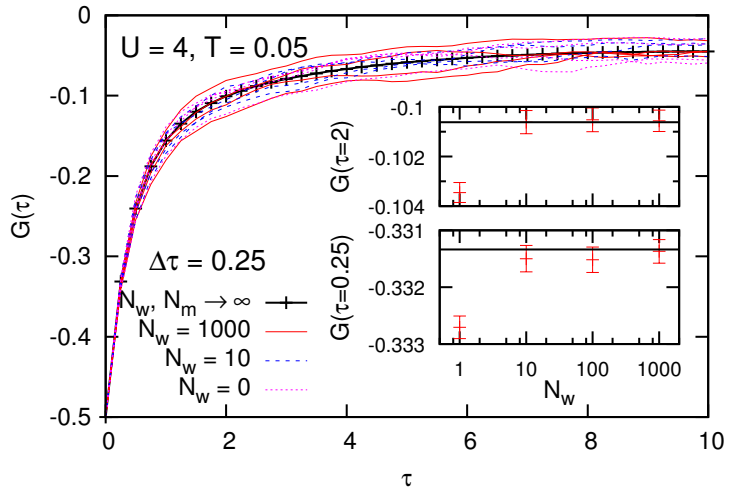


Fig. 5: *Impact of proper warm-up: Green functions estimated from short HF-QMC runs with varying number N_w of warm-up sweeps (see text).*

We also use MPI for the parallel solution of identical impurity problems, with near-perfect scaling up to some 20–30 MPI tasks (and reasonable scaling beyond 100 tasks). These limits could easily be extended by at least a factor of 10 by amortizing the warm-up of the auxiliary field over several (typically 20) DMFT iterations. It should be noted that the averaging over multiple independent solutions also stabilizes the procedure.

2.4 Choice of simulation parameters: discretization and number of sweeps

The important practical question how the simulation parameters should be choses is most easily answered for the number of measurement sweeps, at least on the SIAM level (not taking the DMFT self-consistency into account): since the error scales as N_{meas} for each observable (including the Green function), it can be determined from the desired precision.

A much harder question is the appropriate value of N_{warm}. A too small value should be avoided at all costs since it will lead to a systematic bias which is near-impossible to detect in the result data of a single run.[12] On the other hand, a too large value wastes resources. We typically use $N_{\text{warm}} \geq 1000$; in high-precision runs, we devote 10% of the sweeps of each serial runs to equilibration (possibly an overkill which, however, costs only 5% in statistical precision). In order to quantify the impact of the warm-up at least for one test case, we have performed a large number of independent simulations at fixed $\Delta\tau$ and for fixed hybridization function $\mathcal{G}$ with N_{warm} ranging from 1 to 1000. In order to see the effect, these runs have to be short; we have chosen $N_{\text{meas}} = 100$. This, on the other hand, gives rise to very large fluctuations in the measured Green functions as shown in the main panel in Fig. 5. At this level, a possible effect of N_{warm} is hidden in the noise. Averaging over a large number (here 1000) of realizations, however, allows

[12] However, one might keep track of the relative probabilities of the visited configurations (or rather of their logarithm) and derive a cutoff from the number of sweeps needed for reaching a typical probability level for the first time.

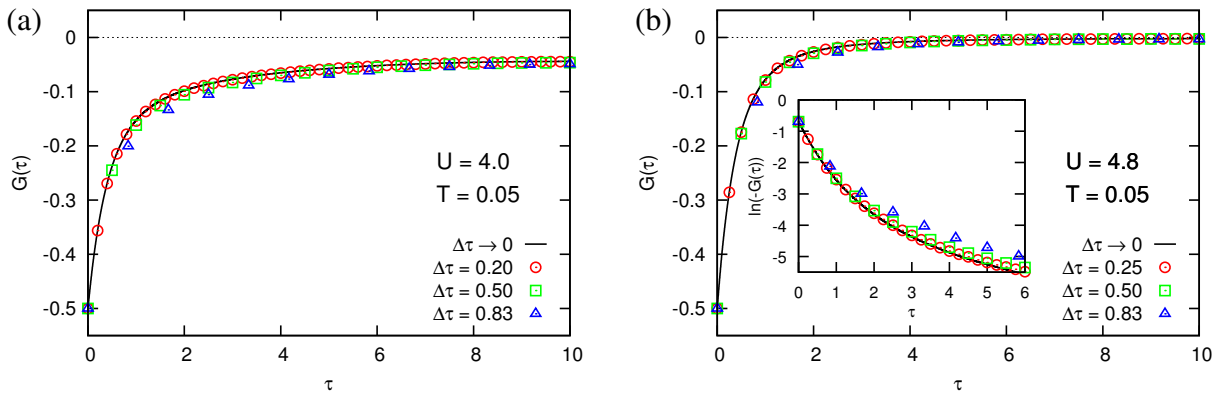


Fig. 6: *Impact of the Trotter discretization on HF-QMC estimates of the Green function at* $n = 1$ *and* $T = 0.05$ *for* $U = 4.0$ *(a) and* $U = 4.8$ *(b), respectively. Only results for* $\tau \leq \beta/2$ *are shown; the rest follows from particle-hole symmetry* $G(\tau) = G(\beta - \tau)$*. For each case,* $\Delta\tau$*-independent hybridization functions (corresponding to exact DMFT solutions obtained from multigrid HF-QMC) have been used.*

to characterize the probability distributions associated with the selected 100 sweeps (offset by 0, 10, 100, or 1000 warm-up sweeps from the random initialization of the Markov chain). The surprising result, show in the inset for the Green function at two representative grid points, is that 10 warm-up sweeps suffice for removing any detectable equilibration bias. One may suspect that the equilibration times are much longer for insulating or ordered phases; otherwise, significant savings in computer time would appear possible. Fig. 5 also illustrates that imprecise HF-QMC estimates of Green functions can violate causality, i.e., be non-monotonous.

For a sufficiently large number of sweeps, the error of the resulting Green functions is typically dominated by the systematic Trotter error. Such systematic shifts, consistent with a quadratic dependence on $\Delta\tau$, are seen in Fig. 6 for hybridization functions corresponding to metallic and insulating DMFT solutions, respectively. At the same time, the grids vary for each discretization which (at least at first sight) seems to limit the low-τ resolution for large values of $\Delta\tau$. The inset in Fig. 6b shows that the relative Trotter error increases for $\tau \to \beta/2$ in the insulating case. Still, the qualitative shape of the Green functions remains unchanged even for extremely large values of $\Delta\tau$ in both cases and it is hard to derive an intrinsic limit for reasonable values of $\Delta\tau$ on this level.

3 Achieving DMFT self-consistency

In the main part of this section, we will discuss two aspects of the quantum Monte Carlo (QMC) algorithm for the DMFT self-consistency problem: the discrete Fourier transformations between imaginary time and Matsubara frequencies and the search strategy for solutions of the DMFT equations. Both aspects are particularly relevant, e.g., for the detection of the Mott transition and the associated coexistence region.

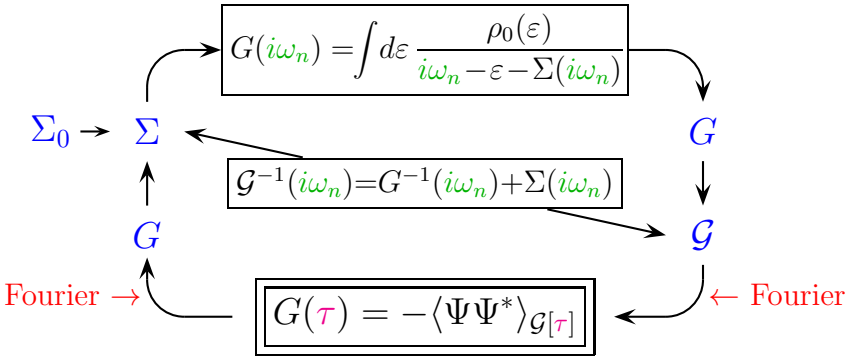


Fig. 7: *Schematic DMFT self-consistency scheme showing the two Fourier transformations needed per iteration.*

3.1 Fourier transformation and splining strategies

A QMC simulation within the DMFT framework consists of a simultaneous solution of two principal equations: the lattice Dyson equation (2) and the defining expression for the impurity Green function (3). The IPT (iterative perturbation theory [6, 37]) and QMC solutions of the impurity problem are formulated in imaginary time, i.e., the bath Green function $\mathcal{G}$ is needed as function of τ and the result is expressed as $G(\tau)$. In contrast, the Dyson equation is formulated (and is local) in the frequency domain, here for Matsubara frequencies $i\omega_n = i(2n+1)\pi T$. Therefore, two Fourier transformations (from frequency to imaginary time and vice versa) per self-consistency cycle are necessary (as shown in Fig. 7), which for G read

$$G(i\omega_n) = \int_0^\beta \mathrm{d}\tau \, \mathrm{e}^{i\omega_n \tau} \, G(\tau) \tag{32}$$

$$G(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} \mathrm{e}^{-i\omega_n \tau} \, G(i\omega_n) \, . \tag{33}$$

Note that (33) implies antiperiodicity of $G(\tau)$ for translations β (since $\mathrm{e}^{i\omega_n \beta} = -1$) and allows for a discontinuity of $G(\tau)$ (at $\tau = 0$) since the number of terms is infinite. The spectral representation of G

$$G(\omega) = \int_{-\infty}^{\infty} \mathrm{d}\omega' \frac{A(\omega')}{\omega - \omega'}; \qquad A(\omega) = -\frac{1}{\pi} \mathrm{Im}\, G(\omega + i0^+)$$

implies a decay of $G(i\omega_n)$ as $1/i\omega_n$ for $|n| \to \infty$. Furthermore, $G(i\omega_n)$ is purely imaginary when $G(\tau) = G(\beta - \tau)$ as in the case of half band filling on a bipartite lattice (with symmetric DOS).

Discretization problem, naive Fourier transformation

Numerically, however, the integral in (32) needs to be discretized and the Matsubara sum in (33) has to be truncated. Since the numerical effort in QMC scales with the number $\Lambda = \beta/\Delta\tau$ of

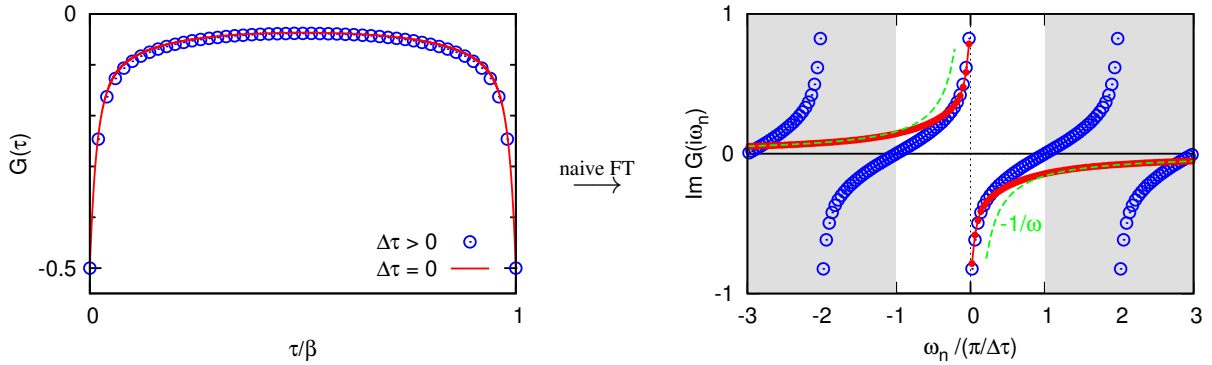


Fig. 8: *Illustration of the Fourier transformation (FT) from imaginary time to Matsubara frequencies, based on HF-QMC data for $U = 4$ and $T = 0.04$. The naive discrete FT (open circles) shows oscillatory behavior with poles or zeros at multiples of the Nyquist frequency. Advanced spline-based methods (filled circles and solid line), in contrast, recover the correct large-frequency asymptotics (dashed line).*

discretized time slices at least[13] as Λ^3, this method is presently restricted to $\Lambda \lesssim 400$. Typically, 200 time slices and less are used. A naive discrete version of the Fourier transform,

$$\tilde{G}(i\omega_n) = \Delta\tau\Big(\frac{G(0)-G(\beta)}{2} + \sum_{l=1}^{\Lambda-1} \mathrm{e}^{i\omega_n\tau_l}\, G(\tau_l)\Big); \quad \tau_l := l\Delta\tau \tag{34}$$

$$\tilde{G}(\tau_l) = \frac{1}{\beta}\sum_{n=-\Lambda/2}^{\Lambda/2-1} \mathrm{e}^{-i\omega_n\tau_l}\, G(i\omega_n)\,, \tag{35}$$

fails for such a coarse grid. The problems are that the Green function $\tilde{G}(\tau)$ estimated from a finite Matsubara sum cannot be discontinuous at $\tau = 0$ (as required analytically for $G(\tau)$ and also for $\mathcal{G}(\tau)$) while the discrete estimate $\tilde{G}(i\omega_n)$ oscillates with periodicity $2\pi i\Lambda/\beta$ instead of decaying for large frequencies. This implies a large error of $G(i\omega_n)$ when $|\omega_n|$ approaches or exceeds the Nyquist frequency $\pi\Lambda/\beta$ as illustrated in Fig. 8. Both (related) effects would make the evaluation of the corresponding self-consistency equations pointless. In particular, in the naive scheme, the self-energy diverges near the Nyquist frequency. Finally, the sum in (35) is numerically somewhat unstable since $\tilde{G}(\tau)$ also oscillates between the grid points τ_l.

Splining method

Fortunately, there is physical information left that has not been used in the naive scheme: $G(\tau)$ is known to be a smooth function in the interval $[0, \beta)$. In fact, it follows from (52) and (53) that $G(\tau)$ and all even-order derivatives are positive definite and, consequently, reach their maxima at the edges $\tau = 0$ and $\tau = \beta$. This knowledge of "smoothness" can be exploited in an interpolation of the QMC result $\{G(\tau_l)\}_{l=0}^{\Lambda}$ by a cubic spline.[14] The resulting functions may

[13] In practice, the scaling is even worse on systems with a hierarchy of memory systems (i.e. caches and main memory) of increasing capacity and decreasing speed.

[14] A spline of degree n is a function piecewise defined by polynomials of degree n with a globally continuous $(n-1)^{\text{th}}$ derivative. For a natural cubic spline, the curvature vanishes at the end points. The coefficients of the

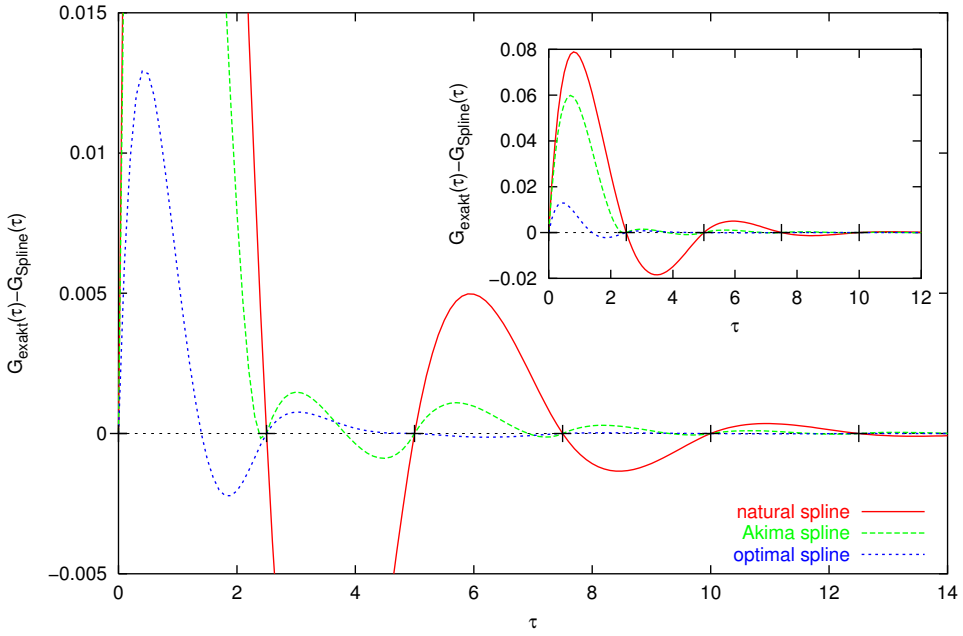


Fig. 9: *Difference between exact noninteracting Green function (for semi-elliptic DOS and $\beta = 100$) and cubic spline approximations to the discretized function for $\Lambda = 40$. The strong oscillations observed for the natural spline and for the Akima spline are due to their unphysically vanishing second derivatives (at $\tau = 0$). An optimal choice of the boundary condition reduces the discrepancies by an order of magnitude and doubles the oscillation frequency.*

then either be used for oversampling, i.e., for generating $G(\tau)$ on a finite grid or for piecewise direct analytic Fourier transforms. In both cases, $G(i\omega_n)$ can be calculated for a much larger frequency range than before.

The simplest practical approach in this context consists of a direct interpolation of the discrete QMC data by a cubic spline (with a continuous second derivative) as implemented by Krauth [1]. While this step suffices for closing the self-consistency equations without any further adjustments (like Ulmke's smoothing trick, see below), it leads to nonanalytic behavior of the self-energy near and beyond the Nyquist frequency [38]. This problem can be traced back to the fact that the natural spline chosen by Krauth is inadequate for this problem: By definition, the second derivative of a natural spline vanishes at its boundaries. However, all even derivatives of the true Green function are maximal at the edges of the interval $[0, \beta]$. The resulting misfit leads to unphysical ringing as illustrated in Fig. 9. Here, the noninteracting Green function for a semi-elliptic DOS is chosen as an example since it can be computed with arbitrary precision. Furthermore, moderate interactions do not lead to qualitative changes in the Green function so that the example is representative. It is clearly seen that a fit of a discrete set of data points (here for $\Lambda = 40$ and $\beta = 100$, i.e., a large discretization $\Delta\tau = 2.5$) using a natural cubic spline with continuous second derivative leads to a large error oscillating with the Nyquist frequency. While the initial misfit at $\tau = 0$ is similarly large for a natural Akima spline, the oscillation decays significantly faster due to jumps in the second derivative for this

spline fitting a given data set can be easily computed by solving a linear equation.

fit. The discrepancies can be reduced by about an order of magnitude and a smooth curve can be obtained by choosing an appropriate boundary condition (short-dashed line) which has here been obtained by a numerical minimization procedure for the variation of the second derivatives (see [38]). It is possible to avoid the minimization by computing the correct second derivative $G^{(2)}(0)$ analytically (for arbitrary interaction U):

$$\frac{d^2G(\tau)}{d\tau^2}\Big|_{\tau=0+} = -\frac{1}{2}\Big(1+\frac{U^2}{4}\Big); \tag{36}$$

in any case, however, a cubic spline interpolation of the full Green function suffers from the problem that derivatives of fourth and higher order vanish on segments of the splines while they can be large for the true Green function, in particular at $\tau = 0$.[15]

Splining the difference Green function using a model self-energy

A related method for improving on the natural spline scheme is to interpolate not $\{G(\tau_l)\}_{l=0}^{\Lambda}$, but only the difference with respect to some reference Green function obtained from, e.g., plain second order perturbation theory or IPT. This approach, discussed in a general context by Deisz [39], is implemented in Jarrell's code [5]. In addition, the high-frequency part may be directly stabilized by supplementing the QMC estimates with IPT (using low-pass/high-pass filters).
In the following, we will describe a stable and accurate method which achieves excellent high-frequency behavior without requiring any filters [38, 40, 41]. It is based on the following exact high-frequency expansion for the self-energy [42]:[16]

$$\Sigma_\sigma(\omega) = U\,(\langle\hat{n}_{-\sigma}\rangle - \frac{1}{2}) + U^2\,\frac{\langle\hat{n}_{-\sigma}\rangle(1-\langle\hat{n}_{-\sigma}\rangle)}{\omega} + \mathcal{O}(\omega^{-2}). \tag{37}$$

One possible choice of a model self-energy with this asymptotic behavior (including the ω^{-1} term) which is nonsingular (and purely imaginary) on the imaginary axis is given by the following two-pole approximation [38, 40, 41].

$$\Sigma_{\text{model},\sigma}(\omega) = U\,(\langle\hat{n}_{-\sigma}\rangle - \frac{1}{2}) + \frac{1}{2}U^2\,\langle\hat{n}_{-\sigma}\rangle\big(1-\langle\hat{n}_{-\sigma}\rangle\big)\big(\frac{1}{\omega+\omega_0} + \frac{1}{\omega-\omega_0}\big). \tag{38}$$

While the quality of the low-frequency part of this fit could be tuned by adjusting the parameter ω_0, it does not depend very sensitively on it as long as it is not much larger than the bandwidth or the Nyquist frequency. By evaluating Σ_{model} on the imaginary axis, the corresponding Green function G_{model} can be computed for an arbitrary number of Matsubara frequencies. Consequently, the Fourier transformation to imaginary time is unproblematic (when the "free" term $1/(i\omega_n)$ is taken care of analytically). Thus, the difference of the Green functions can be accurately evaluated at all time slices. Since the second derivative of the transformed model Green function at $\tau = 0$ exactly reproduces that of the true Green function, the difference is well

[15] This also implies that the boundary value (36) may not be the best choice for the whole interval $\tau \in [0, \Delta\tau]$.

[16] The multi-band generalization [40] contains additional pair occupancy terms which have to be computed numerically even at fixed band filling.

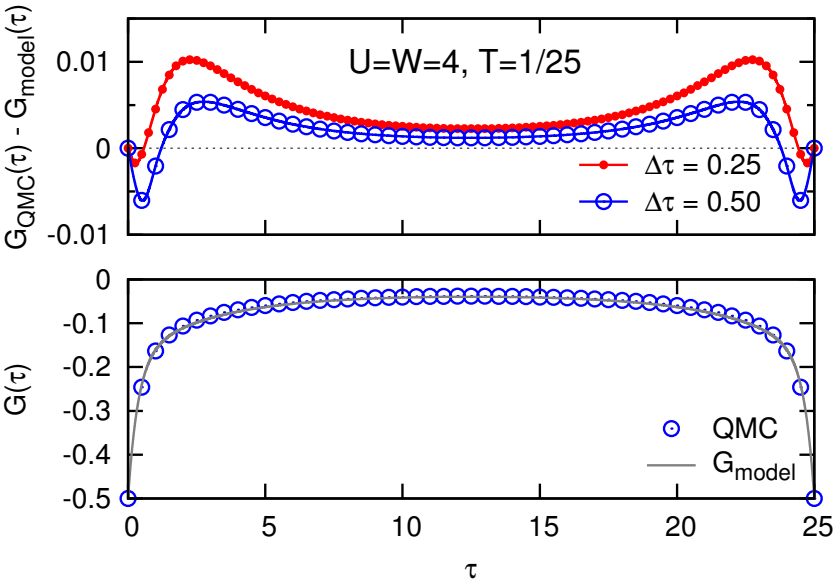


Fig. 10: *Example for model based Green function splining method: discrete Green functions obtained from QMC are complemented by (continuous) reference Green functions (lower panel); the difference is generally small and has vanishing second derivatives at the boundaries, i.e. is well approximated by a natural cubic spline (upper panel).*

represented by a natural spline. The Green function is obtained as a function of Matsubara frequencies by Fourier transforming the oversampled spline and adding the Matsubara-frequency model Green function to the result.

We note that the concept of the method described here of supplementing QMC data with high-frequency information derived from a weak-coupling approach is similar to Jarrell's QMC implementation [5]. The latter is, however, less stable since its Fourier transformation relies on a numerical IPT calculation. At least for a symmetric DOS at half filling, our method is unconditionally stable for arbitrarily large frequencies without the need for bandpass filters. This is illustrated in Fig. 11: The raw HF-QMC data (dashed lines) is indistinguishable from the (numerically) exact result (solid line) already for $|\omega_n| \gtrsim 5$ and converges quickly in the low-frequency region for $\Delta\tau \to 0$. In contrast, continuous-time methods achieve results without systematic bias, however with significant noise at large frequencies, especially in the hybridization variant.

Since each of the approaches discussed so far generates a number of Matsubara frequencies which is much greater than the number of time slices, the inverse Fourier transform (then for $\mathcal{G}$) is unproblematic: the rounding-off near $\tau = 0$ is already small for the first grid point $\tau = \Delta\tau$ while at the end points $\tau = 0$ and $\tau = \beta$, the numerical Fourier transforms can be shifted by the value $1/2$ to exactly cancel the rounding effect.

Ulmke's smoothing trick

A completely different approach was taken by Ulmke [44]. Here, the number of Matsubara frequencies is chosen equal to the number of time slices. In order to enforce the correct analytical

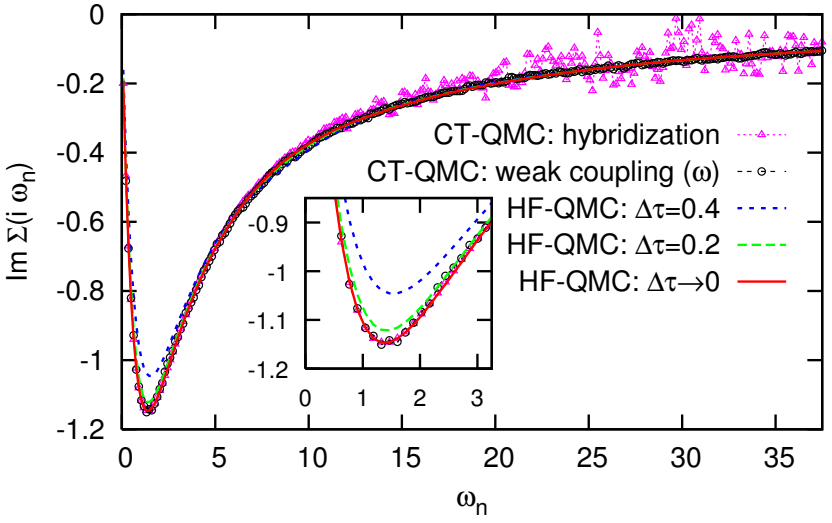


Fig. 11: *(Color online) Imaginary part of self-energy on the imaginary axis for* $T = 1/45$ *as estimated by CT-QMC [43] using the hybridization expansion (triangles) or the weak-coupling expansion (circles) and by Hirsch-Fye QMC using splines for difference Green functions (dashed lines for* $\Delta\tau = 0.4, \Delta\tau = 0.2$*, solid line for* $\Delta\tau \to 0$*).*

behavior of the Fourier transforms, a "smoothing" trick is employed: the naive Fourier transform is combined with an approximate correcting transformation which approaches the identity for $\Delta\tau \to 0$. While this method recovers the high-frequency behavior of $G(i\omega_n)$ and $\Sigma(i\omega_n)$ for $|\omega| \leq \pi/\Delta\tau$ about at the level of Krauth's code, it (unnecessarily) introduces errors at small Matsubara frequencies in insulating phases which may prevent the detection of coexistence regions. This artifact can be cured by introducing a frequency cutoff in the transformation [38]; however, the resulting method is inferior to advanced splining schemes and also less elegant.

3.2 Search for solutions, hysteresis, overrelaxation

An important point not discussed so far is the initialization of the self-consistency cycle in a HF-QMC run for some given set of physical parameters (e.g. U, T, μ) and discretization $\Delta\tau$. Since in most cases the results and also self-energy and Green function will continuously depend on the parameters, one will almost initialize the simulation with a self-energy previously obtained for a related parameter set. One possibility is a sequence of runs for increasing or decreasing values of a physical parameters such as the interaction U. In the vicinity of first-order phase transitions, which show up in the DMFT due to its mean-field character as coexistence regions, such hysteresis runs may be used for determining the phase boundaries. In principle, one could also explore hysteresis in the temperature domain; this is, however, difficult for frequency-based state variables (i.e. $\Sigma(i\omega_n)$) since the Matsubara frequencies depend on the temperature.
In principle much can be gained from hysteresis runs at constant physical parameters, in which the Trotter discretization $\Delta\tau$ is initially chosen quite large; due to the scaling of the effort with $\Delta\tau^{-3}$ convergency can then be obtained rather cheaply. Subsequent runs at intermediate values of $\Delta\tau$, each initialized from the previous run, will then require only few iterations (which are

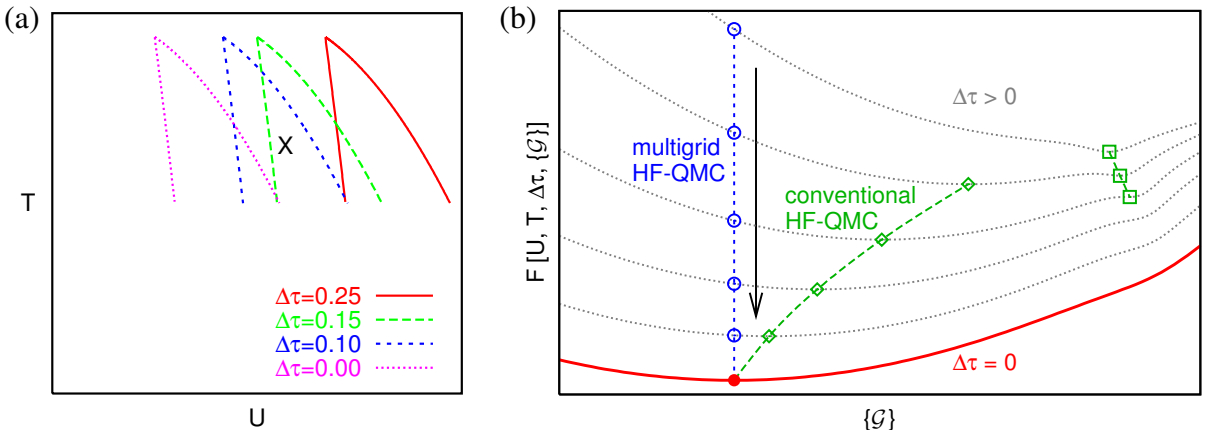


Fig. 12: *(a) Scenario for the effect of a $\Delta\tau \to 0$ extrapolation on a metal-insulator coexistence region, namely a shift to smaller U at roughly constant shape. (b) Schematic view of Ginzburg-Landau free energy in multidimensional space of hybridization functions $\{\mathcal{G}\}$. The fixed points of conventional HF-QMC (diamonds, squares) are adiabatically connected to the exact fixed point (full circle) only for small $\Delta\tau$. In contrast, the multigrid method [45] solves all impurity problems at the $\Delta\tau = 0$ fixed point (circles).*

still relatively cheap). For a sufficiently large number of steps, about 5 iterations will then suffice at the most costly target discretization. This strategy is, however, not without danger, especially near phase transitions: as the location of coexistence regions will depend on $\Delta\tau$ and may, e.g., shift towards smaller interactions with decreasing $\Delta\tau$ as illustrated in Fig. 12a, a phase that is stable for some physical parameter set (cross) at low values of $\Delta\tau$ may not be stable at high $\Delta\tau$ (and vice versa), so it cannot be obtained in such $\Delta\tau$ hysteresis run.

A more abstract view on this problem is given in Fig. 12b: each DMFT fixed point can be understood as the minimum of a generalized Ginzburg-Landau free energy functional in the (high-dimensional) space of hybridization functions (full line). In the presence of Trotter errors, this functional will in general shift and also change its shape, in particular the number of relative minima. Such phase transition as a function of $\Delta\tau$ clearly limit the possibilities for hysteresis schemes as discussed above and also for $\Delta\tau$ extrapolation to be discussed in the Sec. 4.

We should also stress that the convergence of a given iteration scheme (such as shown in Fig. 1) to its fixed point is far from trivial; e.g., a reversed execution of the same scheme would have the same fixed point, but the roles of stable and unstable fixed points would be reversed. [38] However, the rate of convergence for a stable iteration scheme can be improved by overrelaxation where the new solution is replaced by a linear combination with the old solution: $f_{\text{new}} \longrightarrow \alpha f_{\text{new}} + (1-\alpha) f_{\text{old}}$ with $\alpha > 0$ (while $0 < \alpha < 1$ in an under-relaxation scheme could be used for making a previously unstable scheme stable). While such strategies can save computer time they are to be used with care close to phase transitions.

4 Computing observables

So far, we have collected all essential concepts for achieving self-consistent DMFT solutions on the basis of the HF-QMC impurity solver; these solutions are characterized by the resulting self-energies $\Sigma(i\omega_n)$ or, equivalently, by the lattice and bath Green functions $G(i\omega_n)$ and $\mathcal{G}(i\omega_n)$, i.e., high-dimensional vectors in Matsubara space, or their Fourier transforms. Each of them depends on the physical parameters, on the Trotter discretization $\Delta\tau$, and, within coexistence regions, on the initialization.
We will now discuss the derivation of other properties from HF-QMC calculations, characterize the errors of such estimates and show how the systematic Trotter bias can be removed (or at least drastically reduced) by extrapolation.

4.1 Observables: mass renormalization, energetics, and spectra

The characterization of correlation properties and associated phase transitions is greatly simplified if one can find a scalar variable which plays (more or less) the role of an order parameter. In the cases of magnetic or orbital order, an order parameter is trivially obtained from the spin, orbital and/or sublattice resolved densities. The situation is more complicated at metal-insulator transitions, at least at finite temperatures. However, the quasiparticle weight Z to be discussed below often appears as a useful characteristic. In addition, we will explain how to compute the energetics (i.e., kinetic and interaction energy), which are obviously relevant for all thermodynamic properties, and illustrate the derivation of spectra.

Quasiparticle weight

One quantity closely related to the self-energy (which quantifies interaction effects) is the quasiparticle weight also known as mass renormalization factor Z,[17] which is defined in terms of the real-frequency self-energy,

$$Z = \frac{m}{m^*} = \frac{1}{1 - \frac{\partial}{\partial\omega}\mathrm{Re}\,\Sigma(\omega)\big|_{\omega=0}} \,. \tag{39}$$

In the context of QMC simulations, one usually approximates this quantity in the discrete form

$$Z \approx \frac{1}{1 - \frac{\mathrm{Im}\,\Sigma(i\omega_1)}{\pi T}} \,. \tag{40}$$

Due to the rules for complex derivatives (i.e., the Cauchy/Riemann equation $\partial\mathrm{Re}\, f(z)/\partial\mathrm{Re}\, z = \partial\mathrm{Im}\, f(z)/\partial\mathrm{Im}\, z$), both definitions agree in the limit $T \to 0$ as long as the Luttinger theorem is fulfilled, i.e., for $\mathrm{Im}\,\Sigma(\omega = 0) = 0$. For $U \to 0$, no mass renormalization takes place so that $Z = 1$. With increasing U, Z decreases until the quasiparticle peak (in the spectrum) vanishes (cf. Fig. 14). Extending the Fermi liquid picture, one usually associates the disappearance of the quasiparticle peak with $Z \approx 0$. Z in the definition (39), however, is not positive near a

[17] This should not be confused with the coordination number also denoted by Z.

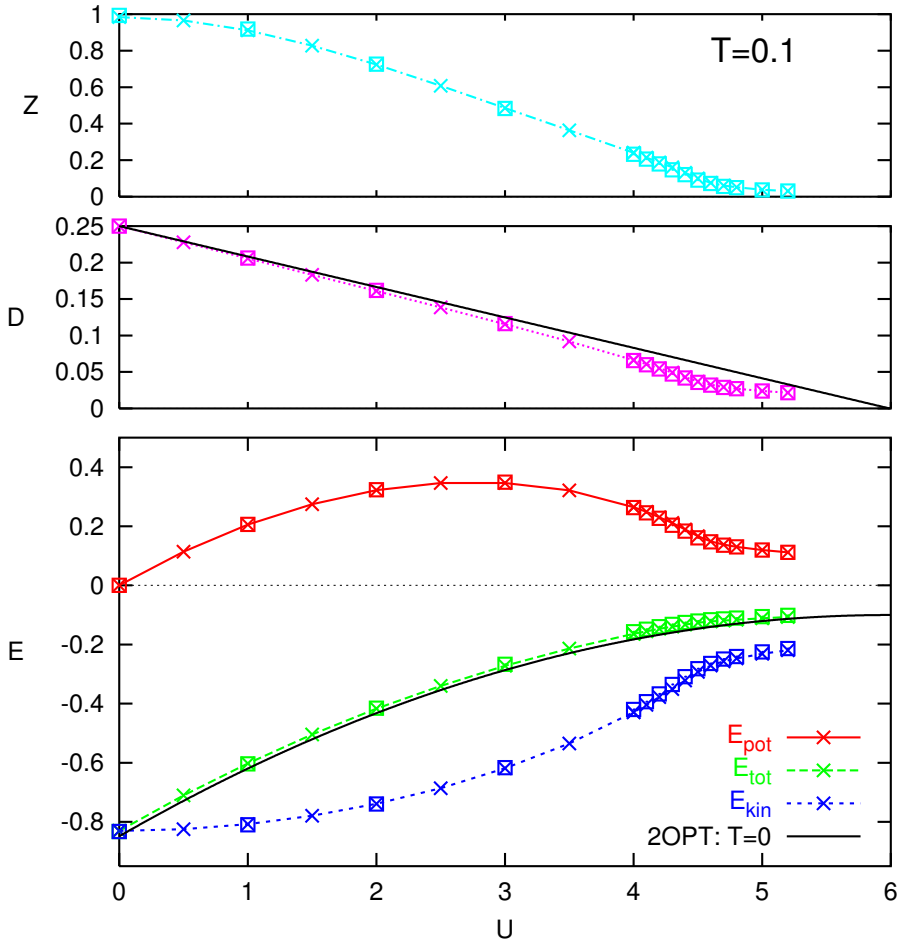


Fig. 13: *Quasiparticle weight Z (discrete estimate), double occupancy D, and energy contributions for the 1-band Hubbard model on the Bethe lattice (bandwidth $W = 4$) at $T = 0.1$ in the paramagnetic phase. Crosses (connected with lines) denote QMC results for $\Delta\tau = 0.2$, squares are for $\Delta\tau = 0.125$. For comparison, results of second order perturbation theory (2OPT) are shown for the total energy and the double occupancy for $T = 0$ (solid black lines).*

metal-insulator transition so that its interpretation as a quasiparticle weight breaks down. In contrast, the discrete version (40) always leads to positive Z and may therefore appear more physical. In any case, Z loses its theoretical foundation outside the Fermi liquid phase where it remains only a heuristic indicator of a metal-insulator transition. In the uppermost part of Fig. 13, Z is shown for the relatively high temperature $T = 0.1$. A rapid change of slopes indicates a transition or crossover near $U \approx 4.7$.

Energy

Within the DMFT, the energy per lattice site is given as [1, 8]

$$E = \lim_{\eta\to 0^+} T\sum_{n,\sigma}\int_{-\infty}^{\infty} \mathrm{d}\epsilon \frac{\mathrm{e}^{i\omega_n\eta}\,\epsilon\rho(\epsilon)}{i\omega_n - \epsilon - \Sigma_\sigma(i\omega_n)} + \frac{1}{2}T\sum_{n,\sigma}\Sigma_\sigma(i\omega_n)G_\sigma(i\omega_n)\,. \tag{41}$$

Note the convergence factor $\mathrm{e}^{i\omega_n\eta}$ which is essential in order to get the correct result. Obviously, such a term is difficult to handle numerically; in practice, η may be replaced, e.g., by the time discretization parameter $\Delta\tau$, which also determines the cutoff frequency in the infinite sum. This approximation can be avoided by evaluating the noninteracting part separately. For the kinetic energy, this implies

$$E_{\text{kin}} = \lim_{\eta\to 0^+} 2T \sum_{n=-\infty}^{\infty} \mathrm{e}^{i\omega_n\eta} \int_{-\infty}^{\infty} d\epsilon\, \epsilon\rho(\epsilon)\, \frac{1}{i\omega_n - \epsilon + \mu - \Sigma(i\omega_n)} \tag{42}$$

$$= 2\int_{-\infty}^{\infty} d\epsilon \frac{\epsilon\rho(\epsilon)}{\mathrm{e}^{\beta(\epsilon-\mu)}+1} + 2T \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} d\epsilon\, \epsilon\rho(\epsilon) \left(G_\epsilon(i\omega_n) - G^0_\epsilon(i\omega_n)\right) \tag{43}$$

$$\approx 2\int_{-\infty}^{\infty} d\epsilon \frac{\epsilon\rho(\epsilon)}{\mathrm{e}^{\beta(\epsilon-\mu)}+1} + 2T \sum_{n=-L/2+1}^{L/2} \int_{-\infty}^{\infty} d\epsilon\, \epsilon\rho(\epsilon) \left(G_\epsilon(i\omega_n) - G^0_\epsilon(i\omega_n)\right) \ , \tag{44}$$

where we have assumed the paramagnetic case. Here, the interacting and noninteracting "momentum-dependent" Green functions read

$$G_\epsilon(i\omega_n) = \frac{1}{i\omega_n - \epsilon + \mu - \Sigma(i\omega_n)}\,; \qquad G^0_\epsilon(i\omega_n) = \frac{1}{i\omega_n - \epsilon + \mu}\,. \tag{45}$$

In (44), the terms in the Matsubara sum fall off at least as $1/\omega^2$, which makes it well-defined also without convergence factor. At the same time, the truncation error is reduced significantly. The complementary ingredient to the energy is the double occupancy D with $E = E_{\text{kin}} + UD$. In the context of QMC calculations, this observable is best calculated directly from Wick's theorem (i.e. as $\langle n_{i\uparrow} n_{i\downarrow}\rangle$ or the corresponding expression in Grassmann variables) when sampling over the auxiliary field. The overall behavior of D and E in the Hubbard model can be read off (for $T = 0.1$) from the middle and lower parts of Fig. 13, respectively. For small U, the kinetic energy increases quadratically while D and, consequently, E_{pot} and E increase linearly. The potential energy reaches a maximum below $U = 3$. A region of strong curvature of D, E_{pot}, and E_{kin} near $U = 4.6$ gives a rough indication of the metal-insulator crossover. The total energy E, however, hardly shows any anomalies at this scale. Note also that the solutions for D and E are quite close to the results of plain zero-temperature second-order perturbation theory. The agreement actually becomes better for low-temperature QMC data, extrapolated to $T \to 0$ (not shown). The offset of the curves for E gives (for not too large U) an indication of the specific heat $c_V = \mathrm{d}E/\mathrm{d}T$ which is linear within the Fermi-liquid phase and is, in general, best evaluated by fitting the temperature dependence of E.[18]

Susceptibilities

The direct evaluation of the compressibility (as of most other susceptibilities) is numerically costly since it requires the QMC computation of 2-particle vertex functions. The formalism is omitted here; it can be found, e.g., in [1, 46].

[18] In Fermi liquid phases, the linear coefficient γ of the specific heat may also be obtained via the quasiparticle weight Z (extrapolated to $T = 0$).

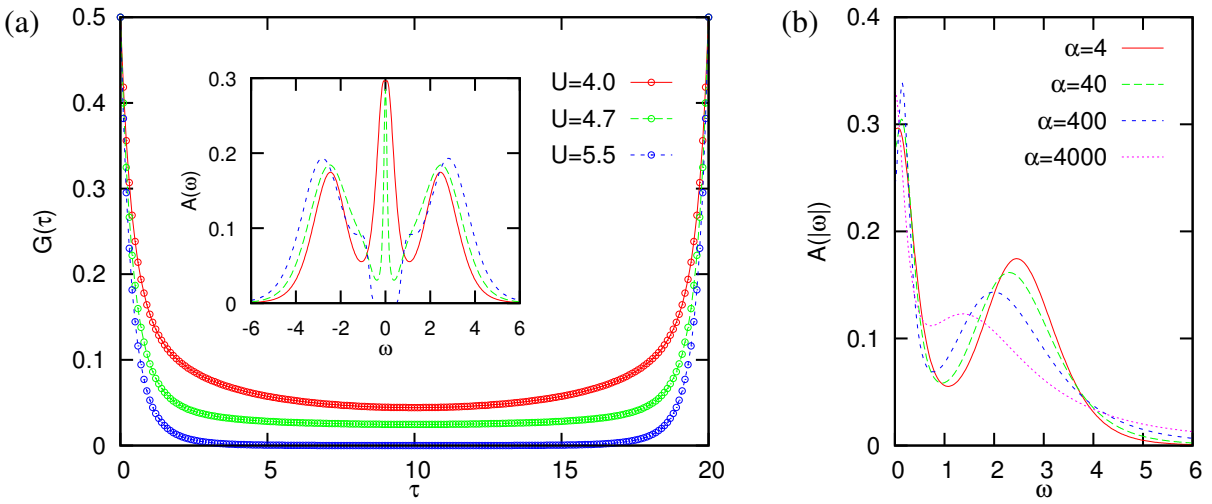


Fig. 14: *Example for analytic continuation of Green functions using the maximum entropy method (MEM). (a) While the Green functions (Bethe lattice, $T = 0.05$) are featureless at all interactions and differ mainly by their value and curvature near $\tau = \beta/2$, the spectra derived using MEM (inset) show clearly the narrowing and disappearance of the quasiparticle peak with increasing interaction U. (b) Impact of the weighing factor α on the MEM result for $U = 4.0$.*

Spectra

Spectra and optical conductivity data also shed light on systems near an MIT and are essential input for quantitative comparisons with experiments. Unfortunately, an analytic continuation is needed for imaginary-time based algorithms (such as the HF-QMC method) which is inherently unstable. In App. B, we discuss the maximum entropy method (MEM) which regularizes the procedure by a constraint on the smoothness of the spectrum (quantified by the entropy function). As seen in Fig. 14 the essential correlation physics is much more transparent on the level of the spectra than on that of the imaginary-time Green function. Thus, even significant efforts in MEM related questions appear well spent.

4.2 Estimation of errors and extrapolation $\Delta\tau \to 0$

As we have already mentioned, any "raw" estimate of an observable measured at some iteration using HF-QMC contains various sources of errors: (i) the statistical MC error, (ii) the convergency error associated with incomplete convergence of the self-consistency cycle, and (iii) the discretization error stemming both from the Trotter decoupling and the approximate Fourier transformations discussed in Sec. 2 and Sec. 3, respectively.

The first two errors can be treated at equal footing when we compute the observables at each iteration and extract averages from their "traces" (i.e. their functional dependence as a function of iteration number) as depicted in Fig. 15 a for the double occupancy D. We see that the measurements fluctuate significantly for each value of $\Delta\tau$; the amplitude of such fluctuations can be controlled by the number of sweeps. Also apparent is a gradient in the initial iterations for the largest discretization shown, $\Delta\tau = 0.20$. It is clear that such lead-in data should not

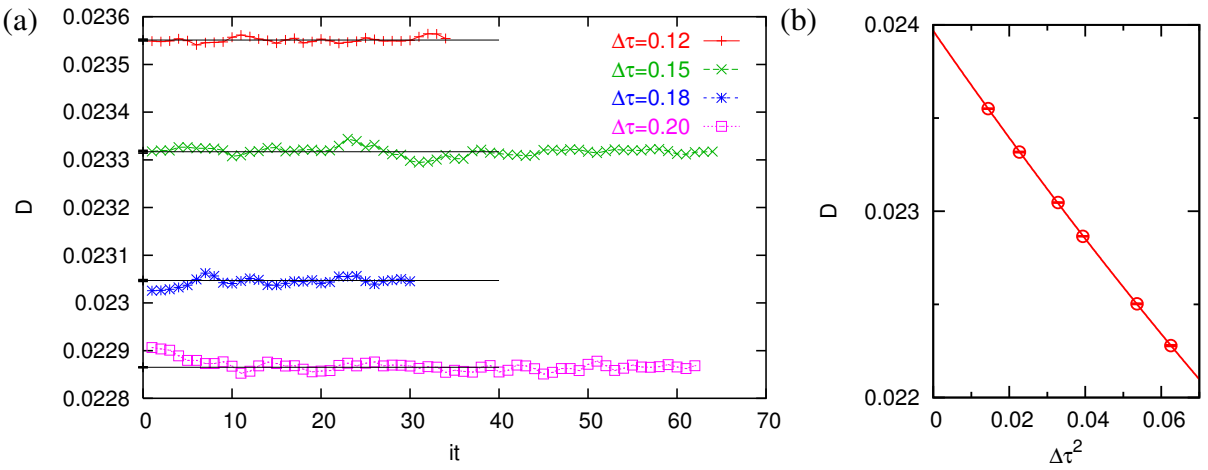


Fig. 15: *Steps in the HF-QMC based computation of the double occupancy at $U = 5$, $T = 0.04$: (a) estimates can be calculated as averages with statistical error bars from the analysis of time series (traces), taking autocorrelation into account, for each value of $\Delta\tau$. (b) Numerically exact results are obtained in a second step using extrapolation by least-squares fits.*

be included in averages. Moreover, all curves show significant autocorrelation which has to be taken into account for error analysis. Taking these issues into account, standard analysis techniques for time series yield raw HF-QMC results for each value of the discretization plus an error bar which takes statistical and convergency error into account. Such data is shown in Fig. 15 b as a function of the squared discretization. Evidently, the discretization dependence is very regular; a straightforward extrapolation using standard least-square fit methods (here with the 3 free parameters corresponding to the orders $\Delta\tau^0$, $\Delta\tau^2$, and $\Delta\tau^4$) essentially eliminates this systematic error, i.e. reduces the total error by two orders of magnitude.[19]

Due to the cubic scaling of the effort with the number of time slices, the total cost of achieving extrapolated results for some grid range is dominated by the smallest discretization. Thus the possibility of extrapolation comes (with the use also of coarser grids) at no significant cost; in contrast, DMFT convergence can be accelerated by using the $\Delta\tau$ hysteresis technique outlined above. Taking all of this into account, HF-QMC with extrapolation can be competitive with the recently developed continuous-time QMC solvers, as demonstrated in Fig. 16: at fixed total computer time, this method achieves the highest precision [47] in energy estimates for a test case established in [43].

Note that the use of insufficiently converged solutions is potentially a very significant source of errors. It is important to realize that in principle measurements have to be performed exactly at the solution of the self-consistency equations, i.e., for the exact bath Green function. Averages over measurements performed for different impurity models corresponding to approximate solutions do not necessarily converge to the exact answer in the limit of an infinite number of models (i.e., iterations) and measurements.[20] Still, the most important practical point when

[19] The reader may note a similarity to the MC example shown in Fig. 3.

[20] Trivially, a measurement of the free energy F itself (using a suitable impurity solver) is a good example. Since F is minimal for the true solution, all measurements taken for approximate solutions will be too large. The correct answer can, therefore, not be approached by averaging over many measurements, but only by reducing the

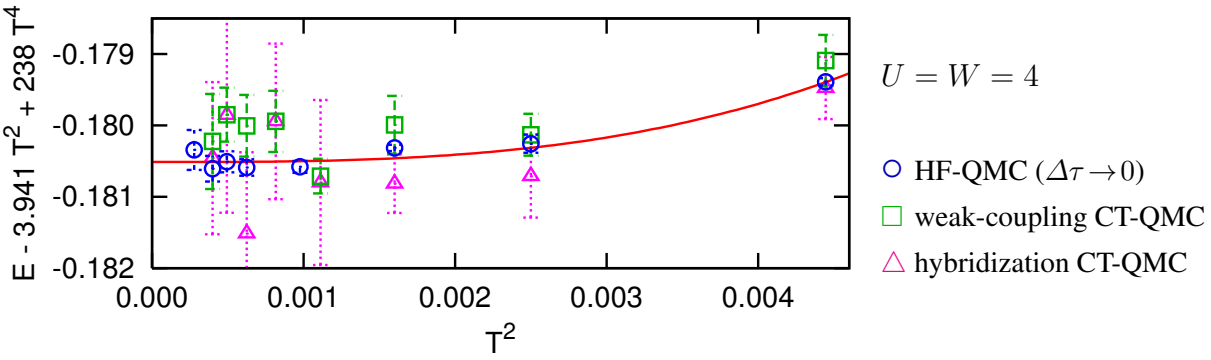


Fig. 16: *Comparison of energy estimates of the half-filled Hubbard model obtained within DMFT using different QMC based impurity solvers at fixed computational effort, with (approximate) leading terms subtracted. HF-QMC is competitive, here even more efficient, (only) after careful extrapolation $\Delta\tau \to 0$ [47].*

computing observables is that only runs are included in averages which are close to the solution in comparison to the asymptotic statistical error. For higher precision, the number of sweeps (and not only the number of measurements) must be increased.

Let us, finally, stress that discretization errors can vary greatly between different observables and, due to the DMFT self-consistency, also between points in parameter space. This is clearly seen in Fig. 17: for the kinetic energy, the quadratic contribution to the Trotter error is significant only in the strongly correlated metallic phase; it is negligible in the insulator! This shows that rules for acceptable values of $\Delta\tau$ sometimes mentioned in the literature (e.g. $\Delta\tau U \lesssim 1$) are of limited value; in fact, they cannot follow from the Trotter decomposition since this becomes exact both in the limits of weak and strong coupling. In the case of the double occupancy, the quadratic term is surprisingly small close to the phase transition (and has a maximum in the absolute value at $U \approx 3.5$). However, in both cases the quartic terms are significant and irregular close to the phase transition which will in practice limit the reliability of extrapolations. Interestingly, some of the errors cancel in the total energy which makes numerically exact estimates of this observables particularly accessible for HF-QMC.

5 Conclusion and outlook

Hopefully, these lecture notes give a good first overview of the most important generic aspects of Hirsch-Fye QMC based method for calculating electronic properties within the dynamical mean-field theory. However, they are certainly not complete in any respect: important topics such as the minus-sign problem occurring in general multi-band and cluster DMFT calculations could only be touched upon; also, important literature is missing.

The topics that we would have liked to discuss in an (yet unwritten) section on advanced aspects also include recent methodological developments such as the multigrid HF-QMC approach [45, 48] and our HF-QMC based implementation of real-space dynamical mean-field

deviations from the exact solution.

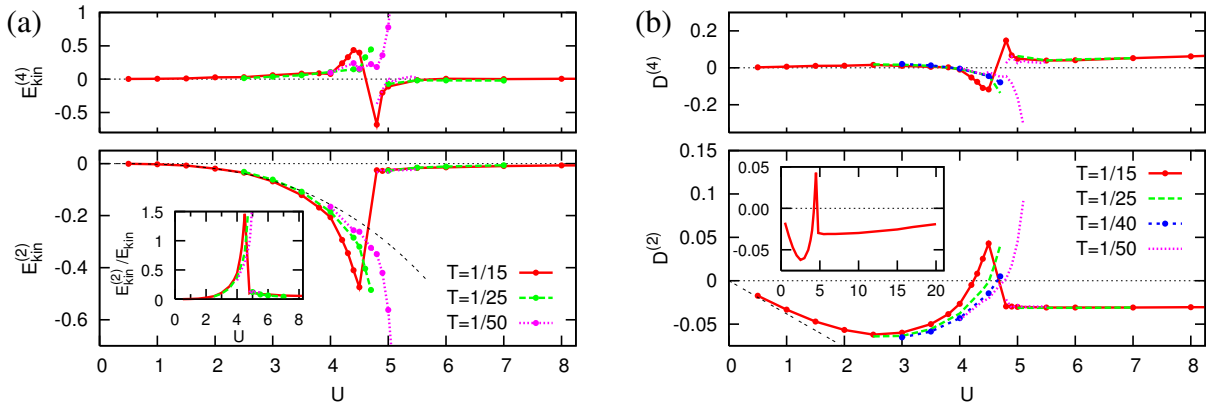


Fig. 17: *(a) Coefficients of Trotter errors in HF-QMC estimates of kinetic energy. Inset: relative coefficients. (b) Trotter coefficients for double occupancy [47].*

theory (RDMFT) [49, 50] as well as their applications, e.g., in the context of cold atoms on optical lattices. It would also be tempting and potentially fruitful to explore the similarities between the HF-QMC approach and the determinantal QMC method [33] for Hubbard model studies in finite dimensions, in particular its recent application as DMFT impurity solver [51]. However, one has to stop at some point and the material presented so far is probably already overwhelming for students and scientists just entering the field.

This audience should be aware of the fact that the future relevance of the HF-QMC method is controversial; influential groups claim that it has been superseded by the continuous-time QMC methods. These methods are, indeed, conceptually elegant and avoid the systematic discretization bias which can be cumbersome in conventional HF-QMC calculations. We have shown, however, that the HF-QMC method is competitive at least for some classes of problems already in the conventional form [47]; the quasi-continuous time formulation in the multigrid approach eliminates most of the remaining problems. Consequently it seems too early to abandon the general HF-QMC concept.

Acknowledgment

The author thanks E. Gorelik and D. Rost for valuable help in preparing these lecture notes. Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Multi-band Hubbard models

The applicability of the one-band Hubbard model to d or f electron systems is a priori questionable since the partially filled bands correspond to atomic orbitals which are 5-fold and 7-fold degenerate (for each spin direction), respectively. While bands in a lattice are more complicated than orbitals of isolated atoms, the remaining degeneracy can be inferred from symmetry considerations alone. Often it is useful to consider a cubic representation of the angular part of atomic d orbitals,

$$\begin{aligned} &|d_{xy}\rangle \propto (|2,2\rangle - |2,-2\rangle)\,, \quad |d_{yz}\rangle \propto (|2,1\rangle + |2,-1\rangle)\,, \quad |d_{zx}\rangle \propto (|2,1\rangle - |2,-1\rangle) \\ &|d_{x^2-y^2}\rangle \propto (|2,2\rangle + |2,-2\rangle)\,, \quad |d_{3z^2-r^2}\rangle \propto |2,0\rangle, \end{aligned} \tag{46}$$

expressed in terms of eigenfunctions of the angular momentum operator,

$$\boldsymbol{l}^2|l,m\rangle = \hbar^2 l(l+1)|l,m\rangle, \quad l_z|l,m\rangle = \hbar m|l,m\rangle\,. \tag{47}$$

In lattices with cubic symmetry the five d orbitals are energetically split into the t_{2g} orbitals ($|d_{xy}\rangle$, $|d_{yz}\rangle$, $|d_{zx}\rangle$) and the e_g orbitals ($|d_{x^2-y^2}\rangle$, $|d_{3z^2-r^2}\rangle$), which give rise to one threefold degenerate and one twofold degenerate band, respectively. Lower symmetry can lift the remaining degeneracies; e.g., in the trigonal case the t_{2g} orbitals are further split into one nondegenerate a_{1g} and one twofold degenerate e_g^π band. Thus, it is possible that in some d systems only one band crosses or touches the Fermi surface which then justifies the one-band assumption made in (1) and used for the examples in this lecture. In general, however, the inclusion of several orbitals per site is important. An SU(2)-invariant generalization of the Hubbard model where the interaction is still local but the valence band is degenerate then contains additional coupling terms[21] [53, 54]

$$\begin{aligned} \hat{H}_{m\text{-band}} \;=\; & -t \sum_{\langle ij\rangle,\nu\sigma} \hat{c}^\dagger_{i\nu\sigma}\hat{c}_{j\nu\sigma} + U\sum_{i\nu}\hat{n}_{i\nu\uparrow}\hat{n}_{i\nu\downarrow} \\ & + U' \sum_{i;\nu<\nu';\sigma\sigma'} \hat{n}_{i\nu\sigma}\hat{n}_{i\nu'\sigma'} + J \sum_{i;\nu<\nu';\sigma\sigma'} \hat{c}^\dagger_{i\nu\sigma}\hat{c}^\dagger_{i\nu'\sigma'}\hat{c}_{i\nu\sigma'}\hat{c}_{i\nu'\sigma}, \end{aligned} \tag{48}$$

where ν,ν' (with $1 \le \nu \le m$, $1 \le \nu' \le m$) are band indices. The exchange term parameterized by the Hund's rule coupling J can be rewritten as

$$\hat{H}_J = -2J \sum_{i,\nu<\nu'} \left(\hat{\boldsymbol{S}}_{i\nu} \cdot \hat{\boldsymbol{S}}_{i\nu'} + \frac{1}{4}\hat{n}_{i\nu}\hat{n}_{i\nu'} \right) \tag{49}$$

with $\hat{\boldsymbol{S}}_{i\nu} = \frac{1}{2}\sum_{\sigma\sigma'} \hat{c}^\dagger_{i\nu\sigma}\boldsymbol{\tau}_{\sigma\sigma'}\hat{c}_{i\nu\sigma'}$being the spin operator for orbital ν at site i and the Pauli matrices $\boldsymbol{\tau}_{\sigma\sigma'}$. In quantum Monte Carlo (QMC) simulations, the spin-flipping terms implicit in (49)

[21] Here, an on-site pair hopping term which only contributes when one orbital is doubly and another singly occupied is neglected [52].

lead to a numeric (minus-sign) problem. Therefore, one here usually replaces the Heisenberg interaction part of (49) by an Ising-type interaction at the cost of breaking the SU(2) symmetry [55, 52]. Since $\hat{s}^z_{i\nu}\hat{s}^z_{i\nu'} = \sigma\sigma'\hat{n}_{i\nu\sigma}\hat{n}_{i\nu'\sigma'}/4 = (2\delta_{\sigma\sigma'} - 1)\,\hat{n}_{i\nu\sigma}\hat{n}_{i\nu'\sigma'}/4$, one can write this modified multi-band Hubbard Hamiltonian as

$$\hat{H}^z_{m\text{-band}} = \sum_{\langle ij\rangle,\nu\sigma} \epsilon_{\boldsymbol{k}\nu}\hat{c}^\dagger_{\boldsymbol{k}\nu}\hat{c}_{\boldsymbol{k}\nu} + U\sum_{i\nu}\hat{n}_{i\nu\uparrow}\hat{n}_{i\nu\downarrow} + \sum_{i;\nu<\nu';\sigma\sigma'}(U' - \delta_{\sigma\sigma'}J)\,\hat{n}_{i\nu\sigma}\hat{n}_{i\nu'\sigma'}. \tag{50}$$

The interaction U between electrons within each orbital is always larger than the interorbital density-density interaction U'. The smaller exchange coupling J can trigger ferromagnetic and (possibly coexisting) orbital order.

A HF-QMC simulation for this type of models proceeds in principle as in the single-band case; however, one auxiliary field is needed for each of the $m(2m-1)$ pair interactions.

B Maximum entropy method

Within the DMFT, all single-particle properties can be expressed in terms of the local single-particle spectral function (also called "full" or "interacting" density of states) which is proportional to the imaginary part of the local retarded Green function,

$$A(\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,G(\omega + i0^+)\,. \tag{51}$$

The spectral function $A(\omega)$ is accessible experimentally: measured (angular averaged) photoemission spectra (PES) can under certain simplifying assumptions be identified with $A(\omega)$ multiplied by the Fermi function $n_{\mathrm{f}}(\omega-\mu)$. Correspondingly, inverse photoemission spectra or X-ray absorption spectra (XAS) can be identified with $A(\omega)$ multiplied by $1 - n_{\mathrm{f}}(\omega-\mu)$. Furthermore, nonlocal spectral functions and the optical conductivity $\sigma(\omega)$ can, within the DMFT, be calculated from $A(\omega)$.

In QMC calculations, however, the Green function G (and thus the spectral function $A(\omega)$) cannot be directly computed on the real axis. Instead, real-time dynamical information has to be extracted from imaginary-time data $G(\tau)$ (or, equivalently, from the Fourier transformed Matsubara-frequency data $G(i\omega_n)$) via analytic continuation. This is in principle possible through inversion of the spectral representation for $G(\tau)$, i.e.,

$$G(\tau) = \int_{-\infty}^{\infty}\mathrm{d}\omega\; K(\tau,\omega-\mu;\beta)\;A(\omega)\,, \tag{52}$$

$$K(\tau,\omega;\beta) := \frac{\exp(-\tau\omega)}{1+\exp(-\beta\omega)}\,, \tag{53}$$

but poses an ill-conditioned problem since $G(\tau)$ is only measured on a grid $\tau_l = l\Delta\tau$ (where $\Delta\tau = \beta/\Lambda$) and since the kernel (53) becomes exponentially small for generic values of τ at large absolute frequencies $|\omega|$ as illustrated in Fig. 18. For a symmetric problem, i.e., symmetric noninteracting DOS and $n = 1$, the integral in (52) can be restricted to positive frequencies when the symmetrized fermion kernel

$$K_s(\tau,\omega;\beta) = K(\tau,\omega;\beta) + K(\tau,-\omega;\beta) \tag{54}$$

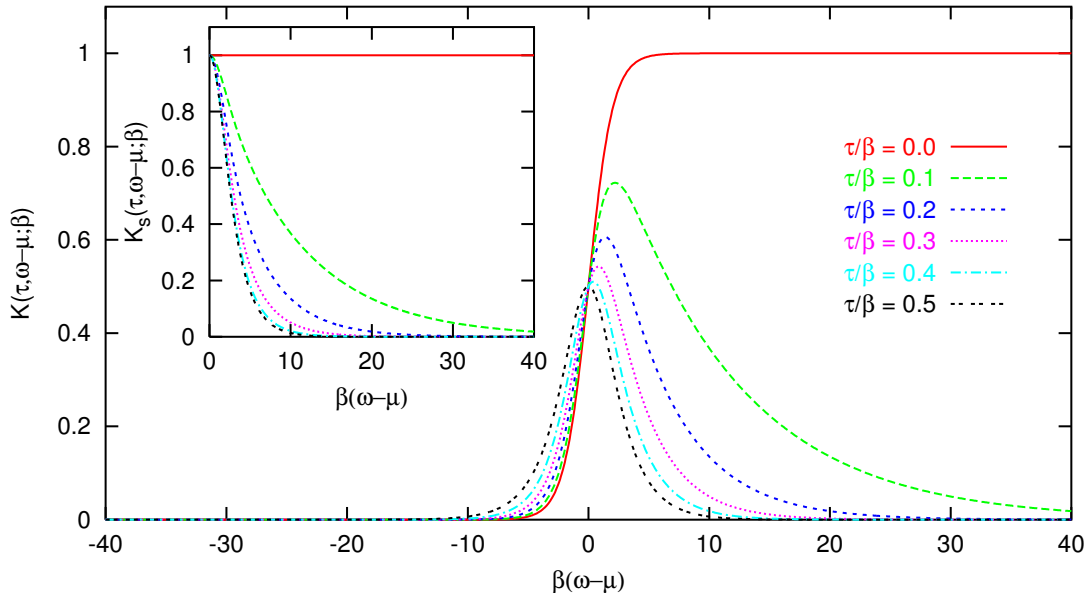


Fig. 18: *General fermion kernel (53) for an equidistant set of values of $\tau/\beta \leq 1/2$. Except for $\tau = 0$, large frequencies are suppressed exponentially. Inset: symmetrized fermion kernel (54).*

shown in the inset of Fig. 18 is used. We stress that a very small number $\Lambda = \mathcal{O}(10)$ of time slices poses a more serious limitation for obtaining reliable spectra with good resolution than the exponential nature of the kernel since the number of degrees of freedom which can be reliably resolved in a spectrum is obviously much smaller than the number of data points $\{G(l\Delta\tau)\}$ and since there always exists an infinite number of spectra which correspond to the same data. Still, in general, the resolution is much better at smaller frequencies and, according to Nyquist's theorem, essentially no information can be obtained from QMC for $\omega > \Lambda\pi/\beta = \pi/(\Delta\tau)$. Before we address the full analytic continuation problem and introduce the maximum entropy method, we collect some useful relations (denoting the m^{th} derivative as $G^{(m)}$),

$$G(\beta) = n, \qquad G(0^+) = 1 - n, \tag{55}$$

$$G^{(m)}(0) + G^{(m)}(\beta) = (-1)^m \langle(\omega-\mu)^m\rangle_{A(\omega)} \tag{56}$$

$$G(\beta/2) \approx \frac{\pi}{\beta}\,\overline{A(\omega)}\big|_{|\omega-\mu|\lesssim\pi/\beta}\,. \tag{57}$$

Since the filling given by (55) is known in the symmetric case, the value of $G(0)$ then provides no useful information which is also seen in the inset of Fig. 18. Equation (56) also shows the loss of high-frequency information from the discretization of imaginary time: for a finite grid the error in estimating derivatives $G^{(m)}$ increases rapidly with order m; thus, the determination of high order moments $\langle(\omega-\mu)^m\rangle_{A(\omega)}$ of the spectrum is in general an ill-posed problem. At low temperatures, $G(\beta/2)$ gives a hint as to the weight of a quasiparticle peak or the existence of a gap via (57) since its value is proportional to the value of the spectral function near the Fermi energy, averaged over an inverse hyperbolic cosine with width π/β.

First attempts to address the analytic continuation problem for QMC data included least-squares fits, Padé approximants, and regularization (for references, see the pedagogical and concise

review by Jarrell [56]). Least-squares fits of spectra approximated as a set of box functions are inherently unstable. Padé approximations for $G(i\omega_n)$ only work well for very precise data (e.g., in the context of Eliashberg equations), but not for QMC. Regularization of the kernel (53) tends to produce overly smeared-out spectra. What is needed instead is a regularization of the solution $A(\omega)$ that only shows features which are supported by the data, but is as smooth as possible otherwise. This is essentially the idea of the maximum entropy method (MEM) of finding the most probable spectrum compatible with the data.

The MEM is a very general approach for reconstructing continuous, positive semidefinite functions (i.e., densities or spectra) from incomplete, noisy, and possibly oversampled[22] data. It has a long history in the context of image reconstruction in such diverse disciplines as radio aperture synthesis, optical deconvolution, X-ray imaging, structural molecular biology, and medical tomography (see [58] and references therein). A preliminary study of its usefulness in the context of the analytic continuation problem [59] was soon followed by applications to spectra [60, 61] and correlation functions [62]. An efficient and general formulation of a MEM algorithm for analytic continuation of (oversampled) QMC data was then given by Gubernatis [63], closely following the general state-of-the-art approach by Bryan [57].

For simplicity, we will first assume that the QMC simulation produces N_d measurements of the observables $G_l \equiv G(l\Delta\tau)$, $0 \leq l < \Lambda$ which are independent both in computer time and in imaginary time, i.e., without significant autocorrelation between subsequent measurements or between adjacent imaginary time slices.[23] If we further disregard systematic errors and assume some "true" spectrum A [which corresponds to a discretized Green function G_l via (52)] as well as a Gaussian distribution of statistical errors, the probability distribution for the observed averaged values $\bar{G} \equiv \{\bar{G}_l\}$ (where $\bar{G}_l = \sum_{i=1}^{N_d} G_l^i$) is

$$P(\bar{G}|A) \propto \mathrm{e}^{-\frac{1}{2}\chi^2}; \quad \chi^2 = \sum_{l=0}^{\Lambda-1} \frac{\left(\bar{G}_l - G_l\right)^2}{\sigma_l^2} \,. \tag{58}$$

Here, σ_l can be estimated from the data alone, $\sigma_l^2 \approx \sum_{i=1}^{N_d} \left(\bar{G}_l - \bar{G}_l^i\right)^2 / (N_d(N_d-1))$. Obviously, the likelihood function $P(\bar{G}|A)$ is not directly helpful; in order to find the most probable spectrum given the measured data, we need the converse probability $P(A|\bar{G})$. These probabilities are related by Bayes' theorem,

$$P(A|\bar{G})\, P(\bar{G}) = P(\bar{G}|A)\, P(A) \,. \tag{59}$$

Since the data $\bar{G}$ is constant in the search for an optimal A, the associated probability $P(\bar{G})$ drops out of the problem. For the probability $P(A)$ of a spectrum $A(\omega)$ in absence of data, an entropic ansatz is made where prior knowledge can also be incorporated by choosing an

[22] Data is called *oversampled* when measurements of different data points (here: for different τ) are not statistically independent, but correlated [57]. In this case, the number of "good" degrees of freedom (entering the error statistics) is reduced.

[23] Typically, one of the "measurements" of a set $\{G_l\}$ is generated by binning a macroscopic number of measurements within QMC. Autocorrelation in computer time, i.e., between different bins vanishes in the limit of infinite bin size.

appropriate (positive semidefinite) default model $m(\omega)$,

$$P(A) = \mathrm{e}^{\alpha S[A(\omega),m(\omega)]}\,. \tag{60}$$

Here, α is a numerical parameter while S is a generalized Shannon-Jaynes entropy,

$$S[A,m] = \int \mathrm{d}\omega \Big(A(\omega) - m(\omega) - A(\omega) \ln \big(A(\omega)/m(\omega) \big) \Big)\,. \tag{61}$$

For a constant default model (within some finite frequency range), the entropic form (61) clearly favors smooth spectra. This is also true for a general smooth default model. It also enforces positivity of A and pushes the solution towards the (normalized) default model in absence of data. From (58), (59), and (60), the posterior probability can be read off as

$$P(A|\bar{G},m,\alpha) = \mathrm{e}^{\alpha S[A,m] - \frac{1}{2}\chi^2[\bar{G},A]}\,. \tag{62}$$

The balance between a tight match of data and a high entropy is calibrated by the Lagrange parameter α which may be chosen so that $\chi^2 = \Lambda$ (historic MEM). Alternatively, one may use the value of α with the highest probability $P(\alpha|\bar{G},A,m)$ which can approximately be calculated within the method (classic MEM). Given the QMC data, a default model, a representation of the spectrum (i.e., a possibly inhomogeneous grid of ω_j of frequencies for which A is going to be computed), and a starting guess for α, a simple MEM program thus both searches for the spectrum $\{A(\omega_j)\}$ with maximum probability $P(A|\bar{G},m,\alpha)$ for given α using, e.g., the Newton-Raphson method and, in an outer loop, searches for the best value of α.
The former procedure can be stabilized using a singular value decomposition (SVD) of the kernel:

$$K = V \Sigma U^T \tag{63}$$

$$\Sigma = diag(\sigma_1, ..., \sigma_s)$$

$$\text{where } \sigma_1 > ... > \sigma_s > 0.$$

Here U, V are orthogonal matrices. Typically, most of the singular values σ_i are equal to zero (within machine precision). The columns of U^T, restricted to $\sigma_i \neq 0$, then span the same space as the columns of K. Projecting all related quantities to this new most smaller space (so-called singular space) a stable search can be performed. In practice, the width of the Newton-Raphson steps has to be restricted using a Levenberg-Marquardt-Parameter. An example for the application of this method is shown in Fig. 14.
Alternatively to this deterministic approach, one may also use Markov chain Monte Carlo updates directly in the image space (i.e. on the level of the spectrum) with simulated annealing, as implemented, e.g., by Sandvik. This stochastic approach is however, quite sensitive to the choice of the frequency discretization, requires smoothing runs, and is computationally more costly.

References

[1] A. Georges, G. Kotliar, W. Krauth, and M. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[2] J.E. Hirsch and R.M. Fye, Phys. Rev. Lett. **56**, 2521 (1986)

[3] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)

[4] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989)

[5] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992)

[6] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)

[7] V. Janiš and D. Vollhardt, Int. J. Mod. Phys. B **6**, 731 (1992)

[8] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971)

[9] F.A. Berezin, *The method of second quantization* (Academic Press, New York, 1966)

[10] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)

[11] H. Keiter and J.C. Kimball, Phys. Rev. Lett. **25**, 672 (1970)

[12] N.E. Bickers, D.L. Cox, and J.W. Wilkins, Phys. Rev. B **36**, 2036 (1987)

[13] T. Pruschke and N. Grewe, Z. Phys. B **74**, 439 (1989)

[14] T. Pruschke, D.L. Cox, and M. Jarrell, Phys. Rev. B **47**, 3553 (1993)

[15] N.E. Bickers, D.J. Scalapino, and S.R. White, Phys. Rev. Lett. **62**, 961 (1989)

[16] N.E. Bickers and D.J. Scalapino, Ann. Physik **193**, 206 (1989)

[17] N.E. Bickers and S.R. White, Phys. Rev. B **43**, 8044 (1991)

[18] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)

[19] H.R. Krishnamurthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1003 (1980)

[20] T.A. Costi, A.C. Hewson, and V. Zlatić, J. Phys. Cond. Matter **6**, 2519 (1994)

[21] R. Bulla, T.A. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

[22] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)

[23] E. Gull, P. Werner, O. Parcollet, and M. Troyer, Europhys. Lett. **82**, 57003 (2008)

[24] M.J. Rozenberg, X.Y. Zhang, and G. Kotliar, Phys. Rev. Lett. **69**, 1236 (1992)

[25] A. Georges and W. Krauth, Phys. Rev. Lett. **69**, 1240 (1992)

[26] M. Ulmke, V. Janiš, and D. Vollhardt, Phys. Rev. B **51**, 10411 (1995)

[27] G. Kotliar, S. Savrasov, K. Haule, V. Oudovenko, O. Parcollet, and C. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[28] H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959)

[29] M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976)

[30] J.E. Hirsch, Phys. Rev. B **28**, 4059 (1983)

[31] K. Held, *Untersuchung korrelierter Elektronensysteme im Rahmen der Dynamischen Molekularfeldtheorie* (Ph.D. thesis, Universität Augsburg, 1999)

[32] W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, New York, 1987)

[33] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D **24**, 2278 (1981)

[34] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953)

[35] P.K.V.V. Nukala, T.A. Maier, M.S. Summers, G. Alvarez, and T.C. Schulthess, Phys. Rev. B **80** (2009)

[36] E. Gull, P. Staar, S. Fuchs, P.K.V.V. Nukala, M.S. Summers, T. Pruschke, T.C. Schulthess, and T.A. Maier, Phys. Rev. B **83** (2011)

[37] X.Y. Zhang, M.J. Rozenberg, and G. Kotliar, Phys. Rev. Lett. **70**, 1666 (1993)

[38] N. Blümer, *Mott-Hubbard Metal-Insulator Transition and Optical Conductivity in High Dimensions* (Ph.D. thesis, Universität Augsburg, 2003)

[39] J.J. Deisz, D.W. Hess, and J.W. Serene, in *Recent Progress In Many Body Theories*, vol. 4 (Plenum, New York, 1995)

[40] C. Knecht, *Numerical and analytical approaches to strongly correlated electron systems* (Ph.D. thesis, Universität Mainz, 2006)

[41] C. Knecht, N. Blümer, and P. van Dongen, Phys. Rev. B **72**, 081103 (2005)

[42] M. Potthoff, T. Wegner, and W. Nolting, Phys. Rev. B **55**, 16132 (1997)

[43] E. Gull, P. Werner, A.J. Millis, and M. Troyer, Phys. Rev. B **76** (2007)

[44] M. Ulmke, *Phasenübergänge in stark korrelierte Elektronensystemen* (Ph.D. thesis, KFA Jülich, 1995)

[45] N. Blümer, arXiv (2008)

[46] N. Blümer, *Quanten-Monte-Carlo-Untersuchungen von magnetischen Phasen im Hubbard-Modell mit Nächst-Nachbar-Wechselwirkungen* (Diploma thesis, RWTH Aachen, 1996)

[47] N. Blümer, Phys. Rev. B **76**, 205120 (2007)

[48] E. Gorelik and N. Blümer, Phys. Rev. A **80**, 051602 (2009)

[49] E. Gorelik, I. Titvinidze, W. Hofstetter, M. Snoek, and N. Blümer, Phys. Rev. Lett. **105**, 065301 (2010)

[50] N. Blümer and E.V. Gorelik, Comp. Phys. Commun. **182**, 115 (2011)

[51] E. Khatami, C.R. Lee, Z.J. Bai, R.T. Scalettar, and M. Jarrell, Phys. Rev. E **81** (2010)

[52] K. Held and D. Vollhardt, Eur. Phys. J. B **5**, 473 (1998)

[53] P.W. Anderson, Phys. Rev. **124**, 41 (1961)

[54] L. Dworin and A. Narath, Phys. Rev. Lett. **25**, 1287 (1970)

[55] Y. Motome and M. Imada, J. Phys. Soc. Jap. **66**, 1872 (1997)

[56] M. Jarrell, in D.J. Scalapino (eds.), *Numerical Methods for Lattice Quantum Many-Body Problems* (Addison Wesley, Reading, 1997)

[57] R.K. Bryan, Eur. Biophys. J **18**, 165 (1990)

[58] J. Skilling and R.K. Bryan, Mon. Not. R. Astr. Soc. **211**, 111 (1984)

[59] R.N. Silver, D.S. Sivia, and J.E. Gubernatis, Phys. Rev. B **41**, 2380 (1990)

[60] R.N. Silver, J.E. Gubernatis, D.S. Sivia, and M. Jarrell, Phys. Rev. Lett. **65**, 496 (1990)

[61] M. Jarrell, D.S. Sivia, and B. Patton, Phys. Rev. B **42**, 4804 (1990)

[62] J. Deisz, M. Jarrell, and D.L. Cox, Phys. Rev. B **42**, 4869 (1990)

[63] J.E. Gubernatis, M. Jarrell, R.N. Silver, and D.S. Sivia, Phys. Rev. B **44**, 6011 (1991)