The Dynamical Cluster Approximation with Quantum Monte Carlo Cluster Solvers *

M. Jarrell,† A. Macridin, K. Mikelsons, and D.G.S.P. Doluweera

Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA

J.E. Gubernatis

Theoretical Division, Los Alamos National Laboratory, USA

Abstract

We present a pedagogical discussions of the dynamical mean field (DMFA) and dynamical cluster (DCA) approximations and associated Monte Carlo and entropy-based methods of Bayesian data analysis. The DMFA and DCA methods are developed as coarse-graining approximations and the relationship between the cluster and lattice problems are detailed. The Hirsch-Fye and continuous time Quantum Monte Carlo (QMC) algorithms are used to solve the cluster problem. The algorithms are discussed, together with methods for efficient measurements and the modifications required by the self-consistency of the DMFA/DCA. Then, several principles of Bayesian data analysis are presented. When coupled with information theory, this analysis produces a precise and systematic way to analytically continue Matsubara-time QMC results to real frequencies. We show how to use Bayesian inference to qualify the solution of the continuation and optimize the inputs. Besides developing the Bayesian formalism, we also present a detailed description of the data qualification, sketch an efficient algorithm to solve for the optimal spectra, give cautionary notes where appropriate, and present two detailed case studies to demonstrate the method.

PACS numbers: 02.60.Pn, 71.27.+a, 89.70.Cf, 02.70.Hm, 02.70.Ss, 02.70.Uu, 07.05.P

* This article will appear in the **"Lectures on the Physics of Strongly Correlated Systems XII"**, AIP Conference Proc., Eds. A. Avella and F. Mancini

Contents

I. INTRODUCTION

Some of the most exotic properties of materials, including high-temperature superconductivity, magnetism, and heavy Fermion and non-Fermi liquid behaviors, are due to strong electronic correlations. The materials which display these properties are characterized by either narrow electronic bands or compact orbitals with large angular momentum in the valence shell. In either case, the potential energy associated with some of these electronic degrees of freedom is of similar magnitude or larger than their electronic kinetic energy (bandwidth), which invalidates conventional perturbative approaches. Thus, we resort to the construction of simplified models to study these systems.

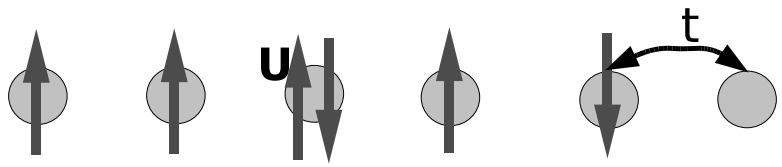

FIG. 1: Cartoon of the Hubbard model, characterized by a single band with near-neighbor hopping t, and local repulsion U.

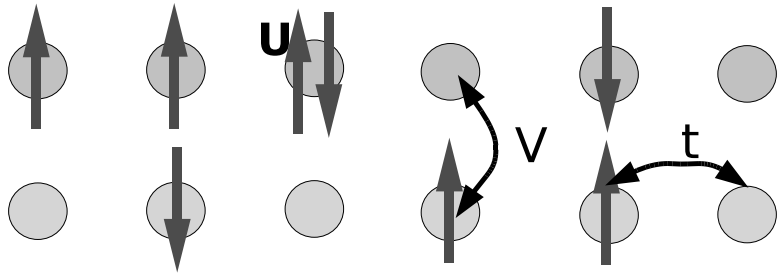

FIG. 2: Cartoon of the periodic Anderson model, characterized by two bands one with near-neighbor hopping t with a local hybridization V to a second band with local repulsion U.

For example, the Hubbard model[1] is the simplest model of a correlated electronic lattice system. Both it and the $t-J$ model are thought to at least qualitatively describe some of the properties of transition metal oxides, and high temperature superconductors[2]. The periodic Anderson model along with various Kondo lattice models have been proposed to describe both the actinide and lanthanide heavy fermion systems and the Anderson insulators. The Holstein model incorporates the essential physics of strongly interacting electrons and phonons. All of these model Hamiltonians contain at least two major ingredients: a local interaction term and a non-local hopping term. For example, the Hubbard model Hamiltonian is

$$H = -t \sum_{\langle j,k\rangle\sigma} (c^\dagger_{j\sigma} c_{k\sigma} + c^\dagger_{k\sigma} c_{j\sigma}) + \epsilon \sum_j (n_{j\uparrow} + n_{j\downarrow}) + U \sum_j (n_{j\uparrow} - 1/2)(n_{j\downarrow} - 1/2)\,, \qquad (1)$$

where $c^\dagger_{j\sigma}$ ($c_{j\sigma}$) creates (destroys) an electron at site j with spin σ, $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$, t is the nearest neighbor hopping which sets the unit of energy and U is the on-site Coulomb repulsion between the electrons. The periodic Anderson model (PAM) Hamiltonian is

$$H = -t \sum_{<ij>\sigma} (d^\dagger_{i\sigma} d_{j\sigma} + d^\dagger_{j\sigma} d_{i\sigma}) + V \sum_{i\sigma} (d^\dagger_{i\sigma} f_{i\sigma} + f^\dagger_{i\sigma} d_{i\sigma}) + \frac{U}{2} \sum_{i\sigma} (n^f_{i,\sigma} - \frac{1}{2})(n^f_{i,-\sigma} - \frac{1}{2}) \qquad (2)$$

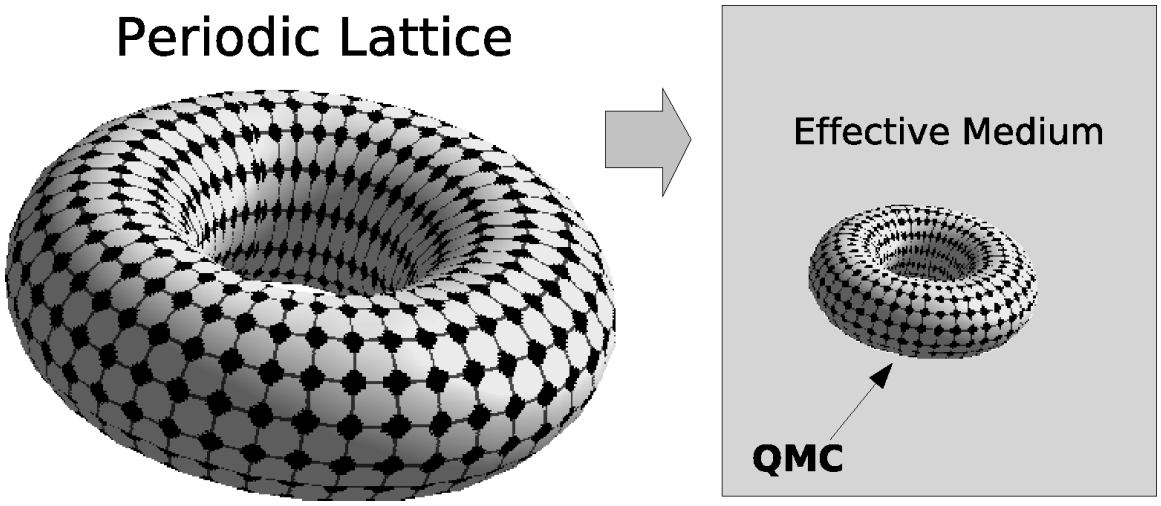

FIG. 3: Quantum cluster approaches, like the DMFA and DCA, map the infinite lattice problem onto a self-consistently embedded cluster problem.

where $d_{i\sigma}$ and $f_{i\sigma}$ ($d^\dagger_{i\sigma}$ and $f^\dagger_{i\sigma}$) destroy (create) a d- and f-electron on site i with spin σ, U is the screened Coulomb-matrix element for the localized f-states and V characterizes the mixing between the two subsystems, f and respectively d orbitals.

However, except for special limits, even such simplified models like Eq. 1 cannot be solved exactly. For example, for the Hubbard model, no exact solutions exist except in one dimension, where the knowledge is in fact rather complete[3]. The periodic Anderson model is only solvable in the limit where the orbital degeneracy diverges[4], and the Holstein model is only solvable in the Eliashberg-Migdal limit where vertex corrections may be neglected. Clearly a new approach to these models is needed if nontrivial exact solutions are desired.

Metzner and Vollhardt suggested such a new approach[5–7] based on an expansion in $1/d$ about the point $d = \infty$ to study these strongly correlated lattice models. The resulting formalism neglects dynamical intersite correlations while retaining the important local dynamical correlations. The resulting formalism is called the Dynamical Mean Field Approximation (DMFA) since it may be employed in any dimension, but is only exact on infinite dimensional lattices. In finite dimensions, the Dynamical Cluster Approximation (DCA) is used to study systematic non-local corrections to the DMFA[8,9]. Quantum cluster approaches such as the DMFA and DCA work by mapping an infinite periodic lattice onto a self-consistently embedded cluster problem, as illustrated in Fig. 3. Correlations up to the cluster size are treated explicitly, while those at longer length scales are treated in a mean

field. The DMFA/DCA cluster problem may be solved by a variety of methods; however, Quantum Monte Carlo (QMC) is the first numerically exact method employed[10] and remains the most powerful and adaptable method.

In this article, we will present a pedagogical discussion of a complete suite of QMC-based formalisms and algorithms for the DMFA and DCA. In Sec. II we will first rederive the DMFA as a course-graining approximation, extend this logic to derive the DCA, and then describe how physical quantities are calculated in this formalism. In Sec. III we will discuss two powerful QMC algorithms used to solve the embedded cluster problem of the DMFA/DCA. Finally, in Sec. IV we will discuss entropy-based Bayesian data analysis and its use to analytically continue the Matsubara time or frequency QMC data to real frequencies.

II. THE DYNAMICAL MEAN FIELD AND CLUSTER APPROXIMATIONS

A. The Dynamical mean-field approximation

The DMFA is a local approximation which was used by Kuramoto in perturbative calculations as a simplification of the k-summations which render the problem intractable[11]. But it was after the work of Metzner and Vollhardt[5] and Müller-Hartmann[6] who showed that this approximation becomes exact in the limit of infinite dimension that it received extensive attention. In this limit, the spatial dependence of the self-energy disappears, retaining only its variation with time. Please see the reviews by Pruschke *et al*[12] and Georges *et al*[13] for a more extensive treatment.

In this section, we will show that it is possible to re-interpret the DMFA as a course graining approximation. For a two-dimensional lattice, this is equivalent to averaging, or coarse-graining, the Green's functions used to calculate the irreducible diagrammatic insertions over the Brillouin zone.

Müller-Hartmann[6] showed that this coarse-graining becomes exact in the limit of infinite-dimensions. For Hubbard-like models, the properties of the bare vertex are completely characterized by the Laue function Δ which expresses the momentum conservation at each vertex. In a conventional diagrammatic approach

$$\begin{aligned}\Delta(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4) &= \sum_{\mathbf{r}} \exp\left[i\mathbf{r}\cdot(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4)\right] \\ &= N\delta_{\mathbf{k}_1+\mathbf{k}_2,\mathbf{k}_3+\mathbf{k}_4}\end{aligned} \tag{3}$$

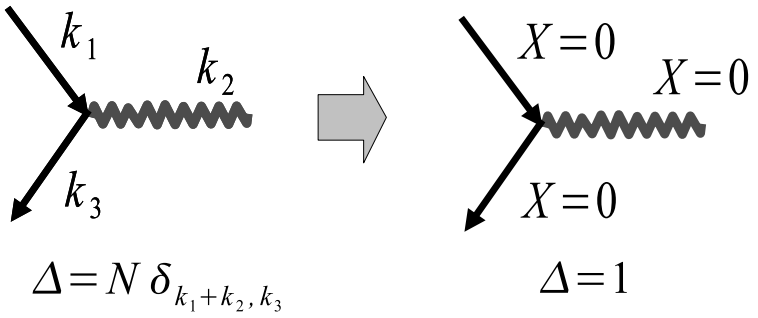

FIG. 4: The Laue function Δ, which described momentum conservation at a vertex (left). In the DMFA, $\Delta = 1$, so momentum conservation is neglected for compact graphs (right) so that we may freely sum over the momentum labels leaving only local propagators and interactions.

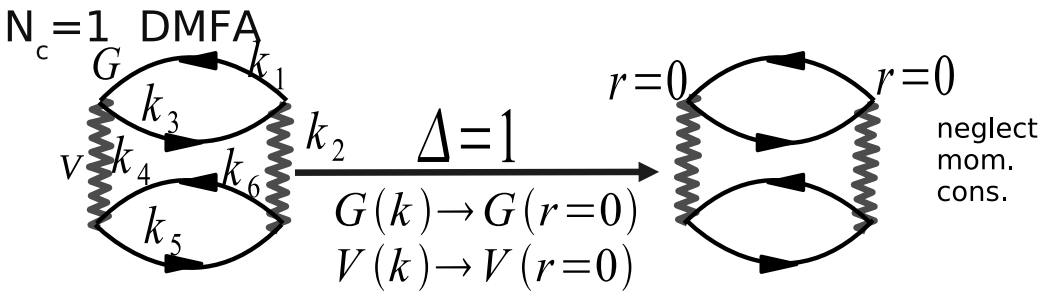

FIG. 5: The second order contribution the generating functional Φ. As we apply the DMFA coarse-graining approximation, Eq. 4, Φ becomes a functional of the local Green's function and interaction.

where $\mathbf{k}_1$ and $\mathbf{k}_2$ ($\mathbf{k}_3$ and $\mathbf{k}_4$) are the momenta entering (leaving) each vertex through its legs of G. However as the dimensionality $D \to \infty$ Müller-Hartmann showed that the Laue function reduces to[6]

$$\Delta_{D\to\infty}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = 1 + \mathcal{O}(1/D) \quad . \tag{4}$$

The DMFA assumes the same Laue function, $\Delta_{DMFA}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = 1$, even in the context of finite dimensions. More generally, for an electron scatting from an interaction (boson) pictured in Fig. 4, $\Delta_{DMFA}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 1$. Thus, the conservation of momentum at internal vertices is neglected. We may freely sum over the internal momentum labels of each Green's function leg and interaction leading to a collapse of the momentum dependent contributions leaving only local terms.

This argument may then be applied to the generating functional Φ. It is the sum over

all closed connected compact graphs constructed from the dressed Green's function G and the bare interaction. The self energy Σ_σ is obtained from a functional derivative of Φ, $\Sigma_\sigma = \delta\Phi/\delta G_\sigma$, and the irreducible vertices $\Gamma_{\sigma\sigma'} = \delta\Sigma_\sigma/\delta G'_\sigma$. The second order contribution to Φ for a Hubbard-like model is illustrated in Fig. 5, becomes a functional of the local interaction and Green's function. The self energy Σ may be obtained from a functional derivative of Φ with respect to the Green's function G, which effectively breaks one of the Green's function lines.

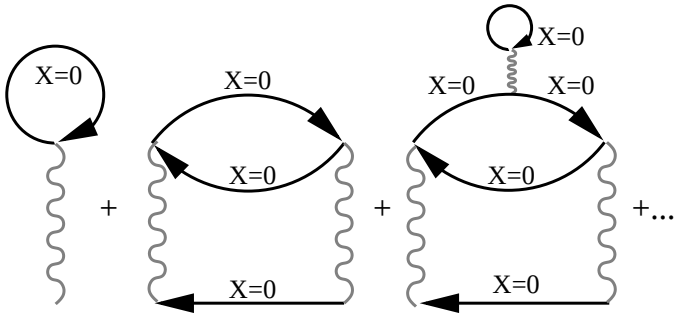

FIG. 6: The DMFA self energy. Note that it contains local self energy corrections See, e.g., the third graph. To prevent overcounting these contributions, the local self energy must be excluded, c.f., Eq. 5 from the Green's function line used in most cluster solvers.

The perturbative series for Φ, Σ and the irreducible vertices Γ in the DMFA are identical to those of the corresponding impurity model, so that conventional impurity solvers may be used. However, since most impurity solvers can be viewed as methods that sum all the graphs, not just the skeletal ones, it is necessary to exclude $\Sigma(i\omega_n)$ from the local propagator $\mathcal{G}$ input to the impurity solver in order to avoid overcounting the local self-energy $\Sigma(i\omega_n)$ ($i\omega_n = (2n+1)\pi T$ is the Matsubara frequency).

$$\mathcal{G}(i\omega_n)^{-1} = G(i\omega_n)^{-1} + \Sigma(i\omega_n) \tag{5}$$

where $G(i\omega_n)$ is the full local Green's function. Hence, in the local approximation, the Hubbard model has the same diagrammatic expansion as an Anderson impurity with a bare local propagator $\mathcal{G}(i\omega_n;\Sigma)$ which is determined self-consistently.

An algorithm constructed from this approximation is the following: (i) An initial guess for $\Sigma(i\omega_n)$ is chosen (usually from perturbation theory). (ii) $\Sigma(i\omega_n)$ is used to calculate the

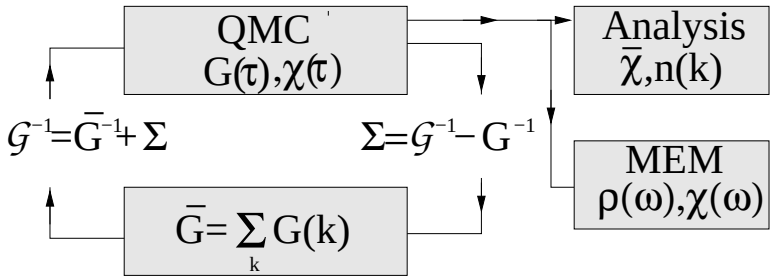

FIG. 7: The DMFA algorithm. QMC is used as a cluster solver. Once convergence is reached, $G = \bar{G}$, and the irreducible quantities are used in the analysis and Maximum Entropy Method (MEM) codes to calculate the phase diagram and spectra, respectively.

corresponding local Green's function

$$G(i\omega_n) = \int d\eta \frac{\rho^0(\eta)}{i\omega_n - \eta - \epsilon - \Sigma(i\omega_n)}\,, \tag{6}$$

where ρ^0 is the non-interacting density of states. (iii) Starting from $G(i\omega_n)$ and $\Sigma(i\omega_n)$ used in the second step, the host Green's function $\mathcal{G}(i\omega_n)^{-1} = G(i\omega_n)^{-1} + \Sigma(i\omega_n)$ is calculated which serves as bare Green's function of the impurity model. (iv) Starting with $\mathcal{G}(i\omega_n)$, the local Green's function $G(i\omega_n)$ is obtained using the Quantum Monte Carlo method (or another technique). (v) Using the QMC output for the cluster Green's function $G(i\omega_n)$ and the host Green's function $\mathcal{G}(i\omega_n)$ from the third step, a new $\Sigma(i\omega_n) = \mathcal{G}(i\omega_n)^{-1} - G(i\omega_n)^{-1}$ is calculated, which is then used in step (ii) to reinitialize the process. Steps (ii) - (v) are repeated until convergence is reached. In step (iv) the QMC algorithm of Hirsch and Fye[14,15] may be used to compute the local Green's function $G(\tau)$ or other physical quantities in imaginary time. Local dynamical quantities are then calculated by analytically continuing the corresponding imaginary-time quantities using the Maximum-Entropy Method (MEM)[16].

B. The Dynamical Cluster Approximation

In this section, we will review the formalism which leads to the dynamical cluster approximation. Here, we first motivate the fundamental idea of the DCA which is coarse-graining, we then describe the mapping to an effective cluster problem and discuss the relationship between the cluster and lattice at the one and two-particle level.

1. Coarse-Graining

Like the DMFA, the DCA may be intuitively motivated with a coarse-graining transformation. In the DMFA, the propagators used to calculate Φ and its functional derivatives were coarse-grained over the entire Brillouin zone, leading to local (momentum independent) irreducible quantities. In the DCA, we wish to relax this condition and systematically restore momentum conservation and non-local corrections. Thus, in the DCA, the reciprocal space of the lattice (Fig. 8) which contains N points is divided into N_c cells of identical linear size Δk. The coarse-graining transformation is set by averaging the Green's function within each cell. If $N_c = 1$ the original lattice problem is mapped to an impurity problem, and we recover the DMFA. If N_c is larger than one, then non-local corrections of length $\approx \pi/\Delta k$ to the DMFA are introduced. Provided that the propagators are sufficiently weakly momentum dependent, this is a good approximation. If N_c is chosen to be small, the cluster problem can be solved using conventional techniques such as QMC. This averaging process also establishes a relationship between the systems of size N and N_c. A simple and unique choice which will be discussed in Sec. II B 2 is to equate the irreducible quantities (self energy, irreducible vertices) of the cluster to those in the lattice.

2. A diagrammatic derivation

This coarse graining procedure and the relationship of the DCA to the DMFA is illustrated by a microscopic diagrammatic derivation of the DCA starting again from the Baym generating functional Φ[19]. The DCA systematically restores the momentum conservation at internal vertices of Φ relinquished by the DMFA. The Brillouin-zone is divided into $N_c = L^D$ cells of size $\Delta k = 2\pi/L$ (c.f. Fig. 8 for $N_c = 8$). Each cell is represented by a cluster momentum $\mathbf{K}$ in the center of the cell. We require that momentum conservation is (partially) observed for momentum transfers between cells, i.e., for momentum transfers larger than Δk, but neglected for momentum transfers within a cell, i.e., less than Δk. This requirement can be established by using the Laue function[9]

$$\Delta_{DCA}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = N_c \delta_{\mathbf{M}(\mathbf{k}_1)+\mathbf{M}(\mathbf{k}_2),\mathbf{M}(\mathbf{k}_3)+\mathbf{M}(\mathbf{k}_4)} \quad , \tag{7}$$

where $\mathbf{M}(\mathbf{k})$ is a function which maps $\mathbf{k}$ onto the momentum label $\mathbf{K}$ of the cell containing $\mathbf{k}$ (see, Fig. 8). This choice for the Laue function systematically interpolates between the

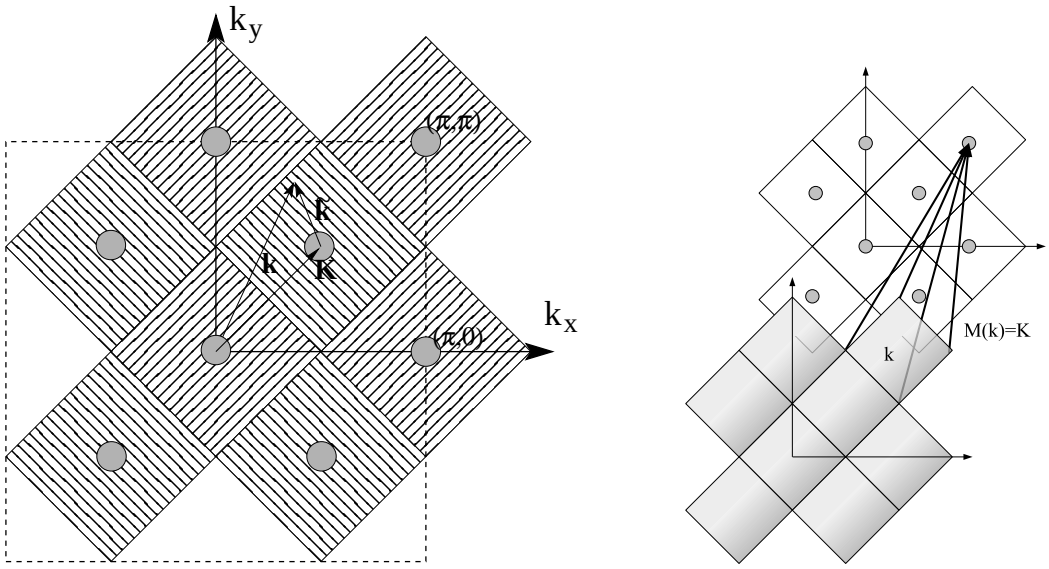

FIG. 8: (left) Coarse-graining cells for $N_c = 8$ (differentiated by alternating fill patterns) that partition the first Brillouin Zone (dashed line). Each cell is centered on a cluster momentum $\mathbf{K}$ (filled circles). (right) To construct the DCA cluster (e.g. for $N_c = 8$) we map a generic $\mathbf{k}$ to the nearest cluster point $\mathbf{K} = \mathbf{M}(\mathbf{k})$ so that $\tilde{\mathbf{k}} = \mathbf{k} - \mathbf{K}$ remains in the cell around $\mathbf{K}$.

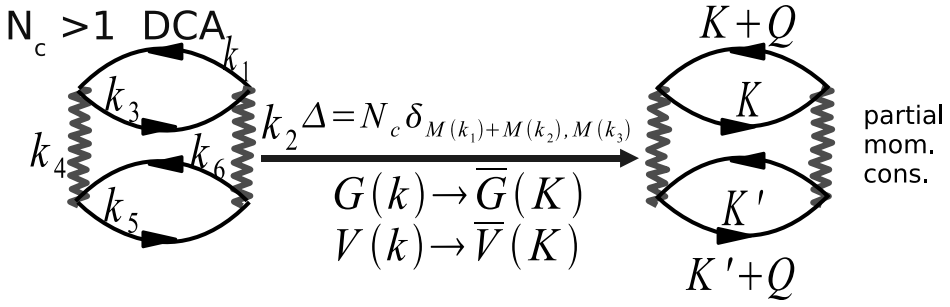

FIG. 9: A second-order term in the generating functional of the Hubbard model. Here the undulating line represents the interaction U, and on the LHS (RHS) the solid line, the lattice (coarse-grained) single-particle Green's functions. When the DCA Laue function is used to describe momentum conservation at the internal vertices, the momenta collapse onto the cluster momenta and each lattice Green's function is replaced by the coarse-grained result.

exact result, Eq. 4, which it recovers when $N_c \to N$ and the DMFA result, Eq. 4, which it recovers when $N_c = 1$. With this choice of the Laue function the momenta of each internal leg may be freely summed over the cell.

This is illustrated for the second-order term in the generating functional in Fig. 9. Each

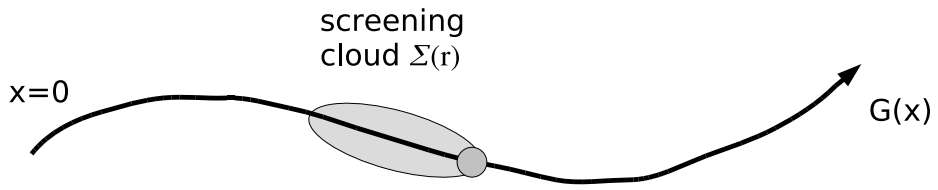

FIG. 10: Screening of a propagating particle. The single particle Green's function, which describes the quantum phase and amplitude the particle accumulates, is poorly approximated by a small cluster calculation. Its self energy, which describes generally short ranged screening processes, can be well approximated by a small cluster calculation.

internal leg $G(\mathbf{k})$ in a diagram is replaced by the coarse–grained Green's function $\bar{G}(\mathbf{M}(\mathbf{k}))$, defined by

$$\bar{G}(\mathbf{K}) \equiv \frac{N_c}{N} \sum_{\tilde{\mathbf{k}}} G(\mathbf{K} + \tilde{\mathbf{k}}) \quad , \tag{8}$$

where N is the number of points of the lattice, N_c is the number of cluster $\mathbf{K}$ points, and the $\tilde{\mathbf{k}}$ summation runs over the momenta of the cell about the cluster momentum $\mathbf{K}$ (see, Fig. 8). The diagrammatic sequences for the generating functional and its functional derivatives are unchanged; however, the complexity of the problem is greatly reduced since $N_c \ll N$.

As with the DMFA, the coarse-graining approximation will be applied to only the compact part of the free energy, Φ, and its functional derivatives. Physically, this is justified by the fact that irreducible terms like the self energy are short ranged, while reducible quantities like G must be able to capture long length and time scale physics. This is motivated in Fig. 10. As the particle propagates from the origin to space-time location x, the quantum phase and amplitude it accumulates is described by the single-particle Green's function $G(x)$. Consequently if x is larger than the size of the DCA cluster, then $G(x)$ is poorly approximated by the cluster Green's function. However, the self energy Σ describes the many-body processes that produce the screening cloud surrounding the particle. These processes are generally short ranged in a strongly correlated many-body system, so the self energy is often well approximated by the cluster quantity. Formally, we have justified this elsewhere by exploring the Δk-dependence of the compact and non-compact parts of the free energy[17]. The generating functional is the sum over all of the closed connected compact diagrams, such as the one shown in Fig. 9.

The corresponding DCA estimate for the free energy is

$$F_{DCA} = -k_B T \left(\Phi_c - \mathrm{Tr} \left[\Sigma_\sigma G_\sigma \right] - \mathrm{Tr} \ln \left[-G_\sigma \right] \right) , \tag{9}$$

where Φ_c is the cluster generating functional. The trace indicates summation over frequency, momentum and spin. F_{DCA} is stationary with respect to $\mathbf{G}_\sigma$ when

$$\frac{-1}{k_B T} \frac{\delta F_{DCA}}{\delta G_\sigma(\mathbf{k})} = \Sigma_{c\sigma}(M(\mathbf{k})) - \Sigma_\sigma(\mathbf{k}) = 0, \tag{10}$$

which means that $\Sigma(\mathbf{k}) = \Sigma_{c\sigma}(M(\mathbf{k}))$ is the proper approximation for the lattice self energy corresponding to Φ_c. The corresponding lattice single-particle propagator is then given by

$$G(\mathbf{k}, z) = \frac{1}{z - \epsilon_{\mathbf{k}} - \epsilon - \Sigma_c(M(\mathbf{k}), z)} . \tag{11}$$

A similar procedure is used to construct the two-particle quantities needed to determine the phase diagram or the nature of the dominant fluctuations that can eventually destroy the quasi-particle. This procedure is a generalization of the method of calculating response functions in the DMFA[10,18]. In the DCA, the introduction of the momentum dependence in the self-energy will allow one to detect some precursor to transitions which are absent in the DMFA; but for the actual determination of the nature of the instability, one needs to compute the response functions. These susceptibilities are thermodynamically defined as second derivatives of the free energy with respect to external fields. $\Phi_c(G)$ and $\Sigma_{c\sigma}$, and hence F_{DCA} depend on these fields only through G_σ and the bare G^0_σ. Following Baym[19] it is easy to verify that, the approximation

$$\Gamma_{\sigma,\sigma'} \approx \Gamma_{c\sigma,\sigma'} \equiv \delta\Sigma_{c\sigma} / \delta G_{\sigma'} \tag{12}$$

yields the same estimate that would be obtained from the second derivative of F_{DCA} with respect to the applied field. For example, the first derivative of the free energy with respect to a spatially homogeneous external magnetic field h is the magnetization,

$$m = \mathrm{Tr} \left[\sigma G_\sigma \right] . \tag{13}$$

The susceptibility is given by the second derivative,

$$\frac{\partial m}{\partial h} = \mathrm{Tr} \left[\sigma \frac{\partial G_\sigma}{\partial h} \right] . \tag{14}$$

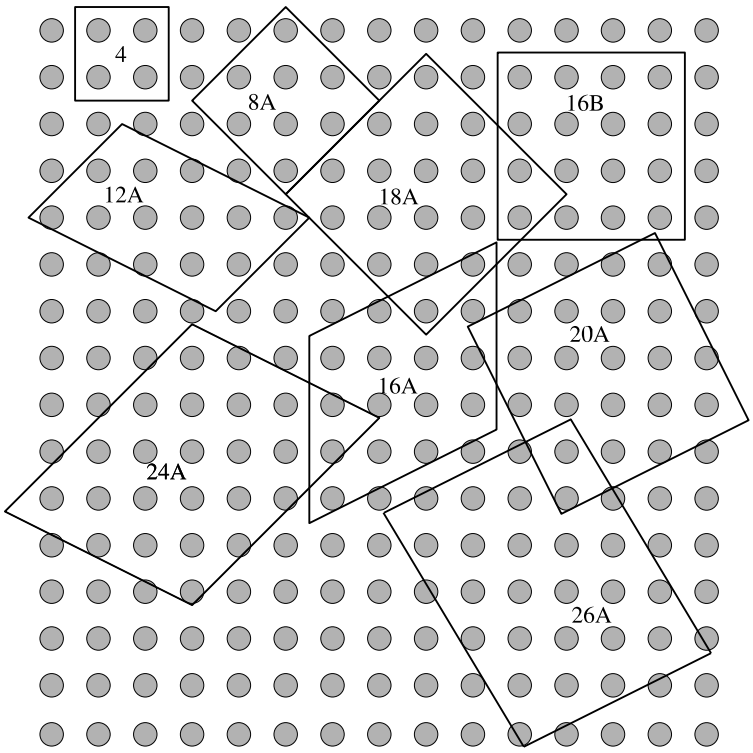

FIG. 11: A variety of cluster geometries which may be used to tile a two-dimensional square lattice.

We substitute $G_\sigma = \left(G_\sigma^{0-1} - \Sigma_{c\sigma}\right)^{-1}$, and evaluate the derivative,

$$\frac{\partial m}{\partial h} = \mathrm{Tr}\left[\sigma \frac{\partial G_\sigma}{\partial h}\right] = \mathrm{Tr}\left[G_\sigma^2 \left(1 + \sigma \frac{\partial \Sigma_{c\sigma}}{\partial G_{\sigma'}} \frac{\partial G_{\sigma'}}{\partial h}\right)\right]. \tag{15}$$

If we identify $\chi_{\sigma,\sigma'} = \sigma \frac{\partial G_{\sigma'}}{\partial h}$, and $\chi_\sigma^0 = G_\sigma^2$, collect all of the terms within both traces, and sum over the cell momenta $\tilde{\mathbf{k}}$, we obtain the two–particle Dyson's equation

$$\begin{aligned} 2(\bar{\chi}_{\sigma,\sigma} \; - \; \bar{\chi}_{\sigma,-\sigma}) & \\ = \; 2\bar{\chi}_\sigma^0 + 2\bar{\chi}_\sigma^0 \left(\Gamma_{c\sigma,\sigma} - \Gamma_{c\sigma,-\sigma}\right) \left(\bar{\chi}_{\sigma,\sigma} - \bar{\chi}_{\sigma,-\sigma}\right) . & \end{aligned} \tag{16}$$

We see that again it is the irreducible quantity, i.e., the vertex function, for which cluster and lattice correspond.

3. Cluster Selection

The geometry of the DCA cluster is relevant for systematic studies of phase transitions. All cluster geometries which can be used to tile the lattice without gaps are valid (c.f. Fig. 11). In this approximation, the lattice has long range order once the correlation length ξ of the order reaches the linear cluster size, since then the lattice tiled with such clusters

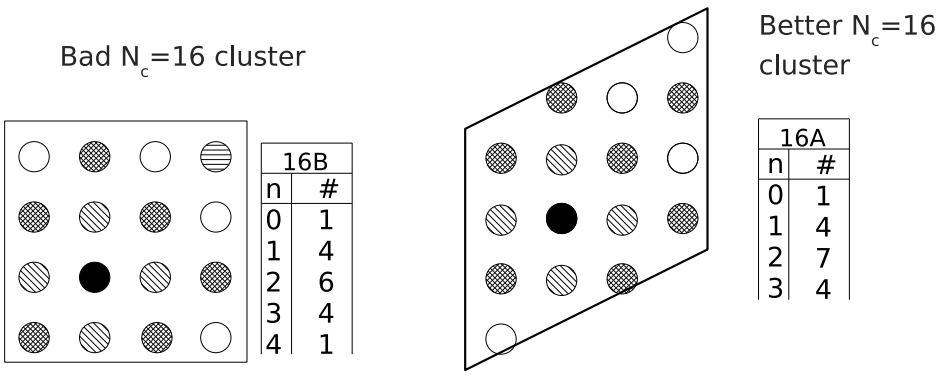

FIG. 12: Two sixteen ($N_c = 16$) site periodic clusters. Here, the neighbors to the black site are shown different patterns.

would be ordered. A scaling ansatz for T_c which captures these ideas is[20]

$$\xi\left(T_c(N_c)\right) = N_c^{1/D}\,. \tag{17}$$

So for a typical transition, with $\xi\left(T\right) \propto |T - T_c^*|^{-\nu}$ (where T_c^* is the transition temperature of the thermodynamic lattice), the scaling formula is $T_c^* = T_c(N_c) - A(N_c)^{-1/D\nu}$, where A, ν, and T_c^* are fit to the data $T_c(N_c)$. However, clearly this approach will not apply to clusters with strange geometries, like e.g., rods in a 2D system.

Ideal clusters should properly represent all length scales up to $N_c^{1/D}$, and no others. Consider the 16-site clusters shown in Fig. 12, the number of neighbors in each neighbor shell are shown in the adjacent tables[21]. On the lattice, the nth shell has 4n neighbors. Both 16A and 16B have complete near neighbor (n=1) shells with four near neighbors. However, the conventional square cluster 16B, has 6 (not 8) sites in the n=2 shell, has 4 (not 12) in the n=3 shell, and even on site in the n=4 shell. The cluster 16A is far better, with no site in the n=4 shell and a nearly complete n=2 shell.

Betts[21] explored different cluster geometries, and realized that there are few clusters with the same geometry (point group) as the lattice, but far more with lower symmetry, that are often superior. He selected clusters based on neighbors in a given shell, symmetry, and squareness. Since for a given cluster size, one incomplete shell is usually inevitable, he classified the imperfection of each cluster by the number of sites missing in nearer shells, rather than the number in an incomplete one and the number of sites in farther shells. Thus, 16B has an imperfection of three while 16A has an imperfection of one. The imperfection number is the most important selection criteria.

C. Calculation of Physical Properties

Most experiments measure quantities which we can express theoretically as reducible one or two-particle Green's functions. As discussed above, the appropriate way to calculate these quantities is to first extract the corresponding irreducible quantity from the cluster calculation, and then use it to calculate the reducible quantity. For example, to calculate the single-particle Green's function (relevant for angle-resolved photoemission spectroscopy) we first extract the cluster self energy and use the Dyson equation to construct the lattice Green's function. To calculate the phase diagram, we calculate the irreducible vertices in the different scattering channels Γ, and insert them into the Bethe-Salpeter equations for the lattice. In this subsection we will provide more details about the relationship between the lattice and cluster two-particle Green's functions and describe how a lattice susceptibility may be calculated efficiently.

a. Particle-hole channel As a specific example, we will describe the calculation of the two-particle Green's function

$$\begin{aligned}\chi_{\sigma,\sigma'}(q,k,k') &= \int_0^\beta \int_0^\beta \int_0^\beta \int_0^\beta d\tau_1 d\tau_2 d\tau_3 d\tau_4 \\ &\times\ e^{i((\omega_n+\nu)\tau_1-\omega_n\tau_2+\omega_{n'}\tau_3-(\omega_{n'}+\nu)\tau_4)} \\ &\times\ \langle T_\tau c^\dagger_{\mathbf{k}+\mathbf{q}\sigma}(\tau_1)c_{\mathbf{k}\sigma}(\tau_2)c^\dagger_{\mathbf{k}'\sigma'}(\tau_3)c_{\mathbf{k}'+\mathbf{q}\sigma'}(\tau_4)\rangle \quad ,\end{aligned}$$

where we adopt the conventional notation[22] $k = (\mathbf{k}, i\omega_n)$, $k' = (\mathbf{k}, \omega'_n)$, $q = (\mathbf{q}, \nu_n)$ and T_τ is the time ordering operator.

$\chi_{\sigma,\sigma'}(q,k,k')$ and $\Gamma_{\sigma,\sigma'}(q,k,k')$ are related to each other through the Bethe-Salpeter equation (Fig. 13):

$$\begin{aligned}\chi_{\sigma,\sigma'}(q,k,k') &= \chi^0_{\sigma,\sigma'}(q,k,k') + \chi^0_{\sigma,\sigma''}(q,k,k'') \\ &\times\ \Gamma_{\sigma'',\sigma'''}(q,k'',k''')\chi_{\sigma''',\sigma'}(q,k''',k') \qquad (18)\end{aligned}$$

where $\Gamma_{\sigma,\sigma'}(q,k,k')$ is the two-particle irreducible vertex which is the analogue of the self-energy, $\chi^0_{\sigma,\sigma'}(q,k,k'')$ is the non-interacting susceptibility constructed from a pair of fully-dressed single-particle Green's functions. As usual, a summation is to be made for repeated indices.

We now make the DCA substitution $\Gamma_{\sigma,\sigma'}(\mathbf{q},\mathbf{k},\mathbf{k}') \rightarrow \Gamma_{c\sigma,\sigma'}(\mathbf{q},\mathbf{M}(\mathbf{k}),\mathbf{M}(\mathbf{k}'))$ in Eq. 18 (where frequency labels have been suppressed). Note that only the bare and dressed two-particle Green's functions χ depend upon the momenta $\tilde{\mathbf{k}}$ within a cell. Since χ and χ^0 in

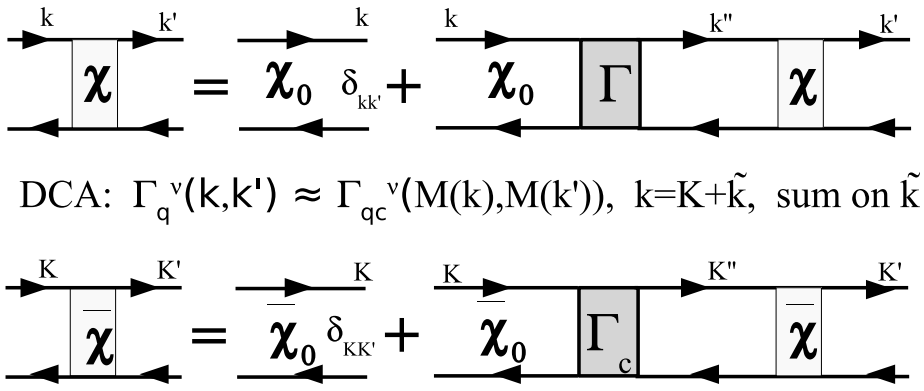

FIG. 13: The Bethe-Salpeter equation in the DCA. We approximate the lattice irreducible vertex Γ^ν by the Γ_c^ν from the DCA cluster and coarse-grain over the $\tilde{\mathbf{k}}$. The remaining equation is a function of the cluster $\mathbf{K}$ only and may be solved by inversion.

the product on the RHS of Eq. 18 share no common momentum labels, we may freely sum over the momenta $\tilde{\mathbf{k}}$ within a cell, yielding

$$\begin{aligned}\bar{\chi}_{\sigma,\sigma'}(q,K,K') &= \bar{\chi}^0_{\sigma,\sigma'}(q,K,K') + \bar{\chi}^0_{\sigma,\sigma''}(q,K,K'') \\ &\times\ \Gamma_{c\sigma'',\sigma'''}(q,K'',K''')\bar{\chi}_{\sigma''',\sigma'}(q,K''',K')\,. \end{aligned} \tag{19}$$

By coarse-graining the Bethe-Salpeter equation, we have greatly reduced its complexity; each of the matrices above is sufficiently small that they may be easily manipulated using standard techniques.

In contrast with the single-particle case where the coarse-grained quantities are identical to those of the cluster, $\chi_{c\sigma,\sigma'}(q,K,K')$ is not equal to $\bar{\chi}_{\sigma,\sigma'}(q,K,K')$. This is because the self-consistency is made only at the single-particle level. Unlike the single particle case where both $\Sigma(K)$ and $\bar{G}(K)$ are directly calculated, neither $\Gamma_{\sigma,\sigma'}(q,K,K')$ nor the coarse-grained susceptibility $\bar{\chi}_{\sigma,\sigma'}(q,K,K')$ are calculated during the self-consistency. Instead, the coarse-grained non-interacting susceptibility $\bar{\chi}^0_{\sigma,\sigma'}(q,K,K')$ is calculated in a separate program after the DCA converges using the following relation

$$\begin{aligned}\bar{\chi}^0_{\sigma,\sigma'}[(\mathbf{q},i\nu_n);(\mathbf{K},i\omega_n);(\mathbf{K}',i\omega'_n)] &= \delta_{\sigma,\sigma'}\delta_{\mathbf{K},\mathbf{K}'}\delta_{\omega_n,\omega'_n} \\ \times\frac{N_c}{N}\sum_{\tilde{\mathbf{k}}} G_\sigma(\mathbf{K}+\tilde{\mathbf{k}},i\omega_n)G_\sigma(\mathbf{K}+\tilde{\mathbf{k}}+\mathbf{q},i\omega_n+\nu_n) \quad .\end{aligned} \tag{20}$$

The corresponding cluster susceptibility is calculated by the cluster solver and the vertex

function is extracted by inverting the cluster two-particle Bethe-Salpeter equation

$$\chi_{c\sigma,\sigma'}(q,K,K') = \chi^0_{c\sigma,\sigma'}(q,K,K') + \chi^0_{c\sigma,\sigma''}(q,K,K'') \times\Gamma_{c\sigma'',\sigma'''}(q,K'',K''')\chi_{c\sigma''',\sigma'}(q,K''',K') \,. \tag{21}$$

If we combine Eqs. 21 and 19, then the coarse-grained susceptibility may be obtained after elimination of $\Gamma(q,K,K')$ between the two equations. It reads

$$\bar{\chi}^{-1} = \chi_c^{-1} - \chi_c^{0^{-1}} + \bar{\chi}^{0^{-1}} \,, \tag{22}$$

where, for example, $\bar{\chi}$ is the matrix formed from $\bar{\chi}_{\sigma,\sigma'}(q,K,K')$ for fixed q. The charge (ch) and spin (sp) susceptibilities $\chi_{ch,sp}(q,T)$ are deduced from $\bar{\chi}$

$$\chi_{ch,sp}(q,T) = \frac{(k_BT)^2}{N_c^2} \sum_{KK'\sigma\sigma'} \lambda_{\sigma\sigma'}\bar{\chi}_{\sigma,\sigma'}(q,K,K') \quad , \tag{23}$$

where $\lambda_{\sigma\sigma'} = 1$ for the charge channel and $\lambda_{\sigma\sigma'} = \sigma\sigma'$ for the spin channel.

b. Particle-particle channel The calculation of susceptibilities in the particle-particle channel is essentially identical to the above. The exception to this rule occurs when we calculate susceptibilities for transitions to states of lower symmetry than the lattice symmetry. For example, in order to obtain the pair function of the desired symmetry (s,p,d), the two-particle Green's function must be multiplied by the corresponding form factors $g(\mathbf{k})$ and $g(\mathbf{k}')$. In the study of the Hubbard model below, we will be particularly interested in $g(\mathbf{k}) = 1$ (s wave), $g(\mathbf{k}) = cos(k_x) + cos(k_y)$ (extended s wave) and $g(\mathbf{k}) = cos(k_x) - cos(k_y)$ ($d_{x^2-y^2}$ wave). These symmetries have been evoked as possible candidates for the superconducting ground state.

These factors modify the Bethe-Salpeter equations

$$\begin{aligned} g(\mathbf{k})\chi(q,k,k')g(\mathbf{k}') &= g(\mathbf{k})\chi^0(q,k,k')g(\mathbf{k}') \\ &+ \; g(\mathbf{k})\chi^0(q,k,k'') \times \Gamma(q,k'',k''') \times \chi(q,k''',k')g(\mathbf{k}') \,. \end{aligned} \tag{24}$$

where

$$\begin{aligned} \chi(q,k,k') \;=\; & \int_0^\beta \int_0^\beta \int_0^\beta \int_0^\beta d\tau_1 d\tau_2 d\tau_3 d\tau_4 \\ & \times \; e^{i((\omega_n+\nu)\tau_1 - \omega_n\tau_2 + \omega_{n'}\tau_3 - (\omega_{n'}+\nu)\tau_4)} \\ & \times \; \langle T_\tau c^\dagger_{\mathbf{k}+\mathbf{q}\sigma}(\tau_1) c^\dagger_{-\mathbf{k}-\sigma}(\tau_2) c_{-\mathbf{k}'-\sigma}(\tau_3) c_{\mathbf{k}'+\mathbf{q}\sigma}(\tau_4)\rangle \quad . \end{aligned} \tag{25}$$

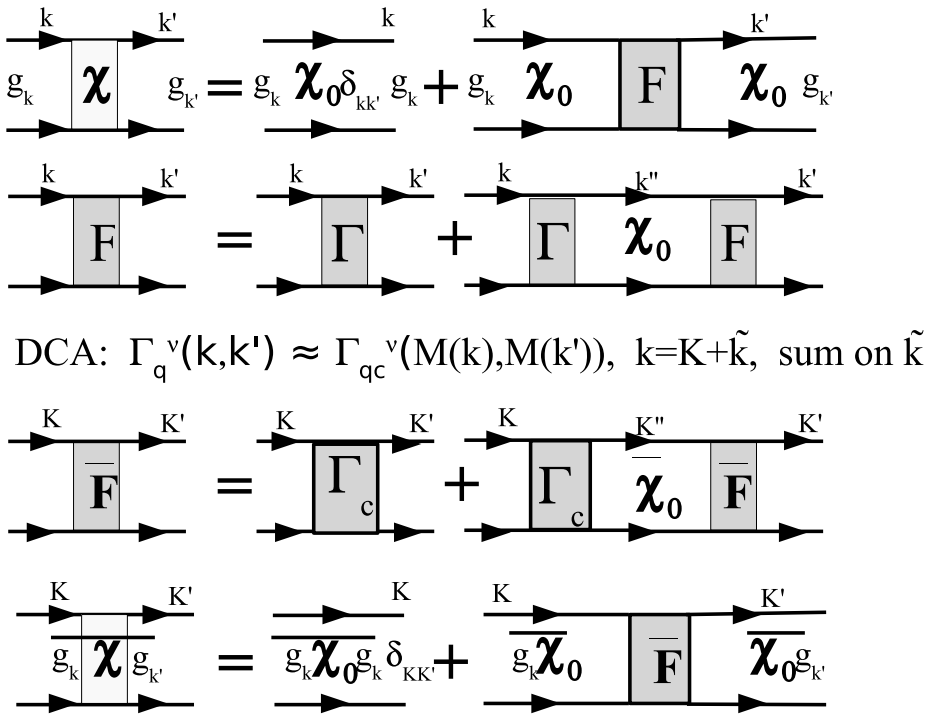

FIG. 14: Calculation of particle-particle projected susceptibilities. Often we want to calculate a projected particle-particle susceptibility (e.g. d-wave, with $g_{\mathbf{k}} = \cos(k_x) - \cos(k_y)$). Here the Bethe-Salpeter equation is rewritten in terms of the irreducible vertex F. We approximate the lattice irreducible vertex Γ^ν by the Γ_c^ν from the DCA cluster and coarse-grain over the $\tilde{\mathbf{k}}$. Then the projected bare bubbles are calculated, and the remaining equation is a function of the cluster $\mathbf{K}$ only and may be solved by inversion.

On the LHS, we have dropped the spin indices since we will consider only opposite-spin pairing. Eq. 24 cannot be easily solved if it is coarse-grained, since this will partially convolve $\chi(q,k,k')$ with *two* factors of g on the LHS and *one* factor on the RHS. Hence for the pairing susceptibilities, or for any situation where non-trivial form factors must be used, we use the equivalent equation involving the reducible vertex F (instead of the irreducible vertex Γ)

$$\begin{aligned} g(\mathbf{k})\chi(q,k,k')g(\mathbf{k}') &= g(\mathbf{k})\chi^0(q,k,k')g(\mathbf{k}') \\ &+ g(\mathbf{k})\chi^0(q,k,k'') \\ &\times F(q,k'',k''')\chi^0(q,k''',k')g(\mathbf{k}')\,, \end{aligned} \tag{26}$$

where

$$F(q,k,k') = \Gamma(q,k,k') + \chi^0(q,k,k'')\Gamma(q,k'',k''')\chi^0(q,k''',k') + \cdots \quad (27)$$

We define

$$\Pi_{g,g}(q,k,k') = g(\mathbf{k})\chi(q,k,k')g(\mathbf{k}') \quad (28)$$

$$\Pi^0_{g,g}(q,k,k') = g(\mathbf{k})\chi^0(q,k,k')g(\mathbf{k}') \quad (29)$$

$$\Pi^0_{g}(q,k,k') = g(\mathbf{k})\chi^0(q,k,k')\,. \quad (30)$$

The remaining steps of the calculation are similar to the particle-hole case. We invert the cluster particle-particle Bethe-Salpeter equation with $g = 1$ for the cluster, in order to extract Γ_c. We then coarse-grain Eq. 27, and use Γ_c to calculate the coarse-grained $\bar{F} = \Gamma_c \left(1 - \bar{\chi}^0\Gamma_c\right)^{-1}$. We then coarse-grain Eq. 26, and use the coarse-grained $\bar{F}$ to calculate the coarse-grained $\bar{\Pi}_{g,g}$

$$\bar{\Pi}_{g,g}(q,K,K') = \bar{\Pi}^0_{g,g}(q,K,K') + \bar{\Pi}^0_{g}(q,K,K'')\bar{T}_2(q,K'',K''')\bar{\Pi}^0_{g}(q,K''',K')\,. \quad (31)$$

The pairing susceptibility of a desired symmetry is given by

$$P_g(q,T) = \frac{(k_B T)^2}{N_c^2}\sum_{K,K'}\bar{\Pi}_{gg}(q,K,K') \quad . \quad (32)$$

D. Summary

Coarse-graining methods are used to derive the DMFA, and DCA which map the lattice onto a self-consistently embedded cluster problem. The DMFA is a local approximation, while the DCA incorporates systematic non-local corrections. Irreducible quantities from the cluster are used to calculate reducible lattice quantities.

III. QUANTUM MONTE CARLO ALGORITHMS FOR THE QUANTUM CLUSTER PROBLEM

A. Introduction

The Hirsch-Fye and Continuous time Quantum Monte Carlo (QMC) algorithms are powerful and adaptable methods which may be used to study models of impurity clusters embedded in a host. As such, they are ideal cluster solvers for the embedded impurity problem at the heart of the Dynamical Mean Field and Dynamical Cluster Approximations.

We will sketch the Hirsch-Fye QMC in Sec. III B and the weak coupling continuous time QMC algorithm in Sec. III C. Since both methods can be considered as expansions about the free electron limit, Wick's theorem applies to both, so similar methods for making and conditioning measurements described in Sec. III D can be used for both HFQMC and CTQMC.

B. Hirsch-Fye QMC

To derive the Hirsch-Fye algorithm, we start with the Hubbard Hamiltonian (1)

$$H = H_0 + H_1, \tag{33}$$

where H_0 is the noninteracting part of the Hamiltonian

$$H_0 \;=\; -t \sum_{\langle j,k\rangle\sigma} (c^\dagger_{j\sigma} c_{k\sigma} + c^\dagger_{k\sigma} c_{j\sigma}) + \epsilon \sum_j (n_{j\uparrow} + n_{j\downarrow}) \tag{34}$$

and H_1 describes the local interaction on a cluster $\mathcal{C}$ (or impurity).

$$H_1 = U \sum_{i\in\mathcal{C}} (n_{i\uparrow} - \frac{1}{2})(n_{i\downarrow} - \frac{1}{2}) \; . \tag{35}$$

The summation in Eq.34 is taken over the entire lattice whereas in Eq.35 it is taken only over the cluster $\mathcal{C}$ sites (see Fig.III B). Our derivation of the equations follows the one presented by Hirsch and Fye for the impurity problem[14,23], but we extend the derivation to treat the embedded cluster in the DMFA and DCA[15].

By dividing the imaginary time (the inverse of the temperature) $\beta = \frac{1}{T}$ into L slices,

$$\Delta\tau = \frac{\beta}{L} \; , \tag{36}$$

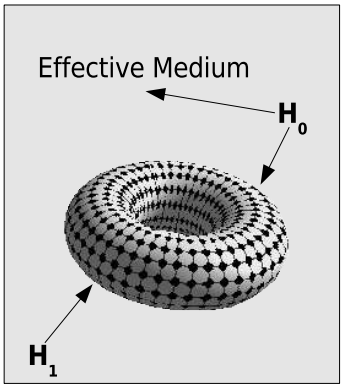

FIG. 15: Hamiltonian breakup for the Hirsch-Fye derivation. H_0 describes both the host and non-interacting degrees of freedom of the cluster (impurity), while H_1 describes the interactions on the cluster.

the partition function can be written as

$$Z = Tr(e^{-\beta H}) = Tr\left(\prod_{l=1}^{L} e^{-\Delta\tau H}\right) . \tag{37}$$

From the Suzuki-Trotter formula

$$e^{-\Delta\tau H} = e^{-\Delta\tau H_0/2} e^{-\Delta\tau H_1} e^{-\Delta\tau H_0/2} + \mathcal{O}(\Delta\tau^3) , \tag{38}$$

one derives

$$Z = Tr(e^{-\beta H}) \approx Tr\left(\prod_{l=1}^{L} e^{-\Delta\tau H_0/2} e^{-\Delta\tau H_1} e^{-\Delta\tau H_0/2}\right) , \tag{39}$$

which has leading errors proportional to $\Delta\tau^2$ (since the Suzuki-Trotter formula was applied L times and $L \sim 1/\Delta\tau$). Then due to the periodic property of the trace, it is easy to see that this is the same as the Trotter decomposition

$$Z = Tr(e^{-\beta H}) \approx Tr\left(\prod_{l=1}^{L} e^{-\Delta\tau H_0} e^{-\Delta\tau H_1}\right) , \tag{40}$$

with leading errors still proportional to $\Delta\tau^2$.

We introduce the identity operator in the occupation number basis

$$I = \sum_{m} |m\rangle\langle m| \tag{41}$$

between exponents of operators at adjacent imaginary time slices. The partition function becomes

$$\begin{aligned} Z = \sum_{m_L, m_{L-1}, .. m_1} & \langle m_L | e^{-\Delta\tau H_0} e^{-\Delta\tau H_1} | m_{L-1} \rangle \langle m_{L-1} | e^{-\Delta\tau H_0} e^{-\Delta\tau H_1} | m_{L-2} \rangle \\ & ... \langle m_1 | e^{-\Delta\tau H_0} e^{-\Delta\tau H_1} | m_L \rangle \end{aligned} \tag{42}$$

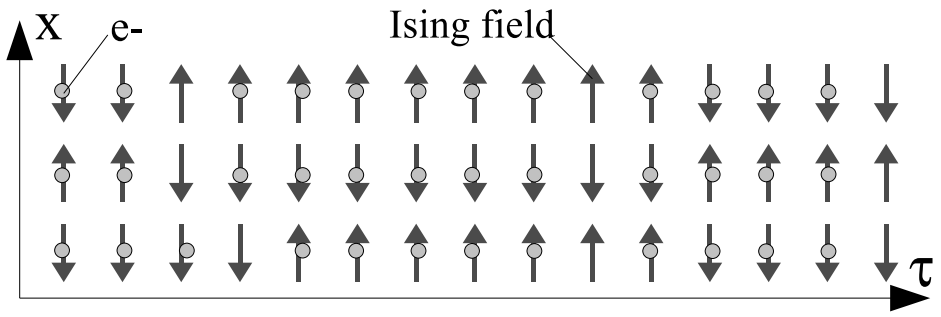

FIG. 16: The Hirsch-Hubbard-Stratonovich transformation Eq. 43 maps an interacting systems of electrons onto a system of non-interacting electrons interacting with a time and space dependent Ising-like fields which coupling to the z-component of the electron spin.

With the identity due to Hirsch[24]

$$e^{-\Delta\tau U\left(n_{i\uparrow}n_{i\downarrow}-\frac{1}{2}(n_{i\uparrow}+n_{i\downarrow})\right)}=\frac{1}{2}\sum_{s_i=\pm 1}e^{\alpha s_i(n_{i\uparrow}-n_{i\downarrow})}\ , \tag{43}$$

with

$$\cosh\alpha=e^{\Delta\tau U/2}\ . \tag{44}$$

It is possible to introduce an auxiliary binary (± 1) field (called Hirsch-Hubbard-Stratonovich field, or HHS field) at every cluster site and at every time point. Thereby, the interacting problem described by Eq. 42 is replaced by a summation over all possible auxiliary field configurations of noninteracting terms. Eq. 43 can be proved by applying both the left and right hand sides on the four possible vectors (empty site, one electron up, one electron down and double occupied site) which span the local Hilbert space.

For each HHS field configuration $\{s\}$, we have a noninteracting problem composed of electrons scattering off a space and time dependent Ising-like field (c.f. Fig. 16). Therefore, by defining

$$O_\sigma(s)=\begin{pmatrix} I & 0 & 0 & ... & e^{-\Delta\tau K}e^{V_\sigma^L(s)} \\ e^{-\Delta\tau K}e^{V_\sigma^1(s)} & I & 0 & ... & 0 \\ 0 & e^{-\Delta\tau K}e^{V_\sigma^2(s)} & I & 0 & ... \\ ... & & & & \\ ... & ... & ... & e^{-\Delta\tau K}e^{V_\sigma^{L-1}(s)} & I \end{pmatrix}\ , \tag{45}$$

with

$$H_0=\sum_{i,j,\sigma}c_{i\sigma}^\dagger K_{ij}c_{j\sigma} \tag{46}$$

and

$$V_\sigma^l(s)_{ij} = \begin{cases} \alpha\sigma s_{il}\delta_{i,j} & i \in \mathcal{C} \\ 0 & i \notin \mathcal{C} \end{cases} , \tag{47}$$

where s_{il} represents the value of HHS field at site i and time l, it is possible to write Eq. 42 as

$$Z = Tr_{\{s\}} \left[\det O_\uparrow(s) \ \det O_\downarrow(s)\right] \tag{48}$$

. Note that every matrix element shown in Eq. 45 is in fact a $N \times N$ sub-block, N being the dimension of the K matrix defined in Eq. 46. Thus O is a $(N \times L) \times (N \times L)$ matrix.

The matrix O is the inverse of the one-particle Green's function

$$G_\sigma^{-1}(s) = O_\sigma(s) \ . \tag{49}$$

By defining

$$T_{il;jl'} = \delta_{l-1,l'}\delta_{i,j} \tag{50}$$

and

$$V_\sigma^l(s)_{il;jl'} = \begin{cases} \alpha\sigma s_{il}\delta_{l,l'}\delta_{i,j} & i \in \mathcal{C} \\ 0 & i \notin \mathcal{C} \ or \ j \notin \mathcal{C} \end{cases} , \tag{51}$$

Eqs. 45 and 49 can be written as

$$G_\sigma^{-1} = I - Te^{-\Delta\tau K}e^{V_\sigma} \ . \tag{52}$$

In Eq. 52 and in the following equations we omit writing the explicit HHS dependence for simplicity reasons. By multiplying Eq. 52 at right with e^{-V_σ} (which is diagonal) the following equation is obtained

$$G_\sigma^{-1}e^{-V_\sigma} = e^{-V_\sigma} - Te^{-\Delta\tau K}. \tag{53}$$

Eq. 53 is used to establish a relation between the Green's functions G' and G which correspond to two different field configurations $\{s\}$ and respectively $\{s'\}$

$$G_\sigma'^{-1}e^{-V_\sigma'} - G_\sigma^{-1}e^{-V_\sigma} = e^{-V_\sigma'} - e^{-V_\sigma} \ . \tag{54}$$

Using

$$A^{-1} - B^{-1} = C \Longleftrightarrow A = B - BCA \ , \tag{55}$$

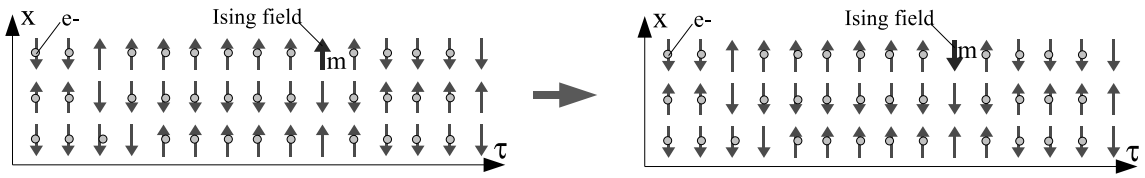

FIG. 17: The original Hirsch-Fye algorithm involved local flips of the HHS fields.

the following equation is obtained

$$e^{V'_\sigma} G'_\sigma = e^{V_\sigma} G_\sigma + e^{V_\sigma} G_\sigma (e^{-V_\sigma} - e^{-V'_\sigma}) e^{V'_\sigma} G'_\sigma \tag{56}$$

. After some easy manipulations, Eq. 56 can be written as

$$G'_\sigma = G_\sigma + (G_\sigma - I)(e^{V'_\sigma - V_\sigma} - I) G'_\sigma \ . \tag{57}$$

Another useful equation is

$$G_\sigma G'^{-1}_\sigma = I - (G_\sigma - I)(e^{V'_\sigma - V_\sigma} - I) \ , \tag{58}$$

obtained by multiplying Eq. 57 at right with G'^{-1}_σ. Eq. 58 will be used in the QMC process for calculating the transition probability form one configuration to another and Eq. 57 for updating to the new configuration when the transition is accepted.

The QMC algorithm implies generating different field configurations with a probability proportional to their weight, given by (see Eq. 48)

$$W(s) = \det G^{-1}_\uparrow(s) \det G^{-1}_\downarrow(s). \tag{59}$$

In the QMC process, the HHS field configuration is updated by proposing local (in both the cluster site and time index) flips of the HHS field (Fig. 17). We consider such two configurations (s and s') that differ only by a flip at point m, where m denotes both a cluster site and a time (i.e. $m \equiv il, i \in \mathcal{C}$). The ratio between the weights of these configurations is

$$R = \frac{W(s')}{W(s)} = \frac{\det G^{-1}_\uparrow(s') \det G^{-1}_\downarrow(s')}{\det G^{-1}_\uparrow(s) \det G^{-1}_\downarrow(s)}. \tag{60}$$

The ratio R determines the acceptance probability of the new configuration, according to either the Metropolis[25] or the heat bath rule. If the proposed configuration was accepted, the new Green's function should be updated accordingly to Eq. 57. This implies the following

$$G'_{pn} = G_{pn} + (G_{pm} - \delta_{pm})(e^{\sigma\alpha(s'_m - s_m)} - 1) G'_{mn} \tag{61}$$

$$G'_{mn} = G_{mn} + (G_{mm} - 1)(e^{\sigma\alpha(s'_m - s_m)} - 1)G'_{mn} \tag{62}$$

$$G'_{mn} = \frac{G_{mn}}{1 - (G_{mm} - 1)(e^{\sigma\alpha(s'_m - s_m)} - 1)} \quad , \tag{63}$$

which results in

$$G'_{pn} = G_{pn} + \frac{(G_{pm} - \delta_{pm})(e^{\sigma\alpha(s'_m - s_m)} - 1)}{1 - (G_{mm} - 1)(e^{\sigma\alpha(s'_m - s_m)} - 1)} G_{mn} \quad . \tag{64}$$

$$R = \frac{W(s')}{W(s)} = \prod_\sigma \det G'^{-1}_\sigma(s') \det G_\sigma(s) = \prod_\sigma [1 - (G_{\sigma mm} - 1)(e^{\sigma\alpha(s'_m - s_m)} - 1)]. \tag{65}$$

Eq. 65 results directly from Eq. 58. Note that only the Green's function defined between clusters sites is required for the QMC procedure (Fig 18). This is also enough for measuring the cluster observables and for determining the irreducible quantities as self-energy, Σ, and two-particle vertices, Γ's.

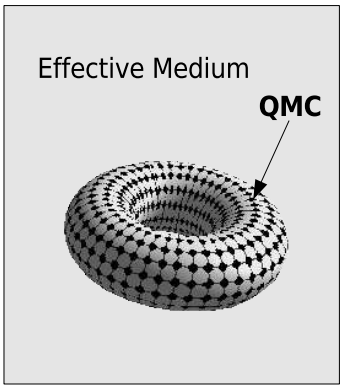

FIG. 18: Whereas the HF Hamiltonian has degrees of freedom on the cluster and in the effective medium, the algorithm may be restricted to interacting (cluster) sites only.

In order to initialize the QMC process, first the HHS fields are set to zero and G_σ is set to $\mathcal{G}_\sigma$. Here, $\mathcal{G}_\sigma$ is the noninteracting Green's function (i.e., the one obtained when $U = 0$) if we mean to simulate a finite size cluster, or it is the cluster excluded Green's function (i.e. the one obtained when we subtract the self energy from the cluster) if we mean to simulate an embedded cluster. Afterwards, by turn, the value of the HHS fields are changed to 1 or -1 at all space-time points and the Green's function is updated in accordance to Eq. 64. The obtained Green's function is used to initialize the QMC procedure. A "sweep" of the QMC procedure consists of proposing a flip of the HHS field for every time slice, calculating (Eq. 65) the ratio R between the weights of the proposed and the present configuration and accepting the flip according to the Metropolis or the heat-bath algorithm. After a certain number of warm-up sweeps through the space-time points (usually between twenty and one hundred), the system reaches equilibrium and the measurements can start. It is a good

idea to consider a few update sweeps between the measurements, in order to eliminate the correlation between them.

Numerical round-off errors accumulate in the Green's function during the updating process, due to the repeated use of Eq. 64. In order to eliminate them, after a certain number of iterations the Green's function should be refreshed. This is done first by setting G_σ to $\mathcal{G}_\sigma$ and afterwards updating it using Eq. 64 repeatedly until it corresponds to the present HHS field configuration.

1. Combining HFQMC with Quantum Cluster Methods

The Hirsch-Fye algorithm naturally produces Green's functions in Matsubara (imaginary) time. However, the self-consistency cycle of quantum cluster methods like the DMFA and DCA require imaginary frequency Green's functions. Thus the algorithm requires Fourier transforms from both the imaginary frequency to the imaginary time and from the imaginary time to the imaginary frequency. The transform from frequency to time is given by

$$G(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} G(i\omega_n)\ e^{-i\omega_n \tau} \tag{66}$$

and the inverse one, from time to frequency, by

$$G(i\omega_n) = \int_0^\beta d\tau\ G(\tau)\ e^{i\omega_n \tau} \quad . \tag{67}$$

The Matsubara frequencies for the fermionic Green's function are defined as

$$\omega_n = \frac{(2n+1)\pi}{\beta} \quad . \tag{68}$$

Due to the finite $\Delta\tau$, the numerical implementation of these transformations requires special care. We will discuss both cases next.

a. Transform from frequency to time In Eq. 66 the frequency summation is taken from minus to plus infinity. The contribution at large (both positive and negative) frequency is important, thus a truncation of the sum is not possible. The trick is to subtract and add back a function which has at large frequency the same asymptotic behavior as the Green's function. For example, if $f(i\omega_n)$ fulfills this condition, Eq. 66 can be approximated by

$$G(\tau) \approx \frac{1}{\beta} \sum_{n=-n_c}^{n_c} (G(i\omega_n) - f(i\omega_n)) e^{-i\omega_n \tau} + \frac{1}{\beta} \sum_{n=-\infty}^{\infty} f(i\omega_n)\ e^{-i\omega_n \tau}\ , \tag{69}$$

where n_c is a cutoff number chosen large enough to have negligible numerical errors.

In order to determine the Green's function behavior at large frequency we integrate Eq. 67 by parts as suggested in[26]

$$\begin{aligned} G(i\omega_n) &= \tfrac{1}{i\omega_n} G(\tau) e^{i\omega_n \tau} \mid_0^\beta - \tfrac{1}{i\omega_n} \int_0^\beta G'(\tau) e^{i\omega_n \tau} d\tau = \\ &= \tfrac{1}{i\omega_n} G(\tau) e^{i\omega_n \tau} \mid_0^\beta - \tfrac{1}{(i\omega_n)^2} G'(\tau) e^{i\omega_n \tau} \mid_0^\beta + \mathcal{O}((i\omega_n)^{-3}) \ . \end{aligned} \tag{70}$$

We find that, for cutoff values of about ≈ 500, it is necessary to consider the asymptotic behavior of G up to second order in ω_n^{-1}. If only the first order term in ω_n^{-1} is considered, in order to get negligible numerical errors the cutoff number n_c should be of order 10^6. Thus, the second order term in Eq. 70 has a major practical importance and therefore $f(i\omega_n)$ should be taken as

$$f(i\omega_n) = \frac{a_1}{i\omega_n} + \frac{a_2}{(i\omega_n)^2} \tag{71}$$

with

$$a_1 = -G(\beta^-) - G(0^+) = G(0^-) - G(0^+) \tag{72}$$

and

$$a_2 = G'(\beta^-) + G'(0^+) = -G'(0^-) + G'(0^+) \ . \tag{73}$$

The last sum in Eq. 69 can be easily calculated by using the following relations

$$\frac{1}{\beta} \sum_{n=-\infty}^{\infty} \frac{e^{-i\omega_n \tau}}{i\omega_n} = \begin{cases} \frac{1}{2} & \tau < 0 \\ -\frac{1}{2} & \tau > 0 \end{cases} \tag{74}$$

and

$$\frac{1}{\beta} \sum_{n=-\infty}^{\infty} \frac{e^{-i\omega_n \tau}}{(i\omega_n)^2} = \begin{cases} -\frac{1}{4}(\beta + 2\tau) & \tau < 0 \\ -\frac{1}{4}(\beta - 2\tau) & \tau > 0 \end{cases} \ . \tag{75}$$

The first of these relations (Eq. 74) is the well known Fourier transform of a noninteracting Green's function[22]

$$\frac{1}{\beta} \sum_{n=-\infty}^{\infty} \frac{e^{-i\omega_n \tau}}{i\omega_n - E} = \begin{cases} \frac{1}{e^{\beta E}+1} e^{-E\tau} & \tau < 0 \\ -(1 - \frac{1}{e^{\beta E}+1}) e^{-E\tau} & \tau > 0 \end{cases} \ . \tag{76}$$

with the energy pole $E = 0$. Eq. 75 results from differentiating Eq. 76 with respect to E and setting $E = 0$ afterwards.

The coefficients a_1 and a_2 in Eq. 71 can also be easily determined. Starting with the Green's function definition

$$G_{ij}(\tau) = -T_\tau \langle c_j(\tau) c_i^\dagger \rangle \ , \tag{77}$$

where i and j are space or momentum indices and using Eq. 72, a_1 becomes

$$a_1 = \langle c_i^\dagger c_j \rangle + \langle c_j c_i^\dagger \rangle = \delta_{ij} \ . \tag{78}$$

Considering

$$\frac{dc_j}{d\tau} = [H, c_j] \tag{79}$$

the coefficient a_2 is

$$a_2 = -\langle c_i^\dagger [H, c_j] \rangle - \langle [H, c_j] c_i^\dagger \rangle \ , \tag{80}$$

and can be evaluated to

$$a_2 = K_{ij} + K_{ij}^{HF} \ . \tag{81}$$

K_{ij} was defined in Eq. 46 and represent the bilinear (noninteracting) part of the Hamiltonian. K_{ij}^{HF} results from the interacting part of the Hamiltonian when the corresponding Hartree-Fock factorization is done.

b. Transform from time to frequency The difficulties associated with this transform are related with the fact that in the QMC process the Green's function is calculated and stored only in a discrete set of points τ_i evenly spaced by $\Delta\tau$. This implies a periodic Fourier transform with the period equal to the Nyquist critical frequency

$$\omega_c = \frac{\pi}{\Delta\tau} \ , \tag{82}$$

which is evidently unphysical, since the Green's function goes as ω_n^{-1} at large frequency (see Eq. 70). Besides that, due to the fact that part of the spectral density at high frequency ($\mid \omega \mid > \omega_c$) is translated into the low frequency domain ($\mid \omega \mid < \omega_c$), a phenomenon called *aliasing*[27], the Fourier transform is inaccurate even at frequencies smaller than the Nyquist frequency.

In order to cure[28] the pathology discussed above we again take advantage of our knowledge of the Green's function behavior at high frequency. A function which has the right large frequency asymptotic behavior is the Green's function G_p obtained from second order perturbation theory (or any other method which becomes exact at high frequency). The Fourier transform can be written as

$$G(i\omega_n) = G_p(i\omega_n) + \int_0^\beta d\tau (G(\tau) - G_p(\tau)) e^{i\omega_n \tau} \ . \tag{83}$$

It is important that the last term in Eq. 83 does not produce spurious high frequency contributions. Therefore before integration the Akima spline[29] is used to interpolate $G(\tau) -$

$G_p(\tau)$. The Akima spline produces smooth curves and therefore acts as a low-pass filter and eliminates the high frequency noise.

C. Continuous Time Quantum Monte Carlo

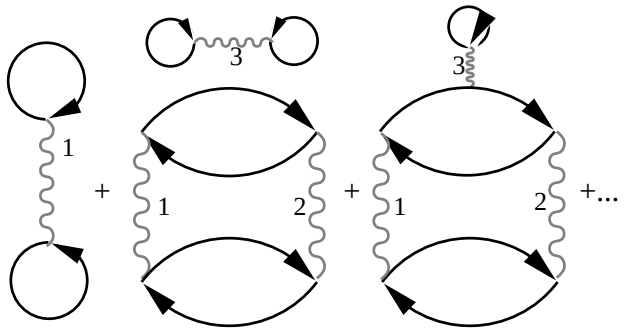

FIG. 19: In contrast to HFQMC which employs a path-integral formalism, CTQMC uses QMC to stochastically sample the graphs in the partition function.

In this subsection, we will derive a Continuous time QMC (CTQMC) algorithm for the Hubbard model, Eq. 1 following closely the derivation by Rubtsov[30]. In contrast to the path integral formalism of HFQMC, CTQMC employs the same tricks used to derive Feynman-Dyson perturbation theory (the interaction representation, the time-ordered S-matrix expansion, Wick's theorem, etc.) to stochastically generate diagrams for the partition function (Fig. 19).

Starting with

$$\exp\left(-\beta(H_0+H_1)\right) = \exp\left(-\beta H_0\right) S(\beta)\,, \tag{84}$$

where

$$S(\beta) = T_\tau \exp\left(-\int_0^\beta H_1(\tau)d\tau\right)\,, \tag{85}$$

where T_τ is the time-ordering operator, and

$$H_1(\tau) = e^{\tau H_0} H_1 e^{-\tau H_0} \tag{86}$$

is the quartic part of H in the interaction picture.

Specializing now on the Hubbard model, Eq. 1, the partition function becomes

$$Z = Z_0 \sum_k \frac{(-U)^k}{k!} \int\cdots\int d1\cdots dk \quad \langle T_\tau n_\uparrow(1)\cdots n_\uparrow(k)\rangle_0 \langle T_\tau n_\downarrow(1)\cdots n_\downarrow(k)\rangle_0 \tag{87}$$

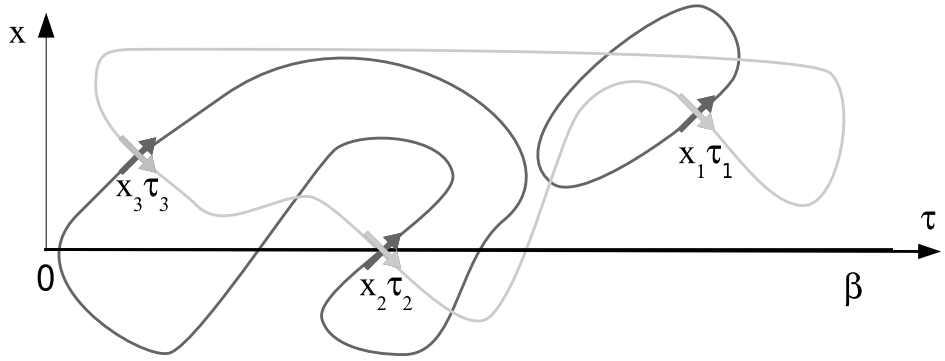

FIG. 20: An example of a third-order ($k = 3$) CTQMC graph. Many graphs may be drawn to third order in U, one example is shown.

where $1 \rightarrow (x_1, \tau_1)$, etc. and $\langle\rangle_0$ indicates a thermodynamic average with respect to $\exp(-\beta H_0)$. Since H_0 is noninteracting, we can apply the Wick's theorem to evaluate the expectation values in Eq. 87 (see, for instance,[22]), by evaluating and summing over all closed Feynman graphs, see for example Fig. 20. For each order in k, there are $k!$ graphs that can be drawn for each of the up and down electrons. Note that the graphs can include both connected and disconnected parts, as shown in the figure. According to Wick's theorem, each graph corresponds to a product of non-interacting Green's functions G^0 (i.e. Green's functions from the solution of H_0 only) with arguments determined by the vertex labels, etc., and sign determined by the number of line crossings. For example, the graph shown is

$$G^0(2,3)G^0(3,2)G^0(1,1)G^0(3,2)G^0(2,1)G^0(1,3) \tag{88}$$

where G^0 with the same argument are to be interpreted as having the right time argument slightly greater. I.e., $G^0(1,1) = G^0(x_1\tau_1; x_1\tau_1+)$. There will be 36 such graphs for each $k = 3$ vertex configuration, and the number of graphs will increase with order like $k!^2$. Fortunately, we may conveniently represent them as the product of up and down determinants

$\det D_\uparrow(k) \det D_\downarrow(k)$, where

$$D_\sigma(k) = \begin{pmatrix} G^0_\sigma(1,1) & G^0_\sigma(1,2) & \cdots & G^0_\sigma(1,k) \\ G^0_\sigma(2,1) & G^0_\sigma(2,2) & \cdots & G^0_\sigma(2,k) \\ G^0_\sigma(3,1) & G^0_\sigma(3,2) & \cdots & G^0_\sigma(3,k) \\ . & . & . & \\ . & . & . & . \\ . & . & . & . \\ G^0_\sigma(k,1) & G^0_\sigma(k,2) & \cdots & G^0_\sigma(k,k) \end{pmatrix} \tag{89}$$

where, in lieu of an external field the bare Green's function does not need a spin label, $G^0_\sigma(k,2) = G^0(k,2)$. Note that $D_\sigma(k)$ is a $k \times k$ matrix.

In CTQMC, the sum of integrals in the Eq. 87 is evaluated using an importance sampling Monte Carlo algorithm. Suppose the cluster has N sites and $L = \beta/\Delta\tau$ time locations (in the end we may let $\Delta\tau \to 0$). We will consider Monte Carlo moves of adding vertices at a random location in space-time or subtracting existing vertices. By combining such moves, we can clearly access all possible vertex configurations. Consider two configurations, one with k and the other with $k+1$ vertices, then the weights W of these two configurations are given by the integrand of Eq. 87, or, in terms of the $D_\sigma(k)$,

$$\begin{aligned} W_k &= (-\Delta\tau U)^k \det D_\uparrow(k) \det D_\downarrow(k) \\ W_{k+1} &= (-\Delta\tau U)^{k+1} \det D_\uparrow(k+1) \det D_\downarrow(k+1)\,. \end{aligned} \tag{90}$$

If we let the probability to add and remove a vertex be P_a and P_r respectively, where $P_a + P_r = 1$, then the detailed balance condition requires that

$$P_a \frac{1}{L}\frac{1}{N} W_k P_{k\to k+1} = P_r \frac{1}{k+1} W_{k+1} P_{k+1\to k}\,. \tag{91}$$

Here $\frac{1}{L}\frac{1}{N}$ is the probability to choose a position in time and space for the vertex you intend to add while $\frac{1}{k+1}$ is the probability to choose one vertex you intend to remove of from the existing $k+1$ ones. If we accept or reject this change using a Metropolis algorithm[25], then we need the Metropolis ratio R.

Suppose we propose to add vertex, then

$$R = P_{k\to k+1}/P_{k+1\to k} = -\frac{U\beta N}{k+1} \prod_\sigma \det D_\sigma(k+1) / \prod_\sigma \det D_\sigma(k)\,. \tag{92}$$

We need an efficient way to calculate the ratio of determinants

$$\begin{aligned} \det D_\sigma(k+1)/\det D_\sigma(k) &= \det D_\sigma(k+1)M_\sigma(k) \\ &= \det\left(I + (D_\sigma(k+1) - D_\sigma(k))M_\sigma(k)\right) \end{aligned} \tag{93}$$

where $M_\sigma(k) = D_\sigma(k)^{-1}$, and we must pad the matrices $M_\sigma(k)$ and $D_\sigma(k)$ with an extra row and column in the $k+1$ location with all zeroes, except for a 1 in the $k+1$, $k+1$ location.

$$\det\left(I + (D_\sigma(k+1) - D_\sigma(k))M_\sigma(k)\right) = \det \tag{94}$$
$$\begin{pmatrix} 1 & 0 & \cdots & G^0_\sigma(1,k+1) \\ 0 & 1 & \cdots & G^0_\sigma(2,k+1) \\ 0 & 0 & \cdots & G^0_\sigma(3,k+1) \\ . & . & . & \\ . & . & . & . \\ . & . & . & . \\ G^0_\sigma(k+1,i)M(k)_{i,1} & G^0_\sigma(k+1,i)M(k)_{i,2} & \cdots & G^0_\sigma(k+1,k+1) \end{pmatrix}$$

A cofactor expansion (first on the last column and then on the last row) yields

$$\det D_\sigma(k+1)/\det D_\sigma(k) = G^0_\sigma(k+1,k+1) - G^0_\sigma(k+1,i)M(k)_{ij}G^0_\sigma(j,k+1) \tag{95}$$

with an implied sum over repeated indices. Clearly, the matrices M are important for the update, and they, not D are stored. The change is accepted if R is greater than a random number between zero and one. Then, we need an equation used to update M which may be derived using the Inversion by Partitioning described in Numerical Recipes[27].

$$M_\sigma(k+1) = \begin{pmatrix} . & . & . & -\lambda^{-1}L_{1k+1} \\ . & . & . & -\lambda^{-1}L_{2k+1} \\ . & {M_\sigma}' & . & -\lambda^{-1}L_{3k+1} \\ . & . & . & \\ . & . & . & . \\ -\lambda^{-1}R_{k+1,1} & -\lambda^{-1}R_{k+1,2} & \cdots & -\lambda^{-1} \end{pmatrix} \tag{96}$$

where the first k by k section is filled by the matrix

$${M_\sigma}'_{ij} = M(k)_{ij} + L_{ik+1}\lambda^{-1}R_{k+1,j} \tag{97}$$

and

$$R_{ij} = G^0_\sigma(i,n)M(k)_{nj} \;\; L_{ij} = M(k)_{in}G^0_\sigma(n,j) \tag{98}$$

and λ is given by Eq. 95.

If we propose to remove vertex at location n, then to determine whether to accept the move, we need

$$\det D_\sigma(k-1)/\det D_\sigma(k) = \det\left((M_\sigma(k) - M_\sigma(k-1))D_\sigma(k-1) + I\right) . \tag{99}$$

Again, using a cofactor expansion, it is easy to show that this is just $M_{\sigma nn}$. Then if

$$R = \frac{-k}{U\beta N}\prod_\sigma \det D_\sigma(k-1)/\prod_\sigma \det D_\sigma(k) \tag{100}$$

is greater than a random number between zero and one, we accept the change and must update M_σ using

$$M_{\sigma ij}(k-1) = M_{\sigma ij}(k) - M_{\sigma in}(k)M_{\sigma nj}(k)/M_{\sigma nn}(k) . \tag{101}$$

One natural question is to ask whether this QMC is convergent, which would require that the average value of k be finite.

$$\begin{aligned}
\langle k\rangle &= \frac{Z_0}{Z}\sum_0^\infty \frac{(-1)^k}{k!}k\int_0^\beta d\tau_1\cdots d\tau_k \left\langle T_\tau H_1(\tau_1)\cdots H_1(\tau_k)\right\rangle_0 \\
&= -\frac{Z_0}{Z}\sum_0^\infty \frac{(-1)^k}{k!}k\int_0^\beta d\tau_1\cdots d\tau_k d\tau \left\langle T_\tau H_1(\tau_1)\cdots H_1(\tau_k)H_1(\tau)\right\rangle_0 \\
&= -\int_0^\beta d\tau \left\langle H_1(\tau)\right\rangle \\
&= -\beta\left\langle H_1\right\rangle
\end{aligned} \tag{102}$$

where the brackets $\langle\rangle$ denote the average with respect to full interacting Hamiltonian (Eq. 33), which is also the average value of Monte Carlo measurements. The last line in Eq. 103 follows if H_1 is independent of τ. Since $\langle H_1\rangle$ is extensive, this means that $\langle k\rangle = -\beta N\langle h_1\rangle$ where h_1 is the average potential energy per site.

As will be described below in Sec. III D 1, the minus sign problem emerges when the Metropolis ratio R is not positive definite. It is easy to see from Eqs. 92 and 100 that when $U < 0$ there is no minus sign problem since the up and down determinants the same and real, so their product is positive. However, the minus sign problem can be severe when $U > 0$ since the sign of the sampling weight will fluctuate as k changes by ± 1. Some tricks can

be used to control the sign problem. For example, at half filling where there is particle-hole symmetry, we rewrite the interaction as

$$H_1 = U \sum_i \left(n_{\uparrow i} - \frac{1}{2} \right) \left(n_{\downarrow i} - \frac{1}{2} \right) \tag{103}$$

then the particle-hole transformation $c_{j\downarrow} \rightarrow c^\dagger_{j\downarrow}$ changes the sign of U eliminating the sign problem. Away from half filling, this trick is useless. However, here we introduce an additional Ising-like auxiliary field $s_i = \pm 1$ on each site and rewrite the interaction as[31]

$$H_1 = \frac{U}{2} \sum_{i,s_i} \left(n_{\uparrow i} - \frac{1}{2} - s_i \alpha \right) \left(n_{\downarrow i} - \frac{1}{2} + s_i \alpha \right) . \tag{104}$$

At least in 0 or 1 dimension, the choice $\alpha > 1/2$ eliminates the sign problem. The cost is that the Ising field is now associated with each point in space-time and must be sampled using the QMC algorithm (Fig. 21).

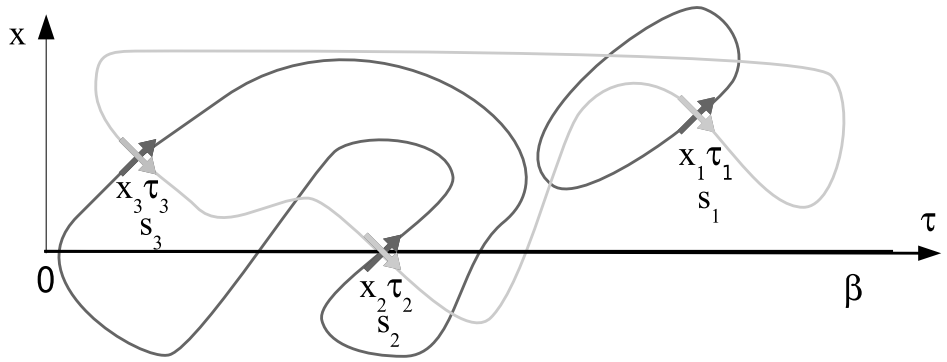

FIG. 21: An example of a third-order ($k = 3$) CTQMC graph with the Rubtsov auxiliary field used to suppress the minus sign configuration. Now each vertex also carries a field $s_i = \pm 1$ which must be updated along with the vertex locations in the QMC.

Since Wick's theorem applies to the configurations of CTQMC, most measurements are simply accomplished by forming the appropriate Wicks contractions of the Green's function

$$G_\uparrow(i,j) = -\left\langle T_\tau c_\uparrow(i) c^\dagger_\uparrow(j) \right\rangle . \tag{105}$$

corresponding to a configuration of vertices (and auxiliary fields when appropriate).

$$G_\uparrow(i,j) = -\frac{Z_0}{Z} \sum_k \frac{(-U)^k}{k!} \int \cdots \int d1 \cdots dk \quad \left\langle T_\tau c_\uparrow(i) c^\dagger_\uparrow(j) n_\uparrow(1) \cdots n_\uparrow(k) \right\rangle_0 \left\langle T_\tau n_\downarrow(1) \cdots n_\downarrow(k) \right\rangle_0 . \tag{106}$$

However, this formula is very similar to the one we encountered when we added a vertex. In fact,

$$G_\uparrow(i,j) = \det D_\uparrow(k+1)/\det D_\uparrow(k)\,, \tag{107}$$

where the $k \times k$ matrix $D_\uparrow(k)$ is given by Eq. 89 and

$$D_\uparrow(k+1) = \begin{pmatrix} {G_\uparrow}^0(1,1) & {G_\uparrow}^0(1,2) & \cdots & {G_\uparrow}^0(1,k) & {G_\uparrow}^0(1,j) \\ {G_\uparrow}^0(2,1) & {G_\uparrow}^0(2,2) & \cdots & {G_\uparrow}^0(2,k) & {G_\uparrow}^0(2,j) \\ {G_\uparrow}^0(3,1) & {G_\uparrow}^0(3,2) & \cdots & {G_\uparrow}^0(3,k) & {G_\uparrow}^0(3,j) \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ {G_\uparrow}^0(k,1) & {G_\uparrow}^0(k,2) & \cdots & {G_\uparrow}^0(k,k) & {G_\uparrow}^0(k,j) \\ {G_\uparrow}^0(i,1) & {G_\uparrow}^0(i,2) & \cdots & {G_\uparrow}^0(i,k) & {G_\uparrow}^0(i,j) \end{pmatrix}. \tag{108}$$

We obtain the familiar form for

$$G_\uparrow(i,j) = {G_\uparrow}^0(i,j) - {G_\uparrow}^0(i,p) {M_\uparrow}_{pq} {G_\uparrow}^0(q,j)\,. \tag{109}$$

Of course, a similar equation holds for $G_\downarrow$.

All other measurements are then formed from Wick's contractions of these Green's functions, as described in standard many-body texts[22]. This is worth illustrating for one example, such as the transverse component of the spin susceptibility

$$\chi^\pm(i,j) = -\left\langle T_\tau c_\uparrow^\dagger(i) c_\downarrow(i) c_\downarrow^\dagger(j) c_\uparrow(j) \right\rangle\,. \tag{110}$$

The form for this is strikingly similar to what we did above for $G_\uparrow(i,j)$, the main difference being that we now have both spin components

$$\chi^\pm(i,j) = \frac{Z_0}{Z} \sum_k \frac{(-U)^k}{k!} \int d1 \cdots dk \quad \left\langle T_\tau c_\uparrow^\dagger(i) c_\uparrow(j) n_\uparrow(1) \cdots n_\uparrow(k) \right\rangle_0 \left\langle T_\tau c_\downarrow^\dagger(j) c_\downarrow(i) n_\downarrow(1) \cdots n_\downarrow(k) \right\rangle_0\,. \tag{111}$$

So the estimator is

$$\chi^\pm(i,j) \;=\; \left\langle \left({G_\uparrow}^0(j,i) - {G_\uparrow}^0(j,p) {M_\uparrow}_{pq} {G_\uparrow}^0(q,i) \right) \left({G_\downarrow}^0(i,j) - {G_\downarrow}^0(i,p) {M_\downarrow}_{pq} {G_\downarrow}^0(q,j) \right) \right\rangle_{QMC}\,. \tag{112}$$

Similar estimators may be formed from $G_\uparrow(i,j)$ and $G_\downarrow(i,j)$ for the longitudinal spin and pair susceptibilities. These will be discussed below.

In order to initialize the QMC process, the number of vertices can be simply set to zero. Subsequent Monte Carlo steps will tend to add vertices until the configuration is thermalized. Alternatively, one can take any other k vertex configuration and use $\mathcal{G}_\sigma$ to construct the D_σ matrices (which are then inverted to obtain M_σ matrices). Here, $\mathcal{G}_\sigma$ is the noninteracting Green's function (the one obtained when $U = 0$) if we mean to simulate a finite size cluster, or its the cluster excluded Green's function (the one obtained when we subtract the self energy from the cluster) if we mean to simulate an embedded cluster.

Numerical round-off errors accumulate in the M_σ matrices during the updating process, Eqs. 97 and 101. In order to eliminate them, after a certain number of iterations they should be refreshed. This may be done by recalculating the D_σ matrices and inverting for the M_σ.

1. Combining CTQMC with Quantum Cluster Methods

Quantum cluster methods generally require the Green's function in terms of Matsubara frequency and wavevector, $G(k, i\omega_n)$. However, unlike HFQMC, there is no difficulty in performing the Fourier transform in CTQMC. In fact, this measurement may be made efficiently directly in terms of Matsubara frequency and wavevector, simply by performing a double Fourier transform of the Green's function estimator in Eq. 109

$$G_\uparrow(K_i) = {G_\uparrow}^0(K_i) - {G_\uparrow}^0(K_i) M_\uparrow(K_i) {G_\uparrow}^0(K_i)\,, \tag{113}$$

where $K = (\mathbf{K}, i\omega_n)$ is now a frequency-wavevector label, and $M_\uparrow$ is

$$M_\uparrow(K_i) = M_\uparrow(\mathbf{K}, i\omega_n) = \sum_{i,j} \exp\left(i\mathbf{K}\cdot(\mathbf{X}_i - \mathbf{X}_j) - i\omega_n(\tau_i - \tau_j)\right) {M_\uparrow}_{ij} \tag{114}$$

where i and j label the space-time locations of the vertices.

D. Making and Conditioning Measurements

The natural byproduct of the QMC algorithm are the Green's functions. These may be used to make measurements of most one and two-particle properties using standard diagrammatic techniques. In doing so, several points must be remembered:

- In the HF QMC algorithm, the Hubbard-Stratonovich transformation reduces the problem to one of free electrons moving in a time-dependent Hubbard-Stratonovich field. In the CTQMC, an interaction representation together with an S-matrix expansion is used. Thus, for each field configuration in the HFQMC or any vertex configuration in the CTQMC, measurements may be formed by summing all allowed Wick's contractions. The full interacting quantity is recovered by QMC averaging this over all configurations.

- It is important to use *all* allowed Wick's contractions, both connected and disconnected, in this series.

- If your Hamiltonian is invariant under translations in space and time, and you are performing a simulation in an unbroken symmetry phase, it is important to average your measurement over all time and space differences in order to achieve the lowest variance estimator.

- It is also important to average over other symmetries of the Hamiltonian, which may not be preserved by the CTQMC or HFQMC algorithms (e.g., spin symmetry).

For example, consider the local impurity magnetic correlation function

$$\begin{aligned} \chi_{ii}(\tau) &\approx \left\langle S_i^+(\tau) S_i^-(0) \right\rangle \\ &\approx \left\langle C_{i\uparrow}^\dagger(\tau) C_{i\downarrow}(\tau) C_{i\downarrow}^\dagger(0) C_{i\uparrow}(0) \right\rangle \\ &\approx \frac{T}{2N} \sum_{i\sigma} \int_0^\beta d\tau' \left\langle G_\sigma(x_i, \tau+\tau'; x_i\tau') G_{-\sigma}(x_i\tau'; x_i\tau+\tau') \right\rangle_{QMC} \end{aligned} \tag{115}$$

where the QMC subscript means that the Monte Carlo average over the Hubbard-Stratonovich fields or vertex configurations is still to be performed. Note that in the last step in Eq. 116 we form all allowed Wick's contractions and average over all equivalent time differences, spins, and sites to reduce the variance of this estimator.

At this point the measurements for CTQMC and HFQMC differ. For CTQMC we essentially have a continuum of Matsubara time, so the time integral in Eq. 116 may be completed very accurately. However, in HFQMC care must be used to reduce the time-step error. Due to time translational invariance, the integral over τ' is not terribly sensitive to $\Delta\tau$ error, so we approximate as a sum using a rectangular approximation. For $\tau > 0$

$$\chi(\tau_l) \approx \frac{1}{2L} \sum_{\sigma,l'} \left\langle G_\sigma(i, ind(l+l'); i, l') G_{-\sigma}(i, l'; i, ind(l+l')) \right\rangle_{QMC} , \tag{116}$$

where $ind(l)$ is the smaller nonnegative value of either l or $l-L$. For $\tau = 0$ the fact that in HFQMC we always store $G_\sigma(l',l') = G_\sigma(\tau_{l'} + 0^+, \tau_{l'})$ requires us to modify the measurement

$$\chi(\tau = 0) \approx \frac{1}{2L} \sum_{\sigma,l'} \langle G_\sigma(l',l') \left(G_{-\sigma}(l',l') - 1\right) \rangle_{QMC} \ . \tag{117}$$

Finally the susceptibility may be calculated by

$$\chi(T) = \int_0^\beta d\tau \chi(\tau) \approx \sum_l sf(l) \Delta\tau \chi(\tau_l) \, , \tag{118}$$

where the Simpson factor $sf(l) = 2\Delta\tau/3$ $(4\Delta\tau/3)$ for odd (even) l is used to reduce the systematic error of the integral.

As a final example, consider the cluster particle-particle Green's function matrix $\chi_c(q, K, K')$ $(K = (\mathbf{K}, i\omega_n))$ which in the cluster space-time takes the form

$$\chi_c(X_1, X_2, X_3, X_4) = \left\langle T_\tau c_\uparrow(X_1) c_\downarrow(X_2) c_\downarrow^\dagger(X_3) c_\uparrow^\dagger(X_4) \right\rangle \ . \tag{119}$$

Here X_i is in the space-(imaginary)time notation $X_i = (\mathbf{X}_i, \tau_i)$, where the points $\mathbf{X}_i$ are on the corresponding reciprocal cluster of $\mathbf{K}$ in real space.

Since the storage associated with this quantity is quite large, it cannot be measured for many times. Thus the measurement in CTQMC and HFQMC are quite similar. First, using Wick's theorem, its value is tabulated for each field configuration and then transformed into the cluster Fourier space. Second, we Monte Carlo average over these configurations. After the first step, the expression for the above two-particle Green's function in the cluster momentum-frequency space becomes

$$\begin{aligned} \chi_c(\mathbf{Q}, i\nu_n; \mathbf{K}, i\omega_n; \mathbf{K}', i\omega_{n'}) \ = \ & \Bigg\langle \sum_{X_1, X_4} e^{iK'X_1} G_{c\uparrow}(X_1, X_4) e^{-iKX_4} \\ & \sum_{X_2, X_3} e^{i(Q-K')X_2} G_{c\downarrow}(X_2, X_3) e^{-i(Q-K)X_3} \Bigg\rangle_{QMC} \ . \end{aligned} \tag{120}$$

where K is the momentum-frequency point $K = (\mathbf{K}, i\omega_n)$.

The sums (integrals) over τ in Eq. 120 require special consideration. Since the Green's functions change discontinuously when the two time arguments intersect, the best applicable integral approximation is the trapezoidal approximation. Using this, we will run into Green's functions $G_c(\mathbf{X}, \tau; \mathbf{X}, \tau)$ with both time and space arguments the same. In the HFQMC algorithm, this is stored as $G_c(\mathbf{X}, \tau^+; \mathbf{X}, \tau)$ (i.e. it is assumed that the first time argument is slightly greater than the second) and in CTQMC, the other time ordering is assumed;

however, if we replaced the equal time Green's function to be the average $\{G_c(\mathbf{X},\tau^+;\mathbf{X},\tau) + G_c(\mathbf{X},\tau;\mathbf{X},\tau^+)\}/2 \;=\; G_c(\mathbf{X},\tau^+;\mathbf{X},\tau) - 1/2 \;=\; G_c(\mathbf{X},\tau^;\mathbf{X},\tau+) + 1/2$ then a trapezoidal approximation of the integrals results. If we call the matrix $\mathbf{G}_c$, with $1/2$ subtracted (added) from its diagonal elements as appropriate for HFQMC (CTQMC), as $\hat{\mathbf{G}}_c$ (note that we can treat one of the three independent momenta involved in χ_c as a variable Q outside the matrix structure), then we can write the two-particle Green's function in a matrix form

$$\begin{aligned}&\chi_{c\imath j}(\mathbf{Q}) = \\ &\left\langle \left(\mathbf{F}^\dagger_{\mathbf{Q}=0}\hat{\mathbf{G}}_{c\uparrow}\mathbf{F}_{\mathbf{Q}=0}\right)_{\imath j}\left(\mathbf{F}^\dagger_{\mathbf{Q}}\hat{\mathbf{G}}_{c\downarrow}\mathbf{F}_{\mathbf{Q}}\right)^*_{ij}\right\rangle_{QMC},\end{aligned} \tag{121}$$

where $(\mathbf{F}_{\mathbf{Q}})_{ij} = \Delta\tau e^{-i(\mathbf{K}_j-\mathbf{Q}).\mathbf{X}_i - i\omega_j\tau_i}$ where we have chosen $\imath$ and j to index the cluster momentum-frequency space.

This measurement may be performed efficiently if the product of three matrices in each set of parenthesis is tabulated as two sequential matrix-matrix products and stored before the direct product between the terms in parenthesis is calculated. When done this way, the calculation time required for this process scales like $(NL)^3$ rather than $(NL)^4$ as would result from a straight-forward evaluation of the sums implicit in Eq. 121. Greater efficiency can be obtained if we perform the Fourier transforms as a two-step process; i.e., first doing the transform in time and then in space. Then the measurement scales like $(N+L)(NL)^2$, of course, these reductions in FLOPs require an increase in memory needed to store the intermediate results.

CTQMC presents the possibility of measuring these two-particle susceptibilities directly in the cluster momentum-frequency space without the need to perform the discrete Fourier transform in Matsubara time presented in Eq. 120. Again, this measurement may be made efficiently again performing a double Fourier transform of the Green's function estimator Eq. 109

$$G_\uparrow(K_i,K_j) = {G_\uparrow}^0(K_i)\delta_{ij} - {G_\uparrow}^0(K_i)M_\uparrow(K_i,K_j){G_\uparrow}^0(K_j)\,, \tag{122}$$

and $M_\uparrow$ is

$$\begin{aligned}M_\uparrow(K_n,K_m) \;&=\; M_\uparrow(\mathbf{K}_n,i\omega_n;\mathbf{K}_m,i\omega_m;) \\ &=\; \sum_{i,j}\exp\left(i\mathbf{K}_n\cdot\mathbf{X}_i - i\omega_n\tau_i)\right)M_{\uparrow ij}\exp\left(-i\mathbf{K}_m\cdot\mathbf{X}_j + i\omega_m\tau_j)\right),\end{aligned} \tag{123}$$

where i and j label the space-time locations of the vertices. The last step is the numerical bottleneck. However,using the methods described above of performing the Fourier

transforming steps and storing the intermediate result, this step may be performed in $\mathcal{O}2(L+N)k^2)$ FLOPs. The estimator of the cluster particle-particle then becomes

$$\chi_c(Q,K,K') = \langle G_\uparrow(K'+Q,K+Q)G_\downarrow(-K',-K)\rangle_{QMC} \ . \tag{124}$$

1. *The minus sign problem*

Except in a few and fortunate situations, the calculation of physical quantities with QMC reduces to the calculation of integrals/sums with an integrand which can take both positive and negative values. This is the origin of the *sign problem* in the Monte Carlo simulations.

For example, the average of the observable A is formally given by

$$\langle A\rangle = \frac{\int W(x)A(x)dx}{\int W(x)dx} \tag{125}$$

where $\int dx$ is a integral over all the possible configurations x of the system of weight $W(x)$. For a classical system $W(x)$ is positive definite and a Monte Carlo process consists of stochastically generating configurations x_i according to their weight $W(x_i)$ and taking

$$\langle A\rangle = \frac{1}{N_d}\sum_{i=1}^{N_d} A(x_i) \ . \tag{126}$$

Where N_d is the number of measurements. For quantum systems $W(x)$ is not necessarily positive definite and therefore cannot be considered a probability distribution. The solution is to rewrite Eq. 125 as

$$\langle A\rangle = \frac{\int \mid W(x) \mid A(x)s(x)dx}{\int \mid W(x) \mid dx}\frac{\int \mid W(x) \mid dx}{\int \mid W(x) \mid s(x)dx} \tag{127}$$

where $s(x)$ is the sign of $W(x)$, i.e.

$$W(x) = \mid W(x) \mid s(x) \ . \tag{128}$$

Thus, we can sample the configurations space with the probability $\mid W(x) \mid$ and

$$\langle A\rangle = \frac{\sum_{i=1}^{N_d} A(x_i)s(x_i)}{\sum_{i=1}^{N_d} s(x_i)} = \frac{\langle As\rangle_{|W|}}{\langle s\rangle_{|W|}} \ . \tag{129}$$

Thus, we may use the absolute value of $|W|$ as the sampling weight as long as we measure the product of the sign and measurement and divide by the average sign. If the average sign $\langle s\rangle_{|W|}$ is finite and not very small, then this approach is successful. However, a problem

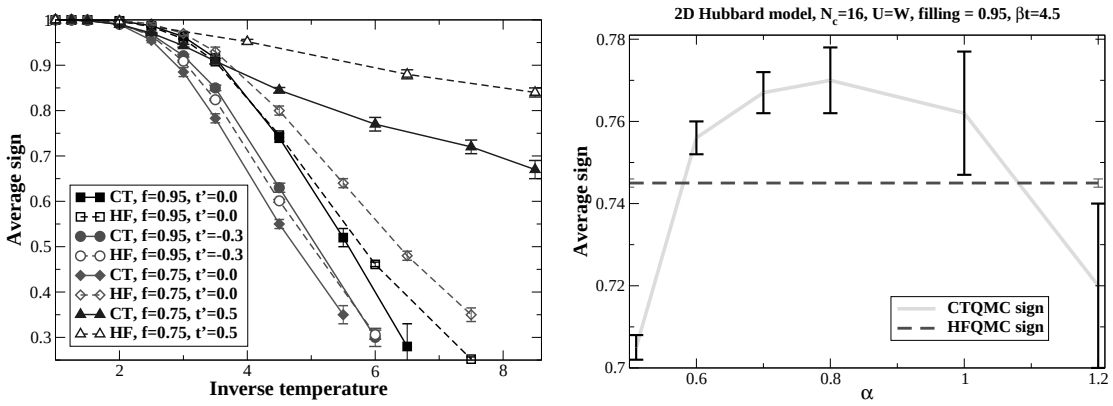

FIG. 22: Comparison of the average sign from a CTQMC and HFQMC DCA simulation of the two-dimensional Hubbard model. (Left) Average sign versus inverse temperature for 4-site cluster with $U = W = 8t$ and different fillings. (Right) the average sign for 16-site cluster with $U = W = 8t$ versus α from Eq. 104 at fixed filling and temperature.

occurs when $\langle s \rangle_{|W|}$ is extremely small, because a small sign strongly amplifies the error bar associated with $\langle A \rangle_{|W|}$ ($\propto \delta s / \langle s \rangle^2$). In fact, for many quantum systems $\langle s \rangle$ is found to fall exponentially to zero with decreasing temperature or increasing cluster size[55]. This makes Monte Carlo calculations impractical at low temperatures.

The average sign obtained from a simulation of the 2D Hubbard model is shown in Fig. 22. The average sign from HFQMC and CTQMC are similar but the CTQMC result is slightly worse. The figure on the right shows that increasing the value of α from Eq. 104 does not help to increase the average sign significantly and it also increases the error bar on the sign measurement due to an increase in the average order $\langle k \rangle$. The error bar increases, since the average order $\langle k \rangle$ increases like α^2 for large α as may be seen from Eq. 103, or here (for some site i).

$$\begin{aligned} \langle k \rangle &= -\beta \langle H_1 \rangle \\ &= -N\beta \left\langle \sum_{s_i} \left(n_{\uparrow_i} - \frac{1}{2} - s_i \alpha \right) \left(n_{\downarrow_i} - \frac{1}{2} + s_i \alpha \right) \right\rangle \end{aligned} \tag{130}$$

On the other hand, the average sign obtained from a simulation of the 1D Hubbard model is shown in Fig. 23. Whereas the average sign from HFQMC depends weakly with cluster size, that from CTQMC actually *increases* strongly with increasing cluster size. Also studied is an additional coupling between chains t_{perp}/t which is a hopping between equivalent sites on adjacent chains. A finite t_{perp}/t causes the average sign to increase strongly.

We find that the average sign depends on many things: fillings, dimensionality model parameters, etc. It is difficult to predict *a priori* whether the average sign will be problematic.

2. CTQMC and real time measurements

In this section, we briefly discuss a possibility to extend CTQMC to measure real time Green's functions or even treat the systems out of thermal equilibrium. We start with the same Hubbard model Hamiltonian [1], which has no time dependent terms, and proceed to derive expressions for partition function and real time Green's functions. In this formalism, the partition function can be expressed as an expansion of contour ordered exponent:

$$Z = Z_0 \sum_{k=0}^{\infty} \frac{(-i)^k}{k!} \int_c dt_1 \dots \int_c dt_k \, \langle T_c H_1(t_1) \dots H_1(t_k) \rangle_0 \ , \tag{131}$$

where

$$H_1(t) = e^{iH_0 t} H_1 e^{-iH_0 t} \tag{132}$$

is the interaction (quartic) part of Hamiltonian in the interaction representation and T_c is the contour-ordering operator[32]. The integrals are performed along the contour c, shown in Fig. 24. By writing out the contributions to the partition function coming from lowest orders in expansion (Eq. 131), one can see that all the terms containing $H_1(t)$ on real parts of contour cancel out exactly and thus the partition function in this case is the same as in Matsubara formalism. We now consider the expansion for the contour ordered Green's function:

$$\begin{aligned} G_c(t,t') &= \left\langle T_c c(t) c^\dagger(t') \right\rangle \\ &= -i \frac{Z_0}{Z} \sum_{k=0}^{\infty} \frac{(-i)^k}{k!} \int_c dt_1 \dots \int_c dt_k \left\langle T_c c(t) c^\dagger(t') H_1(t_1) \dots H_1(t_k) \right\rangle_0 . \end{aligned} \tag{133}$$

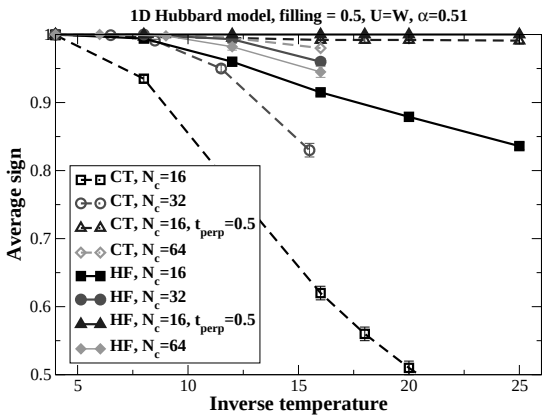

FIG. 23: Comparison of the average sign from a CTQMC and HFQMC DCA simulation of the one-dimensional Hubbard model. The average sign of HFQMC depends weakly on the cluster size; whereas the average sign of CTQMC increases strongly with increasing cluster size.

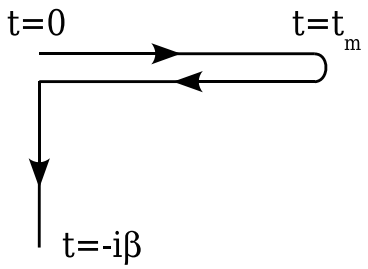

FIG. 24: The contour for measuring the real time Green's function.

In this case, however, the contributions containing $H_1(t)$ on real parts of contour do not cancel out (as one can easily check by writing out $k = 1$ term in the expansion). The application of Wick's theorem yields CTQMC diagrams that contain vertices on both imaginary as well as real parts of contour (See Fig. 25). Thus, the partition function and the Green's function have dominant contributions coming from entirely different regions in the k-dimensional space spanned by time coordinates along the contour. As a consequence, it appears impossible to use the partition function as a sampling weight in Monte Carlo algorithm to measure the Green's function. It is still possible to use the Green's function

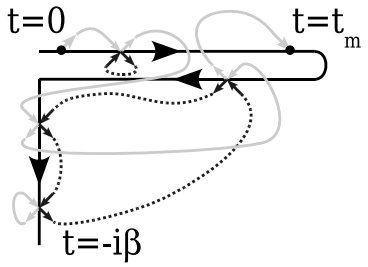

FIG. 25: A $k = 4$ diagram for the contour-ordered spin-up Green's function. The light continuous and darker dotted lines represent the propagators for spin up and spin down electrons, respectively.

itself as a weight in the Monte Carlo importance sampling, and then measure it's value as a distribution in the space of it's arguments (t, t'). Since the Green's function in this case would be a complex number, it's modulus would be taken for the sampling weight instead. The sign problem would now translate into the phase problem, if the average value of the phase of the Green's function gets too small. By looking at Eq. 134 we see that the phase problem is likely to be very severe due to a factor of $(-i)$ attached to every vertex on the real part of the contour (unless one finds again a way to introduce an auxiliary field that

remedies this problem). In any case, the oscillatory nature of real time Green's function implies that the phase problem would most likely restrict measurements of Green's function to small values of its real time arguments.

E. Summary

We describe two QMC algorithms for solving the embedded cluster (impurity) problem used in cluster mean-field methods. In the HFQMC algorithm the interaction term is replaced by a summation over a set of auxiliary fields, with every field configuration defining a noninteracting problem. In the CTQMC algorithm, an expansion in the interaction term in the partition function is employed. The problem is reduced to an infinite sum of integrals with different multiplicity. Measurements, the sign problem and a possible extension of the CTQMC technique to non-equilibrium processes is discussed.

IV. ANALYTIC CONTINUATION OF QUANTUM MONTE CARLO DATA

A. Introduction

Most quantum Monte Carlo (QMC) simulations produce Green's functions $G(\tau)$ of imaginary time $\tau = it$. However, real-frequency results are crucial since most experiments probe dynamical quantities, including transport, densities of states, nuclear magnetic resonance, inelastic scattering, etc. Thus, the inability to extract real-frequency or real-time results from Matsubara (imaginary) time QMC simulations presents a significant limitation to the usefulness of the method. The relation between $G(\tau)$ and $A(\omega) = -\frac{1}{\pi}\mathrm{Im}G(\omega)$ is linear and surprisingly simple

$$G(\tau) = \int d\omega K(\tau, \omega) A(\omega)\,. \tag{134}$$

Nevertheless, inversion is complicated by the exponential nature of the kernel. For example, for a Fermionic single-particle Green's function G, $K(\tau, \omega) = e^{-\tau\omega}/(1 + e^{-\beta\omega})$[33]. For finite τ and large ω the kernel is exponentially small, so that $G(\tau)$ is insensitive to the high frequency features of $A(\omega)$. Equivalently, if we approximate both G and A by equal-length vectors and K by a square matrix, then we find that the determinant of K is exponentially small, so that K^{-1} is ill-defined. Apparently, there are an infinite number of A with very different

characters (many of them not even causal) that yield the same G.

Previous attempts to address this problem include least-squares fits, Padé approximants and regularization. In the least squares method, Schüttler and Scalapino[33] approximated the spectrum with a set of box functions. The location and weight of these functions was determined by minimizing the least-squares misfit between the spectrum and the QMC data. However, as the number of box functions is increased to better resolve features in the spectrum, the fit becomes unstable and noisy. In the Padé method[34], G (or rather its Fourier transform) is fit to a functional form, usually the ratio of two polynomials, which is then analytically continued formally by replacing $i\omega_n \to \omega + i0^+$. This technique works when the data G is very precise, as when analytic continuing Eliashberg equations, or when the fitting function is known *a priori*. However, it is generally unreliable for the continuation of less-precise QMC data to real frequencies. A more useful approach is to introduce regularization to the kernel, so that K^{-1} exists. This method was developed by G. Wahba[35], and employed by White et al.[36] and Jarrell and Biham[37]. They used similar methods to minimize $(G - KA)^2$ subject to constraint potentials which introduce correlations between adjacent points in A and impose positivity. However, these techniques tend to produce spectra A with features which are overly smeared out by the regularization.

With entropy-based methods of Bayesian data analysis, we employ a different philosophy. Using Bayesian statistics, we define the posterior probability of the spectra A given the data G, i.e., $P(A|G) \propto P(A)P(G|A)$ with the prior probability $P(A)$ defined so that A has only those correlations that are required to reproduce the data G. To define the likelihood function $P(G|A)$, we take advantage of the statistical sampling nature of the QMC process. The choice $P(A)$ is based on information theory concepts of entropy. The result is either the most probable $P(A|G)$ or the average of this function over free parameters. The latter has proven to be the most effective. Collectively, these methods are colloquially called Maximum Entropy Methods (MEM).

In this section, we will present a short pedagogical development of the MEM to analytically continue QMC data. A more detailed review has been presented previously[16], and to the extent possible, we will follow the notation used there. This section is organized as follows: In Sec. IV B, we will present the MEM formalism including an efficient way to calculate the spectra when the sign problem is present in the QMC calculations. In Sec. IV C, the Bryan MEM algorithm will be sketched, which has been optimized for this type of problem.

In Secs. IV D and IV E we will illustrate these techniques with the spectra of the periodic Anderson model and Hubbard model respectively. Finally in Sec. IV F, we will conclude.

Throughout this section, we will illustrate the formalism and methods introduced with a Dynamical Field Theory (DMFT)[10,12,13] simulation of the periodic Anderson model (PAM) (2) and a Dynamical Cluster Approximation (DCA)[8,9,38] simulation of the two-dimensional Hubbard model (1). As described in the Sec. III B, the core of both DMFT and DCA algorithms is the Hirsch-Fye impurity algorithm[23]. Here the problem is cast into a discrete path formalism in imaginary time, τ_l, where $\tau_l = l\Delta\tau$, $\Delta\tau = \beta/L$, $\beta = 1/k_BT$, and L is the number of times slices. Matsubara-time Green's functions are measured on this discrete time domain.

B. Formalism

1. *Green's Functions*

If this system is perturbed by an external field which couples to an operator B, then the linear response to this field is described by the retarded Green's function

$$G(t) = -i\Theta(t)\left\langle \left[B(t), B^\dagger(0)\right]_\pm \right\rangle \tag{135}$$

where the negative (positive) sign is used for Boson (Fermion) operators B and $B^\dagger$, and makes reference to the Dirac (anti)commutator. The Fourier transform of $G(t)$, $G(z)$ is analytic in the upper half plane, and its real and imaginary parts are related by

$$G(z) = \int d\omega \frac{\frac{-1}{\pi}\mathrm{Im}G(\omega)}{z-\omega}\,. \tag{136}$$

The Matsubara-frequency Green's function $G(i\omega_n)$ is obtained by letting $z \to i\omega_n$ in Eq. 136. This may be Fourier transformed to yield a relation between the Matsubara-time Green's function produced by the QMC procedure, and $\frac{-1}{\pi}\mathrm{Im}G(\omega)$

$$G(\tau) = \int d\omega \frac{\frac{-1}{\pi}\mathrm{Im}G(\omega)e^{-\tau\omega}}{1 \pm e^{-\beta\omega}}\,. \tag{137}$$

2. *Bayesian Statistics*

We use our QMC algorithm to generate a set $\bar{G}_l^i$ of $i = 1, .., N_d$ estimates for the Green's function at each time slice $\tau_l = (l-1)\Delta\tau$, $l = 1, L$. Since many A correspond to the same

data $\bar{G}$, we must employ a formalism to determine which $A(\omega)$ is the most probable, given the statistics of the data and any *a priori* information that we have about A. To quantify the conditional probability of A given the data, and our prior knowledge, we use Bayesian statistics.

If we have two events a and b, then according to Bayes theorem, the joint probability of these two events is

$$P(a,b) = P(a|b)P(b) = P(b|a)P(a)\,, \tag{138}$$

where $P(a|b)$ is the conditional probability of a given b. The probabilities are normalized so that

$$P(a) = \int dbP(a,b) \quad \text{and } 1 = \int daP(a)\,. \tag{139}$$

In our problem, we search for the spectrum A which maximizes the conditional probability of A given the data $\bar{G}$,

$$P(A|\bar{G}) = P(\bar{G}|A)P(A)/P(\bar{G})\,. \tag{140}$$

Typically, we call $P(\bar{G}|A)$ the likelihood function, and $P(A)$ the prior probability of A (or the prior). Since we work with one set of QMC data at a time, $P(\bar{G})$ is a constant during this procedure, and may be ignored. The prior and the likelihood function require significantly more thought, and will be the subject of the next two subsections.

3. Prior Probability

We can define a prior probability for positive-definite normalizable spectra. For Bosonic Green's functions, we may define positive definite spectra if we redefine the kernel

$$K(\tau,\omega) = \frac{\omega[e^{-\tau\omega} + e^{-(\beta-\tau)\omega}]}{1-e^{-\beta\omega}} \text{ with } A(\omega) = \frac{-1}{\pi\omega}\mathrm{Im}G(\omega) \geq 0 \text{ for Bosons}\,. \tag{141}$$

We modified the kernel to account for the symmetry of the Bosonic data $G(\tau) = G(\beta-\tau)$ and the spectrum $A(\omega) = A(-\omega)$. Note that the kernel is non-singular at $\omega = 0$ and the spectral density $A(\omega)$ is positive definite. For Fermionic Green's functions the spectra are already positive definite

$$K(\tau,\omega) = \frac{\exp(-\tau\omega)}{1+\exp(-\beta\omega)} \text{ with } A(\omega) = \frac{-1}{\pi}\mathrm{Im}G(\omega) \geq 0 \text{ for Fermions}\,. \tag{142}$$

We may also define positive definite spectra for more exotic cases, such as for the Nambu off-diagonal Green's function G_{12}. Since the corresponding spectrum $A_{12}(\omega) = \frac{-1}{\pi}\mathrm{Im}G_{12}(\omega)$ is not positive definite, we enforce positivity by adding a positive real constant b

$$G_{12}(\tau) + b\int d\omega K(\tau,\omega) = \int d\omega K(\tau,\omega)\left(A_{12}(\omega) + b\right) . \tag{143}$$

Here, we may incorporate the symmetry of the spectrum $A_{12}(\omega) = -A_{12}(-\omega)$ and the data $G_{12}(\tau) = -G_{12}(\beta - \tau)$ by modifying the kernel

$$K(\tau,\omega) = \frac{e^{-\tau\omega} - e^{-(\beta-\tau)\omega}}{1 + \exp(-\beta\omega)} . \tag{144}$$

With this kernel, the equation takes the canonical form Eq. 134, if we identify

$$A(\omega) = A_{12}(\omega) + b, \text{ and } G(\tau) = G_{12}(\tau) + b\int d\omega K(\tau,\omega) \tag{145}$$

In each of the Bosonic, Fermionic and Anomalous cases,

$$\int_{-\infty}^{\infty} d\omega A(\omega) < \infty . \tag{146}$$

These positive-definite normalized spectra A may be reinterpreted as probability densities.

Skilling[39] argues that the prior probability for such an unnormalized probability density is proportional to $\exp(\alpha S)$ where S is the entropy defined relative to some positive-definite function $m(\omega)$

$$\begin{aligned} S &= \int d\omega \left[A(\omega) - m(\omega) - A(\omega)\ln\left(A(\omega)/m(\omega)\right)\right] \\ &\approx \sum_{i=1}^{N_f} A_i - m_i - A_i \ln\left(A_i/m_i\right) , \end{aligned} \tag{147}$$

where $A_i = A(\omega_i)d\omega_i$, $i = 1, N_f$. Thus, the prior is conditional on two as yet unknown quantities $m(\omega)$ and α

$$P(A|m,\alpha) = \exp\left(\alpha S\right) . \tag{148}$$

$m(\omega)$ is called the default model since in the absence of data $\bar{G}$, $P(A|\bar{G},m,\alpha) \propto P(A|m,\alpha)$, so the optimal $A = m$. The choice of α will be discussed in Sec. IV B 6.

Rather than try to repeat Skilling's arguments here for the entropic form of $P(A|m,\alpha)$, we argue that this form yields the desired effects:

1. it enforces positivity of A,

2. it requires that A only have correlations which are required to reproduce the data $\bar{G}$, and

3. it allows us to introduce prior knowledge about the the spectra (i.e. exact results at high frequencies) in the default model.

The first effect follows from the form of $P(A|m,\alpha)$, assuming that m is positive definite. The third effect will be discussed in Sec. IV D 5.

To illustrate the second effect, Gull and Skilling use their kangaroo argument[40]. Imagine we have a population of kangaroos. We know that one third of them are left handed and one third have blue eyes. The joint probabilities of left-handedness and eye color may be represented in a contingency table.

	Left Handed	
	T	F
Blue T	p_1	p_2
Eyes F	p_3	p_4

We are given that $p_1 + p_2 = p_1 + p_3 = 1/3$, what is the fraction that are both blue eyed and left handed, p_1? Clearly, there is not enough information to answer this question. We must make some additional assumptions. If we assume that there is a maximum positive correlation between left handedness and blue eyes, then

	Left Handed	
	T	F
Blue T	1/3	0
Eyes F	0	2/3

If these events have a maximum negative correlation, then

	Left Handed	
	T	F
Blue T	0	1/3
Eyes F	1/3	1/3

However, if we are forced to answer this question without the use of further information, a more natural assumption to make is that the events of handedness and eye color are uncorrelated, so that 1/9 of the kangaroos are both blue eyed and left handed.

	Left Handed	
	T	F
Blue T	1/9	2/9
Eyes F	2/9	4/9

This final answer is the one obtained by maximizing the entropy $S = -\sum_{i=1}^{4} p_i \ln p_i$ subject to the Lagrange constraints $\sum_{i=1}^{4} p_i = 1$, $p_1 + p_2 = 1/3$ and $p_1 + p_3 = 1/3$. All other regularization functions yield either positive or negative correlations between handedness and eye color.

To relate this to the analytic continuation problem, imagine that each A_i is an independent event. If we maximize the entropy of A, subject to the constraint of reproducing the data $G = KA$, then the resulting spectrum is the one with the least correlations that is consistent with $\bar{G}$. If we identify a feature in the spectrum as a region of correlated A_i (such as a peak) in deviation from the default model m_i, and such a feature emerges in the spectrum $A(\omega)$ and persists as the data $\bar{G}$ becomes more precise, then we have reason to believe that this feature is real. The choice of any other regularization function would produce artificial features in the data.

4. Likelihood function

The form of the likelihood function is dictated by the central limit theorem, which for the purposes of this section may be illustrated with the following example. Suppose we use our QMC algorithm to generate N_d measurements of the Green's function $\bar{G}_l^i$ (where l is an integer between 1 and L, and i an integer between 1 and N_d). According to the central limit theorem, if each of these measurements is completely independent of the others, then in the limit of large N_d, the distribution of $\bar{G}_l$ will approach a Gaussian, and the probability of a particular value G_l is given by

$$P(G_l) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\chi^2/2}, \tag{149}$$

where $\chi^2 = \frac{1}{\sigma^2}\left(\frac{1}{N_d}\sum_{i=1}^{N_d} \bar{G}_l^i - G_l\right)^2 = \frac{1}{\sigma^2}\left(\left\langle \bar{G}_l \right\rangle - G_l\right)^2$, $\sigma^2 = \frac{1}{N_d(N_d-1)}\sum_i \left(\left\langle \bar{G}_l \right\rangle - \bar{G}_l^i\right)^2$ and the angular brackets indicate an average over the bins of data.

Of course, in the QMC process each of the measurements is not independent of the others. Correlations exist between adjacent measurements ($\bar{G}_l^i$ and $\bar{G}_l^{i+1}$) in the QMC process, and

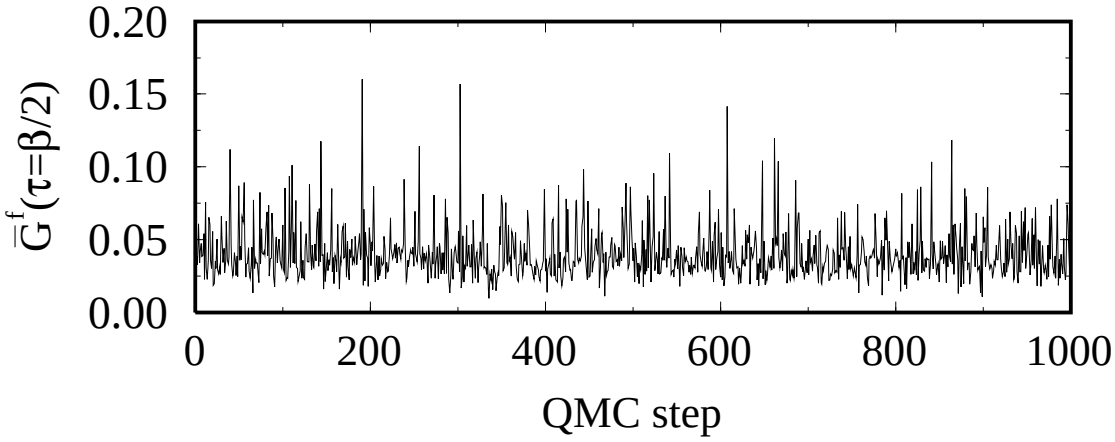

FIG. 26: *Symmetric PAM f-electron local Green's function* $\bar{G}^f(\tau = \beta/2)$ *plotted as a function of the QMC step for* $U = 2$, $V = 0.6$, *and* $\beta = 20$.

between the errors of the Green's function at adjacent time slices ($\bar{G}^i_l$ and $\bar{G}^i_{l+1}$) at the same QMC step. The removal of these correlations is the most critical step in the MEM analytic continuation procedure.

Correlations between adjacent measurements are illustrated in Fig. 26 where measurements of $\bar{G}^f(\tau = \beta/2)$ are plotted versus the QMC step. The data from adjacent QMC steps are correlated and the data are skewed since the Green's function is bounded from below ($\bar{G}^i_l > 0$). As a result the data are not Gaussianly distributed, as shown in Fig. 27(a). Here, a histogram of the data is compared to a Gaussian fit. The deviations from a Gaussian are quantified by the moments of the distribution. The most relevant ones in the present case are the skewness (third moment) and kurtosis (fourth moment) which measure the degree of asymmetry around the mean and the pointedness (or flatness) of the distribution relative to the Gaussian[41]. The data are clearly not Gaussianly distributed, and display significant skew and kurtosis. To deal with this difficulty, we rebin the data. For example, we set $\bar{G}^1_l$ equal to the average of the first 30 measurements, $\bar{G}^2_l$ equal to the average of the next 30 measurements, etc. The distribution of this rebinned data is shown in Fig. 27b. It is well approximated by a Gaussian fit (the solid line).

The chosen bin size (here, 30 measurements) must be large enough so that the bin averages are uncorrelated, but small enough so that sufficient bins remain to calculate the likelihood function. To determine the smallest bin size that yields uncorrelated data we quantify the deviation of the distribution from a Gaussian by measuring moments of the distribution.

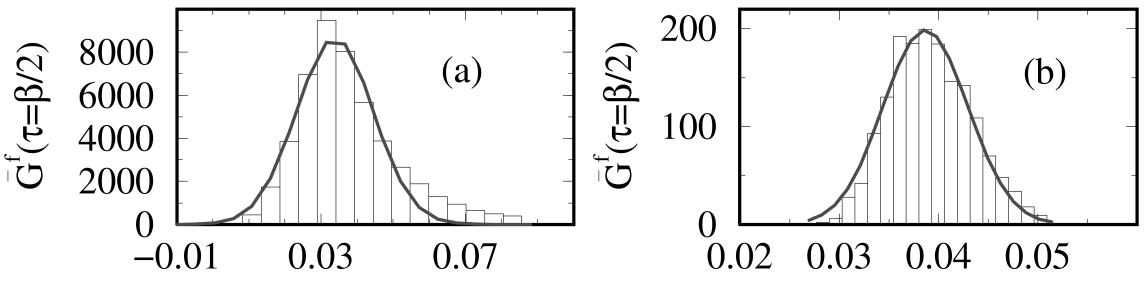

FIG. 27: *Distribution of the data shown in Fig. 26 (a) and after rebinning (b). The solid line is a Gaussian fit. In (b) the data was processed by packing it sequentially into bins of 30 measurements each.*

Of course, because the data are a finite set, each of these measured moments has some standard deviation (proportional to $1/\sqrt{N_{bins}}$). Thus, one way to determine if the skewness and kurtosis of a distribution are acceptably small is to measure these values relative to what is expected from a Gaussian distribution. We will use such relative values.

As the bin size increases, the relative kurtosis and skewness decrease monotonically, indicating the convergence of the distribution to a Gaussian. This behavior is shown in Fig. 28a for the $G(\tau = \beta/2)$ data.

In addition, Fig. 28b shows that the error estimate also converges as the bin size increases. Here, the error estimate is given by

$$\sigma = \sqrt{\left(\left\langle \bar{G}^2 \right\rangle - \langle \bar{G} \rangle^2\right)/(N_{bins} - 1)} \tag{150}$$

where angular brackets indicate an average over the bins of data. Because correlations between successive Monte Carlo measurements always make this error estimate smaller than the actual value, this error estimate should initially increase monotonically with bin size, as shown. This behavior is easily understood by considering a perfectly correlated sample where the data in each bin is identical. Clearly, for this perfectly correlated sample, the error estimate would be zero. As the bins become uncorrelated, the error estimate increases. With independent data and a large number of equally sized bins, eventually $\sigma^2 \sim 1/N_{bins}$. However, with a fixed amount of data, as is typical with a QMC simulation, increasing the bin size decreases N_{bins} proportionally, and the error estimate can saturate as illustrated in Fig. 28b. Thus, the saturation of the error estimate indicates that the correlations between Monte Carlo measurements, i.e., between bin averages, have been removed. The point at

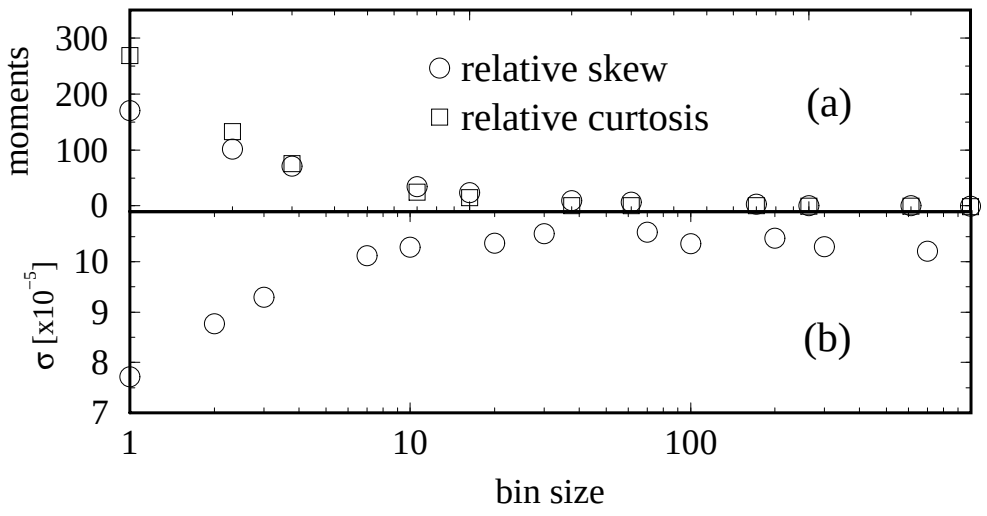

FIG. 28: *Relative kurtosis and skew (a) and error bar (b) of the data shown in Fig. 26 as a function of bin size. Here the total amount of data is fixed, so increasing the bin size decreases N_{bins} proportionately. As the bin size increases to about 30, the relative kurtosis and skew decrease to roughly zero and the error bar saturates, indicating that the bins are uncorrelated samples and that the data has become Gaussianly distributed.*

which saturation occurs in a plot like Fig. 28b provides a useful first estimate of the minimum bin size required to remove correlations between the bins. In general, one should perform this test for the Green's function at all times τ_l; however, we have found it is often sufficient to perform this test at only a few times. For the remainder of this section, we will assume that the bin size is sufficiently large so that both the error estimate and the moments of the distribution have converged to values which indicate that the data are both statistically independent and Gaussian-distributed.

Now, only the errors in the Green's function $\bar{G}$ at adjacent time slices remain correlated. This correlation may be seen by comparing the results from a single measurement with those essentially exact values obtained from averaging over many measurements. Such a comparison is shown in Fig. 29 where if the result from a single measurement differs from the essentially exact result at a certain value of τ, then the results at adjacent values of τ also tend to deviate from the exact results in a similar way. These correlations of the error in Matsubara time are characterized by the covariance

$$C_{lk} = \frac{1}{N_{bins}(N_{bins}-1)} \sum_{j=1}^{N_{bins}} (\left\langle \bar{G}_l \right\rangle - \bar{G}_l^j)(\left\langle \bar{G}_k \right\rangle - \bar{G}_k^j)\,. \tag{151}$$

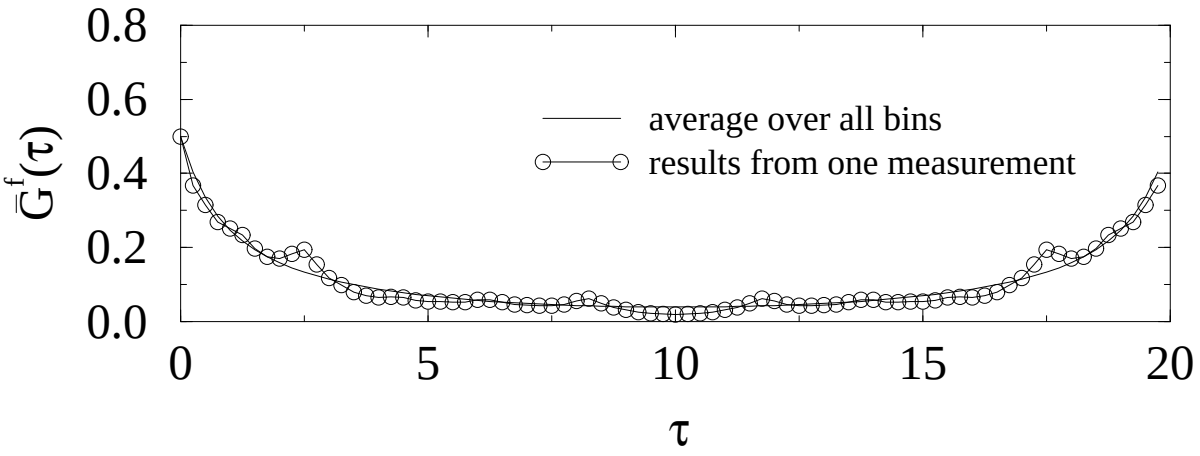

FIG. 29: $\bar{G}^f(\tau)$ *from one measurement compared to* $\bar{G}^f(\tau)$ *obtained from the average over 800 bins of data, each containing 1520 measurements. If the result from a single measurement at a certain point differs from the essentially exact result obtained by averaging over many bins, then the results at adjacent points also deviate from the exact results.*

If C is diagonal, then according to the central limit theorem, the likelihood function is $P(\bar{G}|A) = \exp[-\chi^2/2]$ where

$$\chi^2 = \sum_{l=1}^{L} \left(\frac{\bar{G}_l - \sum_j K_{l,j} A_j}{\sigma_l} \right)^2 , \tag{152}$$

and σ_l^2 are the diagonal elements of C. However, in general, the covariance matrix C_{ij} is not diagonal because errors at different values of τ are correlated. To define a meaningful measure of how well A_i reproduces the data, we must find the transformation $\mathbf{U}$ which diagonalizes the covariance matrix

$$\mathbf{U}^{-1}\mathbf{C}\mathbf{U} = \sigma_i'^2 \delta_{ij} \, . \tag{153}$$

Both the data and kernel are now rotated into this diagonal representation

$$\mathbf{K}' = \mathbf{U}^{-1}\mathbf{K} \qquad \bar{\mathbf{G}}' = \mathbf{U}^{-1}\bar{\mathbf{G}} \, , \tag{154}$$

and each measurement $\bar{G}_i'$ is statistically independent. Therefore, we can use

$$\chi^2 = \sum_{l} \left(\frac{\bar{G}_l' - \sum_j K_{l,j}' A_j}{\sigma_l'} \right)^2 \tag{155}$$

to measure the misfit between the spectrum and the data and to define the likelihood function.

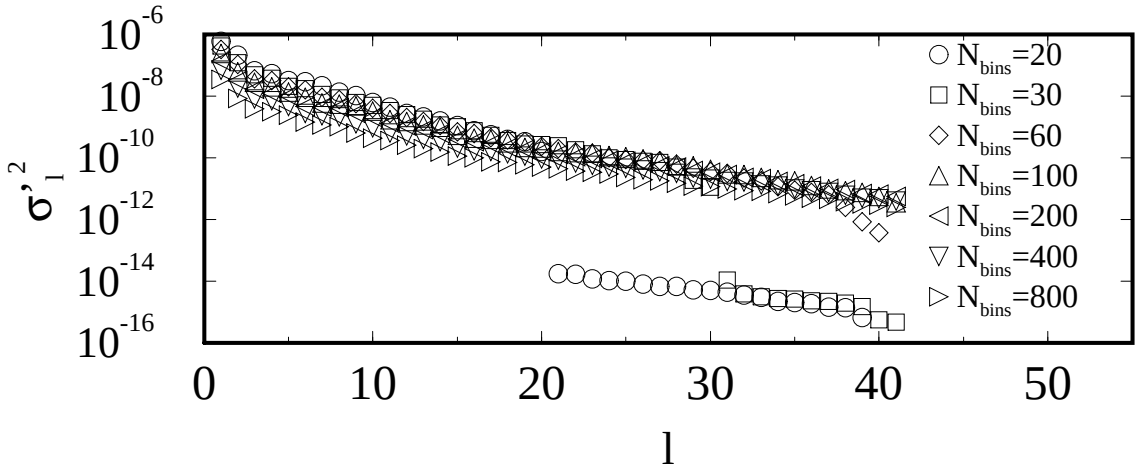

FIG. 30: *Eigenvalue spectra of the covariance matrix of G^f for different numbers of bins of data. Each bin contains 100 measurements and $L = 41$. When $N_{bins} \sim 2L$, σ'^2_l develops a sharp break.*

a. Cautionary Notes We find that proper preparation of the data, removing correlations, is the most critical step in the MEM procedure. If the data are uncorrelated and the covariance is calculated and diagonalized correctly, then the resulting spectra will be reliable (however, for weak data, it will show a significant bias towards the default model). If not, then the Gaussian form of the likelihood function is unjustified and the resulting spectra will generally have spurious features.

However, care must be taken when calculating and diagonalizing the covariance. First, Since the set of data is finite, it is necessary to balance the need of removing the correlations in imaginary-time with the need of removing the correlations between Monte Carlo steps. To remove the correlations in Monte Carlo steps the bin size must be large; however, to calculate the covariance accurately, many bins of data are required. If there are not enough bins of data, then the covariance and (as shown in Fig. 30) its eigenvalue spectrum can become pathological. The reason for this pathology is that when we diagonalize the covariance matrix, we are asking for L independent eigenvectors. We must have enough bins of data to determine these directions so that N_{bins} must be greater than or equal to L. In fact, since the information contained in a given bin of data is not completely independent from the other bins, we must have $N_{bins} > L$. Otherwise, as shown in Fig. 30, where $L = 41$, the eigenvalue spectrum displays a sharp break when $N_{bins} < L$, indicating that only a finite

number of directions, less than L, are resolved. The small eigenvalues after the break are essentially numerical noise and yield artifacts in the spectra. Simply throwing away the small eigenvalues and their associated eigenvectors does not cure the difficulty since the small eigenvalues and eigenvectors contain the most precise information about the solution. Thus, the only reasonable thing to do is to increase the number of bins. Empirically, we find that we need

$$N_{bins} \geq 2L \tag{156}$$

in order to completely remove the pathology of the sharp break in the eigenvalues[42]. Second, as illustrated in Fig. 29 adjacent data in time tend to be highly correlated. These correlations grow as the time step used in the QMC calculation is reduced, making the rows and columns of the covariance more correlated. Eventually, the covariance becomes ill conditioned and cannot be diagonalized. Such severe oversampling of the Green's function data does not provide more information, but a small time step may be useful for other reasons (such as reducing Trotter errors). In this case we can fix the problem by eliminating some fraction of the data (i.e. taking the data from every other time step).

5. Data produced by QMC simulations with sign problem

a. Sign problem in QMC simulations The nature of the sign problem in QMC calculations was discussed in III D 1. When the average sign is small the Monte Carlo calculations become impractical. Even a moderate sign problem, for which the computation of the static quantities described in Eq. 129 is still possible, introduces additional problems in the calculation of the dynamical spectra including especially strong correlations of the data resulting in non-gaussian distributions, such as those shown in Fig. 31 c) and d). However, these cannot be removed by standard techniques due to the strong correlation between the data and the averaged sign of the configurations which produce these data. This makes it essentially impossible to calculate spectra long before the minus sign problem makes the calculation of the static properties impractical. In this section we address this problem and describe a solution which greatly increases the resolution of MEM when calculating spectra from such poorly conditioned data.

Nevertheless, we should mention that we do not attempt to improve the sign problem in general. We only introduce a proper way to compute the dynamical spectra when the

average sign is small, but still large enough to allow the computation of the static quantities.

b. Analysis of QMC data with sign problem The difficulty in obtaining good data for G^i can easily be understood from the measurement process when the sign is not always positive. Then, to obtain the Green's function, we no longer measure G^i but rather the product of it and the sign s of the configuration, $Gs^i \equiv G^i * s$, and the sign s. At the end of the simulation, i.e. after a large number of measurements, we then obtain

$$\langle G_l \rangle = \frac{\langle Gs_l \rangle}{\langle s \rangle} = \frac{\sum_{i=1}^{N_d} Gs_l^i}{\sum_{i=1}^{N_d} s^i} \,. \tag{157}$$

However, besides the means (i.e. the average $\langle G_l \rangle$), in order to employ MEM we need to generate uncorrelated and independent $\bar{G}_l^{(i)}$ points normally distributed around the mean (see Eq. 151). Generating these points is problematic if the sign problem is present, as we discuss below.

First, in order to obtain good estimates of $\bar{G}_l^i$ (i.e. to have a good statistics), we need to average a large number of Gs_l^i and s^i

$$\bar{G}_l^i = \frac{\bar{Gs}_l^{\,i}}{\bar{s}^i} = \frac{\sum_{i=1}^{Bs} Gs_l^i}{\sum_{i=1}^{Bs} s^i} \,. \tag{158}$$

Here $Bs = \frac{N_d}{N_{bins}}$ is the bin size or the number of points in the bin i, thus $\bar{Gs}_l^{\,i}$ and $\bar{s}^i$ are averages over the measurements that form the bin i. This large number of measurements for every $\bar{G}_l^i$ requires large computational effort.

Second, even if the bin size is chosen large enough to produce good estimates of the $\bar{G}_l^i$, the distribution of these will not be Gaussian. This can be easily understood if one notices that the $\bar{G}_l^i$ distribution is a ratio of two normal distributions, $\bar{Gs}_l^{\,i}$ and s^i respectively. However, the ratio distribution of two normal distributions is not a normal distribution. Therefore in order to obtain a normal distribution for $\bar{G}_l^i$ data one has, in the virtue of the central limit theorem, to further rebin the points $\bar{G}_l^i$. Of course, this rebinning will result in an additional computational effort.

Third, as within the same bin i there is a strong correlation between different data points $\bar{Gs}_l^{\,i}$ corresponding to adjacent time slices, there is also a strong correlation between data $\bar{Gs}_l^{\,i}$ and $\bar{s}^i$. The $\bar{G}_l^i$ are a result of a nonlinear operation between two correlated quantities. Thus the average of $\bar{G}_l^i$

$$\overline{G_l} = \frac{1}{N_{bins}} \sum_{i=1}^{N_{bins}} \bar{G}_l^i \tag{159}$$

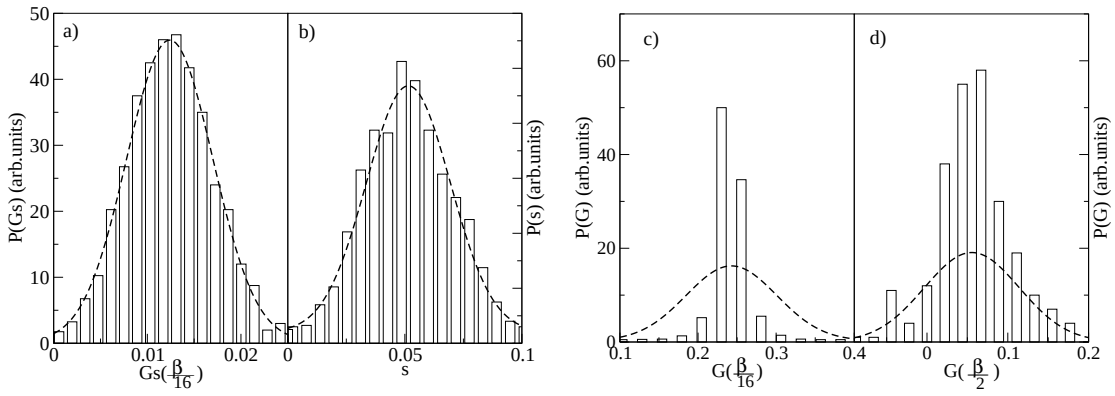

FIG. 31: a) and b) histograms representing the distribution of $\bar{Gs}(\tau = \beta/16)$ and respectively sign $\bar{s}$ for 3000 measurements per bin. c) histogram representing the distribution of $\bar{G}(\tau = \beta/16)$ when a $\bar{G}$ point is obtained as a ratio of 3000 averaged $\bar{Gs}$ and 3000 averaged $\bar{s}$ points. d) histogram representing the distribution of $\bar{G}(\beta/2)$ when every $\bar{G}$ point is obtained by rebinning 30 initial G points obtained as a ratio of 600 averaged $\bar{Gs}$ and $\bar{s}$ points. The dashed lines represent the best Gaussian fit to the data.

will be different from $\langle G_l \rangle$ obtained in Eq. 157. Therefore, one ends with a set of G_l data normally distributed around $\overline{G_l}$ which is not a good estimation for G_l if the correlations are strong.

In order to exemplify the problems discussed above, we employed a QMC based DCA algorithm to produce a very large amount of data for the single-particle Green's function of the two-dimensional Hubbard model (Eq. 1) on a square lattice. We choose $t = 0.25$ so that the bandwidth $W = 2$ and set $U = W$. To make the sign-problem worse, we add a next-nearest neighbor hopping $t' = -0.3t$ to frustrate the lattice. We perform calculations on a 16-site 4×4 cluster at 15% doping, down to temperatures $T = 0.125t$ where we experience a severe sign-problem, $\bar{s} = 0.051$. We calculated 8000 data points (bins) $(\bar{Gs}_l^i, \bar{s}^i)$, and for every data point we averaged 600 QMC measurements.

In Fig. 31 a) and b) we show histograms of the $\bar{Gs}(\tau = \beta/16)$ and respectively $\bar{s}$ when the bin size is increased five times, which corresponds to an average of 3000 measurements per bin. It can be seen that the data are normally distributed to a very good approximation. This is unlike the situation shown in Fig. 31 c) where the $\bar{G}(\tau = \beta/16)$ distribution, obtained as a ratio distribution of the previous $\bar{Gs}(\tau = \beta/16)$ and $\bar{s}$, is shown. The $\bar{G}(\tau = \beta/16)$ data are strongly peaked, being characterized by a large positive kurtosis. Similar distributions

of data are observed (not shown) for the other values of the imaginary time. The worst situation is observed at imaginary times close to $\beta/2$ which affects mainly the low-energy properties of the spectra. In Fig. 31 d) we show the distribution of $\bar{G}(\tau = \beta/2)$ when every point is obtained after rebinning 30 initial $\bar{G}(\tau = \beta/2)$ points, with each initial point being obtained as a ratio of 600 averaged $\bar{G}s$ and $\bar{s}$ points. Even if now a large number of measurements $(18,000)$ is used for producing one data point, the distribution is still not Gaussian. However, this is mainly because we used a small amount of measurements (only 600) for obtaining the initial $\bar{G}$ points (i.e. the ones before rebinning). We found that much better data can be obtained if initially one uses a large number of measurements to calculate the $\bar{G}$ points and afterwards a rebinning of a few successive $\bar{G}$ points is employed. Nevertheless the number of measurements necessary to obtain normal distributed $\bar{G}$ points is much larger than the one required to obtain normal distributed $\bar{G}s$ and $\bar{s}$ points. In our case this number is about five times larger but this value is dependent on the problem considered, being determined by both the magnitude of the correlations and the value and the distribution of the sign.

We propose the following modified approach to MEM for QMC data with sign problem.

i) Rebin $(\bar{G}s_l^i, \bar{s})$ until they become normal distributed. As it was shown in Fig. 31 a) and b) this can be achieved with a relatively small amount of data. *ii)* Employ a rotation in the space (Gs_l, s) in order to obtain the statistically independent vectors. *iii)* Calculate χ^2 in the space spanned by (Gs_l, s). The calculation of χ^2 in a space which now includes the extra sign dimension will be discussed below in Sec. IV B 5 c.

c. Modified likelihood function Denoting $h \equiv (Gs_l, s)$, the *modified* likelihood function is defined as $P[h|A]$, since the measured quantities in the QMC process are the h points (and not G). As we showed in the previous section, for acceptable values of the bin size, the data h are to a good approximation Gaussianly distributed. Therefore, the *modified* likelihood function also will have the usual form as $e^{-\chi^2/2}$, with

$$\chi^2 = \sum_{l,k=1}^{L+1} (\bar{h}_l - h_l(A))[C_h^{-1}]_{lk}(\bar{h}_k - h_k(A)) \tag{160}$$

The new covariance matrix has the dimension $(L+1) \times (L+1)$,

$$C_{hlk} = \frac{1}{N_{bins}(N_{bins}-1)} \sum_{i=1}^{N_{bins}} (\langle h \rangle_l - \bar{h}_l^i)(\langle h \rangle_k - \bar{h}_k^i)\,. \tag{161}$$

The only problem which remains to be solved is finding an equation for $h(A)$, since Eq. 134 only provides a relation for $G(A)$. In order to achieve this we do the following: First we absorb the sign into the spectrum, i.e. we define $\mathcal{A}$ as

$$\mathcal{A}(\omega) = sA(\omega)\,. \tag{162}$$

Instead of searching for a spectrum A which satisfies Eq. 134 we search for $\mathcal{A}$ which satisfies

$$Gs(\tau) = \int K(\tau,\omega)\mathcal{A}(\omega)d\omega\,. \tag{163}$$

Second, we consider the spectrum normalization sum-rule

$$B = \int A(\omega)d\omega, \tag{164}$$

which implies

$$s = \int \frac{1}{B}\mathcal{A}(\omega)d\omega\ . \tag{165}$$

Here B is a constant, equal to one for the the one-particle spectra and equal to the static susceptibility $\chi(T)$ for the two-particle case. We relate the sign fluctuations to the norm of the new spectrum because of the sign s was absorbed into the definition of $\mathcal{A}$.
Both Eq. 163 and Eq. 165 can be written as

$$\langle h_l\rangle = \sum_{\nu=1}^{N_f} K_{hl\nu}\mathcal{A}_\nu, \quad K_{hl\nu} = \begin{cases} K_{i\nu} & l \le L \\ \frac{1}{B} & l = L+1 \end{cases}\ . \tag{166}$$

This is the basic equation which relates h to $\mathcal{A}$ and determines the likelihood function $P[h|A] \equiv P[h|\mathcal{A}]$. MEM will produce the most probable spectrum $\mathcal{A}$ normalized to $\langle s\rangle$ which minimizes the χ^2 function in Eq. (160) subject to the entropy constraint.

We note that for the one-particle case, where $B = 1$, Eq. 165 is equivalent to

$$Gs(0) + Gs(\beta) = s\ . \tag{167}$$

By using Eq. 165 in the calculation of the likelihood function we impose

$$G(0) + G(\beta) = \frac{Gs(0)}{s} + \frac{Gs(\beta)}{s} = 1 \tag{168}$$

at every measurement. Since Eq. 168 results solely from the anti-commutation relation of the one-particle operators it should be satisfied in every possible configuration and implicitly in every measurement. Therefore, this way of implementing the normalization sum-rule is

more natural than the usual way based on Lagrange multipliers where the constraint is globally imposed, i.e. not at every measurement but only for the final Green's function obtained at the end of the QMC process.

For the two-particle case, where $B = \chi(T)$, the sum-rule Eq. 164 is not an independent equation as in the one-particle case, but merely an integration over τ of Eq. 137. Therefore it is essential to treat B as a constant (equal to the final, averaged over all QMC configurations, $\langle\chi(T)\rangle$) and to disregard measurement dependent fluctuations in $\chi(T)$. This way we relate the norm of $\mathcal{A}$ only to the fluctuation of the sign s.

6. Details of the MEM Formalism

We will now construct the formalism to locate the most likely spectrum $\hat{A}$ and set the value of α. The first step is to normalize the likelihood function $P(\bar{G}|A)$ and the prior $P(A|\alpha, m)$. Here it will be necessary to integrate over the space of all spectra A_i. This is done with Gaussian approximations to the integrals. Following Skilling and Bryan[43], we employ a measure $d^N A/\prod_i \sqrt{A_i}$ which amounts to a change of variables to a space where S has no curvature[16].

For example, the normalized prior probability is

$$P(A|\alpha, m) = \frac{1}{Z_S} \exp\left\{\alpha\left(-\sum A_i \ln A_i/m_i - A_i + m_i\right)\right\} \ , \tag{169}$$

where

$$Z_S = \int \frac{d^N A}{\prod_i \sqrt{A_i}} \exp\left\{\alpha\left(-\sum A_i \ln A_i/m_i - A_i + m_i\right)\right\} . \tag{170}$$

The integrand is maximized when $S = 0$, i.e. when $A = m$. We approximate the integral by expanding the argument of the exponent to second order around this maximum, $S \approx \frac{1}{2}\delta A^T \, \nabla\nabla S|_{A=m} \, \delta A = -\frac{1}{2}\delta A^T \{1/m\} \delta A$, where $\{1/m\}$ is the diagonal matrix with finite elements composed of $1/m_i$, and δA is the vector $A - m$.

$$Z_S \approx \int \frac{d^N A}{\prod_i \sqrt{A_i}} \exp\left\{\alpha\left(-\frac{1}{2}\delta A^T \{1/m\} \delta A\right)\right\}. \tag{171}$$

We define a change of variables, so that $dy_i = dA_i/\sqrt{A_i}$ and find

$$Z_S \approx \int d^N y \exp\left\{\alpha\left(-\frac{1}{2}\delta y^T \{m\}^{1/2} \{1/m\} \{m\}^{1/2} \delta y\right)\right\} = (2\pi/\alpha)^{N/2} \tag{172}$$

The likelihood function must also be normalized

$$P(\bar{G}|A) = e^{-\chi^2/2}/Z_L \ , \tag{173}$$

where

$$\chi^2 = \sum_l \frac{\left(\bar{G}'_l - \sum_i K'_{li} A_i\right)^2}{\sigma'^2_l} \ . \tag{174}$$

where K' and $\bar{G}'$ are the kernel and data rotated into the data space where the covariance is diagonal, and σ'^2_l are the eigenvalues of the covariance. If we let $G_l = \sum_i K'_{li} A_i$, then

$$Z_L = \int d^L G \exp\left\{\frac{1}{2}\sum_{l=1}^{L} \frac{\left(\bar{G}'_l - G_l\right)^2}{\sigma'^2_l}\right\} = (2\pi)^{L/2} \prod_l \sigma'_l \ . \tag{175}$$

Using Bayes theorem, we find

$$\begin{aligned} P(A, G|m, \alpha) &= P(G|A, m, \alpha) P(A|m, \alpha) \\ &= P(A|G, m, \alpha) P(G|m, \alpha) \end{aligned} \tag{176}$$

or

$$P(A|G, m, \alpha) \propto P(G|A, m, \alpha) P(A|m, \alpha) = \frac{\exp(\alpha S - \chi^2/2)}{Z_S Z_L} \, . \tag{177}$$

Since the normalization factors Z_S and Z_L are independent of the spectrum, for fixed α and data, the most probable spectrum $\hat{A}(\alpha)$ is the one which maximizes $Q = \alpha S - \chi^2/2$. An algorithm to find this spectrum is discussed in Sec. IV C. However, the question of how to select α and the default model remains.

a. Selection of α The selection of α strongly effects the choice of the optimal spectrum[44] since α controls the competition between S and χ^2. If α is large, then the entropy term is emphasized and the data cannot move the spectrum far from the model. If α is small, then the least square misfit between the spectrum and the data is minimized so that $\chi^2 \ll L$. The numerical error in the QMC data then begins to dominate the solution and the spectra displays random oscillations and noise. Thus, it is important to find a sensible way of selecting α. Typically, α is selected in one of three ways described below.

b. Historic MEM In the historic method[43,45], α is adjusted so that $\chi^2 = L$. The justification for this is that if the spectrum is known and the data was repeatedly measured, then the misfit between the data and the spectrum $\chi^2 = L$ on average. However, the data are only measured once and the spectrum is not known *a priori*. Also, setting $\chi^2 = L$ tends

to under fit the data since good data can cause structure in the spectrum which reduces χ^2 from L. Thus, there is little reason to believe that α can be chosen without input from the data itself.

c. Classic MEM A more appropriate method[44] of setting α is to choose the most probable value, defined by maximizing

$$P(\alpha|\bar{G},m) = \int \frac{d^N A}{\prod_i \sqrt{A_i}} P(A,\alpha|\bar{G},m)\,. \tag{178}$$

The integrand

$$P(A,\alpha|\bar{G},m) = P(A|\bar{G},m,\alpha)P(\alpha) \propto \frac{\exp(\alpha S - \chi^2/2)}{Z_S Z_L} P(\alpha) \tag{179}$$

involves the prior probability of α. Jeffreys[46] argues that since χ^2 and S have different units, α is a scale factor. He asserts that in lieu of prior knowledge, it should have the simplest scale invariant form $P(\alpha) = 1/\alpha$. Thus,

$$P(\alpha|\bar{G},m) = \int \frac{d^N A}{\prod_i \sqrt{A_i}} \frac{\exp(\alpha S - \chi^2/2)}{Z_S Z_L \alpha} = \frac{Z_Q}{Z_S Z_L \alpha}\,. \tag{180}$$

Z_Q is calculated in a similar fashion to Z_S. We expand about the maximum of Q at $A = \hat{A}$ so that $\exp\{\alpha S - \chi^2/2\} \approx \exp\{Q(\hat{A}) + \frac{1}{2}\delta A^T\, \nabla\nabla Q|_{\hat{A}}\, \delta A\} = \exp\{Q(\hat{A}) + \frac{1}{2}\delta A^T\{\frac{1}{2}\, \nabla\nabla\chi^2|_{\hat{A}} - \{\alpha/\hat{A}\}\}\delta A\}$. We again make a Gaussian approximation to the integral, and if λ_i are the eigenvalues of $\frac{1}{2}\{A^{1/2}\}\, \nabla\nabla\chi^2|_{\hat{A}}\, \{A^{1/2}\}$, then

$$P(\alpha|\bar{G},m) = \frac{1}{W_\alpha} \prod_i \left(\frac{\alpha}{\alpha+\lambda_i}\right)^{1/2} \frac{e^{Q(\hat{A})}}{\alpha}\,, \tag{181}$$

where

$$W_\alpha = \int \frac{d\alpha}{\alpha} \prod_i \left(\frac{\alpha}{\alpha+\lambda}\right)^{1/2} e^{Q(\hat{A})}\,. \tag{182}$$

The optimal α, $\hat{\alpha}$ may be determined by the condition

$$\frac{\partial P(\alpha|\bar{G},m)}{\partial \alpha} = 0\,. \tag{183}$$

For strong data, $P(\alpha|\bar{G},m)$ is dominated by the product and $\exp Q(\hat{A})$ so that

$$-2\hat{\alpha}S \approx \sum_i \frac{\lambda_i}{\hat{\alpha}+\lambda_i}\,. \tag{184}$$

Each λ_i which is much greater than $\hat{\alpha}$ contributes one to the sum and hence one to the number of good observations in the data. If the number $N_{good} = -2\hat{\alpha}S$ is large, then $P(\alpha|\bar{G},m)$ is very sharp the spectra corresponding to $\alpha = \hat{\alpha}$ is a good approximation of the spectra which has been properly averaged over $P(\alpha|\bar{G},m)$.

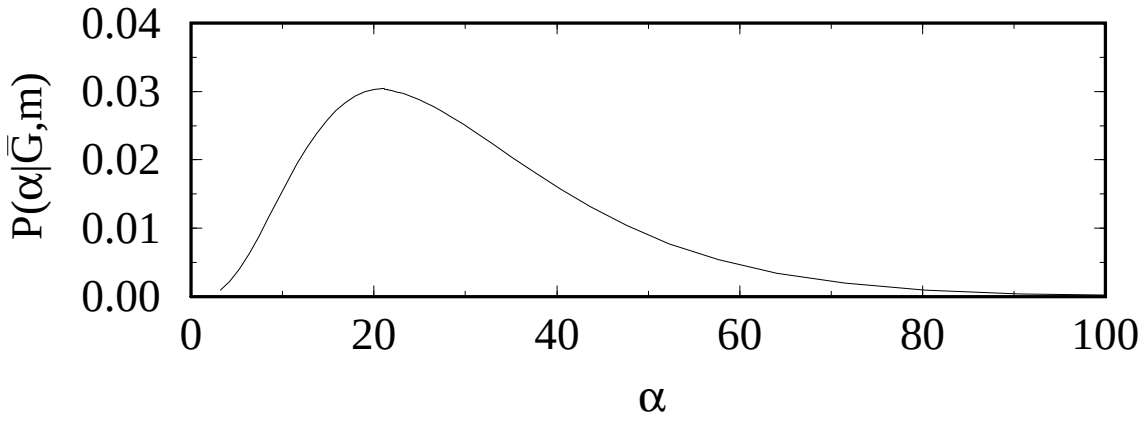

FIG. 32: The posterior probability $P(\alpha|\bar{G},m)$ as a function of α for the periodic Anderson model data presented in Fig. 1. Since $P(G|I)$ is unknown, the magnitude of the ordinate is also unknown. The distribution is wide, so many reasonable values of α exist. The distribution is also skewed, so the value of α at the peak is not representative of the mean.

d. Bryan's Method However, typically we find that $N_{good} \ll L$. Then $P(\alpha|\bar{G},m)$ is a broad and highly skewed distribution. For example, $P(\alpha|\bar{G},m)$ for the data shown in Fig. 26 is plotted in Fig. 32. The distribution is wide, so many reasonable values of α exist. The distribution is also skewed, so the value of α at the peak is not representative of the mean. To deal with this, Bryan[49] calculates the optimal spectrum $\hat{A}(\alpha)$ for each α. The solution is taken to be

$$\bar{A} = \int d\alpha \hat{A}(\alpha) P(\alpha|\bar{G},m)\,. \tag{185}$$

These three MEM methods will produce essentially identical results if the data are uncorrelated and precise. However, when the data are less precise but still uncorrelated, the method suggested by Bryan, averaging $\hat{A}(\alpha)$ weighted by $P(\alpha|G,m)$, generally produces more acceptable results and converges to a good result faster than the classic method and much faster than the historic method as the data is improved. A further advantage of the averaging is that it allows an accurate relative assessment of the posterior probability ($\int_0^\infty d\alpha P(m|G,\alpha)$) of the default model. This information is invaluable in determining which default model yields the most likely A.

e. Cautionary Notes Classic MEM and Bryan's method both rely on the accuracy of Eq. 181 for $P(\alpha|\bar{G},m)$, which is calculated with a Gaussian approximation. This is only

accurate if Q is a sharply peaked function in the space of all images. This approximation clearly fails when $\alpha \to 0$. Here there is no regularization and infinitely many spectra will produce the same Q. In this case, the algorithm can reach a run-away condition where it tends toward small values of α, the approximation for $P(\alpha|\bar{G}, m)$ fails causing the calculation to tend towards ever smaller values of α. This condition is easily identified in the calculation, and it can be cured by choosing a better default model (a Bryan or classic MEM calculation with a *perfect* default model will always tend toward a solution with large α), using the methods described below, or the annealing method described in the example Sec. IV D 5.

7. *Model Selection*

Bayesian statistics may also be employed to select the default model. If we must choose between different models, or set parameters used to define a default model function, then we choose these models or parameters based upon the posterior probability of the model

$$P(m|\bar{G}) = \int d\alpha P(\alpha|m, \bar{G}) P(m) \,. \tag{186}$$

We see no *a priori* reason to favor one default model over an another, so we typically set the prior probability of the model $P(m)$ =constant. Then the integrand in Eq. 186 is given by Eq. 181 so that

$$P(m|\bar{G}) \propto W_\alpha \,. \tag{187}$$

Since the prior probability of the model is unknown, $P(m|\bar{G})$ determines only the relative probability of two models, and by inference the relative probability of their corresponding spectra.

a. Cautionary Notes It can be tempting to try very informative models, such as the uncorrelated spectrum with sharp distinct features. Such default models will often have high posterior probabilities $P(m|\bar{G})$ but should nevertheless be avoided unless one can be *certain* that the sharp features are real. For example, a model with a delta function peak, has a huge amount of information, whereas the information from the QMC data is quite finite an may not be able to correct a wrong delta-function feature in the model. In this respect, again, the annealing technique described later is ideal.

8. Error Propagation

To absolutely qualify the spectrum, we need to assign error bars to it. In the quadratic approximation, the probability of the spectral density is

$$P(A|\bar{G}, m, \alpha) \propto e^{-\frac{1}{2}\delta A^T \cdot \nabla\nabla Q|_{\hat{A}} \cdot \delta A} , \tag{188}$$

thus the covariance of the spectral density is

$$\langle \delta A(\omega) \delta A(\omega') \rangle = - \left(\nabla\nabla Q|_{\hat{A}} \right)^{-1} . \tag{189}$$

It is not possible to assign error bars to each point in the spectral density since this matrix is generally not diagonal. Thus errors between different points are strongly correlated. Also, A_i represents the spectral probability within some region of finite width and hence lacks meaning at a specific value of ω. However, it is possible to assign error bars to integrated functions of the spectral density such as[47],

$$H = \int d\omega A(\omega) h(\omega) , \tag{190}$$

where $h(\omega)$ is an arbitrary function of ω. The error of H may be associated with the covariance of the spectral density $\langle \delta A(\omega) \delta A(\omega') \rangle$

$$\left\langle (\delta H)^2 \right\rangle = \int \int d\omega d\omega' \, h(\omega) h(\omega') \left\langle \delta A(\omega) \delta A(\omega') \right\rangle . \tag{191}$$

The matrix $\nabla\nabla Q|_{\hat{A}}$ is readily available because it is used as the Hessian of the Newton search algorithm typically used to find the optimal spectral density.

a. Cautionary Notes Care should be taken in the interpretation of the error bars, especially if a highly informative default model is used. Suppose for example the data is weak, but a default model in essentially exact agreement with the data is used, then as discussed above, a large α solution will be found corresponding to a Q with small curvature in the space of images, and hence very small error bars. This does not necessarily mean that the resulting spectra is accurate, just that the default model is one of many which is consistent with the weak data. Unless the information in a default model is known to be accurate (such as the spectra from a higher temperature, or one which becomes exact at high frequencies), such highly informative default models should generally be avoided.

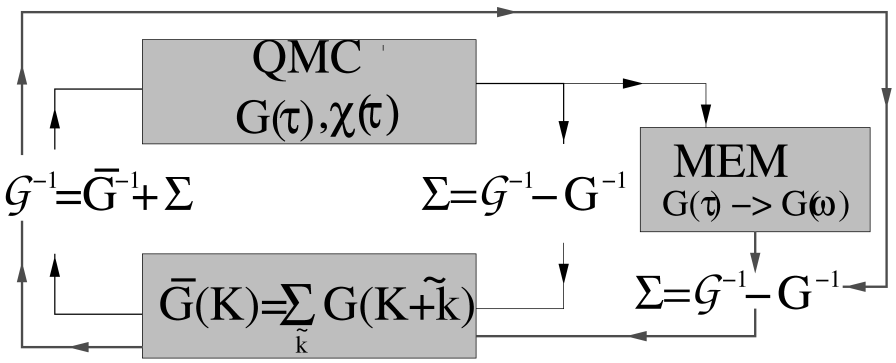

FIG. 33: MEM may be integrated into the DMFA/DCA loop. After each QMC step, MEM is used to calculate the cluster Green's function, and the Dyson equation relating G and $\mathcal{G}$ is used to extract the self energy. This eliminate the need to invert the course-graining equation in order to solve for Σ.

9. Integration of MEM with DMFA/DCA

a. Cautionary Notes A few problems can emerge when MEM is used to analytically continue DMFA and DCA data.

For example, in multiband models, such as the periodic Anderson model Eq. 2 it may be difficult to extract the real-frequency self energy. In such models where only a subset of bands are correlated, we typically integrate the uncorrelated band (in this example, the d-band) out of the action and only sample the Green's functions of the correlated bands (i.e. $G_f(\tau)$). In principle, the self energy may be extracted by inverting the coarse-graining equation. That is, for DMFA solution to the PAM

$$G_f(\omega) = \sum_{\mathbf{k}} G_f(\mathbf{k}, \omega) \tag{192}$$

where the DMFA cluster Green's function $G_f(\omega)$ is calculated with MEM. However, this can be very problematic since there are as many roots to this equation as there are bands, except at very high and low frequency, it can be difficult to disentangle the physical solutions from the unphysical ones. There are two ways to deal with this problem. First, we may directly sample the Green's functions for the uncorrelated bands, using the methods described in the chapter on DMFA and DCA. Second, we may build the real spectra calculation into the DMFA/DCA loop as illustrated in Fig. 33. Using MEM, we analytically continue the cluster DMFA or DCA Green's function. Of course this method requires that the real and

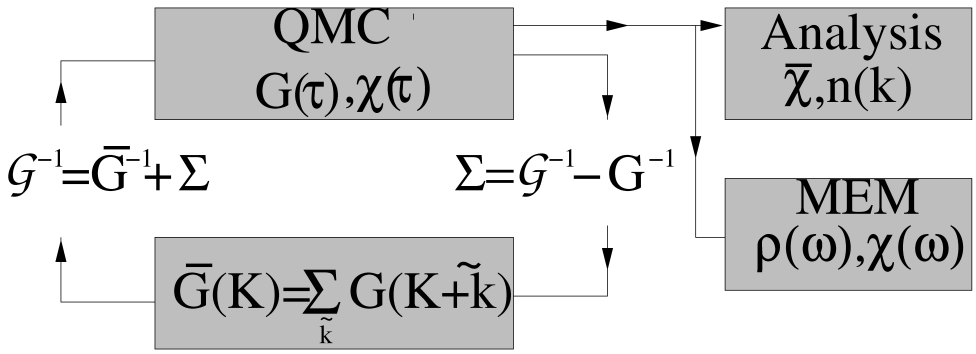

FIG. 34: The DCA algorithm. QMC is used as a cluster solver. Once convergence is reached, $G = \bar{G}$, and the irreducible quantities are used in the Maximum Entropy Method (MEM) codes to calculate the spectra. However, conventional error estimating techniques, only account for the error in the last iteration.

Matsubara frequency self energies used to initialize the DMFA correspond (i.e. you may set $\Sigma = 0$ to initialize or use the self energies from a previous combined DMFA/DCA-MEM run).

Another problem, specific to self consistent methods like DCA and especially DMFA, is that the error estimate for the data described above only accounts for the error from the last DMFA/DCA iteration, but not the error in the host Green's function $\mathcal{G}$ coming from the previous iteration (Fig. 34). The simplest way to deal with this, is to make sure that the penultimate iteration is rather precise. Another way is to sample the DCA procedure. I.e. performing thirty or so statistically independent DMFA/DCA simulations with different random number seeds and then use jackknife or bootstrap[48] statistical methods to calculate the covariance.

C. Bryan's Method: a MEM algorithm

We will now sketch Bryan's numerical algorithm to find the optimal spectrum. For a more detailed description, we refer the reader to his paper[49]. We have found his algorithm to be very appropriate for the numerical analytic continuation problem for two reasons: First, due to the exponential nature of the kernel which relates A to the data $\bar{G}$, we typically have $L \gg N_{good}$. Thus, the problem is usually "oversampled." Bryan tailored his numerical algorithm[49] to this type of problem by working in a reduced space whose dimension is

determined by singular-value-decomposition of the kernel K and is equal to the largest possible number of good singular values (i.e., numerically significant) which may parametrize the solution. The dimension of this space is usually much less than the number of A_i, and we found the computational advantage over methods that use the entire space determined by the number of A_i to be significant. Second, for the analytic continuation problem, the approximation of setting α equal to its optimal value is questionable because of the wide range of reasonably acceptable values of α. Bryan deals with this by calculating a result which is averaged over $P(\alpha|G, m)$.

1. Typical Algorithms

What distinguishes Bryan's numerical algorithm from its predecessors is the way in which the space of possible solutions is searched. Typical algorithms search for an optimal A by stepping through the entire space of A

$$A \to A + \delta A \tag{193}$$

with

$$\delta A = -(\nabla\nabla Q)^{-1}\nabla Q\,. \tag{194}$$

The Hessian $(\nabla\nabla Q)^{-1}$ is

$$(\nabla\nabla Q)^{-1} = (\alpha\nabla\nabla S - \nabla\nabla L)^{-1} = \left(\alpha\{A\}^{-1} - \nabla\nabla L\right)^{-1}\,, \tag{195}$$

where $\{A\}$ is a diagonal matrix with the elements of A along its diagonal. It may conceptually be expanded using the binomial theorem so that $(\nabla\nabla Q)^{-1}$ may be written as a power series in $\{A\}\nabla\nabla L$. Thus, δA may be written as a combination of $\{A\}\nabla Q = \{A\}\,(\alpha\nabla S - \nabla L)$, and powers of $\{A\}\nabla\nabla L$ acting on $\{A\}\nabla S$ and $\{A\}\nabla L$. Each of these vectors defines a direction in which the search can precede. Typically, between three and ten directions are used; however, these directions are often inappropriate for the problem at hand, because as mentioned earlier, the space of all possible solutions is too large for such oversampled data.

2. Singular-Space Algorithm

To alleviate this problem, Bryan performs a singular-value decomposition (SVD) of the kernel K, i.e., $K = V\Sigma U^T$ where U and V are orthogonal matrices and Σ is a diagonal

matrix, and works in the resulting singular space. To see that this space still contains the solution, we consider

$$\nabla L = \frac{\partial F}{\partial A}\frac{\partial L}{\partial F} = K^T \frac{\partial L}{\partial F} \tag{196}$$

where $F = KA$. We see that ∇L lies in the vector space defined by the columns of K^T. We next perform a SVD on K and assume the diagonal elements of Σ are ordered from largest to smallest. The smallest elements are essentially zero (to the numerical precision of the computer) since the kernel is effectively singular. However, s of the elements are assumed finite. Now the vector space spanned by the columns of K^T is the same as the space spanned by the columns of U associated with the non-singular values. Bryan calls this reduced space the *singular space*. Thus, to the precision that can be represented on the computer, $\{A\}\nabla L$ and all of the search directions formed by acting with $\{A\}\nabla\nabla L$ lie in the singular space spanned by the columns of $\{A\}U_s$, where U_s is the singular space projection of U. The only direction not in this space is $\{A\}\nabla S$. Thus, Bryan's algorithm works in *at most* an $s+1$-dimensional subspace of the N-dimensional space of A.

In this singular space, the condition for an extremum of Q, $\nabla Q = 0$, is

$$\alpha\nabla S - \nabla L = 0 \rightarrow -\alpha \ln\left(A_i/m_i\right) = \sum_j K_{ji}\frac{\partial L}{\partial F_j}\,. \tag{197}$$

Thus, the solution may be represented in terms of a vector u

$$\ln\left(A/m\right) = K^T u\,. \tag{198}$$

Unless K is of full rank, so that $s = N$, the components of u will not be independent. However, since K^T and U share the same vector space and since most of the relevant search directions lie in the singular space, Bryan proposes that the solution be represented in terms of U and u as

$$A_i = m_i \exp\sum_n U_{in}u_n\,. \tag{199}$$

Thus, to the precision to which it may be represented on the computer and determined by SVD, the space u must contain the solution defined by $\nabla Q = 0$, and the search can be limited to this s-dimensional space.

Bryan's algorithm precedes by first reducing all the relevant matrices to the singular space. With the definitions $K = V\Sigma U^T$ and $\log(A/m) = Uu$, the condition for an extremum becomes

$$-\alpha U u = U\Sigma V^T\frac{\partial L}{\partial F}\,, \tag{200}$$

or

$$-\alpha u = \Sigma V^T \frac{\partial L}{\partial F} \equiv g\,, \tag{201}$$

where each of these matrices and vectors has been reduced to the singular space. (u is now a vector of order s, Σ is an $s \times s$ diagonal matrix etc.) Bryan then uses a standard Newton's search to find the solution in the singular space, starting from an arbitrary u. The increment at each iteration is given by

$$J\delta u = -\alpha u - g\,, \tag{202}$$

where $J = \alpha I + \partial g/\partial u$ is the Jacobian matrix, I the identity matrix, and

$$\frac{\partial g}{\partial u} = \Sigma V^T \frac{\partial^2 L}{\partial F^2} \frac{\partial F}{\partial A} \frac{\partial A}{\partial u}\,. \tag{203}$$

With the definition $W = \partial^2 L/\partial F^2$ (which is just the diagonal matrix with elements $1/\sigma_l'^2$), $M = \Sigma V^T W V \Sigma$, and $T = U^T A U$. M and T are symmetric $s \times s$ matrices, the Jacobian $J = \alpha I + MT$, and

$$(\alpha I + MT)\,\delta u = -\alpha u - g\,. \tag{204}$$

At each iteration δu must be restricted in size so that the algorithm remains stable. Thus, another parameter μ (a Marquart-Levenberg parameter) is added

$$[(\alpha + \mu) I + MT]\,\delta u = -\alpha u - g \tag{205}$$

and adjusted to keep the step length $\delta u^T T \delta u$ below some the limit

$$\delta u^T T \delta u \leq \sum_i m_i \tag{206}$$

so the search is within the range of validity of a local quadratic expansion of Q.

This search can be made more efficient if Eq. 205 is diagonalized, so that of order s operations are required for each α μ pair. First, we diagonalize T

$$TP = P\Gamma \tag{207}$$

where P is an orthogonal matrix and Γ is diagonal with finite elements γ_i. Then we define

$$B = \{\gamma^{1/2}\} P^T M P \{\gamma^{1/2}\} \tag{208}$$

and solve the second eigenvalue equation

$$BR = R\Lambda \tag{209}$$

where R is orthogonal and Λ the diagonal matrix with finite elements λ_i. Finally, to diagonalize Eq. 205 we define

$$Y = P\{\gamma^{-1/2}\}R\,. \tag{210}$$

Then $Y^{-T}Y^{-1} = T$, and $Y^{-1}MY^{-T} = \Lambda$, so that

$$Y^{-1}\left[(\alpha+\mu)I + MT\right]\delta u = \left[(\alpha+\mu)I + \Lambda\right]Y^{-1}\delta u = Y^{-1}\left[-\alpha u - g\right] \tag{211}$$

which yields s independent equations for $Y^{-1}\delta u$. Again, as these equations are iterated, μ must be adjusted to keep the step length

$$\delta u^T T \delta u = \left|Y^{-1}\delta u\right|^2 \leq \sum_i m_i\,. \tag{212}$$

3. Selection of α

The value α is adjusted so that the solution iterates to either a fixed value of χ^2 (for historic MEM) or to a maximum value of $P(\alpha|G,m)$ given by Eq. 181 (for classic MEM). Then, A is obtained from

$$A_i = m_i \exp\left(\sum_{n=1}^{s} U_{in}u_n\right)\,. \tag{213}$$

Alternatively, Bryan suggests that one may start the algorithm with a large α for which $P(\alpha|\bar{G},m)$ is negligibly small, and then iterate to $\alpha \approx 0$ so that the averaged spectrum may be approximated

$$\langle A\rangle = \int_0^\infty d\alpha P(\alpha|G,m)\hat{A}(\alpha) \tag{214}$$

where $\hat{A}(\alpha)$ is the optimal spectrum (that for which $\nabla Q = 0$) for the value of α specified in the argument. This latter step may be necessary when $P(\alpha|G,m)$ is not a sharply peaked distribution. In fact this is usually the case, as may be seen in Fig. 32.

4. Error Propagation

As discussed in Sec. IV B 8, it is possible to assign error bars to integrated functions of the spectrum $H = \int d\omega A(\omega)h(\omega)$

$$\left\langle (\delta H)^2\right\rangle = \int\int d\omega d\omega'\, h(\omega)h(\omega')\left\langle \delta A(\omega)\delta A(\omega')\right\rangle\,, \tag{215}$$

where

$$\langle \delta A(\omega) \delta A(\omega') \rangle = - \left(\nabla\nabla Q|_{\hat{A}} \right)^{-1} . \tag{216}$$

This is the inverse of the Hessian of the algorithm discussed above. $\nabla\nabla Q|_{\hat{A}}$ and is easily calculated in terms of singular-space quantities

$$- \nabla\nabla Q|_{\hat{A}} = \{1/A\} U Y^{-T} \{\alpha I + \Lambda\} Y^{-1} U^T \{1/A\} . \tag{217}$$

Its inverse

$$- \left(\nabla\nabla Q|_{\hat{A}} \right)^{-1} = \{A\} U Y \left\{ \frac{1}{\alpha + \lambda} \right\} Y^T U^T \{A\} \tag{218}$$

may be used to calculate the error of H, $\sqrt{(\delta H)^2}$ for any α. In principle, one should average the error over $P(\alpha|m, \bar{G})$; however, we find that it is generally adequate to calculate the error of the spectrum at the optimal $\hat{\alpha}$.

We close this section with several practical comments: On a workstation, finding the optimal spectrum by searching in the singular space requires only a few minutes of computer time. This efficiency is in sharp contrast with the amount of computer we needed[50] even on a "supercomputer" for standard Newton algorithms[43] or simulated annealing methods that use the full space of A. We found it essential to use 64 bit arithmetic to obtain stable results. Also, we use LINPACK's[51] singular-value decomposition routine to do the SVD and also to compute any eigenvalues and eigenvectors. The SVD routine in Numerical Recipes[27] and the EISPACK[52] eigenvalue-eigenvector routine RS are not as stable.

D. Case Study I

In this section, we will demonstrate that it is possible to extract spectral densities from the quantum Monte Carlo data that are essentially free from artifacts caused by over fitting to the data and have only small and controllable amounts of statistical error. We will use as an example the electronic spectral densities of the infinite-dimensional periodic Anderson model (PAM). We have already qualified the local Greens function data to remove correlations using the procedure discussed in Sec. IV B 4, so we can begin to process the data to obtain the single-particle density of states spectral function.

For the majority of this section, we will consider particle-hole symmetric data $G(\tau) = G(\beta - \tau)$, and spectra $A(\omega) = A(-\omega)$. This prior information may imposed on the solution by

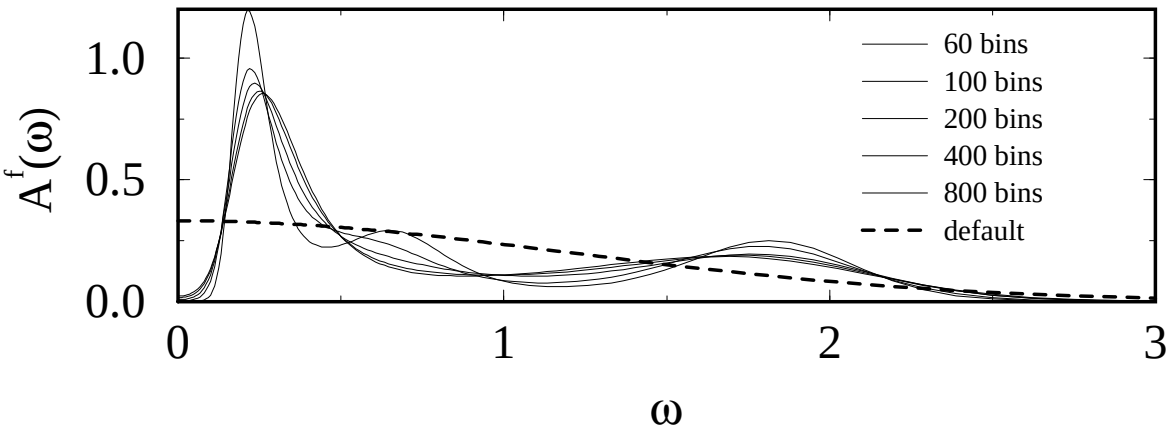

FIG. 35: A sequence of spectral densities generated with increasingly accurate data. Every time the number of bins of data is doubled, the error is reduced by $1/\sqrt{2}$. A Gaussian default model, the dashed line, was used. N_{bins} increases beyond $2L = 82$, spurious structures are quickly suppressed.

constructing a symmetric kernel and default models. We will use three symmetric default models: two non-informative models — the flat model $m(\omega) =$ constant and a simple Gaussian

$$m(\omega) = \frac{1}{\Gamma\sqrt{\pi}} \exp[-(\omega/\Gamma)^2] \tag{219}$$

and also a third one obtained from second-order perturbation theory in U[53,54]. The kernel for symmetric Fermionic Green's functions may be modified to reflect the symmetry and the associated integral restricted to positive frequencies

$$G(\tau) = \int_0^\infty d\omega A(\omega) \frac{e^{-\tau\omega} + e^{-(\tau-\beta)\omega}}{1 + e^{-\beta\omega}} \,. \tag{220}$$

1. *Convergence of the Spectra*

To minimize the effects of statistical error, the accuracy of the data needs to be increased until the spectral density has converged. This is demonstrated in Fig. 35, where the accuracy of the data are improved by increasing the number of bins of data. Here, a Gaussian default model is used whose width $\Gamma = 1.6$ (chosen by an optimization procedure to be discussed below). Each time the number of bins of data is doubled, the accuracy of the data increases by 41%. The spectral densities corresponding to smallest number of bins of data have spurious features associated with over fitting. These features are associated with difficulties in calculating the covariance matrix, as discussed in Sec. IV B 4. As N_{bins} increases beyond

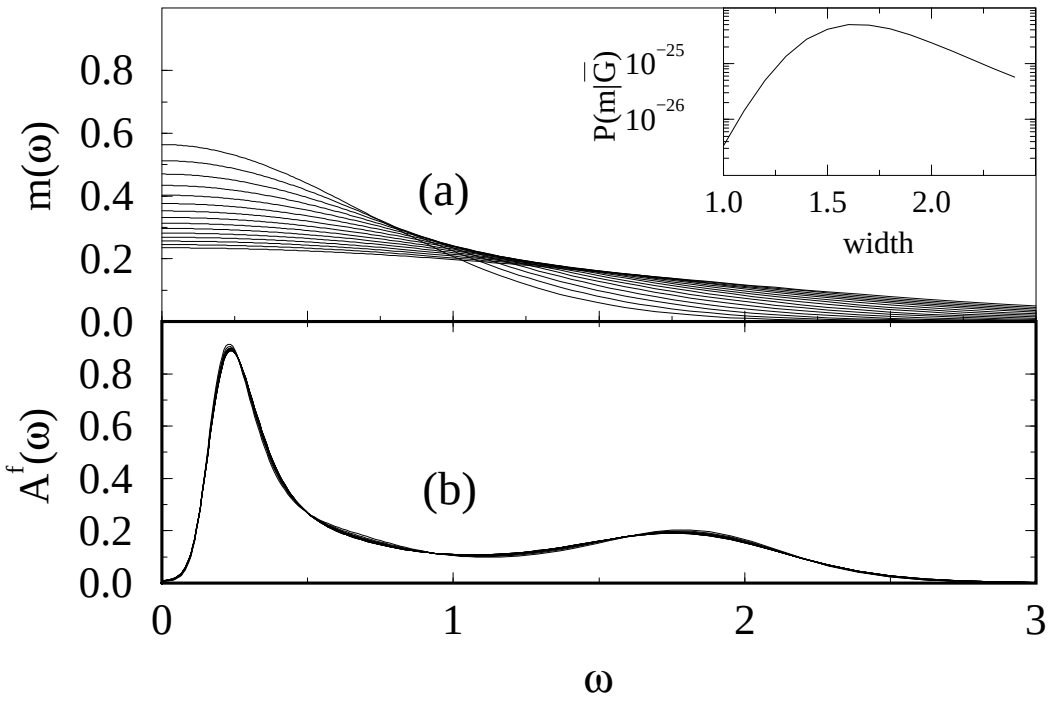

FIG. 36: Dependence of the spectral density upon the default model. The width Γ of the Gaussian default model (a) is varied, producing a series of spectral densities (b). In the inset to (a) is the posterior probability of the default model $P(m|\bar{G})$, produced by integrating the joint probability $P(A, \alpha, m|\bar{G})$ over α and A, is plotted as a function of Γ. The normalization of $P(m|\bar{G})$ is unknown because it depends upon the probability of the data and the prior probability of the default model which are unknown.

$2L = 82$, the spurious structure is quickly suppressed. By the time 800 bins of data have been used, the spectral density appears to be converged to several line widths.

2. Default Model Selection

One may also test the dependence of the spectral density on the default model by changing its parameters or by using different models. The best model is the one with the largest posterior probability, calculated by assuming that the prior probability of the default model is flat, so that $P(A, \alpha, m|\bar{G}) \propto P(A, \alpha|\bar{G}, m)$. Then $P(m|\bar{G})$ is obtained by integrating $P(A, \alpha, m|\bar{G})$ over A and α. The effects of varying the default model parameters are shown in Fig. 36a where the same data set is analytically continued with Gaussian default models whose widths satisfy $1.0 < \Gamma < 2.4$. The posterior probability $P(m|\bar{G})$ of these default

models, shown in the inset, is peaked around $\Gamma \approx 1.6$. (We note that the normalization of $P(m|\bar{G})$ is unknown, since the prior probability of the default model and data are unknown). The resulting spectral densities are shown in Fig. 36b and are found to depend only weakly upon the default model. It is also possible to optimize the perturbation theory default model and hence to optimize the corresponding spectral densities. In the optimization of the default for the PAM spectra, the *df*-hybridization V may be treated as a variational parameter.

3. Error Propagation

In Fig. 37, we compare the optimal spectral densities obtained with the optimal perturbation theory, Gaussian, and flat default models. (The flat default model, with no adjustable parameters, is not optimized.) The posterior probabilities for each result indicate that the perturbation theory default model produces by far the most probable spectral density. However, we note that the qualitative features of the spectral density change little with the default model even though a large variety of default models were used. This independence is one signature of good data!

As a final test of the quality of the spectral density, one can evaluate its error in different intervals of frequency. In Fig. 37, we chose to assign error bars to the integrated spectral density ($h(\omega) = 1$) over different non-overlapping regions. The width of the region centered at each error bar is indicated by the horizontal spread of the error bar, the spectral weight within this region is indicated by the value of the data point, while the estimate of the uncertainty is indicated by the vertical spread. The perturbation theory default model yields the most precise spectra at all frequencies, consistent with the posterior probabilities of the models.

4. Two-Particle Spectra

There are special difficulties associated with the calculation of spectral densities associated with two-particle Green's functions. These difficulties include noisier and more correlated data and the lack of a good default model. The latter problem stems from the traditional difficulties of performing perturbation theory for two-particle properties.

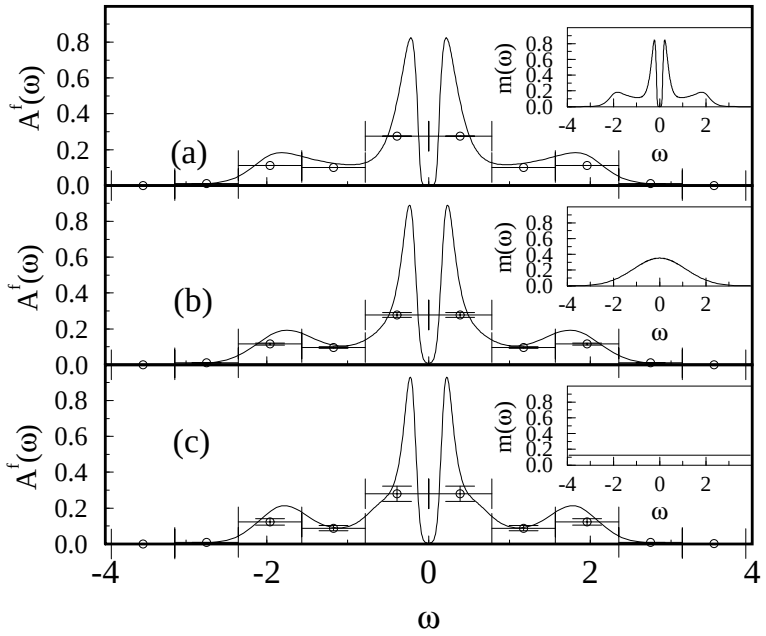

FIG. 37: The f-electron density of states $A^f(\omega)$ generated using (a) a perturbation theory, (b) a Gaussian, and (c) a flat default model. These models are shown as insets to each graph. The data points indicate the integrated spectral weight within 10 non-overlapping regions of width indicated by the horizontal error bar. The vertical error bar indicates the uncertainty of the integrated weight within each region.

As an example, we will analytically continue the local f-electron dynamic spin susceptibility $\chi''(\omega)$ of the symmetric PAM. As discussed in Sec. IV B, the Monte Carlo data $\chi(\tau) = 2\langle S^-(\tau)S^+(0)\rangle$ is related to $\chi''(\omega)$ by

$$\chi(\tau) = \int_0^\infty d\omega \frac{\omega[e^{-\tau\omega} + e^{-(\beta-\tau)\omega}](\chi''(\omega)/\omega)}{1 - e^{-\beta\omega}} \ . \tag{221}$$

To construct a model we will employ an alternative to perturbation theory, and construct a default model from different moments of the spectral function. They will be used as constraints to the principle of maximum entropy. The moments used to generate the default model are

$$\frac{1}{2}\chi(\omega = 0) = \int_0^\infty d\omega(\chi''(\omega)/\omega) \tag{222}$$

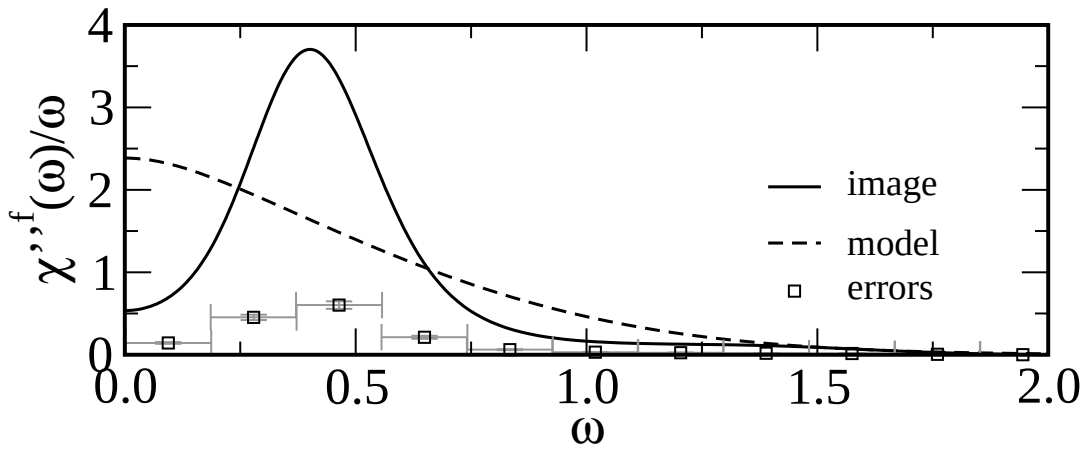

FIG. 38: $\chi''(\omega)/\omega$ for $V = 0.6$, $U = 2$ and $\beta = 20$ for the PAM generated using a default model defined by two moments of the spectral density. The data points indicate the integrated spectral weight within 10 non-overlapping regions of width indicated by the horizontal error bar. The vertical error bar indicates the uncertainty of the integrated weight within each region.

$$\chi(\tau = 0) = \int_0^\infty d\omega \, (\chi''(\omega)/\omega) \, \omega \coth(\beta\omega/2). \tag{223}$$

The (unnormalized) model is then generated by maximizing the entropy subject to these constraints imposed with Lagrange multipliers λ_0 and λ_1 and is easily found to be

$$m(\omega) = \exp[\lambda_0 + \lambda_1 \omega \coth(\beta\omega/2)] \tag{224}$$

where λ_0 and λ_1 are determined by the constraint equations above.

Clearly this procedure may be generalized to utilize an arbitrary number of measured moments and often provides a better default model than perturbation theory. However, as shown in Fig. 38, the final spectral density can differ significantly from the default model when defined in this way. Nevertheless, the error bars indicate that the spectral density is trustworthy.

5. Annealing Method

Occasionally we have reason to calculate a series of spectra for a variety of temperatures (i.e., for the calculation of transport coefficients). If this set is sufficiently dense, then starting from a perturbation theory default at high temperature, we may use the resulting spectra

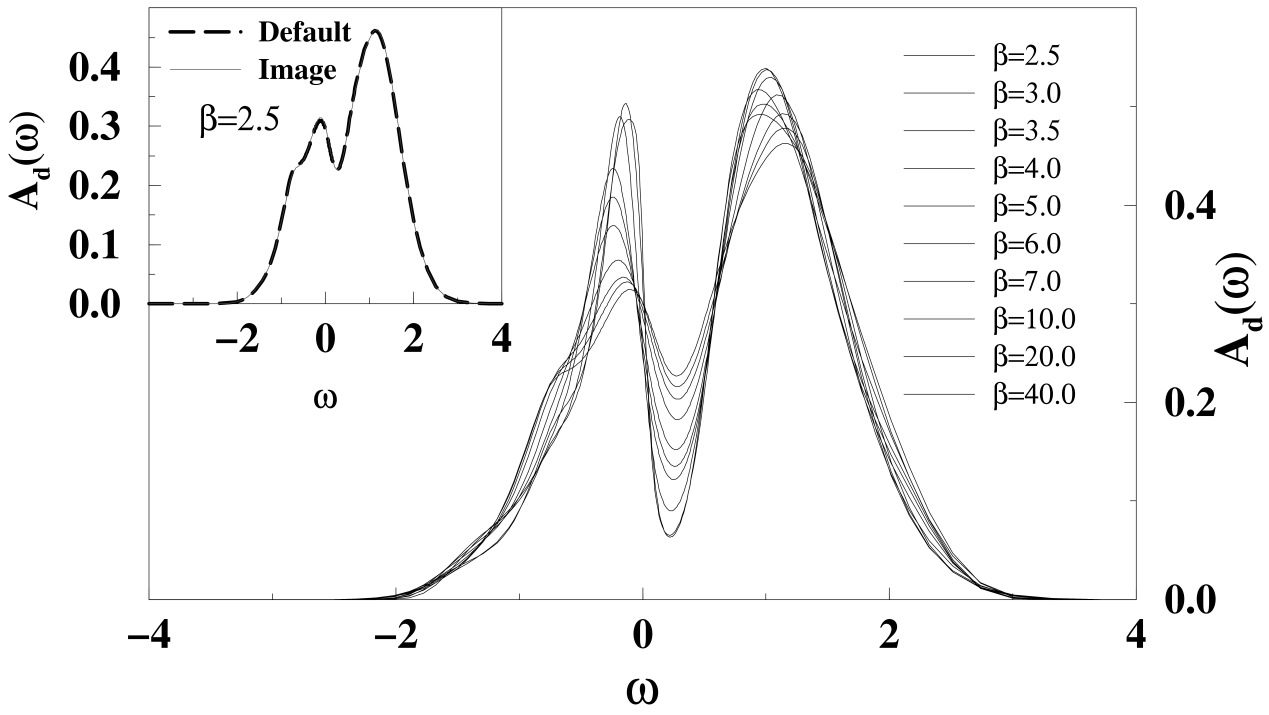

FIG. 39: *The evolution of the d-electron density of states of the asymmetric PAM when* $U = 1.5$, $V = 0.6$ $n_d = 0.6$, *and* $n_f = 1.0$. *At high temperatures, as shown in the inset, the spectra is in essentially exact agreement with second-order perturbation theory. In addition, the d-electron states far from the Fermi surface are weakly renormalized by the strong electronic correlation on the f-orbitals. Thus, as the temperature is lowered, the low-frequency spectra change continuously, whereas the high frequency features change very little.*

as a default model for the next lower temperature. As far as we know, this procedure has no Bayesian justification; however, it has significant physical motivation. At sufficiently high temperatures, perturbation theory often becomes exact. Thus, this annealing procedure may be initialized with an essentially exact result. Furthermore, as the temperature is lowered, we expect the high frequency features of many spectra to freeze out (this is an essential assumption behind the numerical renormalization group method). Thus, the QMC is only required to supply information about the low-frequency features. Since QMC is a discrete sampling procedure in Matsubara time, according to Nyquist's theorem QMC only provides information below the Nyquist frequency $\omega_N = \pi/\Delta\tau$. Thus, the perturbation theory provides the high-frequency information, the QMC the low-frequency information, and MEM provides a natural method for combining these information sources.

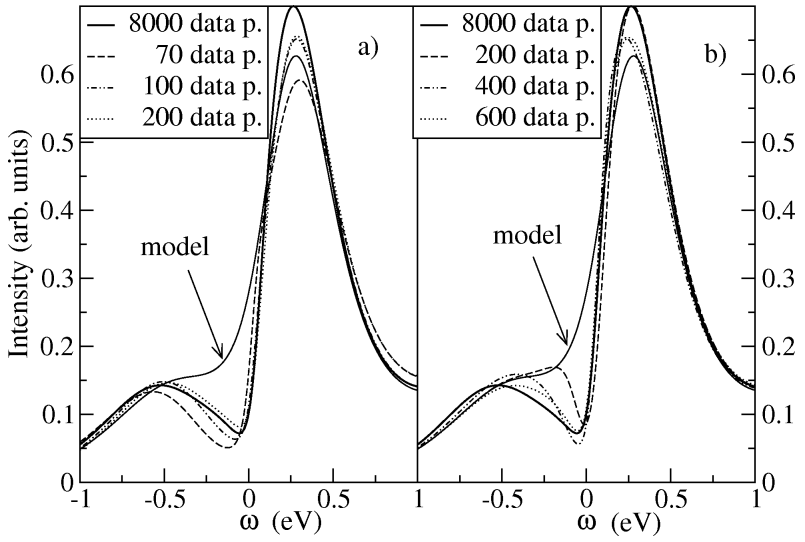

FIG. 40: One-particle spectra at $K = (\pi, \pi/2)$ calculated with different amounts of data using a) the new method and b) the old method.

For example, the evolution of the d-electron density of states of the asymmetric PAM is shown in Fig. 39. At high temperatures, as shown in the inset, the spectra is in essentially exact agreement with second-order perturbation theory. In addition, the d-electron states far from the Fermi surface are weakly renormalized by the strong electronic correlation on the f-orbitals. Thus, as the temperature is lowered, the low-frequency spectra change continuously, whereas the high frequency features change very little.

E. Case Study II

1. Spectra form data with sign problem

Here we present a comparison between the spectra obtained with the conventional (old) approach which does not consider the sign covariance, and the (new) method described in Sec. IV B 5 c. With the old method, Eq. 158 is used to calculate the binned data when the sign is not always one, and Eq. 151 is used to calculate its covariance. In the new method, the sign is incorporated into the data and Eq. 161 is used to calculate the covariance. In this example, we calculate the the one-particle spectrum. At the highest temperature, the model $m(\omega)$ used in the entropy functional is chosen to be a Gaussian function. The annealing procedure is then used to obtain the spectrum at lower temperatures.

In Fig. 40 (a) and (b) we show the one-particle spectra of the Hubbard model at $K = (\pi, \pi/2)$ calculated for different amounts of data with the new and respectively with the old method. In both cases, when a large amount of data is used (8000 data points) the spectrum (thick continuous line) is converged. Moreover the two methods produce the same spectrum. However, notice that with the new method a reasonably good spectrum, i.e., a spectrum close to the converged one, can be obtained with an amount of data as small as 100 data points (see the double-dotted dashed line in Fig. 40 (a)). On the other hand, the old method requires at least 600 data points for a spectrum of comparable quality (see the dotted line in Fig. 40 (b)). Thus, for the case under investigation, we find that the new method reduces the computational cost of calculating the one-particle spectra about six times. For other problems characterized by stronger correlations the improvement is even more significant.

If the sign is not very small, some experience is needed to decide whether to use the old or the new method. Clearly, the old method will fail when the sign is exactly one since the covariance, Eq. 161, will be pathological since the sign has no error. As a rough rule of thumb, we tend to use the old method when the sign is greater than 0.8 and the new method when it falls below this value.

F. Summary

The Maximum Entropy Method is a precise and systematic way of analytically continuing Matsubara-time quantum Monte Carlo results to real frequencies. Due to the exponential nature of the kernel which relates the spectra and the data, there are many A with very different character which correspond to the same $\bar{G}$. With the MEM we employ Bayesian statistics to determine which of these is most probable. Bayesian inference is also used to assign error bars to integrals over the spectrum and optimize the default model.

The posterior probability of the spectrum is given by the product of the prior probability and the likelihood function. The entropic nature of the prior insures that the only correlated deviations from the default model which appear in the spectrum are those which are necessary to reproduce the data. The form of the likelihood function is determined by the central limit theorem, assuming that the data are statistically independent and Gaussianly distributed. Insuring these preconditions is the most critical step in the MEM procedure, and requires that the data be systematically rebinned and that the data and the kernel be

rotated into the space in which the covariance of the data is diagonal.

Once the data has been properly characterized, we calculate the optimal spectrum using Bryan's algorithm which searches for a solution in the reduced singular space of the kernel. Bryan's method is more efficient than conventional techniques which search the entire spectral space. For any search algorithm three different techniques can be employed to set the Lagrange parameter α which determines the relative weight of the entropy and misfit: the historic, classic or Bryan's averaging technique. With precise uncorrelated data, each returns essentially the same spectrum, but with less-precise uncorrelated data, Bryan's technique yields the best results. Also, as the QMC data are systematically improved, images produced with Bryan's technique appear to converge more quickly than those produced by the other techniques.

While the systematic preparation of the data as described in Sec. IV B 4 and the qualification of the spectrum described in later subsections is time-consuming, we believe that it is as important to quality of the final result, as is an accurate MEM code.

Together, the techniques discussed in this section provide a powerful, accurate, and systematic approach to the analytic continuation problem. In each case where we have employed these techniques we have been able to produce spectra that are precise at low frequencies, and free from spurious (unjustified) features at all ω.

1. Steps to ensure a robust MEM calculation.

In this section we have summarized the proper implementation of MEM and given a number of cautionary notes. As a parting discussion, these will be summarized here along with a few other common sense rules of thumb.

1. Rebin your data to remove correlations in QMC time.
2. Generate sufficient bins of data so that $N_{bins} \gtrsim 2L$ where L is the number of time slices used.
3. If DMFA/DCA is used to generate the data, be sure that the error in $\mathcal{G}$ from the previous iteration is negligible.
4. Calculate the covariance of the data making sure that: (1) the eigenvalue spectrum is continuous (if not, increase N_{bins}), and (2) that the covariance matrix is well enough

conditioned to allow it to be diagonalized (if not, the data is oversampled in Matsubara time).

5. Diagonalize the covariance and rotate the data and kernel into the diagonal frame.

6. Choose a good default model, hopefully you can use the annealing technique. Always use a non-informative default model unless you are *certain* that the information in the model is exact.

7. When possible, use Bryan's MEM for marginalizing over α.

8. Systematically improve your data until the calculated spectrum converges.

9. When the annealing method is used, if the temperature step appears large (i.e. the spectrum changes abruptly) you may want to introduce data at additional intermediate temperatures.

10. If the annealing method is not used, try different non-informative default models. A reliable result is independent of the model. You may also want to use the model with the highest posterior probability (calculated when Bryan's method is used).

11. If the sign problem is present in the QMC calculations relate the fluctuations in sign to the norm of the spectra and treat the correlations between the sign and the data by including the sign in the covariance matrix as discussed in Sec. IV B 5.

V. CONCLUSION

We have presented a pedagogical introduction to the dynamical mean field and dynamical cluster approximations, and described two quantum Monte Carlo methods as cluster solvers with the Maximum entropy method used to obtain real frequency spectra. The DMFA is a very powerful method for the study of correlated systems, while DCA incorporates non-local corrections. Both the DMFA and DCA map the lattice onto a self-consistently embedded impurity problem, so that conventional cluster solvers, like Hirsch-Fye or Continuous time QMC, may be used. Both techniques are straightforward, but are limited by the minus sign problem to the study of relatively small clusters at high temperatures. The maximum entropy method is used to affect the Wick rotation of the QMC data from Matsubara time

or frequency to real frequency. Although MEM is a reliable and systematic method, great care must be taken in the application of MEM to ensure that the best possible spectra is found devoid of spurious features. Together, DMFA/DCA with QMC and MEM provide the state-of-the-art for the study of correlated model systems.

We would like to acknowledge useful conversations and fruitful collaborations with J. K. Freericks, B. Goodman, C. Groetsch, E. Gull, M. Hettler, H. R. Krishnamurthy, M. Ma, Th. Maier, R. Scalettar, R. N. Silver, D. Sivia, Th. Pruschke, A. N. Tahvildarzadeh. M. Troyer, and F. C. Zhang. This work was supported by the National Science Foundation grants DMR-0706379 and DMR-0312680, and Department of Energy grant DE-FG02-04ER46129. The work of JEG was supported of the U. S. Department of Energy through the LANL/LDRD program.

---

† Electronic address: jarrellphysics@gmail.com

[1] J. Hubbard, Proc. R. Soc. a**276**, 238(1963); M. C. Gutzwiller, Phys. Rev. Lett. **10**, 159(1963); J. Kanamori, Prog. Theor. Phys. **30**, 257(1963).

[2] Proceedings of the international conference on "Itinerant-Electron Magnetism". Physica B+C**91** (1977); for a review about the theory of the Hubbard model see also D. Vollhardt, Rev. Mod. Phys. 56, 99(1984).

[3] E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968); H. Frahm and V. E. Korepin, Phys. Rev. B**42**, 10533(1990); N. Kawakami and S.-K. Yang, Phys. Rev. Lett. **65**, 2309(1990).

[4] G. Czycholl, J. Mag. Mag. Mat. **47,48**, 350 (1985).

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

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

[7] P. G. J. van Dongen and D. Vollhardt, Phys. Rev. Lett. **65**, 1663 (1990).

[8] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H. R. Krishnamurthy, Phys. Rev. **B 58,** 7475 (1998).

[9] M. H. Hettler, M. Mukherjee, M. Jarrell, and H. R. Krishnamurthy, Phys. Rev. B 61, 12739 (2000).

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

[11] Y. Kuramoto, Springer Series in Sol. State Science, Vol. 62, 152–162 (1985); C. Kim, Y. Ku-

ramoto and T. Kasuya, J. Phys. Soc. Japan **59**, 2414 (1990).

[12] T. Pruschke, M. Jarrell and J .K. Freericks, Adv. in Phys. **42,** 187 (1995).

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

[14] R. M. Fye and J. E. Hirsch, Phys. Rev. B **38**, 433 (1988).

[15] M. Jarrell, Th. Maier, C. Huscroft, and S. Moukouri, Phys. Rev. B 64, 195130/1-23 (2001).

[16] M. Jarrell, and J. E. Gubernatis, Physics Reports Vol. **269** #3, (1996).

[17] K. Aryanpour, M. H. Hettler and M. Jarrell, Phys. Rev. B 65, 153102/1-4 (2002).

[18] V. Zlatić and B. Horvatić, Sol. St. Comm. **75**, 263 (1990).

[19] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287, (1961).

[20] P. Kent, M. Jarrell, T. Maier, and T. Pruschke, Phys. Rev. B 72, 060411 (2005).

[21] D. D. Betts, H. Q. Lin, and J. S. Flynn, Can. J. Phys. 77, 353 (1999); D. D. Betts and G. E. Stewart, Can. J. Phys. 75, 47 (1997).

[22] A. A. Abrikosov, L. P. Gorkov, I. E. Dzyalishinski, *Methods of Quantum Field Theory in Statistical Physics*, (Dover, New York, 1975).

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

[24] J. E. Hirsch, Phys. Rev. B, **28**, 4059 (1983), and **31**, 4403 (1985).

[25] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth. A. H. Teller and E. Teller, J. Chem. Phys. **21**, 1087-1092 (1953).

[26] J. J. Deisz, D. W. Hess and J. W. Serene, *Recent Progress in Many-Body Theories*, edited by E. Schachinger *et al.* (Plenum, New York, 1995), Vol. 4.

[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77*, Second Edition (1992), Cambridge University Press.

[28] M. Jarrell, H. Akhlaghpour, and Thomas Pruschke, *Quantum Monte Carlo Methods in Condensed Matter Physics*, Ed. M. Suzuki, (World Scientific, 1993).

[29] Transactions on Mathematical Software, Vol. **22**, No. 3, p. 362–371 (September, 1996).

[30] A. N. Rubtsov, V. V. Savkin, A. I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005)

[31] F. F. Assaad and T. C. Lang, Phys. Rev. B **76**, 035116 (2007).

[32] M. Wagner, Phys. Rev. B **44**, 6104 (1991); J. Rammer, H. Smith, Rev. Mod. Phys. **58**, 323 (1986).

[33] H.-B. Schüttler and D. J. Scalapino, Phys. Rev. Lett. **55**, 1204 (1985); Phys. Rev. B **34**, 4744 (1986).

[34] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).

[35] G. Wahba, SIAM Journal on Numerical Analysis **14**, 651 (1977).

[36] S. R. White, D.J. Scalapino, R. L. Sugar, and N. E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989).

[37] M. Jarrell and O. Biham, Phys. Rev. Lett. **63**, 2504 (1989).

[38] Th. Maier, M. Jarrell, Th. Pruschke, and M. Hettler, Rev. Mod. Phys. **77** 1027 (2005).

[39] J. Skilling, in *Maximum Entropy and Bayesian Methods* edited by J. Skilling (Kluwer Academic, Dordrecht), p. 45 (1989).

[40] S .F. Gull and J. Skilling, IEE Proceedings **131**, 646 (1984).

[41] For a discussion of kurtosis and skewness, as well as a discussion of the probability that a distribution is Gaussian, see Ref[27] chap. 14.

[42] It is tempting to disregard (i.e., set to 0) the off-diagonal elements of the covariance matrix as an alternative method of alleviating this pathology. Then, the eigenvalues will simply be the well-defined variance of $\bar{G}(\tau)$. However, this procedure neglects the correlations in the error which are clearly visible in Fig. 30 and yeilds an incorrect likelihood function. We have found that this procedure produces unpredictable results, especially when the data quality is marginal.

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

[44] S. F. Gull, in *Maximum Entropy and Bayesian Methods* edited by J. Skilling (Kluwer, Dordrecht), p. 53 (1989).

[45] S. F. Gull and G. J. Daniels, Nature, **272**, 686 (1978).

[46] H. Jeffreys, *Theory of Probability*, (Oxford, Claredon Press, 1939); see also E. Jaynes, IEEE Trans. Sys. Sci. Cyb. Vol. SSC-4 (1993).

[47] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, Kluwer Academic, Dordrecht, p. 455 (1989).

[48] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, 1982); B. Efron and G. Gong, The American Statistician, **37**, 36 (1983).

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

[50] M. Jarrell, J .E. Gubernatis, and R. N. Silver. Phys. Rev. B, **44**, 5347-50 (Sept. 1991).

[51] J. J. Dongarra, C. B. Moler, J. R. Bunch, and G.W. Stewart, *LINPACK User's Guide*,SIAM, Philadelphia, (1979).

[52] B. T. Smith, J. M. Boyle, Y. Ikebe, V. C. Klema, and C. B. Moler, *Matrix Eigensystems Routines — EISPACK Guide* , Springer-Verlag, New York (1976).

[53] M. Jarrell, H. Akhlaghpour, and Th. Pruschke, Phys. Rev. Lett. **70**, 1670 (1993); Phys. Rev. B. **51**, 7429-40 (15, March 1995).

[54] B. Horvatic, D. Sokcevic, and V. Zlatic, Phys. Rev. B **36**, 675 (1987).

[55] S. R. White, D.J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, Phys. Rev. B. **40** 506 (1989); E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994) .