Jyväskylän yliopiston julkaisuarkisto
Jyväskylä University Digital Archive

UNIVERSITY OF JYVÄSKYLÄ

This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.

Author(s): Karlsson, Daniel; van Leeuwen, Robert

Title: Partial self-consistency and analyticity in many-body perturbation theory: Particle number conservation and a generalized sum rule

Year: 2016

Version:

Please cite the original version:

Karlsson, D., & van Leeuwen, R. (2016). Partial self-consistency and analyticity in many-body perturbation theory: Particle number conservation and a generalized sum rule. Physical Review B, 94(12), Article 125124. https://doi.org/10.1103/PhysRevB.94.125124

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Partial self-consistency and analyticity in many-body perturbation theory: Particle number conservation and a generalized sum rule

Daniel Karlsson and Robert van Leeuwen
Department of Physics, Nanoscience Center, P. O. Box 35, FI-40014 University of Jyväskylä, Finland and European Theoretical Spectroscopy Facility, ETSF

(Received 23 June 2016; revised manuscript received 25 August 2016; published 13 September 2016)

We consider a general class of approximations which guarantees the conservation of particle number in many-body perturbation theory. To do this we extend the concept of Φ derivability for the self-energy Σ to a larger class of diagrammatic terms in which only some of the Green's function lines contain the fully dressed Green's function G. We call the corresponding approximations for Σ partially Φ derivable. A special subclass of such approximations, which are gauge invariant, is obtained by dressing loops in the diagrammatic expansion of Φ consistently with G. These approximations are number conserving but do not have to fulfill other conservation laws, such as the conservation of energy and momentum. From our formalism we can easily deduce whether commonly used approximations will fulfill the continuity equation, which implies particle number conservation. We further show how the concept of partial Φ derivability plays an important role in the derivation of a generalized sum rule for the particle number, which reduces to the Luttinger-Ward theorem in the case of a homogeneous electron gas, and the Friedel sum rule in the case of the Anderson model. To do this we need to ensure that the Green's function has certain complex analytic properties, which can be guaranteed if the spectral function is positive-semidefinite. The latter property can be ensured for a subset of partially Φ-derivable approximations for the self-energy, namely those that can be constructed from squares of so-called half diagrams. For the case in which the analytic requirements are not fulfilled we highlight a number of subtle issues related to branch cuts, pole structure, and multivaluedness. We also show that various schemes of computing the particle number are consistent for particle number conserving approximations.

DOI: 10.1103/PhysRevB.94.125124

I. INTRODUCTION

Many-body perturbation theory using Green's functions is a successful and powerful method for studying interacting systems in and out of equilibrium. Its most general nonequilibrium version [1] is routinely used for very diverse situations such as quantum transport (e.g., through single molecules [2]), cold atoms in optical lattices [3], and transient dynamics (e.g., transient photoabsorption [4]). In the limiting case of equilibrium systems it can also be used to study standard photoabsorption and photoemission spectroscopies [5,6]. Moreover the nonequilibrium formalism even in this limiting case has found useful applications to derive approximations that guarantee positive spectral distributions as it allows for a direct expansion for those quantities [7–9].

One of the reasons of the successes of many-body perturbation theory is the use of diagrammatic techniques, where one can sum up diagrams to infinite order in a practical way. However, if partial resummations are used, when one sums over a subset over all diagrams, important conservation laws can be violated. The most important laws are the fulfillment of the continuity equation, momentum and angular momentum conservation, and energy conservation. Approximations which fulfill all of these conservation laws are referred to as conserving [10].

Baym [11] considered approximate generating functionals for the single-particle Green's function G, the Φ functional, from which the self-energy Σ could be constructed via $\Sigma = \delta\Phi/\delta G$. If G is obtained from the self-consistent solution of Dyson's equation for such a Σ, the approximation is called Φ derivable and is automatically conserving. Commonly used Φ-derivable many-body approximations (fully self-consistent) are the Hartree-Fock approximation, the second Born approximation (2BA), the GW approximation, and the T-matrix approximation.

While fully self-consistent schemes have many conceptual advantages, they usually carry a high computation cost, as well as other features that are not desirable, both in and out of equilibrium. In fully self-consistent calculations, it has been shown [12,13] that strong time-dependent fields can yield artificial steady states in finite systems. Partial self-consistency lessened this effect. Another method related to partial self-consistency is the generalized Kadanoff-Baym ansatz (GKBA), in which one approximates certain time-nondiagonal elements of the Green's function and which can be made conserving, and which is also free of artificial damping [14–16].

In equilibrium, especially in the *ab initio* community, the vast majority of calculations are non-self-consistent [6], or has a small degree of partial self-consistency [17,18] (however, see Refs. [19–25] for examples of fully self-consistent GW calculations). These types of approximations tend to yield accurate band gaps and spectral properties. Furthermore, for GW for the electron gas, full self-consistency worsens the spectral features compared to non-self-consistent calculations [20,26], even if the scheme is energy conserving. While these drawbacks will be cured by vertex corrections to GW, there has been a long discussion in the literature of whether one should focus on non-self-consistent vertex corrections, or full self-consistency (see, for example, [27–29]). Moreover, partially self-consistent approximations are expected to yield reasonable energies, especially if one uses the Luttinger-Ward functional for calculating the total energy [30].

 ©2016 American Physical Society

Thus, while the concept of a Φ-derivable approximation is extremely useful, it is quite restrictive. As soon as a scheme is not fully self-consistent, the concept of Φ derivability is lost, together with a convenient way of seeing whether the chosen approximation is particle number conserving. The class of approximations that do conserve particle number is larger than those generated by Φ functionals, however. For example, the partially self-consistent GW_0 approximation (where the screened interaction W is kept fixed to be W_0 during the self-consistency cycle) and the non-self-consistent G_0W_0 approximation are not Φ derivable. However, GW_0 is particle number conserving (first shown for the equilibrium electron gas [26,31], and later more generally [32]) but not fully conserving [33], while G_0W_0 is not particle number conserving [32,34].

Another property that can be violated by summing over subclasses of diagrams is the positivity of the spectral functions of the Green's function and the polarizability. Positivity is an important ingredient for two reasons. First of all, it is related to a probability interpretation of photoemission and photoabsorption processes, which naturally requires positive probabilities. Second, it guarantees correct analyticity and causality properties of the various propagators, which is a necessary condition for performing self-consistent calculations, as the analytic properties in general will deteriorate every iteration cycle otherwise.

The conserving and analytic properties are related but not equivalent. A situation in which both of these properties play a role is in the derivation of sum rules for the particle number, such as the Luttinger-Ward sum rule [35] for the electron gas and the Friedel sum rule [36,37] for the Anderson model. Another occasion is the derivation of the Ward identity and the frequency sum rule for the density response function. Conserving approximations in connection with the Luttinger-Ward sum rule was also studied in the context of the nonperturbative approximation of dynamical mean-field theory in Ref. [38].

In this work we generalize the concept of Φ derivability to partially self-consistent schemes. While the resulting approximations are not fully conserving, a large class of many-body approximations used in the literature is shown to be included in this larger class of approximations. The formalism provides an easy way to see whether an approximation conserves particle number or not. Since we are using the nonequilibrium Green's function formalism, the concept of partially self-consistent Φ functionals is valid both in and out of equilibrium. Using the new formalism, we study particle number sum rules in equilibrium in their most general settings, and highlight the importance of particle conservation. We state sufficient conditions for the sum rules to be valid in a given diagrammatic approximation. In particular, we stress that the sum rules are valid provided that G has the correct analytical properties, and we study the consequences of incorrect analytical properties, including issues related to pole structure and multivaluedness of the logarithm. We exemplify the formalism and sum rules in model systems of quantum transport.

The paper is structured as follows: In the first part of this work, Sec. II, we will introduce the concept of partial Φ derivability. Subsequently in Sec. III we will discuss the complex analytic properties for approximate Green's function and self-energies and the way these properties affect the calculation of the particle number. Then finally, in Sec. IV, we discuss the derivation of a generalized sum rule for the particle number and illustrate the concepts by explicit numerical calculations. We end with our conclusions in Sec. V.

II. PARTIAL Φ DERIVABILITY

Our main object of study is the single-particle Green's function $G(1,2)$, where $1 = (\mathbf{r}_1,\sigma_1,z_1)$ is a collective space-spin-time variable. The Green's function is defined on a time contour γ that has a forward, backward, and Matsubara (imaginary-time) branch [1],

$$G(1,2) = (-i)\frac{\mathrm{Tr}\{\mathcal{T}[e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}\hat{\psi}(1)\hat{\psi}^\dagger(2)]\}}{\mathrm{Tr}\{\mathcal{T}[e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}]\}}, \tag{1}$$

where $\hat{\psi}$ ($\hat{\psi}^\dagger$) are fermionic field destruction (creation) operators, and $\mathcal{T}$ is the time-ordering operator on the contour γ. The time-dependent Hamiltonian $\hat{H}(z) = \hat{H}_0(z) + \hat{H}_{\mathrm{int}}(z)$ is given by

$$\hat{H}_0(z) = \int d\mathbf{x}\, \hat{\psi}^\dagger(\mathbf{x})h(\mathbf{x},z)\hat{\psi}(\mathbf{x}), \tag{2}$$

$$\hat{H}_{\mathrm{int}}(z) = \frac{1}{2}\int d\mathbf{x}d\mathbf{x}'\, \hat{\psi}^\dagger(\mathbf{x})\hat{\psi}^\dagger(\mathbf{x}')v(\mathbf{x},\mathbf{x}',z)\hat{\psi}(\mathbf{x}')\hat{\psi}(\mathbf{x}), \tag{3}$$

where $h(\mathbf{x},z)$ is the single-particle part and $v(\mathbf{x},\mathbf{x}',z)$ is the two-body interaction. The nonequilibrium formalism allows for both time-dependent single-particle potentials, such as electric fields, as well as time-dependent two-body interactions, such as an adiabatic switch-on. To keep the discussion general, the time-dependence will not be specified further at this point. We also define, for convenience, $\mathbf{x} = (\mathbf{r},\sigma)$ and

$$\hat{G}(z_1,z_2) = \int d\mathbf{x}_1 d\mathbf{x}_2\, |\mathbf{x}_1\rangle G(1,2)\langle\mathbf{x}_2| \tag{4}$$

to highlight the temporal dependence without being dependent on the spatial basis.

The reason for considering the contour-ordered G is that general time-dependent systems at finite temperature can be considered, in and out of equilibrium. When the contour times z are on the Matsubara branch, we obtain the Matsubara Green's function as $\hat{G}^M(\tau_1,\tau_2) = \hat{G}(t_0 - i\tau_1, t_0 - i\tau_2)$. For general time-dependent systems, we obtain the lesser $\hat{G}^<$ and greater $\hat{G}^>$ Green's functions from $\hat{G}^<(t_1,t_2) = \hat{G}(t_1^-,t_2^+)$, and $\hat{G}^>(t_1,t_2) = \hat{G}(t_1^+,t_2^-)$, where $z = t^-/t^+$ is on the forward/backward branch. All single-particle quantities can be obtained from the contour-ordered Green's function. For example, if we take $t_1 = t_2$, we obtain the time-dependent density $n(\mathbf{x},t)$ and the time-dependent current density $\boldsymbol{j}(\mathbf{x},t)$ as

$$n(\mathbf{x},t) = -iG^<(\mathbf{x},t;\mathbf{x},t), \tag{5}$$

$$\boldsymbol{j}(\mathbf{x},t) = -\left[\frac{\nabla - \nabla'}{2m}G^<(\mathbf{x},t;\mathbf{x}',t)\right]_{\mathbf{x}'=\mathbf{x}}. \tag{6}$$

The equations of motion for $G(1,2)$ are given by the Kadanoff-Baym equations [1,39],

$$\left[i\frac{d}{dz_1} - \hat{h}(z_1)\right]\hat{G}(z_1,z_2) = \delta(z_1,z_2) + \int_\gamma dz_3\, \hat{\Sigma}(z_1,z_3)\hat{G}(z_3,z_2), \tag{7}$$

$$\left[-i\frac{d}{dz_2} - \hat{h}(z_2)\right]\hat{G}(z_1,z_2) = \delta(z_1,z_2) + \int_\gamma dz_3\, \hat{G}(z_1,z_3)\hat{\Sigma}(z_3,z_2) \tag{8}$$

with the Kubo-Martin-Schwinger (KMS) boundary conditions $\hat{G}(t_0 - i\beta,z_2) = -\hat{G}(t_0,z_2)$ and $\hat{G}(z_1,t_0 - i\beta) = -\hat{G}(z_1,t_0)$. $\hat{\Sigma}(z_1,z_2)$ is the self-energy, which has to be approximated. An equivalent way of writing the equations of motion is given by the Dyson equations,

$$\begin{aligned} G(1,2) &= \tilde{G}_0(1,2) + \iint_\gamma d3d4\, \tilde{G}_0(1,3)\Sigma(3,4)G(4,2), \\ G(1,2) &= \tilde{G}_0(1,2) + \iint_\gamma d3d4\, G(1,3)\Sigma(3,4)\tilde{G}_0(4,2), \end{aligned} \tag{9}$$

where $\tilde{G}_0(1,2)$ is the corresponding Green's function with $\hat{H}_{\text{int}} = 0$ that satisfies the KMS boundary conditions.

In a given approximation to $\hat{\Sigma}$, diagrammatic or not, there is no guarantee that the resulting scheme will be conserving. Since we in this work will focus on the fulfillment of the continuity equation, we here briefly describe it. We subtract Eq. (8) from Eq. (7), and put $2 = 1^+$, where 1^+ denotes $(\mathbf{x}_1,z_1^+)$ and z_1^+ denotes a time infinitesimally later than z_1 on the contour. We then obtain

$$\frac{\partial}{\partial z_1}n(1) + \nabla\cdot\boldsymbol{j}(1) = \int d3[\Sigma(1,3)G(3,1^+) - G(1,3)\Sigma(3,1^+)]. \tag{10}$$

Equation (10) is a continuity equation on the contour, with a source/drain term caused by the interactions. Thus, approximations that guarantee (local) particle conservation fulfill

$$\int d3[\Sigma(1,3)G(3,1^+) - G(1,3)\Sigma(3,1^+)] = 0 \tag{11}$$

for all times, which is then a condition on the approximate $\Sigma(1,2)$. If this condition is fulfilled, the total particle number

$$N = -i\int d\mathbf{x}\, G^<(\mathbf{x},t;\mathbf{x},t) \tag{12}$$

is conserved in time.

Particle number conservation is also relevant in equilibrium situations, as can be seen as follows. Let us consider a finite system in which the particles are noninteracting at t_0 with a switch-on of the interaction for times $t > t_0$. If the continuity equation is fulfilled for all times, the switching of the interaction cannot change the number of particles. In particular, this is true for an adiabatic switch-on of the interaction. Thus, particle conservation is an issue also in equilibrium calculations, a point also stressed in Refs. [32,33]. For example, within a conserving approximation the particle number cannot depend on the bond length of a molecule [33].

Baym [11] introduced a convenient way to generate approximations that are automatically conserving, with the use of a diagrammatically defined functional $\Phi[G]$. If $\Phi[G]$ is invariant under gauge transformations, translations, rotations, and time, the self-energy given via

$$\Sigma[G](1,2) = \frac{\delta\Phi[G]}{\delta G(2,1^+)} \tag{13}$$

will yield a conserving approximation when the Kadanoff-Baym equations, Eq. (7) and Eq. (8), are solved self-consistently with this Σ. In particular, the source term in the continuity equation, Eq. (10), will vanish. The scheme is then said to be Φ derivable.

We now extend the idea of Φ-derivable approximations to include partially self-consistent schemes, and find sufficient criteria for the fulfillment of the continuity equation. We define a partially dressed functional $\Phi[G,G_0]$, in which the Green's function lines can be dressed with either G or additional, fixed, Green's functions G_0. The fixed Green's functions could be outputs from a Hartree-Fock or density-functional theory calculation, but this is in no way necessary. In fact, sets of different fixed Green's functions can be used. We define the corresponding self-energy $\Sigma[G,G_0]$ via

$$\Sigma[G,G_0](1,2) = \left.\frac{\delta\Phi[G,G_0]}{\delta G(2,1^+)}\right|_{G_0\text{ fixed}}. \tag{14}$$

Solving the Kadanoff-Baym equations, Eq. (7) and Eq. (8), with this self-energy defines a partially self-consistent scheme. We will refer to such schemes as partially Φ derivable. Dressing all Green's function lines in Φ with G gives back the original $\Phi[G]$ functional as defined by Baym. Graphically, we obtain $\Sigma[G,G_0]$ from cutting away a G line from all $\Phi[G,G_0]$ diagrams in every possible way. Examples of partially dressed Φ functionals and the corresponding self-energies are shown in Figs. 1–3.

In Fig. 1 we show the example of a ring Φ diagram dressed in various ways. By dressing only one Green's function line with the full G, we obtain non-self-consistent (one-shot, or single-shot) approximations. Combining Fig. 1(a) and Fig. 2(d), we obtain one-shot 2nd Born. Moreover, Fig. 1(a) is part of the G_0W_0 approximation. The G_0W_0 approximation can be obtained by considering all higher order ring diagrams dressed with only one G. If all lines except one are dressed with G, as in Fig. 1(c), we obtain several types of diagrams. In one diagram the baseline in Σ is fully dressed, while in the other diagrams the baseline is not. Any one of the Σ diagrams in Fig. 1(c) is by itself not partially Φ derivable, since closing it with a G line and subsequently differentiating yields also the two other Σ diagrams. The first of the Σ diagrams in Fig. 1(c) is part of the G_0W approximation. It is then readily seen that the G_0W approximation for Σ is not partially Φ derivable. The same can be seen in Fig. 3(b), where we get two nonequivalent classes of Σ diagrams from one Φ diagram. Each separate Σ diagram is not partially Φ derivable, but their sum is.

An important class of diagrams is obtained when each loop in Φ is dressed with either all G or all G_0 separately. We will refer to this as a consistent dressing of the loops. Cutting a G line from consistently dressed Φ diagrams

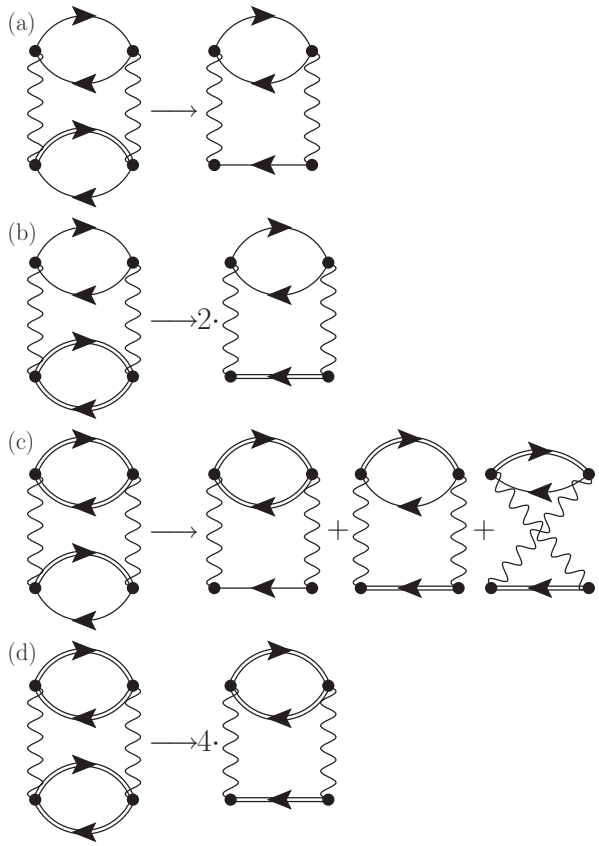

FIG. 1. Examples of partially dressed ring Φ functionals (left) and the corresponding $\Sigma = \frac{\delta\Phi}{\delta G}$ (right). (a) A single partially dressed loop; thus Φ is not gauge invariant and the approximation is not number conserving. This diagram is part of one-shot 2nd Born, as well as G_0W_0. (b) A single, fully dressed loop, which gives a gauge-invariant and thus particle-conserving approximation. This diagram is part of the GW_0 approximation. (c) Partially dressed loops, not gauge invariant. Several types of self-energy diagrams are obtained, where the first is part of the G_0W approximation. (d) Fully dressed, thus conserving. Part of 2nd Born, and GW.

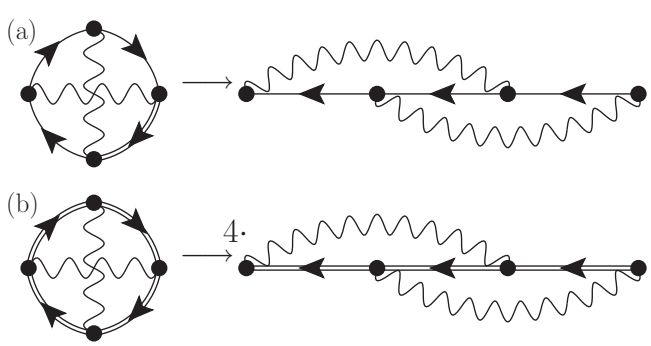

FIG. 2. Examples of partially dressed 2nd-order exchange (2OE) Φ functionals (left) and the corresponding $\Sigma = \frac{\delta\Phi}{\delta G}$ (right). (a) Partially dressed second-order exchange, not gauge invariant. Constitutes one-shot 2nd Born, together with Fig. 1(a). (b) Fully dressed second-order exchange, thus conserving. Constitutes 2nd Born, together with Fig. 1(d). The only way to obtain a number-conserving approximation for these types of diagrams is to fully dress all G lines.

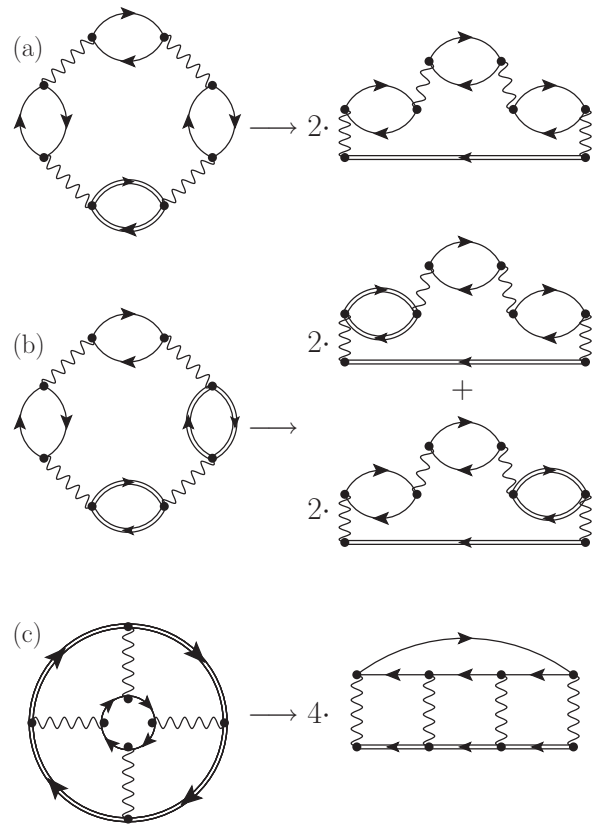

FIG. 3. Examples of partially dressed, gauge-invariant, Φ functionals (left) which fulfill particle number conservation, and their corresponding $\Sigma = \frac{\delta\Phi}{\delta G}$ (right). (a) A 4th-order GW_0 diagram. (b) Dressing two loops yields two nonequivalent classes of diagrams. (c) A 4th-order diagram in the particle-particle T-matrix approximation. Note that we would get the same diagram if we dressed only the inner loop.

yields $\Sigma[G,G_0]$ diagrams with fully dressed baselines. An example of a consistently dressed Φ diagram is shown in Fig. 1(b). This diagram is part of the GW_0 approximation. The remaining terms in the GW_0 approximation can be obtained by considering all higher order ring diagrams with only one loop consistently dressed with G; see also Fig. 3(a). In the next section, we will show that all approximations coming from consistently dressed Φ functionals are number conserving.

In a given Φ-derivable approximation, the diagrammatic structure of $\Phi[G,G_0]$ is similar to the one for the fully dressed $\Phi[G]$,

$$\Phi[G,G_0] = \sum_{nk} c_{nk} \int d1d2\Sigma_k^{(n)}[G,G_0](1,2)G(2,1^+), \quad (15)$$

where c_{nk} is a symmetry prefactor, n is the number of interaction lines, and k labels the Σ diagrams. For a collection of Σ_k diagrams obtained from the same Φ diagram, the number c_{nk} is generally given by one divided by the number of G lines in Φ. The proof of this statement is analogous to the proof that $c_{nk} = 1/2n$ for the case in which all $2n$ G lines are dressed [1]. As an example, in Fig. 3(a) the GW_0 diagram has a prefactor of $1/2$, while the T-matrix diagram has $1/4$. In fact, for GW_0 each Φ diagram has the same prefactor $1/2$, independently of

the order of the diagram. Note that the Feynman rules for Φ include a symmetry prefactor as well, related to the dimension of the symmetry group of each diagram [1].

Number conservation

As was shown by Baym [11] for fully dressed Φ functionals, particle number conservation is guaranteed by the inherent gauge invariance in $\Phi[G]$. Here, we repeat the same steps for partially dressed $\Phi[G,G_0]$ functionals.

Let us consider a change in G of the form $G_\Lambda(1,2) = e^{i\Lambda(1)}G(1,2)e^{-i\Lambda(2)}$, where $\Lambda(t_0) = \Lambda(t_0 - i\beta)$ in order to satisfy the correct boundary conditions. For the exact Green's function, this transformation results from a gauge transformation $\Lambda(1)$ [1]. The Φ functional will change according to $\Phi[G_\Lambda,G_0]$. For a general Φ functional, such as the ones in Figs. 1(a) and 1(c), $\Phi[G_\Lambda,G_0]$ will be different from $\Phi[G,G_0]$, and thus those Φ are not gauge invariant. However, a gauge-invariant Φ is obtained if each loop is consistently dressed, such as in Figs. 1(b) and 1(d), since the phase factors always cancel for each loop separately. We thus have $\Phi[G_\Lambda,G_0] = \Phi[G,G_0]$, and hence $\delta\Phi = 0$ under gauge transformations.

For an arbitrary variation in $G(1,2)$, by definition the corresponding Φ functional will change according to

$$\delta\Phi = \int d1d2\Sigma(1,2)\delta G(2,1^+). \tag{16}$$

For an infinitesimal gauge,

$$\delta G(1,2) = G_\Lambda(1,2) - G(1,2) = i[\Lambda(1) - \Lambda(2)]G(1,2), \tag{17}$$

the variation in Φ becomes

$$\delta\Phi = i\int d1d2[\Sigma(1,2)G(2,1^+) - G(1,2)\Sigma(2,1^+)]\Lambda(1). \tag{18}$$

For gauge-invariant Φ functionals, $\delta\Phi = 0$ for any gauge Λ, and thus we have

$$0 = \int d3[\Sigma(1,3)G(3,1^+) - G(1,3)\Sigma(3,1^+)]. \tag{19}$$

This is the source term which appeared in the continuity equation, Eq. (10). Thus, a gauge-invariant Φ leads to an approximation that conserves the particle number. If the approximate Φ used is not gauge invariant, the source term will in general be nonzero, leading to a violation of the particle number.

Since the resulting Σ from a gauge-invariant Φ has a fully dressed baseline and consistently dressed loops, we can also infer that $\Sigma[G_\Lambda,G_0](1,2) = e^{i\Lambda(1)}\Sigma[G,G_0](1,2)e^{-i\Lambda(2)}$. G_Λ and $\Sigma[G_\Lambda,G_0]$ fulfill the Kadanoff-Baym equations, Eq. (7) and Eq. (8), with a Hamiltonian shifted with the gauge. One can then use the same considerations as in Ref. [1] to show that the Ward identity is satisfied, which then also leads to the fulfillment of the frequency sum rule for the density response function. As in the fully conserving case, this is however only valid under additional assumptions on the correct analytical structure of the response functions [1,40].

An approximate, gauge-invariant Φ functional is guaranteed to yield an approximation that fulfills the continuity equation for all times, independently of the particular shape of the time dependence of the Hamiltonian. For example, the relation is valid for a time-dependent single-particle potential, such as a time-dependent electric bias, relevant for quantum transport. The importance of using particle-conserving approximations is exemplified by non-self-consistent 2nd Born and G_0W_0, which can severely violate the continuity equation in quantum transport [2,41].

For gauge-dependent approximations, particle conservation can be violated already in ground-state calculations. A clear example can be found in Ref. [33], where the H_2 molecule was considered, starting with 2 particles in the noninteracting ground state. After an adiabatic switching, it was found that the particle number depended on the molecular bond length for G_0W_0, while it did not for GW_0.

We also stress that any linear combination of consistently dressed Φ functionals generates particle number conserving approximations. One can then generate particle number conserving approximations when combining many-body methods, such as $GW + T$-matrix [42] and the FLEX approximation [43,44], provided that each Φ diagram is consistently dressed.

Finally, we stress that gauge-invariant partial Φ derivability guarantees number conservation, but not other types of conservation laws ensured by a fully conserving scheme, such as momentum, angular momentum, and energy conservation [11]. For example, the conservation of energy depends on the time invariance of all four Green's function lines joining an interaction line, and as such it is not enough for each loop to be dressed consistently separately. Furthermore, no mention has been made about how large the violations of the conservation laws are, since this also depends on how strongly correlated the system is [34]. In equilibrium, the violation of particle number seems to be small in some systems [32,33], but larger for other systems [34]; see also Sec. IV C. In biased systems in quantum transport, the violation in the particle current can be as large as the current itself [2,41].

III. ANALYTICAL PROPERTIES OF G AND Σ

A. Analytical continuation and causality

In the previous section we discussed particle number conserving approximations. In our discussion the particle number was defined by Eq. (12) using the Green's function in real time. In the following we will discuss only the equilibrium situation. In that case there are at least two others ways to calculate the particle number. The first one is from an integral over the spectral function, as we will discuss in much more detail below. Another way is from the derivative of the grand-canonical potential Ω with respect to the chemical potential μ as

$$N = -\frac{\partial\Omega}{\partial\mu}. \tag{20}$$

Baym showed that for a conserving approximation the latter equation yields the same result as Eq. (12). In his derivation he, however, implicitly assumed some analytic requirements of the Green's function, which are not guaranteed to be fulfilled. In the reminder of the paper we will show that all three ways of calculating the particle number yield the same results for

partially Φ-derivable approximations provided that they preserve the correct complex analytic properties. We will further discuss a generalized sum rule for the particle number, the validity of which depends crucially on the analyticity of the Green's function. For these reasons we will in this section first review the analytic properties of the Green's function and how they relate to the positive semidefiniteness of the spectral function.

For systems in equilibrium at inverse temperature β, the integrals over the γ contour reduce to integrals over the imaginary axis. We can then make use of the Matsubara representation

$$\hat{G}^M(\tau_1,\tau_2) = \frac{1}{-i\beta}\sum_{m=-\infty}^{\infty} e^{-\omega_m(\tau_1-\tau_2)}\hat{G}(\omega_m) \tag{21}$$

with Matsubara frequencies $\omega_m = \frac{2m+1}{-i\beta}$. The exact Green's function can be rewritten using the Lehmann representation [45] as

$$\hat{G}^M(\omega_n) = \int_{-\infty}^{\infty} d\omega' \frac{\hat{A}(\omega')}{\omega_n + \mu - \omega'}, \tag{22}$$

where $\hat{A}(\omega)$ is the spectral function.

We are interested in the analytical continuation of this function to the complex frequency plane. It is not obvious that such a continuation is unique, since $\hat{G}^M(\omega_m)$ is only defined on isolated points. Nevertheless, if we restrict the continuation to be analytic and bounded at infinity, the continuation is unique, and is given by [46]

$$\hat{G}^M(\zeta) = \int_{-\infty}^{\infty} d\omega' \frac{\hat{A}(\omega')}{\zeta + \mu - \omega'}, \quad \mathrm{Im}\,\zeta \neq 0. \tag{23}$$

We use ζ as a complex frequency, while real frequencies are denoted by ω. We note that Eq. (23) immediately implies that $\hat{G}^M$ is analytic away from the real axis. Equation (23) can be used to define the commonly used retarded and advanced Green's functions in their respective half planes as

$$\hat{G}^M(\zeta) = \begin{cases} \hat{G}^R(\zeta+\mu), & \mathrm{Im}\,\zeta > 0, \\ \hat{G}^A(\zeta+\mu), & \mathrm{Im}\,\zeta < 0. \end{cases} \tag{24}$$

The real-frequency retarded and advanced Green's functions are then obtained as $\hat{G}^M(\omega \pm i\eta) = \hat{G}^{R/A}(\omega+\mu)$, where $(\hat{G}^A)^\dagger(\omega) = \hat{G}^R(\omega)$ and η a positive infinitesimal. $\hat{G}^M(\zeta)$ is discontinuous when crossing the real axis, with the discontinuity given by

$$\hat{G}^R(\omega) - \hat{G}^A(\omega) = -2\pi i\hat{A}(\omega), \tag{25}$$

as can be deduced from Eq. (22). The functions $\hat{G}^{R/A}(\zeta)$ are analytic in their respective half planes, which is what we will refer to as the correct analytical properties.

The discussion so far has been independent of the self-energy. However, in practice, we often start from an equation of the self-energy. We therefore want to relate the analytical properties of G to those of Σ. This connection is provided by the Dyson equation, Eq. (9), which in terms of Matsubara notation becomes

$$\hat{G}^M(\omega_m) = \frac{1}{\omega_m + \mu - \hat{h} - \hat{\Sigma}^M(\omega_m)}, \tag{26}$$

where for operators $\hat{A}$ we use the notation $\hat{A}^{-1} = 1/\hat{A}$ [46].

The same considerations for $\hat{G}^M(\zeta)$ also apply to the Matsubara self-energy. Its analytic continuation is

$$\hat{\Sigma}^M(\zeta) = \int_{-\infty}^{\infty} d\omega' \frac{\hat{\Gamma}(\omega')}{\zeta + \mu - \omega'}, \quad \mathrm{Im}\,\zeta \neq 0, \tag{27}$$

where $\hat{\Gamma}(\omega)$ is the rate operator. $\hat{\Sigma}^{R/A}(\zeta)$ are defined as in Eq. (24). Furthermore, $\hat{\Sigma}^M(\omega \pm i\eta) = \hat{\Sigma}^{R/A}(\omega+\mu)$, and $\hat{\Sigma}^R(\omega) - \hat{\Sigma}^A(\omega) = -2\pi i\hat{\Gamma}(\omega)$. The analytic continuations $\hat{G}^M(\zeta)$ and $\hat{\Sigma}^M(\zeta)$ are related by

$$\hat{G}^M(\zeta) = \frac{1}{\zeta + \mu - \hat{h} - \hat{\Sigma}^M(\zeta)}, \quad \mathrm{Im}\,\zeta \neq 0. \tag{28}$$

The discussion so far considered the exact Green's function and self-energy. However, in practice we typically approximate the self-energy and solve the Dyson equation. In a given approximation for Σ the Green's function does not necessarily have the correct analytical properties. For instance, the denominator in Eq. (28) may have zeros away from the real axis in which case $\hat{G}^M(\zeta)$ has poles away from the real axis and consequently a representation as in Eq. (23) does not exist. However, such nonanalyticities are not possible if the approximate self-energy is of the form of Eq. (27) in which the rate operator $\hat{\Gamma}(\omega)$ is positive-semidefinite (PSD). With this we mean that $\langle\varphi|\hat{\Gamma}(\omega)|\varphi\rangle \geqslant 0$ for any one-particle basis function φ. In this case one can show that $\hat{G}^M(\zeta)$ is analytic away from the real axis and has a representation as in Eq. (23) with $\hat{A}(\omega)$ being PSD as well [7–9]. For this reason a diagrammatic perturbation method based on so-called half diagrams has been devised in order to guarantee that the rate function is always PSD [7–9].

For the case in which an approximate form for $\hat{\Sigma}^M$ does give rise to poles for $\hat{G}^M$ a generalized spectral representation, as in Eq. (22), is possible. In Appendix C we show that if $\hat{G}^M$ is analytic except for simple poles away from the real axis we can write

$$\hat{G}^M(\zeta-\mu) = \int_{-\infty}^{\infty} d\omega \frac{\hat{A}(\omega)}{\zeta-\omega} + \sum_l \left(\frac{\hat{\alpha}_l}{\zeta - \xi_l^R} + \frac{\hat{\alpha}_l^\dagger}{\zeta - (\xi_l^R)^*}\right), \tag{29}$$

where $\hat{\alpha}_l$ is the residue matrix, ξ_l^R is the location of a pole in the upper half plane, and the spectral function $\hat{A}(\omega)$ is also in this case given by Eq. (25). If $\hat{G}^M$ is obtained from Eq. (28) then from taking the limit $|\zeta| \to \infty$ (assuming that the approximate Σ is bounded at infinity) we see from

$$\hat{G}^M(\zeta) \to \frac{\hat{\mathbb{1}}}{\zeta} \quad (|\zeta| \to \infty) \tag{30}$$

that

$$\hat{\mathbb{1}} = \int_{-\infty}^{\infty} d\omega \hat{A}(\omega) + \sum_l (\hat{\alpha}_l + \hat{\alpha}_l^\dagger), \tag{31}$$

where $\hat{\mathbb{1}}$ is the unit operator. Thus, the presence of poles can change the norm of the spectral function.

While discussing the analytic properties, we would like to clear up a possible point of confusion regarding the analytic continuation of the retarded and advanced functions

$G^{R/A}(\zeta)$ from the upper/lower half planes in which they are analytic, across the real axis to the other half planes. Such continuations are useful for derivations with the help of Cauchy's theorem and have also been useful for interpreting quasiparticle lifetimes in terms of analytic properties of the analytically continued $G^{R/A}(\zeta)$.

Due to the discontinuity in $G^M(\zeta)$, the analytical continuation from the upper half plane, where $G^M(\zeta) = G^R(\zeta)$, into the lower half plane yields a completely different function than $G^A(\zeta)$. The analytically continued $G^R(\zeta)$ is used in several textbooks (see, for example, Refs. [45,47]) when the quasiparticle peak of the spectral function has a Lorentzian shape. A typical example is the homogeneous electron gas in which case we can regard the spectral function as a scalar function rather than a matrix, as it is diagonal in momentum space. We consider a Lorentzian with width Γ as

$$A(\omega) = \frac{\Gamma}{\pi}\frac{1}{\omega^2 + \Gamma^2}. \tag{32}$$

We obtain $G^R(\zeta) = \frac{1}{\zeta + i\Gamma}$, $\operatorname{Im}\zeta > 0$ from Eq. (23). The analytical continuation is given by the same expression, but defined for all ζ and has a pole in the lower half plane. A useful application of the analytic continuation is the calculation of the retarded Green's function in time space,

$$G^R(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} G^R(\omega) e^{-i\omega t}, \tag{33}$$

by the following procedure. For $t < 0$, $G^R(\zeta)e^{-i\zeta t}$ decays exponentially in the upper half plane, and we can integrate around a closed half circle in the upper half plane. Since $G^R(\zeta)$ has no pole there, $G^R(t) = 0$ for $t < 0$, befitting of a retarded function. This also shows that the existence of poles in the upper half plane will in general violate the causality property of $G^R(t)$. For $t > 0$, we integrate around a half circle in the lower half plane and find

$$G^R(t) = -i\theta(t)e^{-\Gamma t}. \tag{34}$$

Thus, the location of the pole in the lower half plane yields the lifetime of an added particle. This interpretation was entirely based on the Lorentzian form of the quasiparticle peak of the spectral function. However, it is known for the homogeneous electron gas that the moment $\int d\omega\, \omega^2 A(\omega)$ of the spectral function is finite, which precludes a Lorentzian form [48]. A study of the short-time properties of the spectral function [48,49] shows that for this system a more realistic form of the retarded Green's function in real time is

$$G^R(t) = -i\theta(t)\exp\left(-\gamma\frac{t^2}{t+\tau}\right) \tag{35}$$

with γ and τ parameters. This Green's function has a Gaussian short-time behavior and an exponential long-time behavior. We would like to demonstrate that such a short-time behavior can be described by a simple spectral function of Gaussian form, $A(\omega) = \frac{1}{a\sqrt{\pi}}e^{-\frac{\omega^2}{a^2}}$, where $a > 0$ determines the width of the single peak, and concentrate on the analytical properties. From Eq. (23), we obtain

$$G^M(\zeta) = e^{-\frac{\zeta^2}{a^2}}\frac{\pi\operatorname{Erfi}(\zeta/a) + \ln(-\zeta) - \ln(\zeta)}{a\sqrt{\pi}}, \quad \operatorname{Im}\zeta \neq 0, \tag{36}$$

where Erfi is the imaginary error function, analytic in the whole complex plane. As it should, $G^M(\zeta) \to 1/\zeta$ as $|\zeta| \to \infty$. The analytically continued $G^R(\zeta)$ is given by

$$G^R(\zeta) = \frac{\sqrt{\pi}}{a}e^{-\frac{\zeta^2}{a^2}}[\operatorname{Erfi}(\zeta/a) - i], \tag{37}$$

which is analytic for all ζ in the complex plane. Thus, for a Gaussian spectral function, $G^R(\zeta)$ does not have any poles in the lower half plane, as opposed to the case with a Lorentzian spectral function. Moreover, $G^R(\zeta)$ is unbounded in the lower half plane, where $G^R(\zeta) \to -\frac{2i\sqrt{\pi}}{a}e^{-\frac{\zeta^2}{a^2}}$ for $|\zeta| \to \infty$. This affects the calculation for $G^R(t)$ using contour integration. We can close a contour in the upper half plane for $t < 0$ yielding $G^R(t) = 0$, but we cannot close the contour in the lower half plane for $t > 0$, since the integral on the half circle diverges; see Fig. 4. Nevertheless, $G^R(t)$ is well defined for all t, and can be obtained by Fourier-transforming $G^R(\omega)$ directly, yielding

$$G^R(t) = -i\theta(t)e^{-\frac{1}{4}a^2t^2}. \tag{38}$$

We see that, similarly to the Lorentzian case, the width of the spectral function determines the lifetime of an added particle excitation. However, in this case the lifetime is unrelated to the properties of $G^R(\zeta)$ analytically continued to the lower half plane. The main message that we want to give in this section is therefore that the analytic continuations of the retarded and advanced functions beyond their original analytic domains in general can give rise to a richer analytical structure than just simple poles. This not only includes unbounded analytic functions, but, for instance, also functions with algebraic or logarithmic branch cuts.

B. PSD and analyticity

One way of ensuring the correct analytical properties of $G^M(\zeta)$ is to ensure that the chosen diagrammatic approximation retains the PSD property. If the diagrammatic structure has this property, then the correct analytical properties are guaranteed [50]. Recently, it has been shown [7,9] that by choosing diagrams in a specific way, by only keeping diagrams that can be built up from so-called half diagrams, the PSD property and thus the correct analytical properties are guaranteed at zero temperature. Moreover, the PSD properties depend only on the diagrammatic structure, but not on the dressing of the loops provided that we dress with a G coming from a PSD approximation. It is thus important to note that conservingness and PSD are two completely different properties, a point which we feel is not stressed enough in the literature. For example, G_0W_0 is not conserving, but PSD, while the fully dressed second-order exchange approximation [see Fig. 2(b)] is conserving, but not PSD [7]. Thus, we are guaranteed to have the correct analytical properties for the former approximation, but not the latter.

Non-PSD spectral functions have been observed [7,51] in studies of the homogeneous electron gas, when one goes

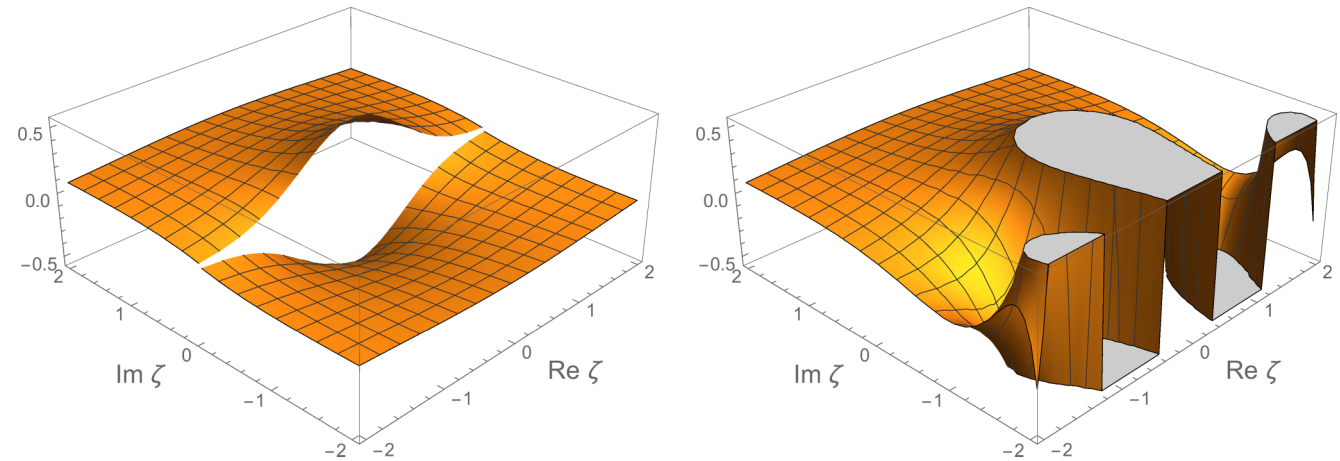

FIG. 4. Left: $-\frac{1}{\pi}\,\mathrm{Im}\,G^M(\zeta)$ for a Gaussian spectral function, Eq. (36), with $a = 1$. $G^M(\zeta) \to 1/\zeta$ for large ζ. Right: Analytically continued $-\frac{1}{\pi}\,\mathrm{Im}\,G^R(\zeta)$ for a Gaussian spectral function, Eq. (37), with $a = 1$. $G^R(\zeta)$ is analytic everywhere and is unbounded in the lower half plane. High and low values have been cut off for the sake of clarity.

beyond GW and takes vertex corrections into account. The lowest order vertex correction is the second-order exchange diagram, where the interaction lines have been dressed with the screened interaction. Moreover, it has also been observed [51] that in this approximation a pole can be created in the lower half plane for $G^A(\zeta)$, violating the analytical properties. Also in finite systems, restricting to certain classes of diagrams can violate positivity for the photoabsorption spectra [40,52].

To elucidate the above considerations, and to illustrate Eq. (29), we consider a very simple model of an interacting self-energy. We assume that the analytically continued $\Sigma^R(\zeta)$ has a single pole on the real axis, with residue $\alpha \in \mathbb{R}$,

$$\Sigma^R(\zeta) = \frac{\alpha}{\zeta - b}. \tag{39}$$

$\Sigma^R(\zeta)$ is analytic in the upper half plane, and $\alpha > 0$ (< 0) corresponds to a positive (negative) definite rate operator $\Gamma(\omega) = -1/\pi\,\mathrm{Im}[\Sigma^R(\omega)]$ on the real axis. The resulting $G^R(\zeta)$ is

$$G^R(\zeta) = \frac{1}{\zeta - \Sigma^R(\zeta)}, \tag{40}$$

which has two poles, at

$$\omega_\pm = \frac{1}{2}(b \pm \sqrt{4\alpha + b^2}). \tag{41}$$

If $4\alpha > -b^2$, $G^R(\zeta)$ is analytic in the upper and lower half planes, but otherwise we obtain two poles in the separate half planes. The imaginary part is

$$\mathrm{Im}[\omega_\pm] = \begin{cases} 0, & 4\alpha > -b^2, \\ \pm\sqrt{-(4\alpha + b^2)}, & 4\alpha < -b^2. \end{cases} \tag{42}$$

If $G^R(\zeta)$ is nonanalytic in the upper half plane, it is impossible to write $G^R(\zeta)$ in the form of Eq. (23). Nevertheless, we can define the spectral function as $A(\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,G^R(\omega)$ due to Eq. (29), which will be the sum of two delta functions for the case in which the poles are on the real axis. Now, however, the integral over the spectral function will depend on the position on the pole. The integral can be obtained by integrating $G^R(\zeta)$ over a semicircle in the upper half plane. In this model, $G^R(\zeta) \to 1/\zeta, \zeta \to \infty$, and thus the half circle integration yields πi. If $4\alpha < -b^2$, then there is a pole ω_+ in the upper half plane, with residue

$$\alpha_1 = \frac{1}{2} - i\frac{b}{2\sqrt{-(4\alpha + b^2)}}. \tag{43}$$

This gives the integral as

$$\int_{-\infty}^{\infty} A(\omega)d\omega = \begin{cases} 1, & 4\alpha > -b^2, \\ 0, & 4\alpha < -b^2. \end{cases} \tag{44}$$

The spectral function is plotted in Fig. 5 for different values of α. As can be seen, if $\alpha > 0$, $A(\omega)$ is positive-definite. If $-b^2 < 4\alpha < 0$, $A(\omega)$ has one negative and one positive peak. In this simple model, we can thus identify three regions:

(1) $\alpha > 0$. The rate operator is positive-definite, and thus the spectral function is positive-definite. This leads to that $G^R(\zeta)$ is analytic in the upper half plane, and $\int d\omega A(\omega) = 1$.

(2) $-b^2 < 4\alpha < 0$. The rate operator is negative definite, and $A(\omega)$ can have both positive and negative weights. $G^R(\zeta)$ is analytic in the upper half plane, and $\int d\omega A(\omega) = 1$.

(3) $4\alpha < -b^2$. The rate operator is negative definite, and $A(\omega) = 0$. There is a pole in the upper half plane for $G^R(\zeta)$ and is of the form of Eq. (29) in the upper half plane:

$$G^R(\zeta) = \frac{\alpha_1}{\zeta - \xi_1^R} + \frac{\alpha_1^*}{\zeta - (\xi_1^R)^*}, \tag{45}$$

where $\xi_1^R = \omega_+$ and $(\xi_1^R)^* = \omega_-$.

Keeping in mind that $\Sigma^R(\zeta)$ is a very simple model of an interacting self-energy, the case $\alpha > 0$ corresponds to when we can write the self-energy as a sum of squares of half diagrams [9,50]. The case $\alpha < 0$ corresponds to when the self-energy is not a sum of squares of half diagrams. However, we see that it is not guaranteed that the correct analytic properties are broken. Taking α as a measure of the correlation effects, for $\alpha < 0$ we find a critical point when the analytical properties are broken. While being a simple model, the same trend can be observed in the Anderson model, discussed in Sec. IV C.

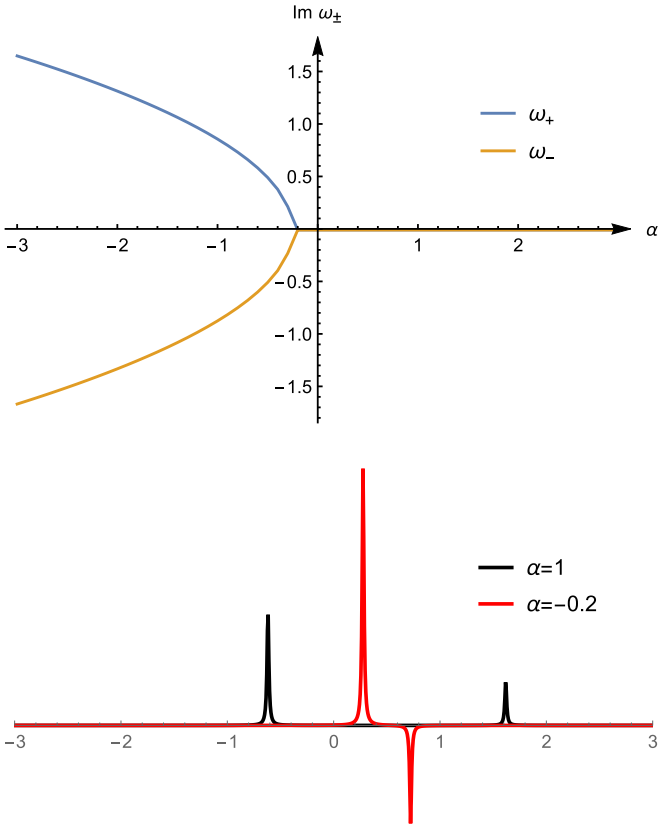

FIG. 5. Upper: The imaginary part of the two poles $\omega_\pm$ of G^R as function of α. The branching occurs at $4\alpha = -b^2$. Lower: The spectral function $A(\omega)$ for $\alpha > 0$, where $A(\omega) > 0$, and $-4b^2 < \alpha < 0$, where $A(\omega)$ can become negative. For $\alpha < -4b^2$, $A(\omega) = 0$, the delta function weights are zero. The parameters are $b = 1$, with a finite broadening.

IV. GAUGE INVARIANCE AND SUM RULES

Gauge-invariant Φ functionals lead to approximations that obey well-known sum rules in equilibrium. Such sum rules have been discussed before (see, e.g., [35,53,54]). Here, we repeat the derivations for approximations that are partially Φ derivable, but we keep a general structure and focus on the importance of the analytical properties.

We consider the number of particles in equilibrium, given by the Matsubara Green's function as

$$N = -i\mathrm{Tr}\{\hat{G}^M(\tau,\tau^+)\} = \frac{1}{\beta}\mathrm{Tr}\left\{\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\hat{G}^M(\omega_n)\right\}$$
$$= \frac{1}{\beta}\mathrm{Tr}\left\{\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\frac{-1}{\hat{\Sigma}^M(\omega_m)+\hat{h}-\mu-\omega_m}\right\}, \quad (46)$$

where $\eta \to 0^+$ after the summation. According to our previous discussion, we can analytically continue $\hat{G}^M(\omega_m)$ to $\hat{G}^M(\zeta)$ when $\mathrm{Im}\,\zeta \neq 0$, for any finite temperature. Here, we wish to point out an issue regarding the analytical properties.

Matsubara sums can be rewritten according to the following rule (see, for example, Ref. [45]), which follows from considering two half circles, in the upper and lower half plane, when $\hat{Q}(\zeta)$ is analytic for $\mathrm{Im}\,\zeta \neq 0$,

$$\frac{i}{\beta}\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\hat{Q}(\omega_m)$$
$$= \lim_{\eta\to 0^+}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}f(\omega)[\hat{Q}(\omega-i\eta)-\hat{Q}(\omega+i\eta)], \quad (47)$$

where $f(\omega) = 1/(e^{\beta\omega}+1)$ is the Fermi function. By taking $\hat{Q} = \hat{G}^M$ and using Eq. (46), we can write N as

$$N = \int_{-\infty}^{\infty}\frac{d\omega}{2\pi i}f(\omega)\mathrm{Tr}\{\hat{G}^A(\omega+\mu)-\hat{G}^R(\omega+\mu)\}$$
$$= \int_{-\infty}^{\infty} d\omega f(\omega-\mu)\mathrm{Tr}\{\hat{A}(\omega)\} \quad (48)$$

provided that $\hat{G}^M(\zeta)$ is analytic when $\zeta \neq 0$. If $\hat{G}^M(\zeta)$ would have simple poles, we must add their residues to Eq. (48) as these poles give an extra contribution in the derivation of Eq. (47). We conclude that if $\hat{G}^M(\zeta)$ is nonanalytic, then $-i\mathrm{Tr}\{\hat{G}^M(\tau,\tau^+)\}$ and $\int_{-\infty}^{\infty} d\omega f(\omega-\mu)\mathrm{Tr}\{\hat{A}(\omega)\}$ can give different results. This shows yet another inconsistency if one works with approximations that do not guarantee the correct analytical properties. Note, however, that no assumption was made on the existence of a generating Φ functional.

Yet another way of obtaining the particle number is via the grand-canonical potential, as was considered by Baym [11] for conserving schemes, via Eq. (20). For a partially Φ-derivable approximation with the correct analytical properties, this gives the same particle number as the two other definitions discussed. For details, see Appendix D.

We now focus on rewriting Eq. (46). Since the continuation of $\hat{G}^M(\omega_m)$ to $\hat{G}^M(\zeta)$ is analytic at the Matsubara frequencies, we can perform derivatives around ω_m. We can thus write Eq. (46) as

$$N = \underbrace{\frac{1}{\beta}\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\mathrm{Tr}\left\{\frac{\partial}{\partial\zeta}\ln[\hat{\Sigma}^M(\zeta)+\hat{h}-\mu-\zeta]\Big|_{\zeta=\omega_m}\right\}}_{I_1}$$
$$+\underbrace{\frac{1}{\beta}\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\mathrm{Tr}\left\{\hat{G}^M(\omega_m)\frac{\partial\hat{\Sigma}^M}{\partial\zeta}\Big|_{\zeta=\omega_m}\right\}}_{I_2}, \quad (49)$$

where we have defined I_1 and I_2, and taken the branch cut of the complex logarithm to be the negative real axis. The relation used for the matrix logarithm,

$$\frac{\partial}{\partial\omega}\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\} = \mathrm{Tr}\left\{\hat{Q}^{-1}(\omega)\frac{\partial}{\partial\omega}\hat{Q}(\omega)\right\}, \quad (50)$$

is valid for all $\hat{Q}(\omega)$ that are smooth and invertible but not necessarily diagonalizable; see Appendix B. The reason for introducing I_1 and I_2 is that at zero temperature I_2 is related to variations of the Φ functional, and is zero for number-conserving approximations. Furthermore, I_1 can be integrated analytically.

Equation (49) simplifies at zero temperature. We will first discuss I_2. For small temperature T, the distance between the Matsubara frequencies, $d\omega = 2\pi i/\beta$, becomes small. In the

limit $T \to 0$, a Matsubara sum can be written as an integral over the upper and lower half plane,

$$\lim_{T\to 0}\frac{1}{\beta}\sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\hat{Q}(\omega_m) = P\int_{-i\infty}^{i\infty}\frac{d\zeta}{2\pi i}\hat{Q}(\zeta), \tag{51}$$

where the principal value of the integral is $P\int_{-i\infty}^{i\infty} = \lim_{\epsilon\to 0^+}\int_{-i\infty}^{-i\epsilon} + \int_{i\epsilon}^{i\infty}$. Equation (51) is valid if $\hat{Q}(\omega_m)$ is not too singular around 0; see Appendix A and also the discussion in Ref. [55]. With this relation, we can write I_2 as $T \to 0$ as

$$I_2 = \int_{-i\infty}^{i\infty}\frac{d\zeta}{2\pi i}\mathrm{Tr}\left\{\hat{G}^M(\zeta)\frac{\partial\hat{\Sigma}^M}{\partial\zeta}\right\}. \tag{52}$$

In Eq. (52) we have made the assumption of integrability and replaced the principal value of the integral by the integral itself. We will here discuss these assumptions. $\hat{G}^M(\zeta)$ is analytic and thus integrable for $\mathrm{Im}\,\zeta \neq 0$, and has a discontinuity around $\zeta = 0$ given by the spectral function at μ, Eq. (25). Thus, assuming that $\hat{A}(\mu)$ is finite, $\hat{G}^M(\zeta)$ is integrable. By the same argument, $\hat{\Sigma}^M(\zeta)$ and $\frac{\partial\hat{\Sigma}^M}{\partial\zeta}$ are integrable for $\mathrm{Im}\,\zeta \neq 0$. From Eq. (27),

$$\left.\frac{\partial\hat{\Sigma}}{\partial\zeta}\right|_{\zeta=0} = -\int_{-\infty}^{\infty} d\omega'\frac{\hat{\Gamma}(\omega')}{(\mu-\omega')^2}, \tag{53}$$

from which we see that if $\hat{\Gamma}(\omega)$ decays quickly enough at μ, $\frac{\partial\hat{\Sigma}}{\partial\zeta}$ is continuous also at $\zeta = 0$. For Fermi liquids, $\hat{\Gamma}(\mu) \sim (\omega-\mu)^2$, for ω close to μ [56], and as such Eq. (53) is integrable. However, for other systems, such as the one-dimensional electron gas [57], or in Mott insulators [58], the rate operator can decay slower or even diverge at μ. Excluding such cases, the integral in Eq. (52) is well defined, and equal to its principal value.

I_2, also referred to as the Luttinger integral, can be related to the Φ functional under special types of variations in G, namely those that shift the Matsubara frequencies. As an example, consider the consistently dressed ring diagram in Fig. 1(b). $\Phi[G,G_0]$ is then given as

$$\begin{aligned}\Phi &= -\frac{i^2}{2}\int d1d2G(1,2)G(2,1)v(1,2)^2G_0(1,2)G_0(2,1)\\ &= -\frac{1}{2\beta^2}\sum_{nmrq}G^M(\omega_n)G^M(\omega_m)v^2G_0^M(\omega_r)G_0^M(\omega_q)\\ &\quad\times\delta_{0,n-m+r-q},\end{aligned} \tag{54}$$

where spatial integration is suppressed in the second line, to focus on the ω dependence. The Kronecker delta represents frequency conservation at every vertex, due to the time locality of the interaction. Φ evaluated at a G which has shifted frequencies, $G^M(\omega_m+\delta\omega)$, [defined via its analytical continuation, Eq. (23), as $G^M(\zeta)|_{\zeta=\omega_m+\delta\omega}$], would in general be different from the unshifted case. For the special case of a shift corresponding to one (or more) Matsubara frequency, i.e., $\delta\omega = \frac{2\pi i}{\beta}$ which gives $\omega_m + \delta\omega = \omega_{m+1}$, we can relabel the sum for m and n in Eq. (54). Since we perform the same relabeling for both m and n, the Kronecker delta does not change, which leaves Φ invariant. If the loop were not consistently dressed, Φ would change since the Kronecker delta would change. The argument works for a general diagram, and thus we see that if the loops are consistently dressed, $\delta\Phi = 0$ when $\omega_m \to \omega_{m+1}$ in G^M. Otherwise, $\delta\Phi \neq 0$ in general.

The reason why consistently dressed Φ diagrams are invariant under frequency shifts is due to the interaction being time local. The loops in Φ can then be considered separately. This type of argument cannot, however, be used for other types of conservation laws. For example, momentum conservation comes from the invariance of Φ under spatial translations in G [11], for interactions that depend only on the interparticle distance. For such a shift in G in Eq. (54), by a variable transform we only shift one index in the interaction, and as such Φ is not invariant. For a Φ diagram to be conserving, we need, in general, full Φ derivability.

Consider again the change in Φ when we make frequency shift in G. This is given by Eq. (16), which can be written using Matsubara Green's functions as

$$\delta\Phi = \sum_{m=-\infty}^{\infty} e^{\eta\omega_m}\mathrm{Tr}\{\hat{\Sigma}^M(\omega_m)\delta\hat{G}^M(\omega_m)\}. \tag{55}$$

Let us fix the shift to $\delta\omega = \frac{2\pi i}{\beta}$, and take the zero-temperature limit of Eq. (55). Here, we cannot use Eq. (51) directly, since the integrand is too singular around 0 (see Appendix A). The limit is

$$\begin{aligned}\delta\Phi \underset{T\to 0}{\to} P\int_{-i\infty}^{i\infty}\mathrm{Tr}\left\{\hat{\Sigma}^M(\omega)\frac{\partial\hat{G}^M}{\partial\omega}\right\}d\omega\\ + \lim_{\epsilon\to 0^+}\mathrm{Tr}\{\hat{\Sigma}^M(0)[\hat{G}^M(i\epsilon)-\hat{G}^M(-i\epsilon)]\}.\end{aligned} \tag{56}$$

By making use of partial integration, we obtain

$$\delta\Phi \underset{T\to 0}{\to} -\int_{-i\infty}^{i\infty} d\zeta\,\mathrm{Tr}\left\{\hat{G}^M(\zeta)\frac{\partial\hat{\Sigma}^M}{\partial\zeta}\right\} = -2\pi i I_2. \tag{57}$$

Thus, at zero temperature, we can relate I_2 to special types of variations of Φ. For a partially Φ-derivable approximation, $I_2 = 0$ if Φ is gauge invariant. Otherwise, $I_2 \neq 0$ in general.

We now derive expressions for I_1 and I_2 involving integrals over real frequencies at zero temperature. Using Eq. (47) we obtain I_2 as

$$I_2 = -2\,\mathrm{Im}\left[\int_{-\infty}^{\mu}\frac{d\omega}{2\pi}\mathrm{Tr}\left\{\hat{G}^R(\omega)\frac{\partial\hat{\Sigma}^R}{\partial\omega}\right\}\right]. \tag{58}$$

Equation (58) can also be derived from Eq. (57) by considering the contour integral around two quarter circles in the upper and lower half plane.

Using Eq. (47) for I_1, partial integration, and $\hat{\Sigma}^M(\pm i\eta) = \hat{\Sigma}^{R/A}(\mu)$, we obtain

$$N = \underbrace{\frac{1}{2\pi i}\mathrm{Tr}\{\ln(\hat{\Sigma}^A(\mu) + \hat{h} - \mu + i\eta) - \ln(\hat{\Sigma}^R(\mu) + \hat{h} - \mu - i\eta)\}}_{I_1} \underbrace{-2\,\mathrm{Im}\left[\int_{-\infty}^{\mu}\frac{d\omega}{2\pi}\mathrm{Tr}\left\{\hat{G}^R(\omega)\frac{\partial\hat{\Sigma}^R}{\partial\omega}\right\}\right]}_{I_2}. \tag{59}$$

The boundary terms for I_1 at $\omega \to -\infty$ give $\mathrm{Im}[\ln(\infty \pm i\eta)] = 0$, which explains our choice of branch and the factor of (-1) in Eq. (46).

In deriving Eq. (59) we have used the relation

$$\int_{\infty}^{\mu}\frac{\partial}{\partial\omega}\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\} = \mathrm{Tr}\{\ln[\hat{Q}(\mu)] - \ln[\hat{Q}(-\infty)]\}. \tag{60}$$

Here one has to assume that during the integration path, none of the eigenvalues $\lambda_k(\omega)$ of $\hat{Q}(\omega)$ cross the branch cut of the logarithm, where we regard $\lambda_k(\omega)$ as a curve in the complex plane parametrized by ω. This can be seen via

$$\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\} = \sum_k \ln[\lambda_k(\omega)], \tag{61}$$

valid for any complex invertible matrix; see Appendix B. If one or more of the eigenvalues traverse the branch cut at $(-\infty,0)$ as a function of ω, a unique logarithm cannot be found, and the argument principle (see, e.g., Ref. [59]) has to be used instead of Eq. (60). Crossing the branch cut changes the logarithm with $2\pi i$ if it is crossed from above, and $-2\pi i$ if it is crossed from below. However, if the rate operator is PSD no eigenvalue can cross the real axis, as we will now show.

A necessary condition for the crossing of the branch cut is $\mathrm{Im}[\lambda_k(\omega)] = 0$. Assume this holds for an eigenvalue $\tilde{\lambda}_k(\omega)$ of $-[\hat{G}^R(\omega)]^{-1}$. Consider the eigenvalue equation (ω-dependence suppressed) $(\hat{\Sigma}^R + \hat{h} - \omega - i\eta)|\tilde{\lambda}_k\rangle = \tilde{\lambda}_k|\tilde{\lambda}_k\rangle$, giving

$$\mathrm{Im}[\langle\tilde{\lambda}_k|\hat{\Sigma}^R|\tilde{\lambda}_k\rangle] - \eta = 0. \tag{62}$$

We write $\hat{\Sigma}^R(\omega) = \hat{\Lambda}(\omega) - \pi i\hat{\Gamma}(\omega)$, where $\hat{\Lambda}(\omega),\hat{\Gamma}(\omega)$ are Hermitian and the rate operator $\hat{\Gamma}(\omega)$ is PSD. We get

$$-2\pi\langle\tilde{\lambda}_k|\hat{\Gamma}|\tilde{\lambda}_k\rangle - \eta = 0. \tag{63}$$

Since $\langle\tilde{\lambda}_k|\hat{\Gamma}|\tilde{\lambda}_k\rangle \geqslant 0$, Eq. (63) cannot be satisfied. Thus, for a PSD approximation, no eigenvalue crosses the real axis, and Eq. (60) is valid.

Equation (59) is a sum rule valid for those approximations that give analytic $\hat{G}^M(\zeta)$ and $\hat{\Sigma}^M(\zeta)$. Furthermore, if the approximation comes from a gauge-invariant Φ, $I_2 = 0$. For example, the PSD approximations GW and GW_0 will fulfill Eq. (59) with $I_2 = 0$, while the PSD approximation G_0W_0 fulfills Eq. (59) with $I_2 \neq 0$ in general. We now give four different examples and limits of the sum rule.

A. Luttinger-Ward theorem

In the case of the homogeneous electron gas, Eq. (59) leads immediately to what is known as the Luttinger-Ward theorem [35]. See Ref. [57] for a more detailed discussion of this important sum rule in the electron gas. For homogeneous systems, the Hamiltonian and self-energy are diagonal in the momentum basis, and $\Sigma^R(\boldsymbol{p},\mu) = \Sigma^A(\boldsymbol{p},\mu)$. Assuming a number-conserving approximation, $I_2 = 0$ and using

$$\ln(x + i\eta) - \ln(x - i\eta) = 2i\pi\theta(-x), \tag{64}$$

where θ is the step function, we obtain the Luttinger-Ward theorem

$$N = 2V\int\frac{d\boldsymbol{p}}{(2\pi)^3}\theta(\mu - \epsilon_{\boldsymbol{p}} - \Sigma^R(\boldsymbol{p},\mu)), \tag{65}$$

where $\epsilon_{\boldsymbol{p}}$ are the single-particle energies, V is the volume, and the factor of 2 accounts for spin. Thus, the Luttinger-Ward theorem is valid for those number-conserving approximations which retain the correct analytical properties.

B. The Friedel sum rule

Equation (59) also allows for a useful relation in a completely different context, namely the Friedel sum rule in quantum transport. We consider a small region in space (e.g., a molecular region), where it is important to consider interaction effects carefully, coupled to macroscopic reservoirs (e.g., metallic leads) where a mean-field description is satisfactory. This implies that the frequency-dependent part of the self-energy is nonzero only in the central region.

The sum rule of Eq. (59) still applies for the total number of particles, but usually we are more interested in the properties of the central region. The trace Tr_C over the central region gives the number of particles N_C in that region. By the use of embedding techniques [60,61] the Green's function in the central region can be written as

$$\hat{G}^{R,A}(\omega) = \frac{1}{\omega - \hat{h}_C - \hat{\Sigma}^{R,A}_{emb}(\omega) - \hat{\Sigma}^{R,A}(\omega) \pm i\eta}, \tag{66}$$

where the embedding self-energy $\hat{\Sigma}_{emb}$ takes the environment into account in an exact way, and $\hat{h}_C$ is the Hamiltonian for the disconnected central region. The trace over the basis of the finite region gives the same type of sum rule as in Eq. (59), provided we replace $\hat{\Sigma}$ by $\hat{\Sigma} + \hat{\Sigma}_{emb}$:

$$N_C = \underbrace{\frac{1}{2\pi i}\mathrm{Tr}_C\left\{\ln\left[\hat{\Sigma}^A(\mu) + \hat{\Sigma}^A_{emb}(\mu) + \hat{h}_C - \mu + i\eta\right] - \ln\left[\hat{\Sigma}^R(\mu) + \hat{\Sigma}^R_{emb}(\mu) + \hat{h}_C - \mu - i\eta\right]\right\}}_{I_1} + I_2. \tag{67}$$

$\hat{\Sigma}^R_{emb}(\omega)$ enters I_2, which now has the form

$$I_2 = -2\,\mathrm{Im}\left[\int_{-\infty}^{\mu}\frac{d\omega}{2\pi}\mathrm{Tr}_C\left\{\hat{G}^R\frac{\partial}{\partial\omega}\left(\hat{\Sigma}^R+\hat{\Sigma}^R_{emb}\right)\right\}\right]. \quad (68)$$

We can thus split the contributions as $I_2 = I_{2,MB} + I_{2,emb}$. Now, $I_2 \neq 0$ in general, even for noninteracting systems. For a number-conserving scheme, $I_{2,MB} = 0$.

I_1 can be written in terms of the eigenvalues $\lambda_k(\omega)$ of $-\hat{G}^R(\omega)$. Since we will make use of this, we write out the form explicitly,

$$\begin{aligned} I_1 &= \frac{1}{2\pi i}\mathrm{Tr}_C\{\ln[-[\hat{G}^A(\mu)]^{-1}] - \ln[-[\hat{G}^R(\mu)]^{-1}]\} \\ &= \frac{1}{2\pi i}\mathrm{Tr}_C\{\ln[-\hat{G}^R(\mu)] - \ln[-\hat{G}^A(\mu)]\} \\ &= \frac{1}{\pi}\,\mathrm{Im}\sum_k \ln[\lambda_k(\mu)]. \end{aligned} \quad (69)$$

To recover the usual formulation of the Friedel sum rule, we take our central system to consist of a single interacting site with energy ϵ, and use the wide-band limit approximation, $\Sigma^{R/A}_{emb}(\omega) = \mp i\Gamma$. Equation (67) becomes

$$N_C = \frac{1}{2\pi i}(\ln[\tilde{\Sigma}^R + i\Gamma] - \ln[\tilde{\Sigma}^R - i\Gamma]) + I_{2,MB},$$

where $\tilde{\Sigma}^R = \Sigma^R(\mu) + \epsilon - \mu$. Using

$$\ln(x+iy) - \ln(x-iy) = \pi - 2\arctan\left(\frac{x}{y}\right), \quad (70)$$

we obtain

$$N_C = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{\Sigma^R(\mu)+\epsilon-\mu}{\Gamma} + I_{2,MB}. \quad (71)$$

By using a gauge-independent Φ, $I_{2,MB} = 0$, and we obtain the well-known Friedel sum rule [36,37]. Thus, we can regard Eq. (67) as a generalized Friedel sum rule.

C. Analytical properties in the Anderson model

In this section, we show explicitly, for the single-impurity Anderson model in the zero-temperature limit, that even conserving approximations can have spectral functions that can become negative. The Anderson model is a simple model in quantum transport, which consists of a single interacting dot in contact to a featureless reservoir with coupling strength Γ. The Hamiltonian of the dot is

$$\hat{H} = \epsilon\sum_\sigma \hat{n}_\sigma + \frac{U}{2}\sum_{\sigma\sigma'}\hat{c}^\dagger_\sigma\hat{c}^\dagger_{\sigma'}\hat{c}_{\sigma'}\hat{c}_\sigma, \quad (72)$$

where $\hat{n}_\sigma$ is the number operator for spin σ and U is the interaction strength. Furthermore, we consider nonmagnetic situations, and as such $G^R_{\sigma\sigma'}(\omega) = \delta_{\sigma\sigma'}G^R(\omega)$. For the Anderson model, $G^R(\omega)$ is

$$G^R(\omega) = \frac{1}{\omega - \epsilon - \Sigma^R(\omega) + i\Gamma}, \quad (73)$$

where Γ is the coupling to the reservoir. As our approximation to $\Sigma^R(\omega)$, we take various 2nd-order diagrams: the single-shot 2nd Born approximation Fig. 1(a) + Fig. 2(a), the single-shot 2nd-order exchange approximation Fig. 2(a), and the self-consistent 2nd-order exchange approximation Fig. 2(b). The 2nd Born approximation is PSD, while the two 2nd-order exchange approximations are not. We write here the explicit expression for $\hat{\Sigma}(z_1,z_2)$ for a general system with local interactions, since we will make use of larger systems in Sec. IV D:

$$\begin{aligned} \Sigma_{kl}(z_1,z_2) &= \delta_{kl}\delta(z_1,z_2)Un_k(z_1) \\ &\quad + c\,U^2G_{kl}(z_1,z_2)G_{lk}(z_2,z_1)G_{kl}(z_1,z_2), \end{aligned} \quad (74)$$

where the indices k,l label the interacting sites. For the Anderson model, $k = l = 1$. The first term is the Hartree-Fock contribution, with $n_k(z_1) = -iG_{kk}(z_1,z_1^+)$. The coefficient c in Eq. (74) is $c = -1$ for the 2nd-order exchange diagram, Fig. 2, $c = 2$ for the ring diagram, Fig. 1, and $c = 1$ for 2nd Born. This simple structure is due to the interaction being space local. Furthermore, since 2nd Born is a PSD approximation, we have immediately that the rate operator of 2nd-order exchange is negative-semidefinite.

The explicit expression for $\hat{\Sigma}^R_{kl}(\omega)$ in equilibrium is obtained from the Langreth rules [36], and Fourier transforming (for more details, see Ref. [62]):

$$\begin{aligned} \Sigma^R_{kl}(\omega) &= \delta_{kl}Un_k + cU^2\iint\frac{d\omega' d\omega''}{(2\pi)^2} \\ &\quad\times\big[G^R_{kl}(\omega')G^<_{lk}(\omega'')G^>_{kl}(\omega-\omega'+\omega'') \\ &\quad + G^<_{kl}(\omega')G^A_{lk}(\omega'')G^<_{kl}(\omega-\omega'+\omega'') \\ &\quad + G^<_{kl}(\omega')G^<_{lk}(\omega'')G^R_{kl}(\omega-\omega'+\omega'')\big], \end{aligned}$$

where the lesser and greater Green's functions in equilibrium are given by the fluctuation-dissipation theorem [1]

$$G^<_{kl}(\omega) = 2\pi i f(\omega)A_{kl}(\omega), \quad (75)$$

$$G^>_{kl}(\omega) = -2\pi i[1-f(\omega)]A_{kl}(\omega). \quad (76)$$

For single-shot 2nd-order exchange, $\Sigma^R(\omega)$ was evaluated using a Green's function from a self-consistent Hartree-Fock calculation. The self-consistent 2nd-order exchange calculations did not converge if we started from the Hartree-Fock Green's functions, however (see below). Instead, as in Ref. [63], we performed self-consistency with smaller values of U, and then starting new calculations with this initial guess.

In Fig. 6, we show $A(\omega)$ for the three approximations. As expected, the 2nd-order exchange approximation yields spectral functions which can become negative. In the one-shot case, one can by increasing U make the spectral function negative in a large region. $\int d\omega A(\omega) = -0.2$ for one-shot 2nd-order exchange, indicating a pole in the upper half plane of $G^R(\zeta)$. For the other approximations, $\int d\omega A(\omega) = 1$. Similarly to the case of α in Sec. III B, we find a crossover (not shown) with respect to U above which a pole moves into the upper half plane. Note also that since the particle number is $N_C = \int_{-\infty}^{\mu} A(\omega)$, it can become negative. However, the particle number defined via the spectral function can be unphysical due to the pole structure; see the discussion of the next section.

We now discuss the Friedel sum rule, Eq. (67), for the Anderson model. The single-shot 2nd Born is PSD and thus fulfills the sum rule, but since the approximation is not particle

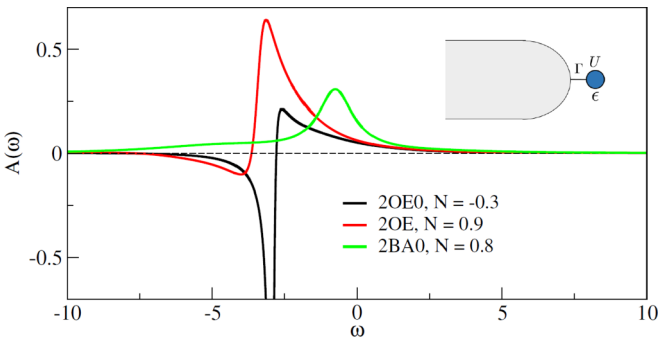

FIG. 6. Spectral functions for the Anderson model (single impurity coupled to a featureless lead). 2OE0: Single-shot 2nd-order exchange diagram, Fig. 2(a). 2OE: Self-consistent 2nd-order exchange, Fig. 2(b). 2BA0: Single-shot 2nd Born, Fig. 1(a) + Fig. 2(a). 2OE0 and 2OE are non-PSD. 2nd Born, however, is PSD. Compare with Fig. 7. The parameters are $U = 6.5, \epsilon = -7$, and $\Gamma = 1$.

conserving, $I_2 \neq 0$. The 2nd-order exchange approximations are not PSD, and are thus not guaranteed to fulfill the sum rule. For our parameters, we found that self-consistent 2OE fulfilled the sum rule with $I_2 = 0$, while for single-shot 2OE, we find that $N_C - I_1 - I_2 = -2$.

To understand the term -2, we show in Fig. 7 $-G^R(\omega)$ as parametrized by ω from $\omega = -\infty$ to $\omega = \mu = 0$. As can be seen, $-G^R(\omega)$ from 2nd Born never crosses the real axis, which is guaranteed by its PSD property. However, $-G^R(\omega)$ obtained from single-shot 2OE crosses the negative real axis, and thus breaks the assumptions in the derivations of the sum rules. Since we change Riemann sheet, but still use the principal branch of the logarithm, we miss an additional term of $2\pi i$. This translates into a term -2 in I_1, as can be seen from Eq. (69). The self-consistent 2OE also cross the real axis, but not at the branch cut. In fact, if we push the parameters further (increase U), analyticity breaks during the self-consistent cycle, and we are unable to reach

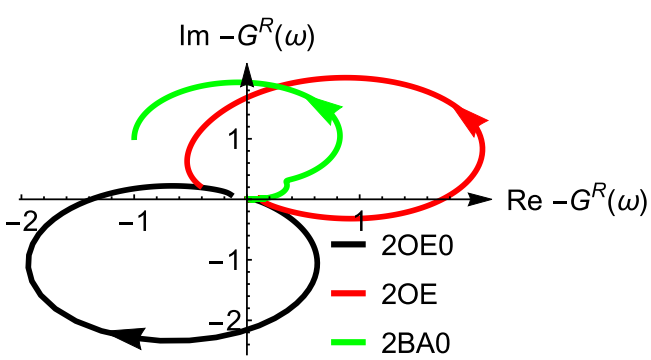

FIG. 7. $-G^R(\omega)$ parametrized with $\omega \in (-\infty, \mu]$, starting from $\omega = -\infty$ where $G^R(-\infty) = 0$, and ending at $\mu = 0$, for the same situation as in Fig. 6. The 2nd-order exchange approximations are not PSD, and thus $-G^R(\omega)$ can cross the real axis. Single-shot 2OE breaks analyticity as well as the validity of the sum rules. Self-consistent 2OE does not. The single-shot 2nd Born approximation is PSD, and thus never crosses the real axis. For illustrative purposes, we have divided G^R from 2OE0 by 3 and multiplied by 2 for 2BA0.

convergence. This again underscores the importance of PSD approximations in self-consistent approximations.

Finally, we note that using only the second-order exchange diagram can seem like a constructed example. However, the same considerations apply if the diagram is dressed with the screened interaction W. The second-order diagram would then correspond to vertex corrections from GW. This approximation can yield a non-PSD spectral function [7], and here we show that the same considerations apply in a model in quantum transport.

D. Example of the generalized sum rule

In the last example, we consider a 3-site linear chain at zero temperature to elucidate the generalized Friedel sum rule, Eq. (67), in a finite region. The Hamiltonian is

$$\hat{H} = \sum_{ij\sigma} h_{ij} \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} + \frac{U}{2} \sum_{i\sigma\sigma'} \hat{c}^\dagger_{i\sigma} \hat{c}^\dagger_{i\sigma'} \hat{c}_{i\sigma'} \hat{c}_{i\sigma}, \tag{77}$$

where $h_{ii} = \epsilon$, $h_{ij} = -1$ for nearest neighbors, and zero otherwise. As contacts, we choose two one-dimensional tight-binding leads with bandwidths of 4, attached to the first and third site, respectively, with coupling strength -1. The 3×3 Green's function matrix for the central region is

$$\hat{G}^R(\omega) = \frac{1}{\omega - \hat{h}_C - \hat{\Sigma}^R(\omega) - \hat{\Sigma}^R_{emb}(\omega) + i\eta}, \tag{78}$$

where the embedding self-energy matrix $\hat{\Sigma}^R_{emb}(\omega)$ corresponding to the tight-binding leads, nonzero only for the first and third site, has an explicit expression; see Refs. [60–62].

We considered four different approximations to $\hat{\Sigma}^R(\omega)$, with various amounts of dressing: the self-consistent ring diagram Fig. 1(d), the partially self-consistent ring diagram Fig. 1(b), the single-shot ring diagram Fig. 1(a), and the single-shot 2nd-order exchange diagram Fig. 2. Of these, the first three approximations are PSD, and as such the total number of particles in the central region N_C is given by the generalized Friedel sum rule, $N_C = I_1 + I_2$, where I_1 can be conveniently written in terms of the eigenvalues $\lambda_k(\omega)$ to $-\hat{G}^R(\omega)$, using Eq. (69),

$$I_1 = \frac{1}{\pi} \sum_{k=1}^{3} \ln[\lambda_k(\mu)], \tag{79}$$

and $I_2 = I_{2,MB} + I_{2,emb}$ is given by Eq. (68). As discussed before, for the fully self-consistent and the consistently dressed partially self-consistent approximations $I_{2,MB} = 0$, while for the single-shot cases $I_{2,MB} \neq 0$, in general. To illustrate the sum rule, we also calculate N_C according to $N_C = \int_{-\infty}^{0} d\omega \mathrm{Tr}\{\hat{A}(\omega)\}$.

To explore a large parameter range, we sweep with the gate voltage ϵ, symmetrically from $\epsilon = -U/2$. N_C, I_1, $I_{2,emb}$, and $I_{2,MB}$ are shown in Fig. 8, for all the different approximations. We note that the correction from $I_{2,emb}$ is small, but becomes larger as we move away from $\epsilon = -U/2$. Note also that the contribution from $I_{2,emb}$ would increase with more sites connected to leads and lead coupling strength. We also see, as expected, that $I_{2,MB} = 0$ for the consistently dressed approximations. For the single-shot case, $I_{2,MB}$ can be very

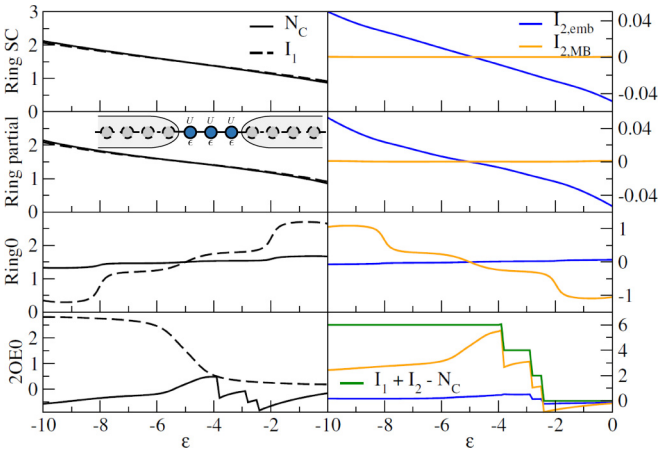

FIG. 8. Number of particles N_C in a 3-site system, for different dressings of ring approximations and single-shot 2nd-order exchange. We show the self-consistent ring approximation [Fig. 1(d)], partially dressed ring diagram [Fig. 1(b)], single-shot ring diagram [Fig. 1(a)], and single-shot 2nd-order exchange [Fig. 2(a)]. $I_{2,emb}$ and $I_{2,MB}$ are the embedding and many-body contributions to the Luttinger integral, Eq. (68), respectively. Note the different scale for I_2 in the non-self-consistent calculations.

large. At $\epsilon = -U/2$, $I_{2,MB} = 0$, but is large away from this point. We plot I_1, which is the resulting particle number one would obtain if I_2 were not taken into account. The magnitude of $I_{2,MB}$ makes N_C and I_1 very different, and in fact I_1 increases dramatically when raising the gate potential. Not taking $I_{2,MB}$ into account gave a similar erroneous trend for the Anderson model using single-shot 2nd Born in Ref. [64], as was also pointed out by Ref. [63]. We also note that in the single-shot ring approximation, N_C increases slightly as ϵ increases. This nonphysical behavior is due to only taking the ring diagram into account. Single-shot 2nd Born (i.e., the ring diagram and the 2nd-order exchange diagram) does not exhibit this nonphysical behavior (not shown).

Finally, in the non-PSD 2nd-order exchange approximation N_C behaves nonphysically and can even become negative. The norm of the spectral function (not shown) also becomes negative for certain parameter values, similarly to the case of the Anderson model. The change of norm indicates poles in the upper half plane for $\hat{G}^R(\zeta)$, according to Eq. (31). We also note that the particle number we calculate is $N_A = \int_{-\infty}^{\mu} d\omega \mathrm{Tr}\{\hat{A}(\omega)\}$. According to the discussion we had above, when there are poles in the upper half plane for $\hat{G}^R(\zeta)$, with residue $a_k^R \neq 0$ and location ξ_k^R, N_A does not have to equal $N = -i\mathrm{Tr}\{\hat{G}^M(\tau,\tau^+)\}$, since we have to add the residues in Eq. (48). For simplicity, we discuss the situation in a single-site case. From Eq. (31), we have in the single-site case that

$$1 = \int_{-\infty}^{\infty} d\omega A(\omega) + 2\,\mathrm{Re}\,a^R. \tag{80}$$

At zero temperature, a pole can only contribute to the particle number if it is located to the left of μ,

$$N = N_A + \theta(\mu - \mathrm{Re}\,\xi^R)2\,\mathrm{Re}\,a^R. \tag{81}$$

Assuming that the jump in $N_A(\epsilon)$ occurs at the same time as when $\mathrm{Re}\,\xi^R(\epsilon)$ equals μ, we found that the jump was canceled by the contribution from the residue, calculated from the norm in Eq. (80). This suggests that N is a more physical quantity than N_A, in general.

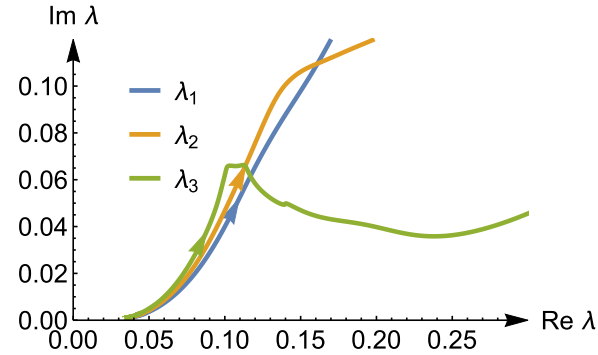

FIG. 9. The three eigenvalues of $-\hat{G}^R(\omega)$ parametrized by ω for the system shown in Fig. 8, for the fully self-consistent ring approximation, Fig. 1(d). ω runs from $-\infty$ to $\omega = \mu = 0$. The parameters are $U = 10$, $\epsilon = -U/2$. Being a PSD approximation, no eigenvalue can cross the real axis. Numerically we find $N_C - I_1 - I_2 = 0$.

Numerically, we find $N_C - I_1 - I_2 = 0$ to a high degree of accuracy for all the ring approximations. For the non-PSD 2nd-order exchange approximation, we find instead that $N_C - I_1 - I_2$ varies, depending on the parameters. This can be understood by studying how the eigenvalues of $-\hat{G}^R(\omega)$ change as ω goes from $-\infty$ to μ. To illustrate, we track $\lambda_k(\omega)$ for the PSD self-consistent ring approximation, shown in Fig. 9, and the non-PSD single-shot 2nd-order exchange approximation, shown in Fig. 10.

The eigenvalues $\lambda_k(\omega)$ of the PSD ring approximation in Fig. 9 do not cross the real axis when we vary ω, as expected. The eigenvalues of the non-PSD approximation in Fig. 10 do, however. For $\epsilon = -U/2$, all three eigenvalues cross the branch cut. This means that, identically to the discussion we had above, for each crossing eigenvalue we obtain an additional factor of 2 from I_1, which explains the result $N_C - I_1 - I_2 = -6$ in Fig. 8 for 2OE0. Depending on ϵ,

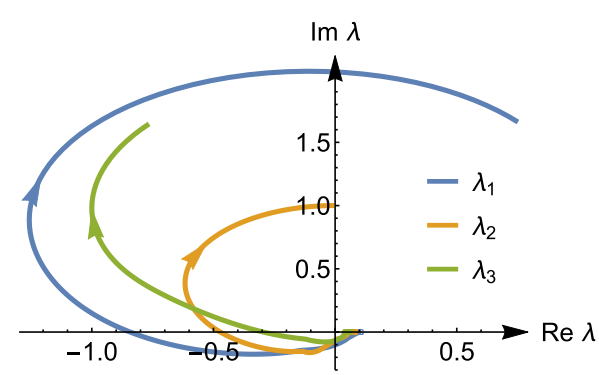

FIG. 10. The three eigenvalues of $-\hat{G}^R(\omega)$ parametrized by ω for the system shown in Fig. 8, for the single-shot 2nd-order exchange, Fig. 2(a). ω runs from $-\infty$ to $\omega = \mu = 0$. The parameters are $U = 10$, $\epsilon = -U/2$. Being a non-PSD approximation, eigenvalues can, and do in this case, cross the real axis. Numerically, we find $N_C - I_1 - I_2 = -6$, since each eigenvalue yields a factor of 2 to I_1.

three, two, one, or no eigenvalues can cross the branch cut, the number of net crossings being $(I_1 + I_2 - N_C)/2$. As a side remark, we noted that eigenvalues could become degenerate for some frequencies (not shown).

V. CONCLUSIONS

We have extended the notion of Φ-derivable schemes to include partially self-consistent and non-self-consistent approximations. For such partially Φ-derivable schemes, we have shown that by dressing the loops in Φ consistently, a gauge-invariant approximation is obtained. This implies a fulfillment of the continuity equation at all times, which in turn implies number conservation. By making use of the nonequilibrium Green's function formalism, the resulting partially Φ-derivable scheme can be applied in equilibrium as well as out of equilibrium, in finite and infinite systems. In equilibrium, partial Φ derivability allows for deriving a generalized sum rule for the particle number, where we have stressed the importance of the correct analytical properties of the Green's functions. If the approximation is not gauge invariant, as is the case for many approximations used in the literature, we obtain another term I_2 in the sum rule. The rule was also applied to systems which allow for a partitioning into smaller subsystems, allowing for a generalized Friedel sum rule. The frequency-dependent embedding self-energies need to be taken into account in I_2.

We have elucidated the known fact that PSD approximations automatically yield correct analytical properties for $G^R(\omega)$. The converse is not true, which we have exemplified in several systems; a spectral function that can become negative can still come from a $G^R(\omega)$ that has the correct analytical properties. We have proven that the validity of the Luttinger-Ward theorem and the Friedel sum rule depend crucially on the correct analytical properties of the Green's function. These considerations show that the simplest vertex corrections to GW can lead to a violation of common sum rules. Extra diagrams need to be added to obtain a PSD spectral function, thus guaranteeing the satisfaction of the sum rules.

While a given approximation can be shown to be number conserving without making use of partial Φ derivability, we believe that this formalism is convenient and has several advantages. It is easy to see whether an approximation will be particle conserving or not, and also if the corresponding sum rules will be fulfilled. Combined with the knowledge of when the approximation has the correct analytical properties, these considerations could serve as a helpful tool in determining which diagrams to choose in a given situation. This could especially be important in, for example, diagrammatic Monte Carlo integrations, where a large number of diagrams can be summed up [7–9].

ACKNOWLEDGMENT

D.K. and R.v.L. would like to thank the Academy of Finland for support under Project No. 267839.

APPENDIX A

In this Appendix, we give more details about the zero-temperature limit of $\delta\Phi$, Eq. (56). We suppress the superscript M for notational convenience. Shifting the Matsubara Green's function with $\delta\omega$, the variation is

$$\delta\hat{G}(\omega_m) = \frac{\partial\hat{G}}{\partial\omega_m}\delta\omega \tag{A1}$$

if $\delta\omega$ is small enough. From considering the spectral representation of $\hat{G}(\omega_m)$, Eq. (23), the requirement is that $\delta\omega < |\omega_m|$. However, in relating I_2 to a variation in Φ, we restricted the frequency shift to $\delta\omega = \frac{2\pi i}{\beta}$, meaning that $\omega_m + \delta\omega = \omega_{m+1}$. Thus, $\delta\omega$ is larger than the Matsubara frequencies $\omega_0 = \frac{\pi i}{\beta}$ and $\omega_{-1} = -\frac{\pi i}{\beta}$.

Moreover, for low temperatures, the variation $\delta\hat{G}(\omega_{-1}) = \hat{G}(\omega_0) - \hat{G}(\omega_{-1})$ yields a finite contribution as $T \to 0$, since ω_0 and ω_{-1} are on different sides of the branch cut at $\mathrm{Im}\,\zeta = 0$. In fact, $\delta\hat{G}(\omega_{-1}) \to \lim_{\epsilon\to 0^+}[\hat{G}(i\epsilon) - \hat{G}(-i\epsilon)] = -2\pi i\hat{A}(\mu)$.

To perform the zero-temperature limit, we thus separate the $m = -1$ and $m = 0$ Matsubara points:

$$\begin{aligned}\sum_{m=-\infty}^{\infty} & e^{\eta\omega_m}\mathrm{Tr}\{\hat{\Sigma}(\omega_m)\delta\hat{G}(\omega_m)\} \\ &= \sum_{\substack{m=-\infty \\ m\neq -1, m\neq 0}}^{\infty} e^{\eta\omega_m}\mathrm{Tr}\left\{\hat{\Sigma}(\omega_m)\frac{\partial\hat{G}}{\partial\omega_m}\right\}\delta\omega \\ &\quad + e^{\eta\omega_0}\mathrm{Tr}\{\hat{\Sigma}(\omega_0)\delta\hat{G}(\omega_0)\} + e^{\eta\omega_{-1}}\mathrm{Tr}\{\hat{\Sigma}(\omega_{-1})\delta\hat{G}(\omega_{-1})\}.\end{aligned} \tag{A2}$$

The sum is now well defined in the $T \to 0$ limit, and we can use Eq. (51), yielding a principal-value integral. The other terms give $\delta\hat{G}(\omega_0) = \hat{G}(\omega_1) - \hat{G}(\omega_0) \to 0$ and $\delta\hat{G}(\omega_{-1}) \to \lim_{\epsilon\to 0^+}[\hat{G}(i\epsilon) - \hat{G}(-i\epsilon)]$. This gives

$$\begin{aligned}\sum_{m=-\infty}^{\infty} & e^{\eta\omega_m}\mathrm{Tr}\{\hat{\Sigma}(\omega_m)\delta\hat{G}(\omega_m)\} \\ &\to P\int_{-i\infty}^{i\infty}\mathrm{Tr}\left\{\hat{\Sigma}(\zeta)\frac{\partial\hat{G}}{\partial\zeta}\right\}d\zeta \\ &\quad + \mathrm{Tr}\{\hat{\Sigma}(0)[\hat{G}(i\epsilon) - \hat{G}(-i\epsilon)]\}.\end{aligned} \tag{A3}$$

Thus, we have derived Eq. (56).

APPENDIX B

We made use of several properties of matrix logarithms in the derivations of the sum rules, which will be more explained here. The key point is that matrix functions are not defined according to their Taylor expansions, since this assumes a radius of convergence. However, we cannot use a spectral representation, since Green's functions are not Hermitian, and as such not guaranteed to be diagonalizable.

We assume that the Green's functions can be written as $L \times L$ matrices. We also assume that the matrices are invertible, which they have to be in order to define a matrix logarithm. We make use of the fact that a general complex matrix can be written in its Jordan normal form, $\hat{Q} = \hat{Z}\hat{J}\hat{Z}^{-1}$, where $\hat{J}$ is block diagonal with M blocks $\hat{J}_i$ corresponding to nondegenerate eigenvalues. Each $m_i \times m_i$ block has the

eigenvalue λ_i of $\hat{Q}$ on the diagonal, with multiplicity m_i. The superdiagonal of each block is equal to 1:

$$\hat{J} = \begin{bmatrix} \hat{J}_1 & & \\ & \ddots & \\ & & \hat{J}_M \end{bmatrix}, \quad \hat{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}. \tag{B1}$$

Matrix functions can then be defined as (see, for example, Ref. [65])

$$f(\hat{Q}) = \hat{Z} f(\hat{J}) \hat{Z}^{-1} = \hat{Z}\mathrm{diag}[f(\hat{J}_i)]\hat{Z}^{-1}, \tag{B2}$$

where

$$f(\hat{J}_i) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!} \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}. \tag{B3}$$

This definition reduces to the usual spectral representation if $\hat{Q}$ is diagonalizable. For matrix logarithms, Eq. (B3) can also be derived by writing $\hat{J}_i = \lambda_i(\mathbb{1} + \hat{J})$, where $\hat{J}$ is lower triangular with 0 on the diagonal. $\ln \hat{J}_i = \ln(\lambda_i)\mathbb{1} + \ln(\mathbb{1} + \hat{J})$ since all eigenvalues of $\hat{J}$ are real [65]. Furthermore, $\ln(\mathbb{1} + \hat{J}) = \sum_{k=1}^{m_i-1}(-1)^{k+1}\frac{\hat{J}^k}{k}$ terminates, since $\hat{J}$ is strictly triangular and thus $\hat{J}^{m_i} = 0$.

Equation (B3) immediately yields Eq. (61),

$$\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\} = \sum_k \ln[\lambda_k(\omega)], \tag{B4}$$

where it is explicit that $\hat{Q}$ must be invertible for the matrix logarithm to exist, since then all $\lambda_k(\omega) \neq 0$.

If we furthermore assume that $\hat{Q}(\omega)$ and the transformation matrices are smooth functions of ω we obtain Eq. (50) as

$$\begin{aligned} &\mathrm{Tr}\left\{\hat{Q}^{-1}(\omega)\frac{\partial}{\partial\omega}\hat{Q}(\omega)\right\} \\ &= \mathrm{Tr}\{J^{-1}(\omega)J'(\omega)\} \\ &= \sum_k \frac{\lambda_k'(\omega)}{\lambda_k(\omega)} = \frac{\partial}{\partial\omega}\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\}. \end{aligned} \tag{B5}$$

These relations can be combined to derive Eq. (60) as

$$\begin{aligned} &\int_a^b \frac{\partial}{\partial\omega}\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\}d\omega \\ &= \sum_k \int_a^b \frac{\lambda_k'(\omega)}{\lambda_k(\omega)}d\omega = \sum_k \int_a^b d\ln[\lambda_k(\omega)]. \end{aligned} \tag{B6}$$

The integrals on the right-hand side are called logarithmic integrals [59]. If the $\lambda_k(\omega)$'s do not cross the branch cut at $(-\infty,0)$, unique primitives can be found for each eigenvalue, yielding Eq. (60):

$$\int_a^b \frac{\partial}{\partial\omega}\mathrm{Tr}\{\ln[\hat{Q}(\omega)]\}d\omega = \mathrm{Tr}\{\ln[\hat{Q}(b)] - \ln[\hat{Q}(a)]\}. \tag{B7}$$

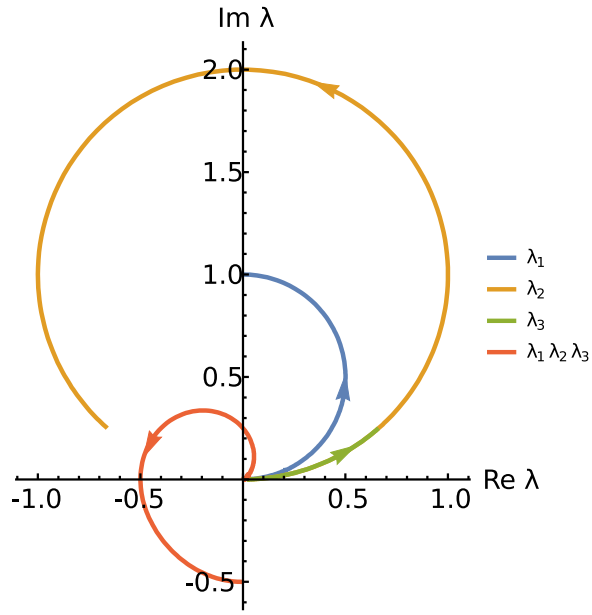

FIG. 11. The eigenvalues $\lambda_k(\omega)$ of $-\hat{G}^R(\omega)$ for the 3-site system in Sec. IV D, with $U = \epsilon = 0$ and using the wide-band limit for simplicity. The curves start at $\omega = -\infty$, for which $\lambda_i(-\infty) = 0$, and end at $\mu = 0$. Neither of the eigenvalues cross the real axis, while their multiplication does.

As a side remark, we note that we can write a logarithmic integral using determinants,

$$\begin{aligned} &\int_a^b \mathrm{Tr}\left\{\hat{Q}^{-1}(\omega)\frac{\partial}{\partial\omega}\hat{Q}(\omega)\right\} \\ &= \int_a^b \frac{\det\{\hat{Q}(\omega)\}'}{\det\{\hat{Q}(\omega)\}}d\omega = \int_a^b d\ln[\det\{\hat{Q}(\omega)\}] \\ &= \int_a^b d\ln[\Pi_k\lambda_k(\omega)]. \end{aligned} \tag{B8}$$

In this case, however, it is not possible to find a unique primitive in general, since the product of $\lambda_i(\omega)$ can cross the branch cut even if the individual $\lambda_i(\omega)$'s do not. This can be easily seen in the case of a noninteracting 3-site model. We consider the model in Sec. IV D, and put $U = 0$, $\epsilon = 0$ and use the wide-band limit approximation. The eigenvalues of $-\hat{G}^R(\omega)$ are $\lambda_1 = -\frac{1}{\omega+i\Gamma}$ and $\lambda_{2,3} = -\frac{2}{(2\omega+i\Gamma\pm\sqrt{8-\Gamma^2})}$, neither of which cross the real axis. The multiplication of them $\lambda_1\lambda_2\lambda_3 = -\frac{1}{(\omega+i\Gamma)(\omega^2+i\Gamma\omega-2)}$ does cross the real axis, at $\omega = \pm 1$. Thus, if $\mu > -1$, we cannot find a unique primitive logarithm in Eq. (B8). The situation is depicted in Fig. 11.

APPENDIX C

In Sec. III, we made use of the spectral representation for $\hat{G}^M(\zeta)$, Eq. (23), and $\hat{\Sigma}^M(\zeta)$, Eq. (27). Using the Lehmann representation, one can show that the exact $\hat{G}^M(\zeta)$ and $\hat{\Sigma}^M(\zeta)$ can be written in this form. In this appendix, we will show that any analytic operator with certain boundedness properties at infinity can be written in a spectral form. Furthermore, we will discuss how the spectral representation changes if the functions are nonanalytic. We first consider scalar functions, and then generalize to operators.

Suppose that we are given a function $f^M(\zeta)$, analytic except for single poles away from the real axis and decaying to zero at infinity, i.e., $f(\zeta) \to 0$ for $|\zeta| \to \infty$. The poles have residue a_k located at ξ_k in the upper and lower half plane. We define functions $f^R(\zeta)$ and $f^A(\zeta)$ such that

$$f^M(\zeta) = \begin{cases} f^R(\zeta), & \operatorname{Im}\zeta > 0, \\ f^A(\zeta), & \operatorname{Im}\zeta < 0. \end{cases} \tag{C1}$$

Consider the integral along a closed half circle in the upper half plane. From the residue theorem, for $\operatorname{Im}\zeta > 0$, we get

$$\frac{i}{2\pi}\oint d\xi\, \frac{f^R(\xi)}{\zeta - \xi} = f^R(\zeta) - \sum_k \frac{a_k^R}{\zeta - \xi_k^R}, \tag{C2}$$

where ξ_k^R is the kth pole of $f^M(\zeta)$ located in the upper half plane, with residue $a_k^R = \lim_{\xi\to\xi_k^R}(\xi - \xi_k^R)f(\xi)$. Since $f(\zeta)$ is bounded at infinity, the integral on the half circle vanishes. Thus, we can write, for $\operatorname{Im}\zeta > 0$,

$$f^R(\zeta) = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{f^R(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^R}{\zeta - \xi_k^R}, \tag{C3}$$

where we have analytically continued $f^R(\omega) = \lim_{\eta\to 0} f^R(\omega + i\eta)$ in order to perform the integration on the real axis. In the same way, for $\zeta < 0$,

$$f^A(\zeta) = -\frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{f^A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^A}{\zeta - \xi_k^A}, \tag{C4}$$

where we also analytically continued $f^A(\zeta)$ to the real axis. Consider now the function

$$f^M(\zeta) = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{f^R(\omega) - f^A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^R}{\zeta - \xi_k^R} + \sum_k \frac{a_k^A}{\zeta - \xi_k^A}. \tag{C5}$$

This function satisfies Eq. (C1), which can be seen by considering $f^M(\zeta) - f^{R/A}(\zeta)$, for $\operatorname{Im}\zeta > 0$ and $\operatorname{Im}\zeta < 0$, respectively. For $\operatorname{Im}\zeta > 0$, for example,

$$f^M(\zeta) - f^R(\zeta) = -\frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{f^A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^A}{\zeta - \xi_k^A} \tag{C6}$$

$$= \frac{i}{2\pi}\oint d\omega \frac{f^A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^A}{\zeta - \xi_k^A} = 0, \tag{C7}$$

where we have closed the contour in the lower half plane.

Defining the spectral function $A(\omega) = \frac{i}{2\pi}[f^R(\omega) - f^A(\omega)]$, we obtain

$$f^M(\zeta) = \int_{-\infty}^{\infty} d\omega \frac{A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^R}{\zeta - \xi_k^R} + \sum_k \frac{a_k^A}{\zeta - \xi_k^A}. \tag{C8}$$

If we add the restriction that $[f^M(\zeta^*)]^* = f^M(\zeta)$, we have that

$$\int_{-\infty}^{\infty} d\omega \frac{A^* - A}{\zeta - \omega} + \sum_k \frac{(a_k^R)^*}{\zeta - (\xi_k^R)^*} - \frac{a_k^R}{\zeta - \xi_k^R} + \sum_k \frac{(a_k^A)^*}{\zeta - (\xi_k^A)^*} - \sum_k \frac{a_k^A}{\zeta - \xi_k^A} = 0. \tag{C9}$$

By integrating around a small circle around a chosen ξ_m^R, which is in the upper half plane, we obtain that

$$2\pi i a_m^R = \sum_k \oint_{\xi_m^R} d\zeta \frac{(a_k^A)^*}{\zeta - (\xi_k^A)^*}. \tag{C10}$$

Since $a_m^R \neq 0$, we must have that $a_m^R = (a_m^A)^*$ and $\xi_m^R = (\xi_m^A)^*$. From Eq. (C9) it then immediately follows that $A(\omega) = A^*(\omega)$. We can then write $f^M(\zeta)$, Eq. (C8), as

$$f^M(\zeta) = \int_{-\infty}^{\infty} d\omega \frac{A(\omega)}{\zeta - \omega} + \sum_k \frac{a_k^R}{\zeta - \xi_k^R} + \sum_k \frac{(a_k^R)^*}{\zeta - (\xi_k^R)^*}. \tag{C11}$$

By considering $f^{R/A}(\omega) = \lim_{\eta\to 0} f^M(\omega \pm i\eta)$ in Eq. (C11), we find that $\mp\frac{1}{\pi}\operatorname{Im} f^{R/A}(\omega) = A(\omega)$. While $A(\omega)$ is real, it is not necessarily positive and can have any sign.

From Eq. (C11) we can see the equivalence between functions analytic for $\operatorname{Im}\zeta \neq 0$ and functions that can be written as an integral over a spectral function. This gives justification to the use of the spectral representation for $\hat{G}^M(\zeta)$, Eq. (23), and $\hat{\Sigma}^M(\zeta)$, Eq. (27), for approximations that guarantee the correct analytical properties. The full justification follows when we consider operators below. When the spectral function is PSD, $-\hat{G}^R(\zeta)$ is a function that maps the upper half plane to itself. This is a so-called Nevanlinna function. Such functions are considered in, e.g., [66], where it is shown that such functions can be written in a spectral representation. However, the notion of spectral function can be generalized to measures, which means that also nondifferentiable spectral functions, such as the ones consisting of delta functions, can be considered.

If $f^M(\zeta)$ is not analytic away from the real axis, but has simple poles, the spectral function can still be defined via the real-frequency retarded or advanced Green's function. We can thus regard Eq. (C11) as a generalized spectral representation. If we would define the particle number as $N_A = \int_{-\infty}^{\mu} A(\omega)$, we would miss the contribution from the poles if they exist, and different definitions of particle number would yield different results.

We now generalize to operators $\hat{F}^M(\zeta)$, which we take to have finite dimension, and impose the condition $[\hat{F}^M(\zeta^*)]^\dagger = \hat{F}^M(\zeta)$. We apply the same considerations as above to each matrix element $F_{ij}^M(\zeta)$. Thus, Eq. (C11) holds for each matrix element, yielding

$$F_{ij}^M(\zeta) = \int_{-\infty}^{\infty} d\omega \frac{A_{ij}(\omega)}{\zeta - \omega} + \sum_k \frac{a_{kij}^R}{\zeta - \xi_{kij}^R} + \sum_k \frac{(a_{kji}^R)^*}{\zeta - (\xi_{kji}^R)^*}, \tag{C12}$$

where $\hat{A}^\dagger(\omega) = \hat{A}(\omega)$. Each matrix element can have their own set of poles. We now define

$$\sum_k \frac{a_{kij}^R}{\zeta - \xi_{kij}^R} = \sum_l \frac{\alpha_{lij}}{\zeta - \xi_l^R}, \tag{C13}$$

in which the sum over l goes over all the poles of all matrix elements of $\hat{F}^M(\zeta)$, and we defined the residue matrix

$$(\hat{\alpha}_l)_{ij} = \alpha_{lij} = \begin{cases} a_{lij}^R, & \text{if } l \text{ correspond to pole in } \hat{F}_{ij}^M, \\ 0, & \text{otherwise.} \end{cases} \tag{C14}$$

Using this rewriting, we obtain the spectral representation in operator form

$$\hat{F}^M(\zeta) = \int_{-\infty}^{\infty} d\omega \frac{\hat{A}(\omega)}{\zeta - \omega} + \sum_l \left(\frac{\hat{\alpha}_l}{\zeta - \xi_l^R} + \frac{\hat{\alpha}_l^\dagger}{\zeta - (\xi_l^R)^*} \right). \tag{C15}$$

We see that even in the presence of poles, we can define the spectral function operator as $\hat{F}^R(\omega) - \hat{F}^A(\omega) = -2\pi i \hat{A}(\omega)$. We have thus shown that analytic operators, whose matrix elements decay to zero at infinity, can be written in a spectral representation. This discussion fully justifies the use of the spectral representation for $\hat{G}^M(\zeta)$, Eq. (23), and $\hat{\Sigma}^M(\zeta)$, Eq. (27), under the assumption of analyticity.

APPENDIX D

The particle number can be obtained from the diagonal of the Green's function, Eq. (12). In this appendix, we will show that we can also obtain the particle number from the grand-canonical potential Ω in equilibrium, and that both definitions agree for a partially Φ-derivable approximation. For the exact case [35], as well as for any conserving approximation [11], Ω is given by the Luttinger-Ward functional, which we write in the Klein form as

$$\beta\Omega[G] = \Phi - \mathrm{Tr}_\gamma\{\hat{\Sigma}\hat{G}\} - \mathrm{Tr}_\gamma\{\ln(-\hat{G}^{-1})\}, \tag{D1}$$

when G is the Green's function of the physical system at hand. Here,

$$\mathrm{Tr}_\gamma\{\hat{\Sigma}\hat{G}\} = \int_\gamma d1d2\, \Sigma(1,2)G(2,1^+) \tag{D2}$$

$$\mathrm{Tr}_\gamma\{\hat{G}\} = \int_\gamma d1\, G(1,1^+) \tag{D3}$$

and the contour is taken to be the Matsubara contour, $\int_\gamma d1G(1,1^+) = -i\int_0^\beta d\tau \int d\mathbf{x}\; G(\mathbf{x},\tau;\mathbf{x},\tau^+)$. Note that the assumption of correct analytical properties is already implicit in Eq. (D1) since the logarithm has to be single-valued [see also the discussion below Eq. (61)].

A partially Φ-derivable scheme which is nonconserving will have ambiguous total energies, since different energy functionals give different values [13]. The Luttinger-Ward functional, however, is variational with respect to G, meaning that $\frac{\delta\Omega}{\delta G} = 0$ when evaluated at a G coming from the Dyson equation from a conserving approximation. Thus, partially Φ-derivable schemes can give results close to the fully conserving scheme [1,21,33,35,67–70]. The Luttinger-Ward functional has recently been evaluated at finite temperature in extended systems [71].

Following Baym [11], we vary Eq. (D1) with respect to the chemical potential. For a partially Φ-derivable approximation, we have that

$$\frac{\partial\Phi}{\partial\mu} = \mathrm{Tr}_\gamma\left\{\hat{\Sigma}\frac{\partial\hat{G}}{\partial\mu}\right\}. \tag{D4}$$

Note that we do not allow G_0 in $\Phi[G,G_0]$ to depend on μ. Equation (D1) becomes

$$\beta\frac{\partial\Omega}{\partial\mu} = -\mathrm{Tr}_\gamma\left\{\frac{\partial\hat{\Sigma}}{\partial\mu}\hat{G}\right\} - \mathrm{Tr}_\gamma\left\{\hat{G}\frac{\partial\hat{G}^{-1}}{\partial\mu}\right\}. \tag{D5}$$

From $\hat{G}^{-1}$ in Matsubara form, we obtain

$$\hat{G}^{-1}(\omega_m) = (\omega_m + \mu)\hat{\mathbb{1}} - \hat{h} - \hat{\Sigma}(\omega_m), \tag{D6}$$

which gives

$$\frac{\partial\hat{G}^{-1}}{\partial\mu} = \hat{\mathbb{1}} - \frac{\partial\hat{\Sigma}}{\partial\mu}. \tag{D7}$$

Equation (D5) simplifies, and we get

$$\beta\frac{\partial\Omega}{\partial\mu} = -\mathrm{Tr}_\gamma\{G\} = i\int_0^\beta d\tau \int d\mathbf{x}\; G(\mathbf{x},\tau;\mathbf{x},\tau^+), \tag{D8}$$

which gives the desired relation

$$\frac{\partial\Omega}{\partial\mu} = -N. \tag{D9}$$

Note that the derivation of Eq. (D9) only makes use of partial Φ derivability, and that the assumption of gauge invariance is not needed.

[1] G. Stefanucci and R. van Leeuwen, *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction* (Cambridge University Press, Cambridge, 2013).
[2] K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[3] N. Schlünzen, S. Hermanns, M. Bonitz, and C. Verdozzi, Phys. Rev. B **93**, 035107 (2016).
[4] E. Perfetto, A.-M. Uimonen, R. van Leeuwen, and G. Stefanucci, Phys. Rev. A **92**, 033419 (2015).
[5] G. Strinati, Riv. Nuovo Cimento **11**, 1 (1988).
[6] F. Bechstedt, *Many-Body Approach to Electronic Excitations: Concepts and Applications* (Springer, Berlin, 2015).
[7] G. Stefanucci, Y. Pavlyukh, A.-M. Uimonen, and R. van Leeuwen, Phys. Rev. B **90**, 115134 (2014).
[8] G. Stefanucci, Y. Pavlyukh, A.-M. Uimonen, and R. van Leeuwen, Phys. Rev. B **93**, 119906(E) (2016).
[9] A.-M. Uimonen, G. Stefanucci, Y. Pavlyukh, and R. van Leeuwen, Phys. Rev. B **91**, 115104 (2015).
[10] G. Baym and L. Kadanoff, Phys. Rev. **124**, 287 (1961).
[11] G. Baym, Phys. Rev. **127**, 1391 (1962).
[12] M. Puig von Friesen, C. Verdozzi, and C.-O. Almbladh, Phys. Rev. Lett. **103**, 176404 (2009).
[13] M. Puig von Friesen, C. Verdozzi, and C.-O. Almbladh, Phys. Rev. B **82**, 155108 (2010).
[14] S. Hermanns, N. Schlünzen, and M. Bonitz, Phys. Rev. B **90**, 125111 (2014).

[15] D. Kremp, T. Bornath, M. Bonitz, and M. Schlanges, Phys. Rev. E **60**, 4725 (1999).
[16] D. Kremp, M. Schlanges, and W.-D. Kraeft, *Quantum Statistics of Nonideal Plasmas* (Springer, Berlin, 2005).
[17] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
[18] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).
[19] H. J. de Groot, P. A. Bobbert, and W. van Haeringen, Phys. Rev. B **52**, 11000 (1995).
[20] B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).
[21] A. Stan, N. E. Dahlen, and R. van Leeuwen, Europhys. Lett. **76**, 298 (2006).
[22] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
[23] F. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, Phys. Rev. B **86**, 081102 (2012).
[24] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).
[25] P. Koval, D. Foerster, and D. Sánchez-Portal, Phys. Rev. B **89**, 155417 (2014).
[26] U. von Barth and B. Holm, Phys. Rev. B **54**, 8411 (1996).
[27] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[28] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics* (Elsevier, Amsterdam, 2000), Vol. 54, pp. 1–218.
[29] F. Aryasetiawan and T. Miyake, J. Comput. Theor. Nanos. **6**, 2451 (2009).
[30] C.-O. Almbladh, U. von Barth, and R. Van Leeuwen, Int. J. Mod. Phys. B **13**, 535 (1999).
[31] B. Holm, Self-Consistency and the *GW* approximation, Ph.D. thesis, Lund University, 1997.
[32] A. Schindlmayr, P. García-González, and R. W. Godby, Phys. Rev. B **64**, 235106 (2001).
[33] A. Stan, N. E. Dahlen, and R. van Leeuwen, J. Chem. Phys. **130**, 114105 (2009).
[34] A. Schindlmayr, Phys. Rev. B **56**, 3528 (1997).
[35] J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).
[36] D. C. Langreth, Phys. Rev. **150**, 516 (1966).
[37] J. S. Langer and V. Ambegaokar, Phys. Rev. **121**, 1090 (1961).
[38] J. Ortloff, M. Balzer, and M. Potthoff, Eur. Phys. J. B **58**, 37 (2007).
[39] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
[40] M. Hellgren and U. von Barth, J. Chem. Phys. **131**, 044110 (2009).
[41] S. Hershfield, J. H. Davies, and J. W. Wilkins, Phys. Rev. B **46**, 7046 (1992).
[42] M. Springer, F. Aryasetiawan, and K. Karlsson, Phys. Rev. Lett. **80**, 2389 (1998).
[43] N. E. Bickers, D. J. Scalapino, and S. R. White, Phys. Rev. Lett. **62**, 961 (1989).
[44] N. Bickers and D. Scalapino, Ann. Phys. **193**, 206 (1989).
[45] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Theory* (Dover Publications, New York, 2003).
[46] G. Baym and N. D. Mermin, J. Math. Phys. **2**, 232 (1961).
[47] E. Gross, E. Runge, and O. Heinonen, *Many-Particle Theory* (Adam Hilger, New York, 1991).
[48] Y. Pavlyukh, J. Berakdar, and A. Rubio, Phys. Rev. B **87**, 125101 (2013).
[49] M. Bonitz, D. Semkat, and H. Haug, Eur. Phys. J. B **9**, 309 (1999).
[50] A.-M. Uimonen, Developments in many-body theory of quantum transport and spectroscopy with non-equilibrium Green's functions and time-dependent density functional theory, Ph.D. thesis, University of Jyväskylä, 2015.
[51] P. Minnhagen, J. Phys. C **7**, 3013 (1974).
[52] D. Sangalli, P. Romaniello, G. Onida, and A. Marini, J. Chem. Phys. **134**, 034115 (2011).
[53] H. Shiba, Prog. Theor. Phys. **54**, 967 (1975).
[54] A. Yoshimori and A. Zawadowski, J. Phys. C **15**, 5241 (1982).
[55] J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).
[56] J. M. Luttinger, Phys. Rev. **121**, 942 (1961).
[57] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).
[58] A. A. Rusakov and D. Zgid, J. Chem. Phys. **144**, 054106 (2016).
[59] T. Gamelin, *Complex Analysis* (Springer, New York, 2003).
[60] P. Myöhänen, A. Stan, G. Stefanucci, and R. van Leeuwen, Europhys. Lett. **84**, 67001 (2008).
[61] J. C. Cuevas and E. Scheer, *Molecular Electronics* (World Scientific Publishing, Singapore, 2010).
[62] D. Karlsson, Analytical and numerical developments in strongly correlated systems: Perspectives from TDDFT and Green's functions, Ph.D. thesis, Lund University, 2014.
[63] J. A. White, Phys. Rev. B **45**, 1100 (1992).
[64] B. Horvatić and V. Zlatić, Phys. Status Solidi (b) **99**, 251 (1980).
[65] N. J. Higham, *Functions of Matrices: Theory and Computation* (Society for Industrial and Applied Mathematics, Philadelphia, 2008).
[66] G. Teschl, *Mathematical Methods in Quantum Mechanics with Applications to Schrödinger Operators*, 2nd ed. (American Mathematical Society, Providence, 2014).
[67] F. Caruso, D. R. Rohr, M. Hellgren, X. Ren, P. Rinke, A. Rubio, and M. Scheffler, Phys. Rev. Lett. **110**, 146403 (2013).
[68] F. Aryasetiawan, T. Miyake, and K. Terakura, Phys. Rev. Lett. **88**, 166401 (2002).
[69] N. E. Dahlen and U. von Barth, Phys. Rev. B **69**, 195102 (2004).
[70] N. E. Dahlen, R. van Leeuwen, and U. von Barth, Phys. Rev. A **73**, 012511 (2006).
[71] A. R. Welden, A. A. Rusakov, and D. Zgid, arXiv:1605.06563.