Non-perturbative conserving approximations and Luttinger's sum rule

Jutta Ortloff, Matthias Balzer, Michael Potthoff

Institut für Theoretische Physik und Astrophysik,
Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

Weak-coupling conserving approximations can be constructed by truncations of the Luttinger-Ward functional and are well known as thermodynamically consistent approaches which respect macroscopic conservation laws as well as certain sum rules at zero temperature. These properties can also be shown for variational approximations that are generated within the framework of the self-energy-functional theory without a truncation of the diagram series. Luttinger's sum rule represents an exception. We analyze the conditions under which the sum rule holds within a non-perturbative conserving approximation. Numerical examples are given for a simple but non-trivial dynamical two-site approximation. The validity of the sum rule for finite Hubbard clusters and the consequences for cluster extensions of the dynamical mean-field theory are discussed.

I. INTRODUCTION

Continuous symmetries of a Hamiltonian imply the existence of conserved quantities: The conservation of total energy, momentum, angular momentum, spin and particle number is enforced by a not explicitly time-dependent Hamiltonian which is spatially homogeneous and isotropic and invariant under global SU(2) and U(1) gauge transformations. For the treatment of a macroscopically large quantum system of interacting fermions, approximations are inevitable in general. Approximations, however, may artificially break symmetries and thus lead to unphysical violations of conservations laws.

Baym and Kadanoff[1,2] have analyzed under which circumstances an approximation for time-dependent correlation functions, and for one- and two-particle Green's functions in particular, respect the mentioned macroscopic conservation laws. They were able to give corresponding rules for a proper construction of approximations, namely criteria for selecting suitable classes of diagrams, within diagrammatic weak-coupling perturbation theory. Weak-coupling approximations following these rules and thus respecting conservation laws are called "conserving". Frequently cited examples for conserving approximations are the Hartree-Fock or the fluctuation-exchange approximation.[1,3,4]

Baym[2] has condensed the method of constructing conserving approximations into a compact form: A conserving approximation for the one-particle Green's function G is obtained by using Dyson's equation $G = 1/(G_0^{-1} - \Sigma)$ with (the free, $U = 0$, Green's function G_0 and) a self-energy $\Sigma = \Sigma_U[G]$ given by a universal functional. Apart from G, the universal functional Σ_U must depend on the interaction parameters U only. Furthermore, the functional must satisfy a vanishing-curl condition or, alternatively, must be derivable from some (universal) functional $\Phi_U[G]$ as $T\Sigma_U[G] = \delta\Phi_U[G]/\delta G$ (the temperature T is introduced for convenience). In short, "Φ-derivable" approximations are conserving.

Φ-derivable approximations have been shown[2] to exhibit several further advantageous properties in addition.

One of these concerns the question of thermodynamical consistency. There are different ways to determine the grand potential of the system from the Green's function which do not necessarily yield the same result when using approximate quantities. On the one hand, Ω may be calculated by integration of expectation values, accessible by G, with respect to certain model parameters. For example, Ω may be calculated by integration of the average particle number, as obtained from the trace of G, with respect to the chemical potential μ. On the other hand, Ω can be obtained as $\Omega = \Phi + \text{Tr} \ln G - \text{Tr}(\Sigma G)$ without integration. A Φ-derivable approximation consistently gives the same result for Ω in both ways.

At zero temperature $T = 0$ there is another non-trivial theorem which is satisfied by any Φ-derivable approximation, namely Luttinger's sum rule.[5,6] This states that the volume in reciprocal space that is enclosed by the Fermi surface is equal to the average particle number. The original proof of the sum rule by Luttinger and Ward[5] is based on the existence of Φ in the exact theory and is straightforwardly transferred to the case of a Φ-derivable approximation. This also implies that other Fermi-liquid properties, such as the linear trend of the specific heat at low T and Fermi-liquid expressions for the $T = 0$ charge and the spin susceptibility are respected by a Φ-derivable approximation.

There is a perturbation expansion[5,7] which gives the Luttinger-Ward functional $\Phi_U[G]$ in terms of closed skeleton diagrams (see Fig. 1). As a manageable Φ-
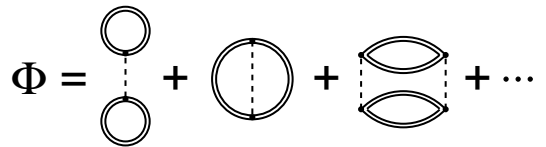

FIG. 1: Diagrammatic representation of the Luttinger-Ward functional $\Phi_U[G]$. Double lines stand for the interacting one-particle Green's function G, dashed lines represent the vertices U.

derivable approximation must specify a (universal) functional $\Phi_U[G]$ that can be evaluated in practice, one usually considers truncations of the expansion and sums up a certain subclass of skeleton diagrams only. This, however, means that the construction of conserving approximations is restricted to the weak-coupling limit.

One purpose of the present paper is to show that it is possible to construct Φ-derivable approximations for lattice models of correlated fermions with local interactions which are *non-perturbative*, i.e. do not employ truncations of the skeleton-diagram expansion. The idea is to employ the self-energy-functional theory (SFT).[8,9,10] The SFT constructs the Luttinger-Ward functional $\Phi_U[G]$, or its Legendre transform $F_U[\Sigma]$, in an indirect way, namely by making contact with an exactly solvable reference system. Thereby, the exact functional dependence of $F_U[\Sigma]$ becomes available on a certain subspace of self-energies which is spanned by the self-energies generated by the reference system.

The obvious question is whether those non-perturbative Φ-derivable approximations have the same properties as the weak-coupling Φ-derivable approximations suggested by Baym and Kadanoff. This requires the discussion of the following points:

(i) *Macroscopic conservations laws.* For fermionic lattice models, conservation of energy, particle number and spin have to be considered. Besides the static thermodynamics, the SFT concept concentrates on the *one-particle* excitations. For the approximate one-particle Green's function, however, it is actually simple to prove that the above conservation laws are respected. A short discussion is given in Appendix A.

(ii) *Thermodynamical consistency.* This issue has already been addressed in Ref. 11. It has been shown that the μ derivative of the (approximate) SFT grand potential (including a minus sign) equals the average particle number $\langle N \rangle$ as obtained by the trace of the (approximate) Green's function. The same holds for *any* one-particle quantity coupling linearly via a parameter to the Hamiltonian, e.g. for the average total spin $\langle S \rangle$ coupling via a field of strength B.

(iii) *Luttinger sum rule.* This is the main point to be discussed in the present paper. There are different open questions: First, it is straightforward to prove that weak-coupling Φ-derivable approximations respect the sum rule as one can directly take over the proof for the exact theory. For approximations constructed within the SFT, a different proof has to be given. Second, it turns out that a non-perturbative Φ-derivable approximation respects the sum rule if and only if the sum rule holds for the reference system that is used within the SFT. As the original and thereby the related reference system may be studied in the strong-coupling regime, this raises the question which reference system does respect the sum rule, i.e. which approximation is consistent with the sum rule. Third, it will be particularly interesting to study reference systems which generate dynamical impurity approximations (DIA)[8,9] and variational cluster approxi-

mations (VCA),[10,12] as these consist of a *finite* number of degrees of freedom. Does the Luttinger sum rule hold for finite systems? Do the DIA and the VCA respect the sum rule? What is the simplest approximation consistent with the sum rule? Note that finite reference systems consisting of a few sites only have been shown[9,13,14,15,16,17,18,19] to generate approximations which qualitatively capture the main physics correctly. Finally, it is important to understand these issues in order to understand whether and how a violation of the sum rule is possible within cluster extensions[20,21,22,23] of the dynamical mean-field theory (DMFT).[24,25,26,27,28] Note that the SFT comprises the DMFT and certain[29] cluster extensions and that possible violations of the sum rule in the two-dimensional lattice models have been reported,[30,31,32] including a study using the dynamical cluster approximation (DCA).[33]

The paper is organized as follows: A brief general discussion of the Luttinger sum rule is given in the next section, and a form of the sum rule specific to systems with a finite number of spatial degrees of freedom is derived. Sec. III clarifies the status of the sum rule with respect to non-perturbative approximations generated within the SFT framework. The results are elucidated by several numerical examples obtained for the most simple but non-trivial non-perturbative conserving approximation in Sec. IV. Violations of the sum rule in finite systems and their consequences are discussed in Sec. V. Finally, Sec. VI summarizes our main conclusions.

II. LUTTINGER SUM RULE

A system of interacting electrons on a lattice is generally described by a Hamiltonian $H(t, U) = H_0(t) + H_1(U)$ consisting of a one-particle part $H_0(t)$ and an interaction $H_1(U)$ with one-particle and interaction parameters t and U, respectively. As a prototype, let us consider the single-band Hubbard model[34,35,36] on a translationally invariant D dimensional lattice consisting of L sites with periodic boundary conditions. The Hamiltonian is given by:

$$ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \frac{U}{2} \sum_{i\sigma} n_{i\sigma} n_{i-\sigma} \,. \tag{1} $$

Here, $i = 1, ..., L$ refers to the sites, $\sigma = \uparrow, \downarrow$ is the spin projection, $c_{i\sigma}$ ($c_{i\sigma}^\dagger$) annihilates (creates) an electron in the one-electron state $|i\sigma\rangle$, and $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$. Fourier transformation diagonalizes the hopping matrix t and yields the dispersion $\varepsilon(k)$. There are L allowed k points in the first Brillouin zone.

Let $G = G_{t,U}$ denote the one-electron Green's function of the model $H(t, U)$. In case of the Hubbard model, its elements are given by $G_{ij}(\omega) = \langle\langle c_{i\sigma}; c_{j\sigma}^\dagger \rangle\rangle_\omega$. In the absence of spontaneous symmetry breaking, the Green's function is spin-independent and diagonal in reciprocal space. It can be written as $G_k(\omega) = 1/(\omega + \mu - \varepsilon(k) - \Sigma_k(\omega))$ where μ is the chemical potential and $\Sigma_k(\omega)$ the

self-energy. We also introduce the notation $\boldsymbol{\Sigma_{t,U}}$ for the self-energy, and $\boldsymbol{G_{t,0}} = 1/(\omega + \mu - \boldsymbol{t})$ for the free (non-interacting) Green's function which exhibits the dependence on the model parameters but suppresses the frequency dependence. Dyson's equation then reads as $\boldsymbol{G_{t,U}} = 1/(\boldsymbol{G_{t,0}^{-1}} - \boldsymbol{\Sigma_{t,U}})$.

The Luttinger sum rule[5,6] states that

$$\langle N \rangle = 2 \sum_{\boldsymbol{k}} \Theta(G_{\boldsymbol{k}}(0)) \tag{2}$$

where $N = \sum_{i\sigma} n_{i\sigma}$ is the particle-number operator, $\langle N \rangle$ its ($T = 0$) expectation value, and Θ the Heaviside step function. The factor 2 accounts for the two spin directions. Since $G_{\boldsymbol{k}}(0)^{-1} = \mu - \varepsilon(\boldsymbol{k}) - \Sigma_{\boldsymbol{k}}(0)$, the sum gives the number of $\boldsymbol{k}$ points enclosed by the interacting Fermi surface which, for $L \to \infty$, is defined via $\mu - \varepsilon(\boldsymbol{k}) - \Sigma_{\boldsymbol{k}}(0) = 0$. In the thermodynamic limit the sum rule therefore equates the average particle number with the Fermi-surface volume (apart from a factor $(2\pi)^D/L$). Note that, as $\Theta(G_{\boldsymbol{k}}(0)) = \Theta(1/G_{\boldsymbol{k}}(0))$, the sum rule Eq. (2) also includes the so-called Luttinger volume[37] which (for $L \to \infty$) is enclosed by the *zeros* of $G_{\boldsymbol{k}}(0)$.

The standard proof of the sum rule can be found in Ref. 5. It is based on diagrammatic perturbation theory to all orders which is used to construct the Luttinger-Ward functional $\Phi_{\boldsymbol{U}}[\boldsymbol{G}]$ as the sum of renormalized closed skeleton diagrams (see Fig. 1). We emphasize that the original proof straightforwardly extends also to finite systems. For $L < \infty$ the sum in Eq. (2) is discrete. Actually, the proof is performed for finite L first, and the thermodynamic limit (if desired) can be taken in the end. The limit $T \to 0$, on the other hand, is essential and is responsible for possible violations of the sum rule (see Sec. V).

Below we need an alternative but equivalent formulation of the sum rule. We start from the following (Lehmann) representation for the Green's function:

$$G_{\boldsymbol{k}}(\omega) = \sum_m \frac{\alpha_m(\boldsymbol{k})}{\omega + \mu - \omega_m(\boldsymbol{k})} \,. \tag{3}$$

Here, $\omega_m(\boldsymbol{k}) - \mu$ are the (real) poles and $\alpha_m(\boldsymbol{k})$ the (real and positive) weights. For real frequencies ω, it is then easy to verify the identity:

$$\Theta(G_{\boldsymbol{k}}(\omega)) = \sum_m \Theta(\omega + \mu - \omega_m(\boldsymbol{k})) - \sum_n \Theta(\omega + \mu - \zeta_n(\boldsymbol{k})) \tag{4}$$

where $\zeta_n(\boldsymbol{k}) - \mu$ is the n-th (real) zero of the Green's function, i.e. $G_{\boldsymbol{k}}(\zeta_n(\boldsymbol{k}) - \mu) = 0$.

For temperature $T = 0$ we have $\langle N \rangle = 2 \sum_{\boldsymbol{k}} \int_{-\infty}^0 d\omega(-1/\pi)\mathrm{Im}G_{\boldsymbol{k}}(\omega + i0^+)$ and thus $\langle N \rangle = 2 \sum_{\boldsymbol{k}} \sum_m \alpha_m(\boldsymbol{k})\Theta(\mu - \omega_m(\boldsymbol{k}))$. Hence, the Luttinger sum

rule reads:

$$2 \sum_{\boldsymbol{k}} \sum_m \alpha_m(\boldsymbol{k})\Theta(\mu - \omega_m(\boldsymbol{k}))$$
$$= 2 \sum_{\boldsymbol{k}} \left(\sum_m \Theta(\mu - \omega_m(\boldsymbol{k})) - \sum_n \Theta(\mu - \zeta_n(\boldsymbol{k})) \right) \,. \tag{5}$$

This form of the sum rule is convenient for the discussion of finite systems with $L < \infty$.

III. SELF-ENERGY-FUNCTIONAL THEORY AND LUTTINGER SUM RULE

Within the self-energy-functional theory (SFT),[8,9,10] the grand potential Ω is considered as a functional of the self-energy:

$$\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma}] = \mathrm{Tr}\ln \frac{1}{\boldsymbol{G_{t,0}^{-1}} - \boldsymbol{\Sigma}} + F_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{6}$$

Here, the trace Tr of a quantity $\boldsymbol{A}$ is defined as $\mathrm{Tr}\boldsymbol{A} \equiv T \sum_n 2 \sum_{\boldsymbol{k}} e^{i\omega_n 0^+} A_{\boldsymbol{k}}(i\omega_n)$ where $i\omega_n = i(2n+1)\pi T$ are the fermionic Matsubara frequencies, and the functional $F_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is the Legendre transform of the Luttinger-Ward functional $\Phi_{\boldsymbol{U}}[\boldsymbol{G}]$. The self-energy functional (6) is stationary at the physical self-energy, $\delta\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma_{t,U}}]/\delta\boldsymbol{\Sigma} = 0$, and, if evaluated at the physical self-energy, yields the physical value for the grand potential: $\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma_{t,U}}] = \Omega_{\boldsymbol{t,U}} \equiv -T \ln \mathrm{tr}\exp(-\beta(H(\boldsymbol{t},\boldsymbol{U}) - \mu N))$ where $\beta = 1/T$.

Comparing with the self-energy functional

$$\Omega_{\boldsymbol{t',U}}[\boldsymbol{\Sigma}] = \mathrm{Tr}\ln \frac{1}{\boldsymbol{G_{t',0}^{-1}} - \boldsymbol{\Sigma}} + F_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \tag{7}$$

of a reference system with the same interaction but a modified one-particle part, i.e. with the Hamiltonian $H(\boldsymbol{t'},\boldsymbol{U})$, the not explicitly known but only $\boldsymbol{U}$-dependent functional $F_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ can be eliminated:

$$\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma}] = \Omega_{\boldsymbol{t',U}}[\boldsymbol{\Sigma}] + \mathrm{Tr}\ln \frac{1}{\boldsymbol{G_{t,0}^{-1}} - \boldsymbol{\Sigma}} - \mathrm{Tr}\ln \frac{1}{\boldsymbol{G_{t',0}^{-1}} - \boldsymbol{\Sigma}} \,. \tag{8}$$

An approximation is constructed by searching for a stationary point of the self-energy functional on the subspace of trial self-energies spanned by varying the one-particle parameters $\boldsymbol{t'}$:

$$\frac{\partial\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma_{t',U}}]}{\partial\boldsymbol{t'}} = 0 \,. \tag{9}$$

Inserting a trial self-energy into Eq. (8) yields

$$\Omega_{\boldsymbol{t,U}}[\boldsymbol{\Sigma_{t',U}}] = \Omega_{\boldsymbol{t',U}} + \mathrm{Tr}\ln \frac{1}{\boldsymbol{G_{t,0}^{-1}} - \boldsymbol{\Sigma_{t',U}}} - \mathrm{Tr}\ln \boldsymbol{G_{t',U}} \,. \tag{10}$$

The decisive point is that the r.h.s. can be evaluated exactly for a reference system which is exactly solvable.

Apart from the free Green's function $\boldsymbol{G}_{\boldsymbol{t},0}$, it involves quantities of the reference system only.

This strategy to generate approximations has several advantages: (i) Contrary to the usual conserving approximations, the exact functional form of $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is retained. Any approximation is therefore non-perturbative by construction. On the level of one-particle excitations, macroscopic conservation laws are respected as shown in Appendix A. (ii) With $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]$ evaluated at the stationary point $\boldsymbol{t}' = \boldsymbol{t}'_{\mathrm{s}}$, an approximate but explicit expression for a thermodynamical potential is provided. As all physical quantities derive from this potential, the approximation is thermodynamically consistent in itself (see Ref. 11 for details). (iii) As different reference systems generate different approximations, the SFT provides a unifying framework that systematizes a class of "dynamic" approximations (see Refs. 29,38 for a discussion).

In the following we discuss the question whether or not a dynamic approximation respects the Luttinger sum rule. For this purpose consider first the $\mathrm{Tr}\ln(\cdots)$ terms in Eq. (10). These can be evaluated using the analytical and causal properties of the Green's functions as described in Ref. 9 (see Eq. (4) therein). Using $-T\ln(1+\exp(-\omega/T)) \to \omega\Theta(-\omega)$ for $T \to 0$ yields:

$$\mathrm{Tr}\ln\frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}}$$
$$= 2\sum_{\boldsymbol{k}}\sum_{m}(\omega_m(\boldsymbol{k}) - \mu)\Theta(\mu - \omega_m(\boldsymbol{k}))$$
$$- 2\sum_{\boldsymbol{k}}\sum_{n}(\zeta_n(\boldsymbol{k}) - \mu)\Theta(\mu - \zeta_n(\boldsymbol{k})) . \qquad (11)$$

Analogously, we have

$$\mathrm{Tr}\ln\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}} = 2\sum_{\boldsymbol{k}}\sum_{m}(\omega'_m(\boldsymbol{k}) - \mu)\Theta(\mu - \omega'_m(\boldsymbol{k}))$$
$$- 2\sum_{\boldsymbol{k}}\sum_{n}(\zeta_n(\boldsymbol{k}) - \mu)\Theta(\mu - \zeta_n(\boldsymbol{k})) . \qquad (12)$$

Note that the reference system is always assumed to be in the same macroscopic state as the original system, i.e. it is considered at the same temperature and, more importantly here, at the same chemical potential μ. Furthermore, it has been used that, by construction of the approximation, the self-energy and hence its poles at $\zeta_n(\boldsymbol{k}) - \mu$ are the same for both, the original and the reference system. This implies that the second terms on the r.h.s. of Eq. (11) and (12), respectively, cancel each other in Eq. (10). Finally, a (large but) finite system ($L < \infty$) and a finite reference system are considered. Hence, the set of poles of the Green's function and of the self-energy as well as sums over $\boldsymbol{k}$ are discrete and finite.

Taking the μ derivative on both sides of Eq. (10) then yields:

$$\frac{\partial\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]}{\partial\mu} = \frac{\partial\Omega_{\boldsymbol{t}',\boldsymbol{U}}}{\partial\mu} - 2\sum_{\boldsymbol{k}}\sum_{m}\Theta(\mu - \omega_m(\boldsymbol{k}))$$
$$+ 2\sum_{\boldsymbol{k}}\sum_{m}\Theta(\mu - \omega'_m(\boldsymbol{k})) . \qquad (13)$$

Here we have assumed the ground state of the reference system to be non-degenerate with respect to the particle number. From the (zero-temperature) Lehmann representation[39] it is then obvious that, within a subspace of fixed particle number, the μ-dependence of the Green's function is the same as its ω-dependence, i.e. $\boldsymbol{G}(\omega) = \widetilde{\boldsymbol{G}}(\omega + \mu)$ with a μ-independent function $\widetilde{\boldsymbol{G}}$. Via the Dyson equation of the reference system, this property can also be inferred for the self-energy and, via the Dyson equation of the original system, for the (approximate) Green's function of the original system. Consequently, the poles of $(\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}})^{-1}$ and of $\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$ are linearly dependent on μ, i.e. $\omega_m(\boldsymbol{k})$ and $\omega'_m(\boldsymbol{k})$ in Eqs. (11) and (12) are independent of μ.

We once more exploit the fact that the self-energy of the original system is identified with the self-energy of the reference system. Using Eq. (4) one immediately arrives at

$$\langle N\rangle = \langle N\rangle' + 2\sum_{\boldsymbol{k}}\Theta(G_{\boldsymbol{k}}(0)) - 2\sum_{\boldsymbol{k}}\Theta(G'_{\boldsymbol{k}}(0)) . \quad (14)$$

This is the final result: The Luttinger sum rule for the original system, Eq. (2), is satisfied if and only if it is satisfied for the reference system, i.e. if $\langle N\rangle' = 2\sum_{\boldsymbol{k}}\Theta(G'_{\boldsymbol{k}}(0))$.

A few remarks are in order. For the reference system, the status of the Luttinger sum rule is that of a general theorem (as long as the general proof is valid); $\langle N\rangle'$ and $G'_{\boldsymbol{k}}(0)$ represent exact quantities. The above derivation shows that the theorem is "propagated" to the original system *irrespective of the approximation that is constructed within the SFT*. This propagation also works in the opposite direction. Namely, a possible violation of the exact sum rule for the reference system would imply a violation of the sum rule, expressed in terms of approximate quantities, for the original system.

Eq. (14) holds for any choice of $\boldsymbol{t}'$. Note, however, that stationarity with respect to the variational parameters $\boldsymbol{t}'$ is essential for the thermodynamical consistency of the approximation. In particular, consistency means that the average particle number $\langle N\rangle = -\partial\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]/\partial\mu$ on the l.h.s. can be obtained as the trace of the Green's function. Stationarity is thus necessary to get the sum rule in the form (5).

There are no problems to take the thermodynamic limit (if desired) on both sides of Eq. (14) (after division of both sides by the number of sites L). The $\boldsymbol{k}$ sums turn into integrals over the unit cell of the reciprocal lattice. For a D-dimensional lattice the $D - 1$-dimensional manifolds of $\boldsymbol{k}$ points with $G_{\boldsymbol{k}}(0) = \infty$ or $G_{\boldsymbol{k}}(0) = 0$ form Fermi or Luttinger surfaces, respectively.

For the above derivation, translational symmetry has been assumed for both, the original as well as the reference system. Nothing, however, prevents us from repeating the derivation in case of systems with reduced (or completely absent) translational symmetries. One simply has to re-interpret the wave vector $\boldsymbol{k}$ as an index which, combined with m, refers to the elements of the diagonalized Green's function matrix $\boldsymbol{G}$. The exact sum rule, Eq. (5), generalizes accordingly. The result (14) remains valid (with the correct interpretation of $\boldsymbol{k}$) for an original system with reduced translational symmetries. It is also valid for the case of a translationally symmetric original Hamiltonian where, due to the choice of a reference system with reduced translational symmetries, the symmetries of the (approximate) Green's function of the original system are (artificially) reduced. A typical example is the variational cluster approximation (VCA) where the reference system consists of isolated clusters of finite size.

IV. TWO-SITE DYNAMICAL-IMPURITY APPROXIMATION

While the Hartree-Fock approximation may be considered as the most simple weak-coupling Φ-derivable approximation, the most simple non-perturbative Φ-derivable approximation is given by the dynamical-impurity approximation (DIA). This shall be demonstrated in the following for the single-band Hubbard model (1) as the original system to be investigated. The DIA is generated by a reference system consisting of a decoupled set of single-impurity Anderson models with a finite number of sites n_s and is known[8] to recover the dynamical mean-field theory in the limit $n_s \to \infty$. As long as the Luttinger sum rule holds for the single-impurity reference system, the DIA must yield a one-particle Green's function and a self-energy respecting the sum rule.

The Hamiltonian of the reference system is $H(\boldsymbol{t'}, \boldsymbol{U}) = \sum_{i=1}^{L} H_i'$ with

$$H_i' = \sum_\sigma \varepsilon_0 c_{i\sigma}^\dagger c_{i\sigma} + \frac{U}{2} \sum_\sigma n_{i\sigma} n_{i-\sigma}$$
$$+ \sum_{k=2}^{n_s} \sum_\sigma \varepsilon_k a_{ik\sigma}^\dagger a_{ik\sigma} + \sum_{k=2}^{n_s} \sum_\sigma V_k (a_{ik\sigma}^\dagger c_{i\sigma} + \text{h.c.}) \,. \tag{15}$$

For a homogeneous phase, the variational parameters $\boldsymbol{t'} = (\{\varepsilon_0^{(i)}, \varepsilon_k^{(i)}, V_k^{(i)}\})$ can be assumed to be independent of the site index i: $\varepsilon_0 \equiv \varepsilon_0^{(i)}$, $\varepsilon_k \equiv \varepsilon_k^{(i)}$, $V_k \equiv V_k^{(i)}$. For the sake of simplicity, we consider the two-site DIA ($n_s = 2$), i.e. a single bath site per correlated site only. In this case there are three independent variational parameters only: the on-site energies of the correlated and of the bath site, ε_0 and $\varepsilon_c \equiv \varepsilon_{k=2}$, respectively, as well as the hybridization strength $V \equiv V_{k=2}$. As the reference
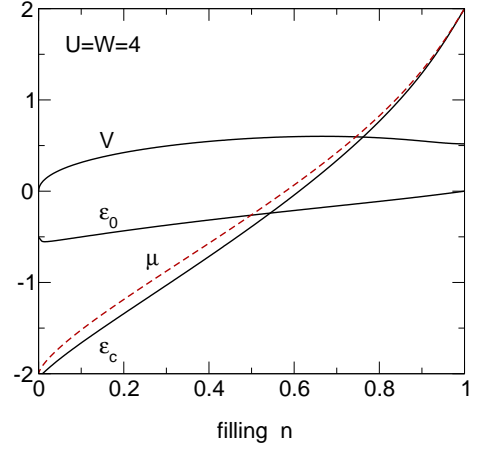

FIG. 2: Filling dependence of the variational parameters at their respective optimized values and of the chemical potential. Calculations for the Hubbard model with a semi-elliptical free density of states of band width $W = 4$ and interaction strength $U = W = 4$ using the two-site DIA.

system consists of replicated identical impurity models which are spatially decoupled, the trial self-energy is local and site-independent, $\Sigma_{ij}(\omega) = \delta_{ij}\Sigma(\omega)$.

Calculations have been performed for the Hubbard model with a one-particle dispersion $\varepsilon(\boldsymbol{k}) = L^{-1} \sum_{ij} e^{-i\boldsymbol{k}(\boldsymbol{R}_i - \boldsymbol{R}_j)} t_{ij}$ such that the density of one-particle energies $D(\varepsilon)$ is semi-elliptic. For $|\varepsilon| \le W/2$,

$$D(\varepsilon) = \frac{1}{L} \sum_{\boldsymbol{k}} \delta(\varepsilon - \varepsilon(\boldsymbol{k})) = \frac{8}{\pi W^2} \sqrt{(W/2)^2 - \varepsilon^2} \,. \tag{16}$$

The free band width is set to $W = 4$. This serves as the energy scale.

The computation of the SFT grand potential is performed as described in Ref. 9. Stationary points of the resulting function $\Omega(\varepsilon_0, \varepsilon_c, V) \equiv \Omega_{\boldsymbol{t}, \boldsymbol{U}}[\boldsymbol{\Sigma}_{\varepsilon_0, \varepsilon_c, V}]$ are obtained via iterated linearizations of its gradient. There is a unique non-trivial stationary point (with $V \ne 0$). Fig. 2 shows the variational parameters at this point as functions of the filling n. For the entire range of fillings, the ground state of the reference system lies in the invariant subspace with $N_{\text{tot}} = \sum_\sigma (c_{i\sigma}^\dagger c_{i\sigma} + a_{i\sigma}^\dagger a_{i\sigma}) = 2$. The parameters as well as the chemical potential are smooth functions of n. We have checked that the thermodynamic consistency condition $n = -L^{-1} \partial\Omega/\partial\mu = \int_{-\infty}^0 \rho(\omega) d\omega$ is satisfied within numerical accuracy. Here

$$\rho(\omega) = D(\omega + \mu - \Sigma(\omega)) \tag{17}$$

is the interacting local density of states (DOS).

At half-filling the values of the optimized on-site energies are consistent with particle-hole symmetry. With $\varepsilon_0 - \mu = -U/2$ and $\varepsilon_c - \mu = 0$ the reference system is in the Kondo regime with a well-formed local moment at the correlated site. The finite hybridization strength V leads, for $U = W$, to a finite DOS $\rho(\omega = 0) > 0$ and thus
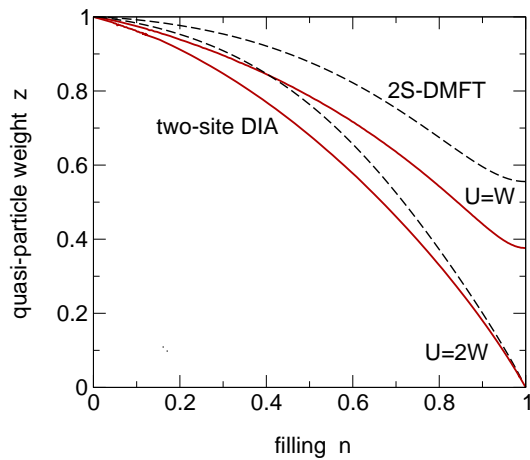
FIG. 3: Quasi-particle weight z as a function of the filling within the two-site DIA (full lines) and the two-site DMFT[40] (dashed lines). Calculations for $U = W$ and $U = 2W$.

to a metallic Fermi liquid as it is expected for the Hubbard model within a (dynamical) mean-field description. Due to the simple structure of the self-energy generated by the two-site reference system, however, quasi-particle damping effects are missing.

Decreasing the filling from $n = 1$ to $n = 0$ drives the reference system more and more out of the Kondo regime. While ε_c stays close to the chemical potential, the on-site energy of the correlated site ε_0 crosses μ close to quarter filling and lies above μ eventually. Note that $\varepsilon_0 = 0$ within the DMFT, i.e. for $n_s \to \infty$, while for finite n_s there is a clear deviation from $\varepsilon_0 = 0$ which is necessary to ensure thermodynamical consistency. For fillings very close to $n = 0$, the grand potential $\Omega_{t,U}[\mathbf{\Sigma}_{\varepsilon_0, \varepsilon_c, V}]$ becomes almost independent of $\mathbf{\Sigma}$. This implies that it becomes increasingly difficult to locate the stationary point with the numerical algorithm used. The slight upturn of ε_0 below $n = 0.01$ (see Fig. 2) might be a numerical artifact.

It is instructive to compare the parameters with those of the two-site DMFT (2S-DMFT).[40] The 2S-DMFT is a simplified version of the DMFT where a mapping onto the two-site single impurity Anderson model is achieved by means of a simplified self-consistency equation. Assuming $\varepsilon_0 = 0$ as in the full DMFT, there are two parameters left (ε_c and V) which are fixed by considering the first non-trivial order in the low- and in the high-frequency expansion of the self-energy and the Green's function in the DMFT self-consistency equation. Although being well motivated, this approximation is essentially *ad hoc*. One therefore has to expect that the 2S-DMFT is thermodynamically inconsistent and exhibits a violation of Luttinger's sum rule. A comparison of the DIA for $n_s = 2$ with the 2S-DMFT is thus ideally suited to demonstrate the advantages gained by constructing approximations within the variational framework of the SFT.

First of all, there are differences in fact. At half-filling

the 2S-DMFT predicts the hybridization to be somewhat larger than the two-site DIA while the value for ε_c is again fixed by particle-hole symmetry. Deviations grow with decreasing filling. Contrary to the two-site DIA, V monotonously increases and is larger in the entire filling range, $\varepsilon_0 = 0$ by construction, and ε_c even diverges for $n \to 0$ within the 2S-DMFT (see Ref. 40). On the other hand, the system is essentially uncorrelated in the limit $n \to 0$. Strong differences in the parameters, which enter the self-energy only, therefore do not necessarily imply strongly different physical quantities. This is demonstrated by Fig. 3 which shows the quasi-particle weight calculated via

$$z = \left(1 - \frac{d\Sigma(\omega = 0)}{d\omega}\right)^{-1} \qquad (18)$$

as a function of the filling. While there are obvious differences when comparing the results from the two-site DIA with those of the 2S-DMFT, the qualitative trend of z is very similar in both approximations. Both approximations also compare well with the full DMFT: There is a quadratic behavior of $z(n)$ for $n \to 1$ in the Fermi-liquid phase ($U = W$) and a linear trend when approaching the Mott phase ($U = 2W$). The critical interaction strength for the Mott transition is found to be $U_c \approx 1.46W$ for the two-site DIA and $U_c = 1.5W$ within the 2S-DMFT. For details on the Mott transition see Refs. 9,40.

In case of a local and site-independent self-energy, the Luttinger sum rule can be written in the form[41]

$$\mu = \mu_0 + \Sigma(\omega = 0), \qquad (19)$$

where μ_0 is the chemical potential of the free ($U = 0$) system at the same particle density. Eq. (19) implies that not only the enclosed volume but also the shape of the Fermi surface remains unchanged when switching on the interaction. Using Eq. (17) this immediately implies[41]

$$\rho(0) = D(\mu_0) = \rho_0(0), \qquad (20)$$

i.e., in case of a correlated metal, the value of the interacting local density of states at $\omega = 0$ is independent of U and thus fixed to the value of the density of states of the non-interacting system at the same filling.

The interacting and the non-interacting DOS are plotted in Fig. 4 for different fillings and for $U = W$ and $U = 2W$. The impurity self-energy of the two-site reference system is an analytical function of ω except for two first-order poles on the real axis. Via Eq. (17) this two-pole structure implies that the DOS consists of three peaks the form of which is essentially given by the non-interacting DOS. At half-filling the three peaks are easily identified as the lower and the upper Hubbard band and the quasi-particle resonance as it is characteristic for a (dynamical) mean-field description.[27] For $U = W$ the resonance still has a significant weight. The weight decreases upon approaching the critical interaction, and the resonance has disappeared in the Mott insulator for
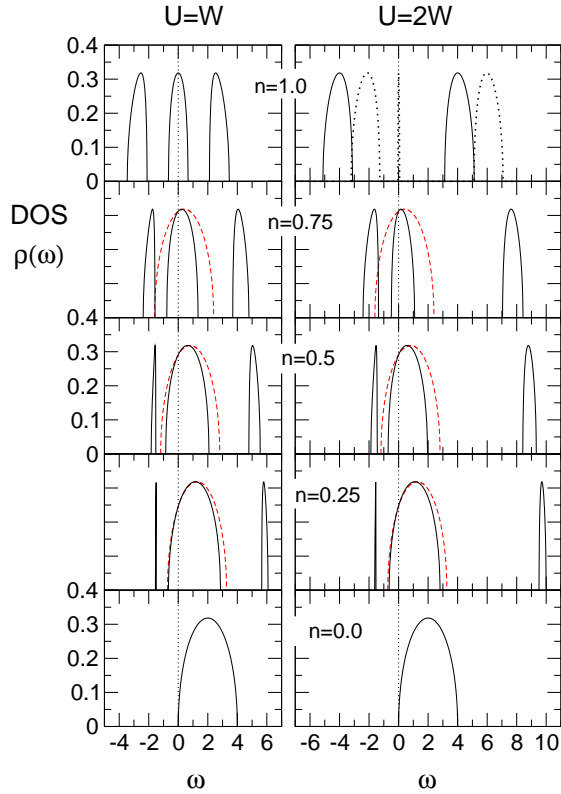
FIG. 4: Interacting local density of states $\rho(\omega)$ (solid lines) for different fillings as indicated. Calculations using the two-site DIA for $U = W$ (left) and $U = 2W$ (right). For $n = 0.25$, $n = 0.5$ and $n = 0.75$ the non-interacting DOS $\rho_0(\omega)$ is shown for comparison (dashed lines). Note that $\rho(0) = \rho_0(0)$. The dotted line for $U = 2W$ in the top panel is the DOS for $n = 0.99$.

$U = 2W$. Hole doping of the Mott insulator is accomplished by the reappearance of the resonance at $\omega = 0$ which preempts the creation of holes in the lower Hubbard band.[42] As can be seen in the spectrum for $n = 0.99$ in the top panel (dotted line), the quasi-particle resonance appears within the Mott-Hubbard gap. With decreasing filling, the upper Hubbard band gradually shifts to higher excitation energies and loses weight. This weight is transferred to the low-energy part of the spectrum. For lower fillings where the Kondo regime has been left, one would actually expect that the quasi-particle resonance disappears by merging with the lower Hubbard band. This, however, cannot be described with the simple two-pole structure of the self-energy. One therefore should interpret the gap around $\omega = -1$ at $n = 0.25$ as an artifact of the approximation. Furthermore, the widths of the Hubbard bands is considerably underestimated as damping effects are missing completely. The filling-dependent spectral-weight transfer across the Hubbard gap as well as the energy positions of the main peaks, however, are in overall agreement with general expectations.[34,43]
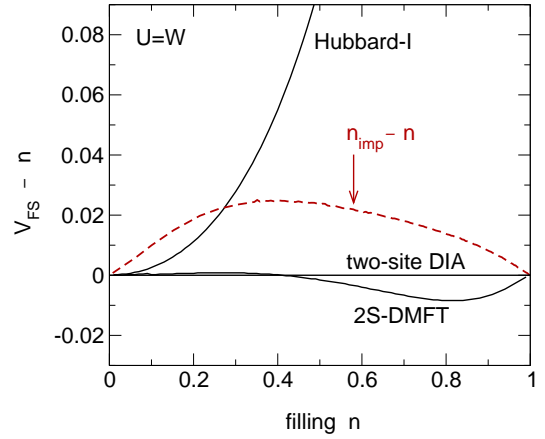
It is worth emphasizing that this simple two-site

FIG. 5: Numerical results for the difference between the volume enclosed by the Fermi surface V_{FS} and the filling n as a function of n for $U = W = 4$. The Luttinger sum rule ($V_{\mathrm{FS}} - n = 0$) is exactly respected by the two-site DIA. Results for the 2S-DMFT and the Hubbard-I approximation are shown for comparison. Dashed line: difference between the filling n and the average occupation of the correlated (impurity) site in the reference system at stationarity for the two-site DIA.

dynamical-impurity approximation exactly fulfills the Luttinger sum rule. In Fig. 4 this can be seen by comparing with the DOS of the non-interacting system (dashed lines). The non-interacting DOS cuts the interacting one at $\omega = 0$ which shows that Eq. (20) is satisfied. Note that this is trivial for $n = 1$ as this is already enforced by particle-hole symmetry. Off half-filling, however, the pinning of the DOS to its non-interacting value at $\omega = 0$ is a consequence of Φ-derivability and thereby a highly non-trivial feature.

In contrast, the 2S-DMFT does show a violation of Luttinger's sum rule which, however, must be attributed to the *ad hoc* nature of the approximation. Fig. 5 shows
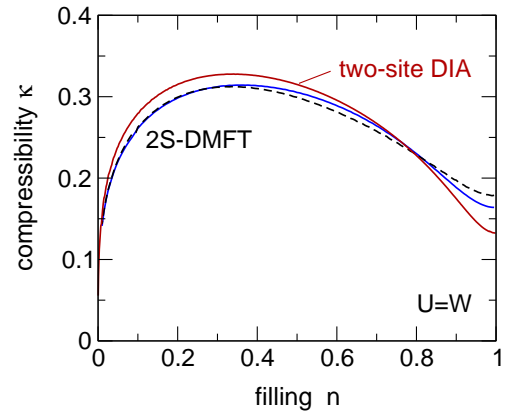

FIG. 6: Filling dependence of the compressibility κ for $U = W$ as obtained within the 2S-DMFT via $\kappa = \partial n / \partial \mu$ (solid line) and via a general Fermi-liquid relation (Eq. (22), dashed line). Using the two-site DIA identical results are obtained for both cases.

the difference between the volume enclosed by the Fermi surface

$$V_{\text{FS}} = \frac{2}{L} \sum_{\boldsymbol{k}} \Theta(\mu - \varepsilon(\boldsymbol{k}) - \Sigma(0)) = 2 \int_{-\infty}^{0} d\varepsilon D(\varepsilon + \mu - \Sigma(0)) \tag{21}$$

and the filling n as a function of the filling. As can be seen, there is an artificial violation of the sum rule for the 2S-DMFT which is of the order of a few per cent while for the Φ-derivable two-site DIA the sum rule is fully respected. Note that, unlike the DMFT and also unlike the simplified 2S-DMFT, the two-site DIA predicts a filling which slightly differs from the average occupation of the correlated impurity site in the reference system (see dashed line in Fig. 5). For a finite number of bath sites n_{s} this appears to be necessary to fulfill the Luttinger sum rule. The figure also shows the result obtained within the Hubbard-I approximation.[34] Here a very strong (artificial) violation of up to 100 % (for n close to half-filling) is obtained. This should be considered as a strong drawback which is typical for uncontrolled mean-field approximations.

There are more relations which, analogously to the Luttinger sum rule, can be derived by means of perturbation theory to all orders[6] in the exact theory and which are respected by weak-coupling conserving approximations. For example, the compressibility, defined as $\kappa = \partial n / \partial \mu$, can be shown to be related to the interacting DOS and the self-energy at the Fermi edge via

$$\kappa = 2\rho(0) \left(1 - \frac{\partial \Sigma(0)}{\partial \mu} \right). \tag{22}$$

Fig. 6 shows that for the 2S-DMFT it makes a difference whether κ is calculated as the μ-derivative of the filling or via Eq. (22). Again, this must be attributed to the fact that the 2S-DMFT is not a Φ-derivable approximation. Contrary, the two-site DIA does respect the general Fermi-liquid property (22) and thus yields the same result in both cases (see Fig. 6).

V. VIOLATION OF LUTTINGER'S SUM RULE IN FINITE SYSTEMS

The preceding section has demonstrated that the two-site DIA satisfies the Luttinger sum rule. According to Eq. (14), we can conclude that the Luttinger sum rule must hold for the corresponding reference system, i.e. for the two-site single-impurity Anderson model. Of course, this can be verified more directly by evaluating Eq. (5). In case of a finite system or a system with reduced translational symmetries, the Green's function is a matrix with elements $G_{\alpha\beta}(\omega)$ where α refers to the one-particle basis states, and the Luttinger sum rule reads:

$$\sum_{k,m} \alpha_m^{(k)} \Theta(\mu - \omega_m^{(k)}) = \sum_{k,m} \Theta(\mu - \omega_m^{(k)}) - \sum_{k,n} \Theta(\mu - \zeta_n^{(k)}). \tag{23}$$

Here the index k labels the elements of the diagonalized Green's function, i.e. Eq. (5) is generalized by replacing $(\boldsymbol{k}, \sigma) \to k$. In case of an impurity model, Eq. (23) actually represents the Friedel sum rule.[44,45] For the two-site single-impurity Anderson model, the different one-particle excitation energies $\omega_m^{(k)} - \mu$, the zeros of the Green's function $\zeta_n^{(k)} - \mu$ and the weights $\alpha_m^{(k)}$ are easily determined by full diagonalization. We find that Eq. (23) is satisfied in the entire parameter space (except for $V = 0$, see below).

Note that a violation of the sum rule occurs when, as a function of a model parameter x, a zero of the Green's function crosses $\omega = 0$ for $x = x_c$. At x_c the number of negative zeros counted by the second term on the r.h.s. changes by one while the first term as well as the l.h.s. remain constant since (unlike a pole) a zero of the Green's function is generically not connected with a change of the ground state (level crossing). This implies that the sum rule would be violated for $x < x_c$ or for $x > x_c$.

The case $V = 0$ is exceptional. Within the two-site DIA this corresponds to the Mott insulator (see Fig. 4, topmost panel for $U = 2W$). For $V = 0$ the reference system consists of two decoupled sites, and the Green's function becomes diagonal in the site index. There is no zero of the local Green's function corresponding to the uncorrelated site. We can thus concentrate on the correlated site where the local Green's function exhibits a zero at $\eta - \mu = \varepsilon_0 + U/2$. In the sector with one electron at the correlated site ($\varepsilon_0 < \mu < \varepsilon_0 + U$), the second term on the r.h.s. changes by *two* at $\mu = \mu_c = \varepsilon_0 + U/2$ because of the two-fold degenerate ground state. In this case Luttinger's sum rule in the form (23) is violated for $\mu < \mu_c$ *and* for $\mu > \mu_c$. This "violation", however, is a trivial one which immediately disappears if the ground-state degeneracy is lifted by applying a weak field term, for example.

Fig. 7 shows a phase diagram of the single-impurity Anderson model with $n_{\text{s}} = 4$ sites as obtained by full diagonalization. The diagram covers the entire range of the total particle number $N = \sum_{\sigma} \langle c_{\sigma}^{\dagger} c_{\sigma} \rangle + \sum_{\sigma} \sum_{k=2}^{n_{\text{s}}} \langle a_{k\sigma}^{\dagger} a_{k\sigma} \rangle$ from $N = 0$ to $N = 2n_{\text{s}} = 8$. A non-degenerate ground state is enforced by applying a small but finite magnetic field. No violation of the Luttinger sum rule is found. We have repeated the same calculation also for $n_{\text{s}} = 10$ using the Lanczos technique.[46] Again, the sum rule is found to be always satisfied (We have performed calculations for different U and bath parameters). This might have been expected as the ($n_{\text{s}} \to \infty$) Anderson model can generally be classified as a (local) Fermi liquid.[47]

The situation is less clear in the case of correlated lattice models such as the Hubbard or the t-J model. For two dimensions there are several numerical studies using high-temperature expansion,[30] quantum Monte-Carlo,[31] extended DMFT,[32,48] and dynamical cluster approximation (DCA)[33] which indicate a violation in the strongly correlated metallic phase close to half-filling. For studies of large clusters or studies directly working in the thermo-
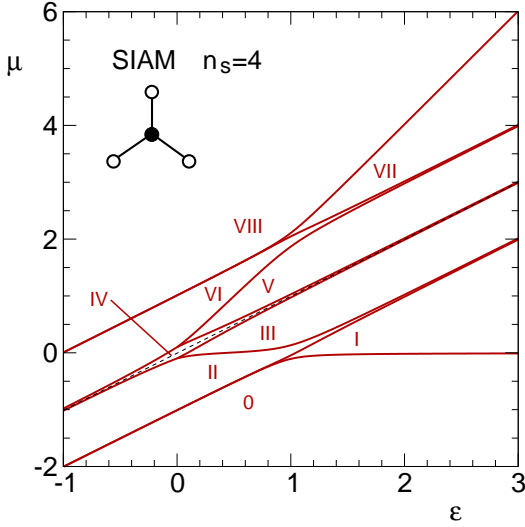
FIG. 7: Phase diagram μ vs. ε of the single-impurity Anderson model with $n_{\mathrm{s}} = 4$ sites. Total particle numbers are indicated by Roman figures. Results have been obtained by full diagonalization for the following model parameters. One-particle energies: $\varepsilon_0 = 0$ (correlated site), $\varepsilon_k = \varepsilon + (k - 3)$ with $k = 2, 3, 4$ (uncorrelated bath sites). Hubbard interaction: $U = 2\varepsilon$. Hybridization strength: $V_k = 0.1$ for $k = 2, 3, 4$. To lift Kramers degeneracy in case of an odd particle number, a weak (ferromagnetic) field of strength $b = 0.001$ is coupled to the local spins. The dashed line marks the particle-hole symmetric case. Luttinger's sum rule is found to be satisfied in the entire parameter space.

dynamic limit, a definite conclusion on the validity of the sum rule is difficult to obtain as finite-temperature or artificial broadening effects etc. must be controlled numerically. Contrary, full diagonalization of Hubbard clusters consisting of a few sites only can provide exact results. While their direct relevance for the thermodynamic limit is less clear, it is important to note that reference systems with a finite number of sites or a finite number of correlated sites provide the basis for a number of cluster approaches within the SFT framework. Via Eq. (14) their properties are transferred to the approximate treatment of lattice models in the thermodynamic limit.

The validity of Eq. (23) has been checked for Hubbard clusters of different size and in different geometries. The μ vs. U phase diagram for an $L = 4$-site open Hubbard chain with nearest-neighbor hopping in Fig. 8 shows a representative example. Again, a small but finite field term is added to avoid a ground-state degeneracy. As the chemical potential, for fixed U, is moved off the particle-hole symmetric point $\mu = U/2$ and exceeds certain critical values (red lines), the particle number N [as obtained from the l.h.s. of Eq. (23)] changes from $N = L$ down to (up to) $N = 0$ ($N = 2L$). A critical μ value indicates a change of the ground state (level crossing) that is accompanied by a change of the ground-state particle number. In the one-particle Green's function this is characterized by a pole $\omega_m^{(k)} - \mu$ crossing $\omega = 0$. The blue lines indicate
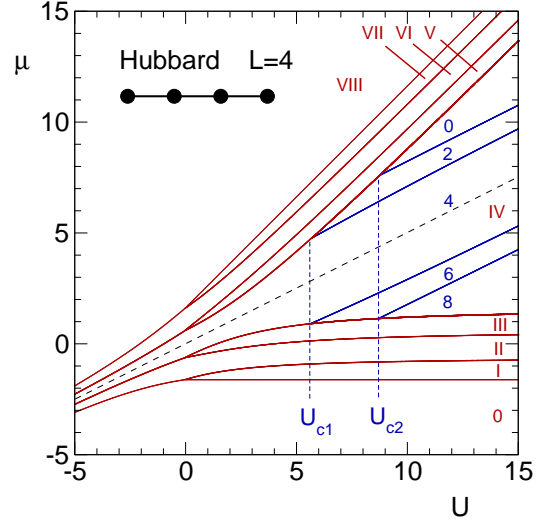

FIG. 8: Phase diagram μ vs. U of the Hubbard model with $L = 4$ sites (open chain) as obtained by full diagonalization. Nearest-neighbor hopping $t = -1$. A weak (ferromagnetic) field of strength $b = 0.01$ is applied to lift Kramers degeneracy. The dashed line marks the particle-hole symmetric case. Particle numbers (l.h.s. of Eq. (23)) are indicated by Roman figures. R.h.s. of Eq. (23): Arabic figures. Luttinger's sum rule is found to be violated for sufficiently strong U. $U_{c1}, U_{c,2}$: critical interactions.

those chemical potentials at which a zero of the Green's function $\zeta_n^{(k)} - \mu$ crosses $\omega = 0$. Whenever this happens the r.h.s. of Eq. (23) changes while the l.h.s. is constant. Fig. 8 shows that this occurs several times in the $N = L$ sector. At the particle-hole symmetric point $\mu = U/2$ the Luttinger sum rule is obeyed while it is violated in a wide region of the parameter space corresponding to half-filling $N = L$. However, a critical interaction strength U_c turns out to be necessary. The value for U_c strongly varies for different cluster sizes and geometries but has always been found to be positive and finite. Note that for $L = 4$ the sum rule is fulfilled for any particle number $N \neq L$. Qualitatively similar results can be found for the $L = 2$-site Hubbard cluster where calculations can be done even analytically. Again, a violation of the sum rule is found in the half-filled sector beyond a certain critical U.

This has already been noticed by Rosch[49] and was used in combination with a strong-coupling expansion to argue that a violation of the sum rule generically occurs for a Mott insulator. Stanescu et al.[50] have shown quite generally that the sum rule is fulfilled when particle-hole symmetry is present (the Luttinger surface is the same as the Fermi surface of the non-interacting system) but violated in the Mott insulator away from particle-hole symmetry. It is interesting to note that these arguments cannot be used to construct a violation of the sum rule within DMFT or for a single-impurity Anderson model: For an (almost) particle-hole symmetric case and model parameters describing a Mott insulator (within DMFT),
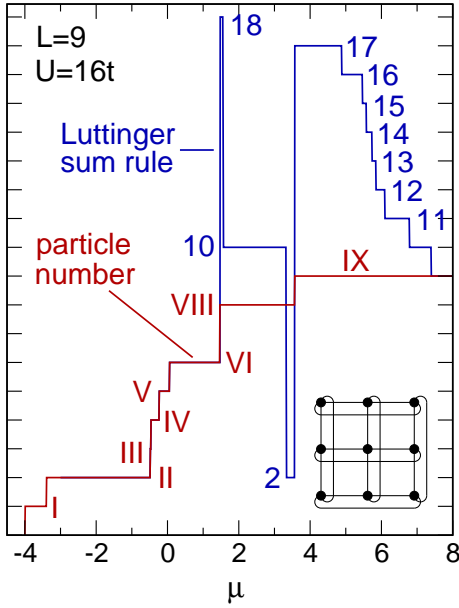
FIG. 9: Ground-state particle number (red Roman figures, l.h.s. of Eq. (23)) and prediction by the Luttinger sum rule (blue Arabic figures, r.h.s. of Eq. (23)) as functions of the chemical potential for a $L = 9$-site Hubbard cluster with periodic boundary conditions. Arabic numbers are only given when different from Roman ones. Calculations using the Lanczos method and a finite but small magnetic field and finite but small on-site potentials to lift ground-state degeneracies.

FIG. 10: The same as Fig. 9 but for 10 sites.

an odd number of sites n_s (with $n_s \to \infty$) must be considered and thus a magnetic field is needed to lift Kramers degeneracy. Even an infinitesimal field, however, leads (at zero temperature) to a finite and even large polarization corresponding a well-formed but unscreened local moment. This polarization is incomplete for any finite U as the DMFT predicts a small but finite double occupancy for a Mott insulator. Still there is a proximity to the fully polarized band insulator which finally results in a weakly correlated state and thus in a situation which is unlikely to show a violation of the sum rule.

We have also considered Hubbard clusters with $L = 9$ and $L = 10$ sites by using the Lanczos technique.[46] Calculations have been performed for different Lanczos depths l_{max} to ensure that the results are independent of l_{max}. Fig. 9 displays an example for $L = 9$ and a highly symmetric cluster geometry with periodic boundary conditions and a well-defined reciprocal space. To lift ground-state degeneracies resulting from spatial symmetries as well as the Kramers degeneracy, small but finite on-site potentials and a small magnetic-field term are included in the cluster Hamiltonian. Fig. 10 shows an example for $L = 10$ sites without any spatial symmetries. Kramers degeneracy for odd N is removed by applying a small magnetic field. With the figures we compare the expressions on the left-hand and the right-hand side of Eq. (23). Obviously, the sum rule is respected in
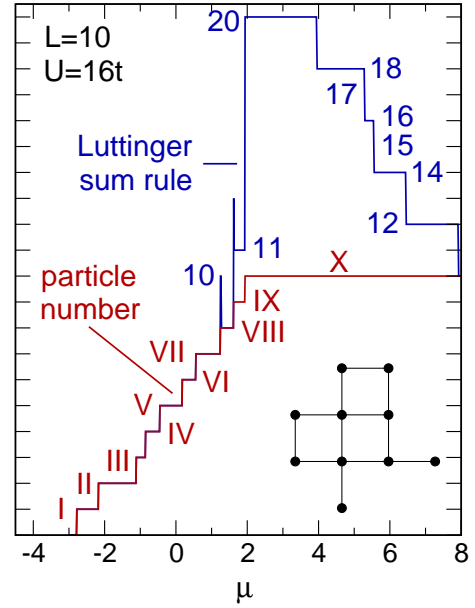
most cases. Violations are seen for half-filling $N = L$, i.e. in the "Mott-insulating phase", which is consistent with Ref. 49. However, the sum rule is also violated in the "metallic phase" close to half-filling, namely for $N = L - 1$ (Fig. 9, $L = 9$) and $N = L - 1, L - 2$ (Fig. 10, $L = 10$). This nicely corresponds to the generally observed trend[30,31,32,33,48] for violations in the slightly doped metallic regime. We have also verified that the sum rule is restored by lowering U.

Fig. 9 and 10 demonstrate that the sum rule is violated in the *whole* μ range corresponding to $N = L - 1$. This is an important point as it shows that it is irrelevant whether the $T = 0$ limit is approached by holding $\langle N \rangle$ fixed and adjusting $\mu = \mu(T)$ or by fixing μ and let $\langle N \rangle = \langle N \rangle(T)$ be T-dependent. A violation of the sum rule is found in both cases.

Kokalj and Prelovšek[51] have demonstrated that violations of the sum rule can also be found for the t-J model on a finite number of sites. Our result provides an explicit example showing that not only for t-J[51] but also for Hubbard clusters a violation can be found when the chemical potential is set to $\mu = \lim_{T \to 0} \mu(T)$ with $\mu(T)$ obtained for given $\langle N \rangle = $ const. Anyway, the original proof[5] does not depend on this choice for μ but appears to work for any μ.

The results raise the question which assumptions used in the original proof of the theorem are violated or where the proof breaks down. Note that the recently proposed alternative topological proof[52] assumes a Fermi-liquid state from the very beginning and thus cannot be applied to a finite system. Using weak symmetry-breaking fields, a more or less trivial breakdown due to ground-state degeneracy has been excluded. An analysis of the ground state of the $L = 2$ and $L = 4$ Hubbard clusters which

are accessible with exact (analytical or numerical) methods has shown that, for model parameters where the sum rule is violated, the interacting ground state can nevertheless be adiabatically connected to the non-interacting one. This excludes level crossing as a potential cause for the breakdown. While we cannot make a definite statement, it appears at least plausible that the violation of the sum rule results from a non-commutativity of two limiting processes, the infinite skeleton-diagram expansion and the limit $T \to 0$.

Using a functional-integral formalism, the Luttinger-Ward functional at finite T can also be constructed in a *non-perturbative* way, i.e. avoiding an infinite summation of diagrams, as has been shown recently.[53] Formally, the Luttinger sum rule can be obtained by exploiting a gauge invariance of the Luttinger-Ward functional [see Ref. 53]:

$$\frac{\partial}{\partial(i\omega_n)} \Phi_U[\boldsymbol{G}(i\omega_n)] = 0 \,. \tag{24}$$

If at all, this invariance can only be shown for $T = 0$ where $i\omega_n$ becomes a continuous variable. Unfortunately, the non-perturbative construction of Φ_U requires a $T > 0$ formalism. Hence, the validity of the sum rule depends on question whether the limit $T \to 0$ commutes with the frequency differentiation. Necessary and sufficient conditions for this assumption are not easily worked out. An understanding of the main reason for the possible breakdown of the sum rule in finite systems, very similar to the case of Mott insulators, is therefore not yet available (see also the discussion in Ref. 49).

VI. CONCLUSIONS

Φ-derivable approximations are conserving, thermodynamically consistent and, for $T = 0$, formally respect certain non-trivial theorems such as the Luttinger sum rule. As the construction of the Luttinger-Ward functional Φ is by no means trivial and may conflict with the limit $T \to 0$ or different other limiting processes, however, the validity of the sum rule may be questioned. Violations of the sum rule can be found in fact for the case of strongly correlated electron systems. For Mott insulators and finite systems in particular, a breakdown is documented easily.

This implies that a general approximation for the spectrum of one-particle excitations (of the one-particle Green's function) may violate the sum rule for *two* possible reasons, namely because (i) the sum rule is violated in the exact theory, or (ii) the approximation generates an artificial violation.

Within the usual weak-coupling conserving approximations, such as the fluctuation-exchange approximation, the sum rule always holds as the formal steps in the general proof of the sum rule can be carried over to the approximation – but with the important simplification of a limited class of diagrams. This also implies that weak-coupling conserving approximations, when applied beyond the weak-coupling regime, might erroneously predict the sum rule to hold.

The present paper has focussed on *non-perturbative* conserving approximations. Non-perturbative approximations, constructed within the framework of the self-energy-functional theory and referring to a certain reference system, are Φ-derivable and consequently respect certain macroscopic conservation laws and are thermodynamically consistent. Whether or not the sum rule holds within the approximate approach, however, cannot be answered generally. We found that Luttinger's sum rule holds within an (SFT) approximation if and only if it holds exactly in the corresponding reference system.

The reference system that leads to the most simple but non-trivial example for a non-perturbative conserving approximation consists of a single correlated and a single bath site. For this two-site system, we have found the sum rule to be valid in the entire parameter space. Consequently, the resulting two-site dynamical-impurity approximation (DIA) – opposed to more *ad hoc* approaches like the two-site DMFT – fully respects the sum rule as could be demonstrated in different ways. In view of the simplicity of the approximation this is a remarkable result. Since the sum rule dictates the low-frequency behavior of the one-particle Green's function, important mean-field concepts, such as the emergence of a quasi-particle resonance at the Fermi edge, are qualitatively captured correctly, even away from the particle-hole symmetric case. This qualifies the two-site DIA for a quick but rough estimate of mean-field physics, including phases with spontaneously broken symmetries.

Full diagonalization and the Lanczos method have been employed to show that also the single-impurity Anderson model with a finite number of $n_\mathrm{s} > 2$ sites respects the sum rule. Consequently, this property is transferred to an n_s-site DIA. For $n_\mathrm{s} \to \infty$ the full dynamical mean-field theory is recovered which is thereby recognized as the prototypical non-perturbative conserving approximation. Clearly, in the case of the DMFT, Φ-derivability is well known[27] and obvious, for example, when constructing the DMFT with the help of the skeleton-diagram expansion.

Using as a trial self-energy the self-energy of a cluster with $L > 1$ correlated sites, generates an approximation where short-range spatial correlations are included up to the cluster extension. These variational cluster approximations provide a first step beyond the mean-field concept. Again, whether or not the sum rule is respected within the VCA depends on the reference system itself. For the $L = 2$ Hubbard cluster, analytical calculations straightforwardly show that violations of the sum rule occur at half-filling, beyond a certain critical interaction strength. In the thermodynamic limit, this would correspond to the Mott-insulating regime. Applying the Lanczos method to larger clusters, has shown, however, that a breakdown of the Luttinger sum rule is also possible for fillings off half-filling. For sufficiently strong U, the sum rule is violated in the whole $N = L - 1$-particle

sector. This would correspond to a (strongly correlated) metallic state in the thermodynamic limit. Whether or not a VCA calculation is consistent with the sum rule, then depends on the set of cluster hopping parameters t' which make the self-energy functional stationary. First VCA calculations[54] for the $D = 2$ Hubbard model at low doping and using clusters with up to $L = 10$ sites do predict a violation in fact.

It is by no means clear *a priori* what happens in a cluster approach using additional bath degrees of freedom as variational parameters, as e.g. in the cellular DMFT.[21,22] The usual periodization of the self-consistent C-DMFT self-energy, however, should be avoided when testing the sum rule as this introduces an additional (though physically motivated) approximation. Instead, Eq. (5) must be used with k re-interpreted as an index referring to the elements of the self-consistent diagonalized lattice Green's function.

Employing the dynamical cluster approximation (DCA)[20] represents an alternative which directly operates in reciprocal space. From a real-space perspective, the DCA is equivalent with the cellular DMFT but applied to a modified model $H = H(t, U) \rightarrow H(\bar{t}, U)$ with modified hopping parameters which are invariant under superlattice translations as well as under translations on the cluster.[29,55] In the limit $L \rightarrow \infty$ the replacement $t \rightarrow \bar{t}$ becomes irrelevant. Analogous to the C-DMFT, the sum rule then holds within the DCA if and only if it holds for the individual cluster at self-consistently determined cluster parameters. Note, however, that this requires that (besides the DCA self-energy) the *modified* hopping $\bar{t}$ instead of the physical hopping has to be considered in the computation of the volume enclosed by the Fermi (Luttinger) surface of the lattice model. This is exactly what is usually done in DCA calculations.

Within this context and in view of the violations found for finite Hubbard clusters, it is possible to understand why a non-perturbative cluster approximation, like the VCA,[54] or a cluster extension of the DMFT, like the DCA,[33] can produce results that are inconsistent with Luttinger's theorem.

Acknowledgments

APPENDIX A: MACROSCOPIC CONSERVATION OF ENERGY, PARTICLE NUMBER AND SPIN

The one-particle Green's function as obtained within an approximation generated by the choice of a reference system respects the macroscopic conservation laws which result from symmetries of the system with respect to continuous transformation groups:

Energy conservation is apparently respected as *by construction* the approximate SFT Green's function depends on a single frequency only, i.e. is invariant under time translations.

Conservation of the total particle number and spin is respected if the approximate G transforms in the same way as the exact Green's function under global U(1) and SU(2) gauge transformations. Consider a general transformation of the form

$$c_\alpha^\dagger \rightarrow \bar{c}_\alpha^\dagger = \sum_\beta S_{\beta\alpha} c_\beta^\dagger \qquad (A1)$$

with unitary S such that the interaction part $H_1(U)$ of the Hamiltonian is invariant (α refers to the states of the one-particle basis). In a diagrammatic approach, the invariance of $H_1(U)$ implies that the corresponding conservation law is respected "locally" at each vertex. Hence, for a conserving approximation in the sense of Baym and Kadanoff, the transformation behavior of the free Green's function is then propagated by the diagram rules to the full Green's function. Consequently, the latter must transform under S in the same way as the exact G, i.e.

$$G_{\alpha\beta} \rightarrow \bar{G}_{\alpha\beta} = \left(S G S^\dagger \right)_{\alpha\beta} . \qquad (A2)$$

Consider now the case of the SFT. One has to show that the approximate Green's function $\bar{G}$ for the transformed system with Hamiltonian $\bar{H}$ is given by $\bar{G} = S G S^\dagger$ if G is the approximate Green's function of the model H. Applying the transformation (A1) to H, one finds $H = H_0(t) + H_1(U) \rightarrow \bar{H} = H_0(\bar{t}) + H_1(U)$ with $\bar{t} = S t S^\dagger$. Again, S is assumed to leave the interaction part invariant.

The Green's function $\bar{G}$ of the transformed model is (approximately) constructed via

$$\bar{G} = \frac{1}{G_{\bar{t},0}^{-1} - \Sigma_{\bar{t}'_s, U}} \qquad (A3)$$

from the free Green's function of the transformed model and the SFT self-energy which is the self-energy of the reference system $H' = H_0(\bar{t}'_s) + H_1(U)$ at the stationary point $\bar{t}'_s$.

For the transformed problem $\bar{H}$, the stationary point $\bar{t}'_s$ is determined from the SFT Euler equation:

$$\sum_{\omega\alpha\beta} \left(\frac{1}{G_{\bar{t},0}^{-1} - \Sigma_{\bar{t}',U}} - G_{\bar{t}',U} \right)_{\omega;\beta\alpha} \frac{\partial (\Sigma_{\bar{t}',U})_{\omega;\alpha\beta}}{\partial t'} = 0 . \qquad (A4)$$

As an *ansatz* to solve the Euler equation we take

$$\bar{t}' = S t'_1 S^\dagger \qquad (A5)$$

with t'_1 to be determined. The transformation law (A2) for the *exact* Green's function of the reference system

is $G_{\overline{t}',U} = G_{St_1'S^\dagger,U} = SG_{t_1',U}S^\dagger$. This also holds for the free Green's function. Using the Dyson equation of the reference system we can deduce $\Sigma_{\overline{t}',U} = S\Sigma_{t_1',U}S^\dagger$. Furthermore, for the free Green's function of the transformed original model we have $G_{\overline{t},0} = SG_{t,0}S^\dagger$. Using these results, we see that Eq. (A4) is equivalent to

$$\sum_{\omega\alpha\beta}\left(\frac{1}{G_{t,0}^{-1} - \Sigma_{t_1',U}} - G_{t_1',U}\right)_{\omega;\beta\alpha}\frac{\partial(\Sigma_{t_1',U})_{\omega;\alpha\beta}}{\partial t'} = 0\,.$$
(A6)

But this is just the Euler equation for the original

model which is solved by $t_1' = t_s'$. Remembering the ansatz made, we now have for the stationary point $\overline{t}_s' = St_s'S^\dagger$. Inserting this into Eq. (A3) gives $\overline{G} = S(G_{t,0}^{-1} - \Sigma_{t_s',U})^{-1}S^\dagger = SGS^\dagger$ which is the desired result.

Note that one has to ensure that the stationary point for the transformed problem $\overline{t}_s' = St_s'S^\dagger$ lies within the space of one-particle parameters characteristic for the reference system. For models with local interaction part and for local (and also global) gauge transformations, however, this is always easily satisfied.

[1] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[2] G. Baym, Phys. Rev. **127**, 1391 (1962).
[3] N. E. Bickers, D. J. Scalapino, and S. R. White, Phys. Rev. Lett. **62**, 961 (1989).
[4] N. E. Bickers and S. R. White, Phys. Rev. B **43**, 8044 (1991).
[5] J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).
[6] J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).
[7] A. A. Abrikosow, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964).
[8] M. Potthoff, Euro. Phys. J. B **32**, 429 (2003).
[9] M. Potthoff, Euro. Phys. J. B **36**, 335 (2003).
[10] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003).
[11] M. Aichhorn, E. Arrigoni, M. Potthoff, and W. Hanke, Phys. Rev. B **74**, 024508 (2006).
[12] C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni, and M. Potthoff, Phys. Rev. B **70**, 245110 (2004).
[13] K. Pozgajcic, preprint cond-mat/0407172.
[14] W. Koller, D. Meyer, Y. Ono, and A. C. Hewson, Europhys. Lett. **66**, 559 (2004).
[15] D. Sénéchal, P.-L. Lavertu, M.-A. Marois, and A.-M. S. Tremblay, Phys. Rev. Lett. **94**, 156404 (2005).
[16] K. Inaba, A. Koga, S.-I. Suga, and N. Kawakami, Phys. Rev. B **72**, 085112 (2005).
[17] K. Inaba, A. Koga, S.-I. Suga, and N. Kawakami, J. Phys. Soc. Jpn. **74**, 2393 (2005).
[18] M. Aichhorn and E. Arrigoni, Europhys. Lett. **72**, 117 (2005).
[19] M. Eckstein, M. Kollar, M. Potthoff, and D. Vollhardt, Phys. Rev. B **75**, 125103 (2007).
[20] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H. R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
[21] G. Kotliar, S. Y. Savrasov, G. Pálsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001).
[22] A. I. Lichtenstein and M. I. Katsnelson, Phys. Rev. B **62**, R9283 (2000).
[23] S. Okamoto, A. J. Millis, H. Monien, and A. Fuhrmann, Phys. Rev. B **68**, 195121 (2003).
[24] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[25] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
[26] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).
[27] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[28] G. Kotliar and D. Vollhardt, Physics Today **57**, 53 (2004).
[29] M. Potthoff and M. Balzer, Phys. Rev. B **75**, 125112 (2007).
[30] W. O. Putikka, M. U. Luchini, and R. R. P. Singh, Phys. Rev. Lett. **81**, 2966 (1998).
[31] C. Gröber, R. Eder, and W. Hanke, Phys. Rev. B **62**, 4336 (2000).
[32] K. Haule, A. Rosch, J. Kroha, and P. Wölfle, Phys. Rev. Lett. **89**, 236402 (2002).
[33] T. A. Maier, T. Pruschke, and M. Jarrell, Phys. Rev. B **66**, 075102 (2002).
[34] J. Hubbard, Proc. R. Soc. London A **276**, 238 (1963).
[35] M. C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963).
[36] J. Kanamori, Prog. Theor. Phys. (Kyoto) **30**, 275 (1963).
[37] I. Dzyaloshinskii, Phys. Rev. B **68**, 085113 (2003).
[38] M. Potthoff, Adv. Solid State Phys. **45**, 135 (2005).
[39] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
[40] M. Potthoff, Phys. Rev. B **64**, 165114 (2001).
[41] E. Müller-Hartmann, Z. Phys. B **76**, 211 (1989).
[42] D. S. Fisher, G. Kotliar, and G. Moeller, Phys. Rev. B **52**, 17112 (1995).
[43] A. B. Harris and R. V. Lange, Phys. Rev. **157**, 295 (1967).
[44] J. S. Langer and V. Ambegaokar, Phys. Rev. **121**, 1090 (1961).
[45] D. C. Langreth, Phys. Rev. **150**, 516 (1966).
[46] H. Q. Lin and J. E. Gubernatis, Comput. Phys. **7**, 400 (1993).
[47] A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge, 1993).
[48] K. Haule, A. Rosch, J. Kroha, and P. Wölfle, Phys. Rev. B **68**, 155119 (2003).
[49] A. Rosch, preprint cond-mat/0602656.
[50] T.D. Stanescu, P. Phillips and Ting-Pong Choy, Phys. Rev. B **75**, 104503 (2007).
[51] J. Kokalj and P. Prelovšek, Phys. Rev. B **75**, 045111 (2007).
[52] M. Oshikawa, Phys. Rev. Lett. **84**, 3370 (2000).
[53] M. Potthoff, Condens. Mat. Phys. **9**, 557 (2006).
[54] M. Balzer, W. Hanke, and M. Potthoff, unpublished (2007).
[55] G. Biroli, O. Parcollet, and G. Kotliar, Phys. Rev. B **69**, 205108 (2004).