

# Sum rules and Ward identities in the Kondo lattice

P. Coleman,<sup>1,2</sup> I. Paul,<sup>1,3</sup> and J. Rech<sup>1,2,3</sup>

<sup>1</sup>*Kavli Institute for Theoretical Physics, Kohn Hall, UCSB Santa Barbara, Santa Barbara, California 93106, USA*

<sup>2</sup>*Center for Materials Theory, Rutgers University, Piscataway, New Jersey 08855, USA*

<sup>3</sup>*SPhT, L'Orme des Merisiers, CEA-Saclay, 91191 Gif-sur-Yvette France*

(Received 3 May 2005; published 26 September 2005)

We derive a generalized Luttinger-Ward expression for the free energy of a many-body system involving a constrained Hilbert space. In the large- $N$  limit, we are able to explicitly write the entropy as a functional of the Green's functions. Using this method we obtain a Luttinger sum rule for the Kondo lattice. One of the fascinating aspects of the sum rule is that it contains two components: one describing the heavy electron Fermi surface, the other, a sea of oppositely charged, spinless fermions. In the heavy electron state, this sea of spinless fermions is completely filled and the electron Fermi surface expands by one electron per unit cell to compensate the positively charged background, forming a “large” Fermi surface. Arbitrarily weak magnetism causes the spinless Fermi sea to annihilate with part of the Fermi sea of the conduction electrons, leading to a small Fermi surface. Our results thus enable us to show that the Fermi surface volume contracts from a large to a small volume at a quantum critical point. However, the sum rules also permit the possible formation of a new phase, sandwiched between the antiferromagnet and the heavy electron phase, where the charged spinless fermions develop a true Fermi surface.

DOI: [10.1103/PhysRevB.72.094430](https://doi.org/10.1103/PhysRevB.72.094430)

PACS number(s): 75.30.Mb, 75.20.Hr, 71.27.+a, 72.15.Qm

## I. INTRODUCTION

Sum rules play a vital role in condensed matter physics. The most famous sum rule—the Luttinger sum rule,<sup>1–3</sup> defines rigorously the volume of the Fermi surface of a Fermi liquid in terms of the density of electrons:

$$2 \frac{v_{FS}}{(2\pi)^D} = n_e, \quad (1)$$

where  $v_{FS}$  is the Fermi surface volume,  $n_e$  is the density of electrons per unit cell, and  $D$  is the dimension. Historically, sum rules have also played an important role in our understanding of strongly correlated systems. In the context of the Kondo effect for example, the Friedel sum rule,<sup>4–6</sup>

$$\sum_{\lambda} \frac{\delta_{\lambda}}{\pi} = \Delta n_e, \quad (2)$$

relating the sum of the scattering phase shifts in channels labeled by  $\lambda$  to the number of bound states  $\Delta n_e$  helped to establish a rigorous foundation for the Abrikosov-Suhl resonance<sup>7–9</sup> that develops in Anderson and Kondo impurity models. Later, Martin<sup>10</sup> applied the Luttinger sum rule to the Anderson lattice model, to argue that heavy electron metals must have a “large Fermi surface” that counts both the conduction electrons and also the localized  $f$  electrons,

$$2 \frac{v_{FS}}{(2\pi)^D} = n_e + 1. \quad (3)$$

Today, there is a renewed interest in sum rules, in connection with models of strongly correlated electrons. For instance, in the context of high-temperature superconductors, which are Mott insulators when undoped, there has been a longstanding debate over whether the “large” Fermi surface predicted by Luttinger's sum rule might be replaced at low doping, by a “small” Fermi surface determined by the num-

ber of doped holes.<sup>11,12</sup> Related issues arise in the context of heavy electron systems, where experimental advances have made it possible to tune through the quantum critical point that separates the heavy electron paramagnet from the local moment antiferromagnet. There have been a number of theoretical speculations that the Fermi surface of the heavy electron material may jump from “large” to “small” at the quantum critical point.<sup>13–15</sup> Recent experimental work, based on de Haas-van Alphen measurements<sup>16</sup> and Hall measurements,<sup>17</sup> provide experimental support for this hypothesis, but the idea has lacked rigorous theoretical support.

The main difficulty in extending sum rules to strongly correlated systems is that the theoretical machinery used and developed by Luttinger and Ward to derive sum rules applies to models with unprojected Hilbert spaces. About eight years ago, Affleck and Oshikawa<sup>18</sup> demonstrated that such sum rules have a more general existence. By using a modification of the Leib-Mattis theorem, Affleck and Oshikawa showed that the “large Fermi surface” that counts both local moments and conduction electrons develops in the one-dimensional  $S=1/2$  Kondo model, even though in this case the ground state is not a Fermi liquid. More recently, Oshikawa<sup>19</sup> has extended this derivation to higher-dimensional Kondo lattices. This work suggests that it ought to be possible to extend the Luttinger-Ward approach to Hamiltonian problems with strong constraints.

In this paper, we show how the original methods of Luttinger and Ward can indeed be extended to strongly correlated models. A key element of this work is the construction of a functional relating the free energy of a strongly interacting system to the Green's functions of the projected Hilbert space. Our approach is based on the use of slave particles, such as a Schwinger boson description for the local moments in a Kondo lattice. We begin with a generalization of the Luttinger-Ward free energy functional, appropriate for systems of interacting bosons and fermions,<sup>20,21</sup> which can be compactly written as

$$F = T \underline{\text{Str}}[\ln(-\mathcal{G}^{-1}) + \Sigma \mathcal{G}] + Y[\mathcal{G}], \quad (4)$$

where  $\underline{\text{Str}}[A] = \text{Tr}[A_B] - \text{Tr}[A_F]$  denotes the supertrace of a matrix containing both bosonic (B) and fermionic (F) components (where the underline notation is used to denote a sum over internal frequencies and a trace over the internal quantum numbers of the matrix).  $\mathcal{G} = (\mathcal{G}_0^{-1} - \Sigma)^{-1}$  is the matrix describing the fully dressed Green's function of all elementary particles and fields entering the Lagrangian, including the slave particles, where  $\Sigma$  is the self-energy matrix and  $\mathcal{G}_0$  the bare propagator of the fields. The quantity  $Y[\mathcal{G}]$  is, diagrammatically, the sum of all closed-loop two-particle irreducible skeleton Feynman diagrams. Figure 1 illustrates the sum of diagrams contributing to  $Y[\mathcal{G}]$  for a Kondo lattice. Variations of  $Y[\mathcal{G}]$  with respect to the Green's function  $\mathcal{G}$  generate the self-energy

$$\delta Y[\mathcal{G}] = -T \underline{\text{Str}}[\Sigma \delta \mathcal{G}] \quad (5)$$

or

$$\Sigma(\omega) = -\beta \frac{\delta Y}{\delta \mathcal{G}(\omega)}, \quad (6)$$

where the use of the supertrace as the measure for functional derivatives avoids the need to introduce a relative minus sign between Fermi and Bose parts of this expression. Diagrammatically, the functional derivative of  $Y$  with respect to  $\mathcal{G}$  corresponds to “cutting” one of its internal lines. If we truncate  $Y$  to some order in  $1/N$ , this relationship determines a conserving Kadanoff-Baym approximation.<sup>23</sup> For example, if we take the generating function for a Kondo lattice illustrated in Fig. 1(c), and differentiate the leading order contri-

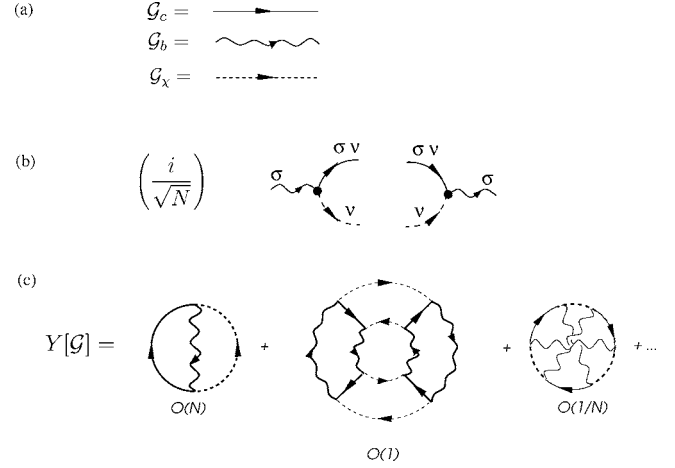


FIG. 1. Illustrating the Luttinger-Ward functional for the Kondo lattice. (a) Fully renormalized propagators for the conduction electrons,  $\chi$  fermions, and Schwinger bosons, where  $\sigma \in [1, N]$  is the spin index and  $\nu \in [1, K]$  the channel index, (b) Interaction vertices showing spin ( $\sigma$ ) and channel ( $\nu$ ) indices and (c) leading skeleton diagrams for the Luttinger-Ward functional with dependence on  $1/N$ . The first diagram involves two loops carrying spin and channel quantum numbers, and one pair of vertices and is consequently of order  $O(KN/N) = O(N)$ . The second diagram has four loops carrying internal quantum numbers and four pairs of vertices, so it is of order  $O(N^2 K^2 / N^4) = O(1)$ . The final diagram has two quantum number loops and three pairs of vertices, and is hence of order  $O(NK/N^3) \sim O(1/N)$ .

bution in the large  $N$  expansion with respect to  $\mathcal{G}$ , we obtain the following self-energies:

$$\begin{aligned} \Sigma_b(\omega) &= -\beta \frac{\delta Y}{\delta \mathcal{G}_b(\omega)} = \text{diagram with cross} = \text{diagram with } \mu \text{ loop} \quad O(K/N), \\ \Sigma_\chi(\omega) &= +\beta \frac{\delta Y}{\delta \mathcal{G}_\chi(\omega)} = \text{diagram with cross} = \text{diagram with } \sigma \text{ loop} \quad O(1), \\ \Sigma_c(\omega) &= +\beta \frac{\delta Y}{\delta \mathcal{G}_c(\omega)} = \text{diagram with cross} = \text{diagram with } \text{dashed loop} \quad O(1/N), \end{aligned} \quad (7)$$

where the cross indicates the line that is eliminated by the functional differential. Each of these terms contains a factor  $O(1/N)$  from the vertices, but the first two self-energies contain summations over the internal channel or spin indices,

elevating  $\Sigma_b$  and  $\Sigma_\chi$  to terms of order  $O(1)$ . In the leading-order large  $N$  approximation,  $\Sigma_c \sim O(1/N)$ , so that a consistent large- $N$  approximation is produced by leaving the conduction electron lines undressed. This provides an alternative

diagrammatic derivation of the Parcollet–Georges approach to the multichannel Kondo model.<sup>24</sup>

Each conserved quantity  $Q$  that commutes with the Hamiltonian leads to a gauge invariance of the underlying quantum fields, such that the action and all physical properties are invariant under the transformation

$$\psi_\zeta(t) \rightarrow e^{i\theta(t)q_\zeta}\psi_\zeta. \quad (8)$$

Here,  $q_\zeta$  is the gauge charge of the field: it is this quantity that controls the charge of any physical excitations associated with the field. Provided this gauge invariance is unbroken, we show that this leads to a Ward identity at zero temperature,

$$\int_{-\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \hat{q} \mathcal{G} \frac{d\Sigma(\omega)}{d\omega} \right] = 0, \quad (9)$$

where the integral runs along the imaginary axis. Luttinger's original work derived the elementary version of this relationship for electrons: here it is seen to hold for interacting species of fields.

Each of these Ward identities leads to a corresponding, generalized Luttinger sum rule given by

$$Q = \frac{1}{\pi} \text{Im Tr} \{ \hat{q} \ln [ -\mathcal{G}_F^{-1}(0 - i\delta) ] \}, \quad (10)$$

where  $\mathcal{G}_F$  is the complete fermionic Green's function, involving all fermions involved in the description of the constrained Hilbert space, including those introduced as slave particles. In impurity systems, the trace over the logarithms reduces to a sum of phase shifts, and the quantity  $Q$  is replaced by  $\Delta Q$ , the change in  $Q$  induced by the impurity. In the limit of infinite bandwidth,  $\Delta Q=0$ , which leads to a multiparticle version of the Friedel sum rule,

$$0 = \sum q_\zeta n_\zeta \left( \frac{\delta_\zeta}{\pi} \right), \quad (11)$$

where  $n_\zeta$  is the spin degeneracy associated with the trace over internal quantum numbers of the field  $\psi_\zeta$ . In lattice systems, the trace over the log becomes the Fermi surface volume associated with the field  $\psi_\zeta$ , so that

$$Q = \sum_\zeta \left[ q_\zeta n_\zeta \frac{v_{FS}(\zeta)}{(2\pi)^D} \right], \quad (12)$$

where  $v_{FS}(\zeta)$  is the Fermi surface volume associated with field  $\psi_\zeta$ . A closely related Fermi surface sum rule has recently been derived by Powell *et al.* for mixtures of fermions and bosons in atom traps.<sup>25</sup>

In this paper we apply these results to the Kondo lattice model. The key to our approach is the adoption of Schwinger bosons for the description of quantum spins. Schwinger bosons offer a key advantage, because they can describe both the antiferromagnetic and the heavy electron ground states of the Kondo lattice.<sup>26</sup> The Kondo effect induces a development of a retarded interaction in the spin singlet channel between the electrons and the spins. Formally, this interaction manifests itself as the mediating field  $\chi_j$  in the Hubbard–Stratonovich decomposition of the interaction<sup>24,27,28</sup>

$$J_K \vec{S}_j \cdot c_{j\alpha}^\dagger \vec{\sigma}_{\alpha\beta} c_{j\beta} \rightarrow \frac{1}{\sqrt{N}} (c_{j\alpha}^\dagger b_{j\alpha}) \chi_j^\dagger + \frac{1}{\sqrt{N}} (b_{j\alpha}^\dagger c_{j\alpha}) \chi_j + \frac{\chi_j^\dagger \chi_j}{J_K}. \quad (13)$$

Here  $\vec{S}_j$  is the spin at site  $j$  and  $c_{j\alpha}^\dagger$  creates an electron at site  $j$ . At long times the  $\chi$  field develops charge dynamics that describe the scaling of the Kondo interaction in both frequency and momentum space. In particular, the  $\chi$ -propagator describes the momentum- and frequency-dependent Kondo interaction,

$$-\mathcal{G}_\chi(\mathbf{k}, \omega) = \left( \frac{1}{J_K} + \Sigma_\chi(\mathbf{k}, \omega) \right)^{-1} = J_K^*(\mathbf{k}, \omega). \quad (14)$$

At each point in momentum space, where the Kondo interaction scales to strong coupling,  $\mathcal{G}_\chi$  develops a pole. In this way, the physics of the Kondo effect is intimately linked<sup>15</sup> to the possible emergence of a positively charged, spinless Fermi field—a “holon.”<sup>29</sup>

Under gauge transformations, the conduction and  $\chi$  fields have opposite gauge charge,

$$q_c = 1, \quad q_\chi = -1. \quad (15)$$

When we apply the Luttinger–Ward procedure, we find that the physical charge density is given by the total volume of the electron Fermi surface, minus the volume of the holon Fermi surface, as follows:

$$n_e = 2 \frac{v_{FS}}{(2\pi)^D} - \frac{v_\chi}{(2\pi)^D}. \quad (16)$$

Here, the volume of the electron Fermi surface is given by the conventional Luttinger formula,

$$\begin{aligned} v_{FS} &= \frac{1}{\pi} \sum_{\mathbf{k}} \text{Im} \ln [ \epsilon_{\mathbf{k}} + \Sigma_c(\mathbf{k}, 0 - i\delta) - \mu + i\delta ] \\ &= \sum_{\mathbf{k}} \Theta(\mu - E_{\mathbf{k}}), \end{aligned} \quad (17)$$

where  $\mu$  is the chemical potential,  $E_{\mathbf{k}} = \epsilon_{\mathbf{k}} + \Sigma_c(\mathbf{k}, E_{\mathbf{k}})$  is the renormalized energy of the heavy electrons. Now  $v_\chi$  is

$$v_\chi = \frac{1}{\pi} \sum_{\mathbf{k}} \text{Im} \ln \left( \frac{1}{J_K} + \Sigma_\chi(\mathbf{k}, 0 - i\delta) \right) = \sum_{\mathbf{k}} \Theta[-J_K^*(\mathbf{k})], \quad (18)$$

where  $J_K^*(\mathbf{k}) = -\mathcal{G}_\chi(\mathbf{k}, 0 - i\delta)$ .

The meaning of the renormalized coupling constant  $J_K^*$  in the paramagnetic phase needs a little discussion. This quantity describes the residual interaction between the electron fluid and any additional spins that are added to the Fermi liquid ground state. If we add an additional Schwinger boson, increasing  $S \rightarrow S + \frac{1}{2}$  at a given site, the additional spin unit remains unscreened, because each channel can only screen one spin unit, and all channels are fully screening. If  $\omega_b = E(n_b + 1) - E(n_b)$  is the energy to adding one additional spin, then the residual interaction between the additional Schwinger boson and the conduction electrons is  $J_K^*(\omega_b)$ . Since the additional spin decouples, it follows that  $J_K^*(\omega)$  has

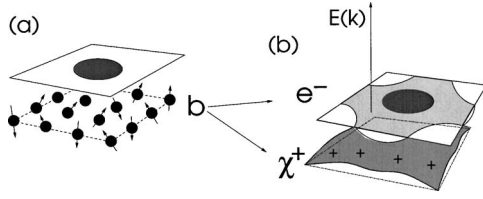


FIG. 2. Schematic diagram illustrating the enlargement of the Fermi surface in the Kondo lattice (a) coupling of local moments to a small Fermi surface; (b) each spin-boson “ionizes” into an electron and a “holon,” producing a large electron Fermi surface and a filled holon Fermi sea.

a zero at  $\omega = \omega_b$ ,  $J_K^*(\omega_b) = 0$ . At higher energies, the residual interaction will become positive, ultimately connecting up to the single-ion scaling behavior  $J_K^*(\omega) \sim 1/\ln(T_K/\omega)$ . Since  $J_K^*$  passes through zero at  $\omega = \omega_b$ , it follows that at lower energies  $\omega < \omega_b$ , and in particular, at zero frequency, the residual coupling must be ferromagnetic, i.e.  $J_K^*(0) < 0$ . This, in turn, means that the Fermi sea of holons is entirely full,  $v_\chi = (2\pi)^D$ , so that

$$n_e = 2 \frac{v_{FS}}{(2\pi)^D} - 1. \quad (19)$$

To preserve the overall charge density, the electron Fermi surface volume is forced to enlarge by one unit per spin to “screen” the finite background density of positively charged holons (Fig. 2).

Let us now consider the antiferromagnet. In this case, the Schwinger boson field condenses, and the  $\chi$  field becomes hybridized with the conduction electrons. In this case, the holon Fermi surface annihilates with the expanded part of the heavy electron Fermi surface, to reveal a single integrated set of Fermi surface sheets with a “small” volume that encloses the total electron count. If the transition from paramagnet to antiferromagnet occurs via a single quantum critical point, then our results indicate that the transition in the Fermi surface volume at the second-order quantum critical point is abrupt.<sup>30</sup>

Perhaps the most interesting aspect of this picture is that it suggests that the background field of holons can develop dynamics. One intriguing possibility is that the holons are liberated when the Fermi surface volume contracts at the quantum critical point. Another more exotic possibility, allowed by the Ward identities, is the formation of a separate *phase* where the holons develop a Fermi surface, so that  $v_\chi = x(2\pi)^D$  is partially filled. This “spin-charge” decoupled phase would be sandwiched between the localized magnetic phase and the fully formed paramagnet. Pépin has recently suggested that the holons might develop a Fermi surface at the quantum critical point.<sup>15</sup> This might provide a mechanism for such a scenario. A good candidate for this phase is the one-dimensional Haldane  $S=1$  chain, coupled to two conducting chains to form a one-dimensional, two-channel Kondo lattice. Here, the conjectured intermediate phase would separate a Haldane gap phase and a large Fermi surface phase, leading to Friedel oscillations with wave vector

$q = 2k_F^{(0)} + (\pi/2)x$ , intermediate between that of the large ( $x=1$ ) and small ( $x=0$ ) Fermi surface.

This paper is divided up into various parts. In the first we develop the Luttinger–Ward free energy expansion for the Kondo lattice. Next, we present a new formula for the entropy of the interacting fluid, expressed exclusively in terms of the Green’s functions of the particles. We then go on to derive the Luttinger sum rule in a general form for a constrained system, which we then apply to the Kondo impurity and lattice model. In the final part, we discuss the general implications of our sum rule in the Kondo lattice.

## II. LUTTINGER WARD EXPRESSION FOR THE KONDO LATTICE

The Luttinger–Ward approach was originally based on perturbation theory in the strength of the Coulomb interaction. In order to develop these methods for strongly interacting systems, we need to take account of the constraints in the theory, requiring a Luttinger–Ward approach appropriate to gauge theories. In this respect, our efforts to apply the Luttinger Ward approach to the Kondo lattice closely parallel recent efforts to extend the Luttinger–Ward approach to quark-gluon plasmas.<sup>21</sup>

Although the Luttinger–Ward functional may be derived nonperturbatively,<sup>31</sup> the usefulness of any derived sum rules rests on the direct relationship between the Green’s functions that enter in the functional and physical excitations. Strongly interacting systems do not generally provide a natural small parameter on which to base a perturbation expansion. Our philosophy is that to develop this control, we need to define a family of large- $N$  models that, in the infinite- $N$  limit, provide a mean-field description of the phases of interest. In the expansion about this limit,  $1/N$  then provides a small parameter. The large- $N$  solutions constructed this way will of course satisfy the sum rules we derive, but the sum rules are expected to have a far greater generality.

In the context of our interest in the Kondo lattice, we adopt a family of models given by

$$\mathcal{H} = \sum_{\vec{k}, \nu, \alpha} \epsilon_{\vec{k}} c_{\vec{k}\nu\alpha}^\dagger c_{\vec{k}\nu\alpha} + \overbrace{\frac{J_K}{N} \sum_{j\nu\alpha\beta} c_{j\nu\alpha}^\dagger c_{j\nu\beta} b_{j\beta}^\dagger b_{j\alpha}}^{H_I} \quad (20)$$

Here, we have adopted the Schwinger boson representation, where  $b_{j\alpha}^\dagger$  creates a Schwinger boson at site  $j$ , with spin component  $\alpha \in [1, N]$ . The combination  $b_{j\alpha}^\dagger b_{j\beta} = S_{\alpha\beta}(j)$  represents the local spin operator and the system is restricted to the physical Hilbert space by requiring that

$$\sum_{\alpha} b_{j\alpha}^\dagger b_{j\alpha} = 2S, \quad (21)$$

at each site  $j$ . The index  $\nu \in [1, K]$  is a channel index that is absent in the physical  $SU(2)$  model, but that is included to create a family of large- $N$  models. The “filled shell” case  $2S=K$  defines a perfectly screened Kondo lattice, and our discussion will focus on this case. Here  $c_{\vec{k}\alpha}^\dagger$  creates a conduction electron of momentum state  $\vec{k}$  and spin  $\alpha \in (1, N)$ .  $c_{j\alpha}^\dagger$



$= (1/\sqrt{\mathcal{N}_s}) \sum_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha}^\dagger e^{-i\vec{k}\cdot\vec{x}_j}$  creates an electron at site  $j$ , where  $\mathcal{N}_s$  is the number of sites in the lattice. The above model is known to have a controlled large- $N$  expansion that contains both magnetic and paramagnetic solutions when  $N$  is taken to infinity with  $k=K/N$  fixed.

To develop a Luttinger–Ward expansion, we shall factorize the interaction in terms of a Grassman field  $\chi_{j\nu}$ ,

$$H_I \rightarrow \sum_{j\nu\alpha} \frac{1}{\sqrt{N}} [(c_{j\nu\alpha}^\dagger b_{j\alpha}) \chi_{j\nu}^\dagger + (b_{j\alpha}^\dagger c_{j\nu\alpha}) \chi_{j\nu}] + \sum_{j\nu} \frac{\chi_{j\nu}^\dagger \chi_{j\nu}}{J_K}, \quad (22)$$

and the Lagrangian becomes

$$\mathcal{L} = \sum_{\vec{k}\alpha\nu} c_{\vec{k}\alpha\nu}^\dagger (\partial_\tau + \epsilon_{\vec{k}}) c_{\vec{k}\alpha\nu} + \sum_{j\alpha} b_{j\alpha}^\dagger (\partial_\tau + \lambda_j) b_{j\alpha} + \sum_{j\nu} \frac{\chi_{j\nu}^\dagger \chi_{j\nu}}{J_K} + \sum_{j\nu\alpha} \frac{1}{\sqrt{N}} [(c_{j\nu\alpha}^\dagger b_{j\alpha}) \chi_{j\nu}^\dagger + (b_{j\alpha}^\dagger c_{j\nu\alpha}) \chi_{j\nu}] - \sum_j 2S\lambda_j. \quad (23)$$

Note that the Grassman fields  $\chi_{j\nu}$  contain no time derivative, and so instantaneously they behave as inert, neutral fields. Here  $\lambda_j$  is the static value of the Lagrange multiplier used to enforce the constraint on each lattice site. At finite temperatures the constraint is only enforced on the average, but in the ground state, the value of the conserved quantities  $n_{bj}$  will be quantized as a step function of the  $\lambda_j$ . In the range  $\lambda_j \in [\lambda_j^-, \lambda_j^+]$  (where  $\lambda_j^\pm$  are the energy gaps between the physical ground state and the states with  $2S \pm 1$  Schwinger bosons at site  $j$ ), the constraint will be precisely satisfied. We shall assume that a translationally invariant choice  $\lambda_j = \lambda$  is always available.

For our sum rules, we need to know comparatively little about the nature of the ground state excitations. However, from the large- $N$  solution of this model, we do know that there are two main classes of solution to the Kondo lattice:

(i) The heavy electron phase, where the Schwinger bosons pair condense in either the Cooper channel ( $\langle \sigma b_{i\sigma}^\dagger b_{j-\sigma}^\dagger \rangle \neq 0$  antiferromagnetic interactions) or particle-hole channel ( $\langle \sigma b_{i\sigma}^\dagger b_{j\sigma} \rangle \neq 0$  ferromagnetic interactions). In field-theory language, this is a “Higg’s phase” in which the local  $U(1)$  symmetry is broken and both the spin and holon excitations can propagate from site to site.

(ii) The magnetic phase, in which the Schwinger bosons individually condense ( $\langle b_j \rangle \neq 0$ ) to develop long-range magnetic order. In this phase, the  $\chi_j$  fields become hybridized with the conduction fields, and  $\mathcal{G}_c$  and  $\mathcal{G}_\chi$  become diagonal members of a single matrix propagator.

In general, we will therefore have to preserve the momentum dependence of the propagators. The bare propagators of the theory are diagonal in momentum, and given by

$$\mathcal{G}_b^{(0)}(\vec{k}, i\nu_n) = (i\nu_n \tau_3 - \lambda \mathbf{1})^{-1},$$

$$\mathcal{G}_c^{(0)}(\vec{k}, i\omega_n) = (i\omega_n - \epsilon_{\vec{k}})^{-1},$$

$$\mathcal{G}_\chi^{(0)}(\vec{k}, i\omega_n) = -J_K, \quad (24)$$

where we have written the bare boson propagator as a two-dimensional Nambu propagator, to allow for the possible introduction of boson pairing, which develops as antiferromagnetic correlations begin to grow. Formally, we will combine these three propagators into a single propagator, given by

$$\mathcal{G}^{-1} = \begin{bmatrix} i\nu_n \tau_3 - \lambda \mathbf{1} & & \\ & i\omega_n - \epsilon_{\vec{k}} & \\ & & -\frac{1}{J_K} \end{bmatrix} - \Sigma, \quad (25)$$

where  $\Sigma$  is a matrix self-energy. Following Luttinger and Ward, we will regard  $\mathcal{G}$  as a variational function, and  $\Sigma$  as a derived quantity.

We now consider the effect of tuning up the strength of the Kondo interaction from zero to  $J_K$ , by replacing  $J_K \rightarrow \alpha J_K$  where  $\alpha \in [0, 1]$ , keeping the chemical potential of the conduction electrons and bosons fixed. Now the partition function is given by

$$Z = \text{Tr}[e^{-\beta H}] = \int \mathcal{D}[c, b, \chi] e^{-\int_0^\beta \mathcal{L} d\tau}, \quad (26)$$

where  $\mathcal{D}[c, b, \chi]$  is the measure of the path integral. If we vary  $\alpha$  inside this expression, we obtain

$$\begin{aligned} \frac{dZ}{d\alpha} &= \int \mathcal{D}[c, b, \chi] \sum_{j\nu} \int_0^\beta d\tau \left( \frac{1}{\alpha^2 J_K} \chi_{j\nu}^\dagger \chi_{j\nu} \right) e^{-\int_0^\beta \mathcal{L} d\tau} \\ &= \beta Z \times \frac{1}{\alpha^2 J_K} \sum_{j\nu} \langle \chi_{j\nu}^\dagger \chi_{j\nu} \rangle, \end{aligned} \quad (27)$$

so that if  $F = -T \ln Z$  is the free energy, then

$$\frac{dF}{d\alpha} = -\frac{1}{(\alpha^2 J_K)} \sum_{j\nu} \langle \chi_{j\nu}^\dagger \chi_{j\nu} \rangle = -\frac{T}{(\alpha^2 J_K)} \text{Tr}[\mathcal{G}_\chi], \quad (28)$$

where  $\mathcal{G}_\chi$  is the Green’s function for the  $\chi$  fermion and  $\text{Tr}[\mathcal{G}_\chi] \equiv K \sum_{i\omega_n, \vec{k}} \mathcal{G}_\chi(\vec{k}, i\omega_n) e^{i\omega_n 0^+}$  denotes a trace over the frequency and momentum of  $\mathcal{G}_\chi$  (here,  $e^{i\omega_n 0^+}$  ensures the normal ordering of the operators). We next consider the expression

$$\tilde{F} = T \text{Str}[\ln(-\mathcal{G}^{-1}) + \Sigma \mathcal{G}] + Y[\mathcal{G}]. \quad (29)$$

We will show how this expression generalizes the Luttinger–Ward free energy functional to a mixture of interacting bosons and fermions. Here, we use the notation

$$\begin{aligned} \text{Str}[A] &\equiv \text{Tr}[A_B] - \text{Tr}[A_F] \equiv \sum_{i\nu_n} \text{Tr}[A_B(i\nu_n)] e^{i\nu_n 0^+} \\ &\quad - \sum_{i\omega_n} \text{Tr}[A_F(i\omega_n)] e^{i\omega_n 0^+} \end{aligned} \quad (30)$$

to denote the “supertrace” over the bosonic and fermionic parts of the matrix  $A$ . The “underline” notation is used to indicate a sum over both the frequency variable of  $A$  and a trace over the internal quantum numbers of  $A$  (such as momentum). The supertrace includes a relative minus sign for

the fermionic component of the matrix, and we have explicitly displayed the trace over the discrete Matsubara Bose [ $i\nu_n = (2n)\pi T$ ] and Fermi [ $i\omega_n = (2n+1)\pi T$ ] frequencies. In the above expression, the self-energy matrix  $\Sigma = \mathcal{G}_0^{-1} - \mathcal{G}^{-1}$ . The quantity  $Y[\mathcal{G}]$  is the sum of all closed, two-particle irreducible skeleton Feynman graphs for the Free energy.

As mentioned earlier, the Luttinger Ward functional  $Y[\mathcal{G}]$  has the property that its variation with respect to  $\mathcal{G}$  generates the self-energy matrix,

$$\delta Y[\mathcal{G}] = -T \text{Str}[\Sigma \delta \mathcal{G}]. \quad (31)$$

The variation of the first term in  $\tilde{F}$  with respect to  $\mathcal{G}$  is given by

$$\delta \{T \text{Str}[\ln(-\mathcal{G}^{-1})]\} = T \text{Str}[-\mathcal{G}^{-1} \delta \mathcal{G}] \quad (32)$$

By using the relation  $\Sigma[\mathcal{G}] = \mathcal{G}_0^{-1} - \mathcal{G}^{-1}$ , the variation in the second term in  $\tilde{F}$  is given by

$$\delta \{T \text{Str}[(\mathcal{G}_0^{-1} - \mathcal{G}^{-1})\mathcal{G}]\} = T \text{Str}[\mathcal{G}_0^{-1} \delta \mathcal{G}], \quad (33)$$

so that the total variation of  $\tilde{F}$  with respect to  $\mathcal{G}$ ,

$$\delta \tilde{F} = T \text{Str} \left[ \left( \overbrace{-\mathcal{G}^{-1} + \mathcal{G}_0^{-1}}^{\Sigma} - \Sigma \right) \delta \mathcal{G} \right] = 0 \quad (34)$$

identically vanishes,

$$\frac{\delta \tilde{F}}{\delta \mathcal{G}} = 0. \quad (35)$$

Now the Hubbard–Stratonovich transformation that we have carried out on the interaction  $H_I$  assures that the only place that the coupling constant  $\alpha J_K$  enters, is in  $[\mathcal{G}_\chi^0]^{-1} = -1/\alpha J_K$ . This means that on the rhs of Eq. (29)  $\alpha$  enters explicitly only through  $\Sigma_\chi = -(\alpha J_K)^{-1} - [\mathcal{G}_\chi]^{-1}$ . Then,

$$\frac{d\tilde{F}}{d\alpha} = \underbrace{\frac{\delta \tilde{F}}{\delta \mathcal{G}} \frac{\partial \mathcal{G}}{\partial \alpha}}_{=0} + \frac{\partial \tilde{F}}{\partial \alpha} = -\frac{T}{(\alpha^2 J_K)} \text{Tr}[\mathcal{G}_\chi(\vec{q}, i\omega_n) e^{i\omega_n 0^+}]. \quad (36)$$

But by comparison with (28), we see that  $d\tilde{F}/d\alpha = dF/d\alpha$  and, since  $F(\alpha=0) = \tilde{F}(\alpha=0)$  in the noninteracting case, the two quantities must be equal for all  $\alpha$ , i.e.,

$$F = T \text{Str}[\ln(-\mathcal{G}^{-1}) + \Sigma \mathcal{G}] + Y[\mathcal{G}]. \quad (37)$$

There are various points to make about this derivation.

(i) The derivation is very general. Its correctness only depends on the stationarity of  $\tilde{F}[\mathcal{G}]$  with respect to variations in  $\mathcal{G}$  and the equivalence between  $dF/d\alpha$  and  $d\tilde{F}/d\alpha$ . This means that the above expression will hold for broken symmetry or “Higg’s phase” solutions that involve off-diagonal components to  $\mathcal{G}$ .<sup>20</sup> In the context of the Kondo lattice, this means that  $\mathcal{G}$  can be extended to include anomalous boson pairing terms that are driven by short-range antiferromagnetic correlations, or alternatively, off-diagonal terms driven by long-range magnetism, which mix the conduction and  $\chi$  fermions. This expression can also be used to describe superconducting states, where the conduction electron and  $\chi$  propagators will contain off-diagonal terms.

(ii) Even though the free energy is a functional of three Green’s functions, there is no overcounting.

(iii) The above Free energy functional can be used as a basis for developing conserving approximations that generalize the Kadanoff–Baym approach to a constrained system.<sup>23</sup> In particular, in the large- $N$  limit, the skeleton graph expansion for  $Y$  truncates at the leading diagram (Fig. 1), providing the basis for a controlled treatment of both the magnetic and the paramagnetic phases of the Kondo lattice.<sup>24,26</sup>

### III. ENTROPY FORMULA

In this section, we derive an approximate formula for the entropy of a Luttinger–Ward system that becomes exact in the large- $N$  limit and in any approximation where the vertex corrections can be neglected. The result, which we shall derive below, is

$$S(T, \mathcal{G}) = \int \frac{d\omega}{\pi} \text{Tr}_{B,F} \left[ \left( \frac{d\hat{n}(\omega)}{dT} \right) \{ \text{Im} \ln[-\mathcal{G}^{-1}(\omega - i\delta)] + \text{Im} \Sigma(\omega) \text{Re} \mathcal{G}(\omega) \} \right], \quad (38)$$

where

$$\hat{n}(\omega) = \begin{bmatrix} n(\omega) \\ f(\omega) \end{bmatrix} \quad (39)$$

is the matrix containing the Bose or Fermi–Dirac distribution functions  $n(\omega) = [e^{\beta\omega} - 1]^{-1}$ ,  $f(\omega) = [e^{\beta\omega} + 1]^{-1}$ . The trace is carried out over the Bose and Fermi components of both expressions. Here  $\mathcal{G}(\omega)$  and  $\Sigma(\omega)$  are the self-energies that have been analytically continued onto the real axis. This expression is extremely useful, for it only involves the low-energy part of the Green’s functions of the particles, and does not involve the functional  $Y[\mathcal{G}]$ . A version of this formula was first quoted in the context of lattice gauge theory by Blaizot *et al.*<sup>21</sup>

Key to our approach is the notion that the Luttinger–Ward functional  $F[\mathcal{G}]$  can be rewritten in terms of the *real* frequency Green’s functions. This has the advantage that one does not have to deal with the temperature dependence of the Matsubara frequencies. The approach that we adopt is, in fact, reminiscent of the Keldysh approach for nonequilibrium systems.<sup>32,33</sup> To preserve the symmetry between bosons and fermions, it proves useful to introduce a kind of “Keldysh” notation, writing

$$\hat{h}(\omega) = \frac{1}{2} + \hat{\eta} \hat{m}(\omega) = \begin{bmatrix} \frac{1}{2} + n(\omega) & \\ & \frac{1}{2} - f(\omega) \end{bmatrix} \quad (40)$$

Use of this function preserves the “supertrace” symmetry of our approach. Consider the first part of the free energy,

$$F_1 = T \text{Str}[\ln(-\mathcal{G}^{-1}) + \Sigma \mathcal{G}]. \quad (41)$$

We replace the summation in this term by an integral over  $\hat{n}(\omega)$  around the imaginary axis,

$$T \text{Str}[A] \rightarrow - \int \frac{d\omega}{\pi} \text{Im Str}[\hat{h}(\omega)A(\omega - i\delta)], \quad (42)$$

where  $\text{Str}[A] = \text{Tr}[A_B] - \text{Tr}[A_F]$  is the supertrace over the spatial and internal quantum numbers of  $A$ . By distorting the contour around the real axis, we obtain

$$F_1 = - \int \frac{d\omega}{\pi} \text{Im Str}[\hat{h}(\omega)\{\ln(-\mathcal{G}^{-1}) + (\mathcal{G}_0^{-1} - \mathcal{G}^{-1})\mathcal{G}\}_{\omega-i\delta}]. \quad (43)$$

If we vary the Green's function, then we obtain

$$\delta F_1 = - \int \frac{d\omega}{\pi} \text{Im Str}[\hat{h}(\omega)\tilde{\Sigma}(\omega)\delta\mathcal{G}]_{\omega-i\delta}, \quad (44)$$

where we have employed the notation  $\tilde{\Sigma} = \mathcal{G}_0^{-1} - \mathcal{G}^{-1}$ . In a similar fashion, when we vary  $\mathcal{G}$  inside  $Y[\mathcal{G}]$ , we obtain

$$\delta Y = \int \frac{d\omega}{\pi} \text{Im Str}[\hat{h}(\omega)\Sigma(\omega)\delta\mathcal{G}(\omega)]_{\omega-i\delta}. \quad (45)$$

The condition that the two terms cancel sets  $\mathcal{G}_0^{-1} - \mathcal{G}^{-1} = \Sigma$ , defining both the real and the imaginary parts of the self-energies in terms of the Green's functions.

We can exploit the stationarity  $\delta F/\delta\mathcal{G}=0$  to simplify the differentiation of the free energy with respect to the temperature. When we differentiate  $F$ , we can neglect the temperature dependence of the spectral functions,

$$S(T) = - \frac{dF}{dT} = - \left. \frac{\partial F}{\partial T} \right|_{\mathcal{G}}. \quad (46)$$

The contribution to the entropy from  $F_1$  is then

$$S_1 = \int \frac{d\omega}{\pi} \text{Im Str} \left( \frac{\partial \hat{h}(\omega)}{\partial T} \{ \ln[-\mathcal{G}^{-1}(\omega - i\delta)] + \Sigma(\omega - i\delta)\mathcal{G}(\omega - i\delta) \} \right). \quad (47)$$

The temperature derivative of the second term  $Y[\mathcal{G}]$  in (37) requires more careful consideration. The general diagrammatic contribution to  $Y$  contains a certain number of frequency summations. When we analytically continue these frequency summations, distorting the contour integrals around the branch cuts of the Green's functions, we pick up “on-shell” contributions from each branch cut of the form

$$h_B(\omega)G_B''(\omega) \quad (48)$$

and

$$h_F(\omega)G_F''(\omega). \quad (49)$$

These combinations are nothing more than the “Keldysh” Green's function well known from nonequilibrium physics.<sup>32,33</sup> In equilibrium, the Keldysh Green's function satisfies the fluctuation dissipation theorem,

$$G_K(\omega) = [G_R(\omega) - G_A(\omega)]2h(\omega) = -4iG''(\omega)h(\omega),$$

where  $G_R$  and  $G_A$  are the retarded and advanced propagators, respectively. In other words,  $h\mathcal{G}''(\omega) \equiv (i/4)G_K(\omega)$ . The point

is that we can formally imagine evaluating  $Y$  in a general steady state, replacing  $G_K \rightarrow -4ig''(\omega)h(\omega)$  at the end of the calculation. It is, however, equally consistent to compute  $Y$  in equilibrium, and substitute  $g''(\omega)h(\omega) \rightarrow g_K$ . The functional  $Y$  is closely related to the expectation value of the interaction energy. If we associate an amplitude  $\sqrt{\alpha}$  with the strength of each vertex, then the expectation value of the interaction energy has the expansion

$$\langle H_I \rangle = \sum \alpha^n Y_n(\mathcal{G}_K, \mathcal{G}_R, \mathcal{G}_A),$$

where  $\mathcal{G}_{K,R,A}$  are the fully renormalized nonequilibrium propagators. The functional  $Y$  is the weighted sum

$$Y[\mathcal{G}_K, \mathcal{G}_R, \mathcal{G}_A] = \sum \frac{\alpha^n}{n} Y_n(\mathcal{G}_K, \mathcal{G}_R, \mathcal{G}_A).$$

In this way, we can absorb all thermal functions into the Green's functions, so that all temperature dependence is entirely contained within the Keldysh Green's functions. To differentiate  $Y$  with respect to temperature, we need to determine  $\delta Y/\delta\mathcal{G}_K$ . Unfortunately,  $Y$  is not a true generating functional for the Keldysh self-energies, and we must be very careful in taking the next step.

Once we know the variation of  $Y$  with respect to the Keldysh Green's function, we can immediately compute the variation with respect to temperature. Now, from (45), we see that we may write

$$\delta Y = \int \frac{d\omega}{\pi} \text{Str}[\Sigma'(\omega)\hat{h}(\omega)\delta\mathcal{G}''(\omega) + \hat{h}(\omega)\Sigma''(\omega)\delta\mathcal{G}'(\omega)]. \quad (50)$$

If we now identify

$$\hat{h}(\omega)\mathcal{G}'' \equiv \frac{i}{4}\mathcal{G}_K,$$

$$\hat{h}(\omega)\Sigma'' \equiv \frac{i}{4}\Sigma_K, \quad (51)$$

as the Keldysh components of the propagator and self-energy, respectively, it is very tempting to write

$$\delta Y = \int \frac{d\omega}{\pi} \text{Str} \left[ \left( \frac{i}{4} \right) \Sigma'(\omega) \delta\mathcal{G}_K(\omega) + \left( \frac{i}{4} \right) \Sigma_K(\omega) \delta\mathcal{G}'(\omega) \right]. \quad (52)$$

This form would always be correct if  $Y$  were a true generating functional for Keldysh propagators. Unfortunately, in the Keldysh approach, the distinction between the “measurement” and “response” vertices means that, in general, the differential of  $Y$  with respect to  $\mathcal{G}_K$  does not generate the real part of the self-energy. Fortunately, this difficulty vanishes in the leading large- $N$  approximation, which is sufficient to generate a wide class of noncrossing approximation schemes. Diagrams involving vertex corrections do not satisfy this relationship, and there are corrections to the above diagram. Interestingly enough, however, our large- $N$  expansion does absorb all the “RPA” diagrams of the interaction lines into explicit propagators, enabling the entropy formula

to be derived for interacting collective modes and fermions.

With this proviso, we write

$$\frac{\delta Y}{\delta \mathcal{G}_K(\omega)} = \frac{i}{4} \Sigma'(\omega),$$

where  $\Sigma' = \frac{1}{2}[\Sigma(\omega - i\delta) + \Sigma(\omega + i\delta)]$ .

It is worth making a short diversion at this point to explicitly demonstrate that this relationship works in the large- $N$  limit. Consider the leading-order form for  $Y$  in the large- $N$  expansion,

$$Y_1 = K \times N \times \left( \frac{1}{\sqrt{N}} \right)^2 T^2 \sum_{k,q} G_c(k) G_\chi(q-k) G_b(q), \quad (53)$$

where  $k \equiv (i\omega_n, \vec{k})$ ,  $q \equiv (i\nu_n, \vec{q})$  is a shortened space-time notation. When we carry out the Matsubara sums in this expression, we obtain

$$Y = -K \int_{-\infty}^{\infty} \frac{d\omega}{\pi} \frac{d\nu}{\pi} \sum_{\vec{k}, \vec{q}} \left\{ G'_c G''_\chi G''_b \left( h_\chi h_b - \frac{1}{4} \right) + G''_c G'_\chi G''_b h_c h_b + G''_c G''_\chi G'_b h_c h_\chi \right\}, \quad (54)$$

where we have used the shorthand notation  $G_c \equiv G_c(k)$ ,  $h_c \equiv h_c(\omega)$ ,  $G_\chi \equiv G_\chi(q-k)$ , etc., which can be rewritten as

$$Y = \frac{K}{4} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{d\nu}{2\pi} \sum_{\vec{k}, \vec{q}} \text{Re} \{ G_{Rc} G_{K\chi} G_{Kb} + G_{Kc} G_{R\chi} G_{Kb} + G_{Kc} G_{K\chi} G_{Rb} + G_{R\chi} G_{Rc} G_{Ab} \}. \quad (55)$$

This last expression can also be derived using the Keldysh formalism to calculate the leading-order expression for  $\langle H_I \rangle$ .<sup>33</sup> The variation of this expression with respect to  $G_K$  gives the leading-order expressions for the boson  $\chi$  and conduction self-energy; thus, for instance,

$$\begin{aligned} \frac{\delta Y}{\frac{i}{4} \delta G_{Kb}(q)} &= -\frac{i}{4} k \sum_{\vec{k}} \int \frac{d\omega_c}{\pi} [G'_c(k) G_{K\chi}(q-k) \\ &\quad + G_{Kc}(k) G'_\chi(q-k)] \\ &= -k \sum_{\vec{k}} \int \frac{d\omega_c}{\pi} [G'_c(k) G''_\chi(q-k) h(q-k) \\ &\quad + G''_c(k) h(k) G'_\chi(q-k)], \end{aligned} \quad (56)$$

where the replacement  $K \rightarrow k = K/N$  occurs because  $Y$  contains the contribution of all  $N$  spin channels of  $G_{Kb}$ , and we are only differentiating with respect to one of them. This result can be independently confirmed using Matsubara techniques.

Let us now return from this digression and continue to calculate the temperature dependence of  $Y$ . When we differentiate the general expression with respect to the temperature, we only need to keep track of how each of the thermal functions changes. This leads to the contribution

$$\begin{aligned} \hat{h}(\omega) \mathcal{G}''(\omega) &= \frac{i}{4} \mathcal{G}_K(\omega) \rightarrow \hat{h}(\omega) \mathcal{G}''(\omega) + \frac{d\hat{h}(\omega)}{dT} \mathcal{G}''(\omega) \delta T \\ &= \frac{i}{4} [\mathcal{G}_K(\omega) + \delta \mathcal{G}_K(\omega)], \end{aligned} \quad (57)$$

where

$$\frac{i}{4} \delta \mathcal{G}_K(\omega) = \frac{d\hat{h}(\omega)}{dT} \mathcal{G}''(\omega) \delta T. \quad (58)$$

We can now combine these results to obtain

$$\begin{aligned} \left. \frac{\delta Y}{\delta T} \right|_{\mathcal{G}} &= \int \frac{d\omega}{\pi} \text{Str} \left[ \frac{\delta Y}{\frac{i}{4} \delta \mathcal{G}_K(\omega)} \left( \frac{d\hat{h}(\omega)}{dT} \mathcal{G}''(\omega) \right) \right] \\ &= \int \frac{d\omega}{\pi} \text{Str} \left[ \Sigma'(\omega) \left( \frac{d\hat{h}(\omega)}{dT} \mathcal{G}''(\omega) \right) \right], \end{aligned} \quad (59)$$

so that

$$S_2 = - \left. \frac{\delta Y}{\delta T} \right|_{\mathcal{G}} = - \int \frac{d\omega}{\pi} \text{Str} \left[ \left( \frac{d\hat{h}(\omega)}{dT} \right) \text{Re} \Sigma(\omega) \mathcal{G}''(\omega) \right]. \quad (60)$$

This is a key element in our proof of the entropy equation.

When we add  $S_1$  and  $S_2$  together,  $S_2$  partially cancels the second term in  $S_1$ , yielding the final answer

$$\begin{aligned} S = S_1 + S_2 &= \int \frac{d\omega}{\pi} \text{Str} \left[ \left( \frac{d\hat{h}(\omega)}{dT} \right) \{ \text{Im} \ln[-\mathcal{G}^{-1}(z)] \right. \\ &\quad \left. + \text{Im} \Sigma(z) \text{Re} \mathcal{G}(z) \}_{z=\omega-i\delta} \right]. \end{aligned} \quad (61)$$

We may replace  $\hat{h} \rightarrow \hat{n}$  and  $\text{Str} \rightarrow \text{Tr}_{B,F}$  to recover (38). There are a few important points to be made about this result.

(i) Our ability to write the entropy in terms of the Green's functions only works because in the leading-order approximation for  $Y$ , there are no hidden collective modes that carry the entropy.

(ii) This approach is general, and can, for example, be applied to the interacting electron gas, and various interacting plasmas, such as the quark gluon plasma.<sup>21</sup> By treating the interaction line as an independent particle (photon), the leading-order approximation generates a generalized RPA-Eliashberg scheme for the self-consistent interaction and electron self-energies. Note that the entropy formula cannot be used in an RPA scheme in which the interaction line is not treated as an independent particle,<sup>22</sup> and it does not work if one includes vertex corrections.

(iii) It may be possible in future work to evaluate a more general expression for  $\delta Y / \delta \mathcal{G}_K(\omega)$ , permitting one to generalize the entropy formula to higher-order approximations for  $Y$ .

#### IV. CONSERVATION LAWS AND SUM RULES

Sum rules are intimately related to the existence of conserved charges. For example, in the Kondo lattice, there are



three independent types of conserved charges: the electric charge (in each channel), the total spin, and the conserved number of bosons at each site. Each conserved charge corresponds to a gauge invariance of the Lagrangian. For example, the electric charge (for channel  $\nu$ ) is

$$Q_\nu = \sum_{\vec{k}, \alpha} c_{\vec{k}\nu\alpha}^\dagger c_{\vec{k}\nu\alpha}, \quad (62)$$

and this quantity is associated with the global gauge transformation,

$$c_{\vec{k}\nu\alpha} \rightarrow e^{i\theta_\nu(\tau)} c_{\vec{k}\nu\alpha}, \quad \chi_{j\nu} \rightarrow e^{-i\theta_\nu(\tau)} \chi_{j\nu}. \quad (63)$$

The opposite sign in the exponents expresses the opposite charge of the two fields. Similarly, the conserved number of bosons at each site,

$$Q_{bj} = \sum_{\alpha} b_{j\alpha}^\dagger b_{j\alpha}, \quad (64)$$

generates the local gauge transformation,

$$b_{j\alpha} \rightarrow e^{i\theta_b(j,\tau)} b_{j\alpha}, \quad \chi_{j\nu} \rightarrow e^{i\theta_b(j,\tau)} \chi_{j\nu}. \quad (65)$$

In phases where these symmetries are unbroken, they give rise to sum rules, which we now derive.

The general form of these gauge transformations is given by

$$\psi_\zeta \rightarrow e^{i\theta(\tau)q_\zeta} \psi_\zeta. \quad (66)$$

We can relate this gauge invariance to the conserved charge by examining how the time-derivative terms in the action behave under this transformation. In general, the time derivative part of the action has the form

$$S_0 = \int_0^\beta d\tau \psi^\dagger \underline{\gamma} \partial_\tau \psi, \quad (67)$$

where the elements  $\gamma_\zeta$  of the diagonal matrix  $\underline{\gamma}$  are unity,  $\gamma_\zeta=1$  for conventional fermions or bosons and zero,  $\gamma_\zeta=0$  for Hubbard–Stratonovich fields without any short-time dynamics, such as the Grassmanian field  $\chi_{j\mu}$  introduced in the Kondo lattice. We can relate  $\underline{\gamma}$  to the frequency dependence of  $\mathcal{G}_0^{-1}$ ,

$$\underline{\gamma} = \frac{d\mathcal{G}_0^{-1}(\omega)}{d\omega}. \quad (68)$$

Under a time-dependent gauge transformation, the change in the action is given by

$$\begin{aligned} \Delta S &= \left( \int_0^\beta d\tau \psi^\dagger e^{-i\theta\hat{q}} \underline{\gamma} \partial_\tau e^{i\theta\hat{q}} \psi \right) - S_0 = i \int_0^\beta d\tau \partial_\tau \theta(\tau) Q(\tau) \\ &= -i \int_0^\beta d\tau \theta(\tau) \partial_\tau Q(\tau), \end{aligned} \quad (69)$$

where we have commuted the diagonal matrices  $\hat{q}$  and  $\underline{\gamma}$  to obtain

$$Q = \psi^\dagger (\hat{q} \underline{\gamma}) \psi. \quad (70)$$

Invariance of the action under time-dependent gauge transformations implies that  $\partial_\tau Q(\tau)=0$ , i.e., the charge is conserved. In this way, we see that  $\hat{q} \underline{\gamma}$  is the single-particle operator associated with the charge  $Q$ . Notice that the full charge operator only depends on those fields with a time derivative. When the system is probed at short times, only these fields carry the charge. However, at long times, it is the gauge charge  $\hat{q}$  that determines the charge of physical low-energy excitations.

Using (70), we can now write the expectation value of the charge  $Q$  in the following compact form:

$$\langle \hat{Q} \rangle = \langle \psi^\dagger \hat{q} \underline{\gamma} \psi \rangle = -\frac{1}{\beta} \text{Str}[\hat{q} \underline{\gamma} \mathcal{G}]. \quad (71)$$

At zero temperature, provided there is a scale to the excitations, then the summations over discrete Matsubara frequencies can be replaced by a continuous integral. In particular,

$$T \text{Str}[A] \rightarrow \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str}[A(\omega)] e^{\omega 0^+}, \quad (72)$$

where  $\text{Str}[A]$  without an underline indicates a supertrace purely over the spatial variables of  $A$ . This enables us to write

$$\langle \hat{Q} \rangle = - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str}[\hat{q} \underline{\gamma} \mathcal{G}(\omega)], \quad (73)$$

where, for clarity, we have temporarily suppressed the convergence factor  $e^{\omega 0^+}$ . Using (68),  $\underline{\gamma} = (d\mathcal{G}_0^{-1})/(d\omega)$ , we then obtain at  $T=0$ ,

$$\langle \hat{Q} \rangle = - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \hat{q} \frac{d\mathcal{G}_0^{-1}}{d\omega} \mathcal{G} \right]. \quad (74)$$

Now let us consider the effect of the gauge transformation on the Free energy. If we carry out a gauge transformation  $\psi_\zeta \rightarrow e^{i\theta(\tau)q_\zeta} \psi_\zeta$ , with  $\theta(\tau) = \Delta\omega \tau$ , then diagrammatically this has the effect of shifting the frequencies associated with each propagator,  $\omega_\zeta \rightarrow \omega_\zeta + q_\zeta \Delta\omega$ , where  $q_\zeta$  is the gauge charge of the propagating particle. In the zero temperature limit, this shift in frequency can be made infinitesimally small,  $\Delta\omega_\zeta \rightarrow \delta\omega_\zeta$ .

Diagrammatically, the conservation of charges corresponds to the existence of closed loops (Fig. 3) in the skeleton diagrams of  $Y$  through which the conserved charge circulates. The gauge transformation causes the frequency variable in each of these closed loops to shift by an amount  $q \Delta\omega$ . This change does not affect the value of  $Y$ , and vanishing of the change in  $Y$  then gives rise to new Ward identities.

When the frequency around the closed loops is shifted, each propagator entering into  $Y$  changes, with  $\delta\mathcal{G}_\zeta(\omega) = (d\mathcal{G}/d\omega)q_\zeta \delta\omega$ , so that the change in  $Y$  at zero temperature is

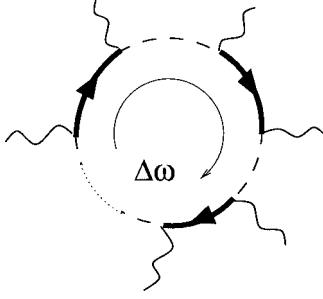


FIG. 3. Illustrating the shift in frequencies around a closed loop in  $Y[\mathcal{G}]$  for the Kondo lattice, generated by the gauge transformation associated with charge conservation.

$$\begin{aligned}\delta Y &= - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str}[\Sigma(\omega) \delta \mathcal{G}(\omega)] \\ &= - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left( \Sigma(\omega) \frac{d\mathcal{G}}{d\omega} \hat{q} \right) \delta\omega = 0,\end{aligned}\quad (75)$$

so that

$$0 = \frac{\delta Y[\mathcal{G}]}{\delta\omega} = \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \Sigma \frac{d\mathcal{G}}{d\omega} \hat{q} \right].$$

Integrating by parts, we obtain the Ward Identity,

$$\int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \frac{d\Sigma}{d\omega} \mathcal{G} \hat{q} \right] = 0. \quad (76)$$

Combining Eqs. (74) and (76), we obtain

$$\begin{aligned}\langle \hat{Q} \rangle &= - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \hat{q} \frac{d}{d\omega} \overbrace{(\mathcal{G}_0^{-1} - \Sigma)}^{\mathcal{G}^{-1}(\omega)} \mathcal{G} \right] e^{\omega 0^+} \\ &= - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \frac{d}{d\omega} \{ \text{Str}[\hat{q} \ln(-\mathcal{G}^{-1})] \} e^{\omega 0^+},\end{aligned}\quad (77)$$

where we have explicitly reinstated the convergence factor. Folding the integration contour around the negative real axis (as dictated by the convergence factor), we then obtain

$$\begin{aligned}\langle \hat{Q} \rangle &= - \int_{-\infty}^0 \frac{d\omega}{\pi} \frac{d}{d\omega} \frac{1}{2i} (\text{Str}\{\hat{q} \ln[-\mathcal{G}^{-1}(\omega - i\delta)]\} \\ &\quad - \text{Str}\{\hat{q} \ln[-\mathcal{G}^{-1}(\omega + i\delta)]\}) \\ &= - \int_{-\infty}^0 \frac{d\omega}{\pi} \frac{d}{d\omega} \text{Im Str}\{\hat{q} \ln[-\mathcal{G}^{-1}(\omega - i\delta)]\} \\ &= - \left[ \frac{1}{\pi} \text{Im Str}\{\hat{q} \ln[-\mathcal{G}^{-1}(\omega - i\delta)]\} \right]_{-\infty}^0.\end{aligned}\quad (78)$$

The presence of the minus sign inside the logarithms here is chosen so that the lower bound of the integral vanishes. We are then able to write

$$\langle \hat{Q} \rangle = - \frac{1}{\pi} \text{Im Str}\{\hat{q} \ln[-\mathcal{G}^{-1}(\omega - i\delta)]\} \Big|_{\omega=0}. \quad (79)$$

As it stands, the trace in this expression involves both bosonic and fermionic components. The latter will give rise to Fermi surface volume contributions in the lattice, and to phase shift terms in the case of impurity models. What about the bosons? General arguments lead us to believe that the bosonic component in the trace vanishes. The important point here is that the right-hand side of this sum counts the number of one-particle states that drop to *negative energy*. Whereas fermions can acquire a negative energy through the formation of a resonance or Fermi surface, bosons cannot—they only condense. If the bosons condense, then the symmetry they are associated with is broken, and we can no longer apply the sum rule. If they do not condense, then we expect the associated phase shifts to be zero. The final form of the sum rule is then

$$\langle \hat{Q} \rangle = \frac{1}{\pi} \text{Im Tr}[\hat{q} \ln(-\mathcal{G}_F^{-1})]_{\omega=0-i\delta}, \quad (80)$$

where, for clarity, we have suppressed the frequency argument of the Green's function.

## V. KONDO IMPURITY MODEL

As a first illustration of the sum rules, let us apply them to the Kondo impurity model. There are two sum rules to consider here, corresponding to the conservation of  $Q_c$  and  $n_b = 2S = K$  in the case of the perfectly screened solution. Let us first consider

$$n_b = \sum_{\alpha} b_{\alpha}^{\dagger} b_{\alpha}. \quad (81)$$

Even though the  $\chi$  fermions do not enter into this conserved charge, they carry the gauge charge, and we have  $(q_c, q_b, q_{\chi}) = (0, 1, 1)$ . Applying the sum rule (79), we deduce

$$\langle n_b \rangle = - \frac{1}{\pi} \text{Im Tr}[\ln(-\mathcal{G}_b^{-1})]_{\omega=i\delta} + \frac{1}{\pi} \text{Im Tr}[\ln(-\mathcal{G}_{\chi}^{-1})]_{\omega=i\delta}, \quad (82)$$

where, for the moment, we have retained the bosonic trace. When we carry out the trace over the spin components of the boson and the channel components of the  $\chi$  fermion, we obtain

$$\langle n_b \rangle = -N \frac{\delta_b}{\pi} + K \frac{\delta_{\chi}}{\pi}, \quad (83)$$

where

$$\begin{aligned}\delta_b &= \text{Im}[\ln(\lambda + \Sigma_b(0 - i\delta) + i\delta)], \\ \delta_{\chi} &= \text{Im} \left[ \ln \left( \frac{1}{J_K} + \Sigma_{\chi}(0 - i\delta) \right) \right],\end{aligned}\quad (84)$$

are the “phase shifts” associated with the boson and holon fields. Note in passing that at zero temperature,  $\Sigma_b(0 - i\delta)$

and  $\Sigma_\chi(0-i\delta)$  are purely real. The argument of the logarithm in  $\delta_b$  is the renormalized chemical potential  $\lambda^* = \lambda + \Sigma_b(0)$  of the boson field. In the “filled shell” configuration  $2S=K$ , the formation of a perfectly screened Kondo singlet generates a gap for the addition of extra Schwinger bosons to the ground state, which forces  $\lambda^* > 0$ , and  $\delta_b = 0$ , as argued previously. It follows that

$$\langle n_b \rangle = K \frac{\delta_\chi}{\pi}. \quad (85)$$

But  $n_b = K$  is part of the constraint, so it follows that  $\delta_\chi = \pi$ . This result can be understood by relating  $\delta_\chi$  to the sign of the renormalized Kondo coupling constant, as follows:

$$\frac{\delta_\chi}{\pi} = \frac{1}{\pi} \operatorname{Im} \left[ \ln \left( \frac{1}{J_K} + \Sigma_\chi(0-i\delta) \right) \right] = \theta(-J_K^*). \quad (86)$$

In this way, we see that  $\delta_\chi = \pi$  corresponds to a residual “ferromagnetic” coupling. This is consistent with our expectations, for when an additional Schwinger boson is added to the Fermi liquid, it increases the impurity spin by one-half unit to form an underscreened Kondo model, where the residual spin coupling is indeed ferromagnetic. By contrast, in an antiferromagnet or spin liquid, the local moments spins pair condense, mutually screening one another. In this situation, an additional Schwinger boson at any one site will now be free to undergo a Kondo effect with the conduction electrons, so the residual spin coupling is antiferromagnetic. These results are also confirmed in the large- $N$  limit.

Let us now turn to the sum rule derived from the conservation of electron charge. Here the conserved quantity is

$$Q_e = \sum_{k,v,\sigma} c_{k,v,\sigma}^\dagger c_{k,v,\sigma}, \quad (87)$$

and the corresponding gauge charges are  $(q_c, q_b, q_\chi) = (1, 0, -1)$ , so the sum rule becomes

$$Q_e = \frac{1}{\pi} \left( \operatorname{Im} \operatorname{Tr} [\ln(-\mathcal{G}_c^{-1})] - \frac{1}{\pi} \operatorname{Im} \operatorname{Tr} [\ln(-\mathcal{G}_\chi^{-1})] \right) \Big|_{\omega=0-i\delta}. \quad (88)$$

Now in an impurity model, it is convenient to subtract off the total charge in the absence of the impurity, given by

$$Q_e^{(0)} = \frac{1}{\pi} \operatorname{Im} \operatorname{Tr} [\ln(-\mathcal{G}_{c0}^{-1})]_{\omega=0-i\delta}, \quad (89)$$

where  $\mathcal{G}_{c0} \equiv \mathcal{G}_c^{(0)}$  is the noninteracting propagator of the conduction electrons. If we subtract this from (88), we can combine the conduction electron traces and replace  $\ln(-\mathcal{G}_c^{-1}) - \ln(-\mathcal{G}_{c0}^{-1}) = \ln(\mathcal{G}_{c0}\mathcal{G}_c^{-1}) = \ln(1 - \mathcal{G}_{c0}\Sigma_c)$ . The change in total charge is then given by

$$Q_e - Q_e^{(0)} = \frac{1}{\pi} \operatorname{Im} \operatorname{Tr} [\ln(1 - \mathcal{G}_{c0}\Sigma_c)]_{\omega=0-i\delta} - \frac{1}{\pi} \operatorname{Im} \operatorname{Tr} [\ln(-\mathcal{G}_\chi^{-1})]_{\omega=0-i\delta}. \quad (90)$$

According to the Anderson-Clogston compensation theorem, the total change in electron charge due to a Kondo or Ander-

son impurity vanishes in the infinite bandwidth limit, and at finite bandwidth the change is of the order of the ratio of the Kondo temperature to the bandwidth  $\Delta Q_c = O(T_K/D)$ , and can be neglected.<sup>34,35</sup> Setting  $\Delta Q_e = Q_e - Q_e^{(0)} = 0$ , we obtain

$$\Delta Q_e = 0 = NK \frac{\delta_c}{\pi} - K \frac{\delta_\chi}{\pi}, \quad (91)$$

where we have identified

$$\delta_c = \operatorname{Im} \operatorname{Tr}_k [\ln(1 - \mathcal{G}_{c0}\Sigma_c)]_{\omega=0-i\delta}, \quad (92)$$

where the trace is purely over momenta. If we expand the logarithm using a power series, we see that the trace over internal momenta in (92) is accomplished by replacing the momentum-dependent Green's function  $\mathcal{G}_c^{(0)}(\vec{k}, \omega)$  by the local electron Green's function  $g_0(\omega) = \Sigma_k \mathcal{G}_c^{(0)}(\vec{k}, \omega)$ , so that

$$\delta_c = \operatorname{Im} \ln[1 - g_0(\omega)\Sigma(\omega)]_{\omega=0-i\delta}. \quad (93)$$

In the large bandwidth limit,  $g_0(\omega-i\delta) \rightarrow i\pi\rho$ , where  $\rho$  is the density of states per spin per channel, so that

$$\delta_c = \operatorname{Im} \ln[1 - i\pi\rho\Sigma(0)]. \quad (94)$$

There is thus a direct link between the phase shift of the  $\chi$  fermion and that of the conduction electrons. Combining this with the earlier result  $\delta_\chi = \pi$ , we obtain

$$\delta_c = \frac{\delta_\chi}{N} = \frac{\pi}{N}. \quad (95)$$

Notice the following.

(i) In the large- $N$  limit, the phase shift identity permits us to relate the conduction electron phase shift  $\delta_c$ , which is  $O(1/N)$  to the  $\chi$  phase shift  $\delta_\chi$ , which is finite in the large  $N$  limit. In this way, the sum rules enable us to study the phase shift and Fermi surface volume changes of the Kondo impurity and lattice, in the large- $N$  limit.

(ii) The development of the conduction electron phase shift does not occur because conduction electron states drop beneath the Fermi sea: It is a consequence of the injection of new quasiparticle states into the Fermi sea as a response to the formation of spinless, charged holons.

## VI. SUM RULE FOR THE KONDO LATTICE

When we come to consider the Kondo lattice, there are a number of additional subtleties that must be entertained. One of the most important aspects of the discussion here concerns the possibility that the extremal solutions to  $F[\mathcal{G}]$  break the  $U(1)$  symmetry of the bosons, to produce “Higg's” phase in which the Schwinger bosons either pair condense, or hybridize between sites. In the large- $N$  solutions to the model, these phases appear to develop as a precursor to the formation of the antiferromagnet, where the bosons themselves condense. The main effect on the sum rules is twofold:

(i) We cannot assume that the  $\chi$  fermion is localized. This implies that we must consider the momentum dependence of the  $\chi$  fermion.

(ii) Once the Schwinger bosons pair condense, the Ward identities associated with the conservation of  $n_b$  no longer

apply, because the Luttinger–Ward functional  $Y[\mathcal{G}]$  is no longer invariant under shifts of the boson frequency.

However, we can still take advantage of the Ward identity associated with charge conservation. Provided that the Schwinger boson is uncondensed, i.e., there is no magnetism, then charge conservation guarantees that

$$\int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Tr}_F \left[ \frac{d\Sigma}{d\omega} \mathcal{G} \hat{q} \right] = 0 \quad (q_c = 1, q_\chi = -1). \quad (96)$$

Now in the paramagnet, the holon and conduction electron fields are unmixed, so we can replace

$$\text{Tr}_F \rightarrow \text{Tr}_c + \text{Tr}_\chi.$$

This is a special property of the paramagnet. By contrast, in the antiferromagnet, the Schwinger bosons are condensed, so the holon and conduction electron fields admix, and both  $\Sigma$  and  $\mathcal{G}$  contain off-diagonal terms so that  $\text{Tr}_F$  remains a single, integral trace over the two admixed fields.

For the paramagnet, we can write

$$\begin{aligned} 0 &= - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Str} \left[ \frac{d\Sigma}{d\omega} \mathcal{G} \hat{q} \right] \quad (q_c = 1, q_\chi = -1) \\ &= \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Tr} \left[ \frac{d\Sigma_c}{d\omega} \mathcal{G}_c \right] - \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \text{Tr} \left[ \frac{d\Sigma_\chi}{d\omega} \mathcal{G}_\chi \right] \\ &= K \int_{-i\infty}^{i\infty} \frac{d\omega}{2\pi i} \int \frac{d^D k}{(2\pi)^D} \left[ N \frac{d\Sigma_c(\vec{k}, \omega)}{d\omega} \mathcal{G}_c(\vec{k}, \omega) \right. \\ &\quad \left. - \frac{d\Sigma_\chi(\vec{k}, \omega)}{d\omega} \mathcal{G}_\chi(\vec{k}, \omega) \right]. \end{aligned} \quad (97)$$

It is this identity that permits us to extend the Luttinger sum rule to the case of the paramagnetic Kondo lattice. In the paramagnetic Kondo lattice, the charge sum rule

$$Q_e = \frac{1}{\pi} \{ \text{Im Tr}[\ln(-\mathcal{G}_c^{-1})] - \text{Im Tr}[\ln(-\mathcal{G}_\chi^{-1})] \} \Big|_{\omega=0-i\delta}, \quad (98)$$

now involves a trace over momentum:

$$\begin{aligned} \frac{Q_e}{K} &= N \sum_{\vec{k}} \frac{1}{\pi} \text{Im} \ln[ \epsilon_{\vec{k}} + \Sigma_c(\vec{k}, 0 - i\delta) + i\delta ] \\ &\quad - \sum_{\vec{k}} \frac{1}{\pi} \text{Im} \ln \left( \frac{1}{J_K} + \Sigma_\chi(\vec{k}, 0 - i\delta) \right). \end{aligned} \quad (99)$$

The first term in this expression is the electron Fermi surface volume,

$$\sum_{\vec{k}} \frac{1}{\pi} \text{Im} \ln[ \epsilon_{\vec{k}} + \Sigma_c(\vec{k}, 0 - i\delta) + i\delta ] = \sum_{\vec{k}} \theta(-E_{\vec{k}}) = \frac{v_{FS}}{(2\pi)^D}, \quad (100)$$

where the region where  $E_{\vec{k}} = \epsilon_{\vec{k}} + \text{Re} \Sigma_c(\vec{k}, E_{\vec{k}})$  is negative, defines the interior of the Fermi surface. The second term can be interpreted in a similar way—the momentum trace over the logarithm of  $\mathcal{G}_\chi^{-1}$ ,

$$\sum_{\vec{k}} \frac{1}{\pi} \text{Im} \ln \left( \frac{1}{J_K} + \Sigma_\chi(\vec{k}, 0 - i\delta) \right) = \sum_{\vec{k}} \theta[-J_K^*(\vec{k})] = \frac{v_\chi}{(2\pi)^D}, \quad (101)$$

can be seen as the volume of the region in momentum space where the effective interaction  $J_K^*(\vec{k}) = -\mathcal{G}_\chi(\vec{k}, 0)^{-1}$  is negative, or ferromagnetic. Following our earlier discussion, the ferromagnetic sign of the residual interaction is a consequence of the fact that additional spins added to the state completely decouple from the Fermi sea. In the simplest scenario,  $J_K^*(\vec{k}) < 0$  for all  $\vec{k}$ , in which case  $v_\chi/(2\pi)^D = 1$ , and the sum rule becomes

$$n_e = \frac{Q_e}{K} = N \frac{v_{FS}}{(2\pi)^D} - 1, \quad (102)$$

where  $n_e$  is the electron density per unit cell, per conduction electron channel. It follows that the total Fermi surface volume expands by one unit per unit cell,

$$N \frac{v_{FS}}{(2\pi)^D} = n_e + 1. \quad (103)$$

Now if there is some region of momentum space where  $J_K^*(\vec{k})$  is not ferromagnetic, it follows that the  $\chi$  fermions will have a Fermi surface, and the excitation spectrum will now involve charged, spinless fermions. In this phase,

$$n_e = \frac{Q_e}{K} = N \frac{v_{FS}^*}{(2\pi)^D} - \frac{v_\chi}{(2\pi)^D}. \quad (104)$$

Such a “spin-charge decoupled” phase could not develop from a Fermi liquid without a phase transition. As long as the heavy electron fluid does not undergo such a phase transition, the Fermi surface must remain large.

Let us now consider what happens in the antiferromagnet. In the general higher-dimensional case, the boson field condenses, causing the conduction and  $\chi$  fields to hybridize to produce a single species. In the antiferromagnet, it is more logical to make a particle-hole transformation of the  $\chi$  field, writing  $\psi_{\vec{k}} = \chi_{-\vec{k}}^\dagger$ . In this case,  $q_\psi = q_c$  have the same charge, and the sum rule becomes simply

$$Q_e = \frac{1}{\pi} \text{Im Tr}_F[\ln(-\mathcal{G}_F^{-1})] \Big|_{\omega=-i\delta}, \quad (105)$$

where  $\mathcal{G}_F$  is the admixed propagator for the combined conduction and  $\psi$  fields. The right-hand side cannot be separated into  $\chi$  and conduction parts, and, as such, defines an admixed set of Fermi surfaces, with an average Fermi surface volume that counts the total charge per unit cell,

$$n_e = N \frac{\langle v_{AFM} \rangle}{(2\pi)^D}. \quad (106)$$

The Fermi surface volume for the antiferromagnet and the paramagnet must then differ by  $1/N$ th of a unit cell. In the special case of  $N=2$ , if the spatial unit cell doubles, this formal change in Fermi surface volume is indistinguishable from the new size of the Brillouin zone. However, even in this case, we can imagine more general classes of antiferro-



magnet, such as an incommensurate helimagnet, where the unit cell size is unchanged, but the total Fermi surface volume must jump.

## VII. SPIN-CHARGE SEPARATION AND QUANTUM CRITICALITY

Although the sum rules do not provide us with any details of the dynamics, they provide stringent constraints on the way the spectrum of excitations can evolve in the Kondo lattice as we approach the magnetic quantum critical point.

One of the fascinating aspects of our sum rule,

$$n_e = N \frac{v_{FS}}{(2\pi)^D} - \frac{v_\chi}{(2\pi)^D}, \quad (107)$$

is that it suggests that a Kondo lattice may develop low-lying, spinless charged fermionic excitations near a quantum critical point. This is clearly a controversial idea, the consequences of which we now explore in this discussion. If we take this idea seriously, then two important points seem to emerge.

(i) In the heavy Fermi liquid, spinons have “ionized” into electrons and a background filled sea of holons,

$$b_\sigma \leftrightarrow e_\sigma^- + \chi^+ \quad (108)$$

It is this ionization process that lies behind the expansion of the Fermi surface and the sum rule.

(ii) There is an intimate link between the Kondo interaction and the formation of holons. The link between the propagator of the holons and the renormalized Kondo interaction,

$$-\mathcal{G}_\chi(\vec{q}, \omega) = \frac{1}{\frac{1}{J_K} + \Sigma_\chi(\vec{q}, \omega)} = J_K^*(\vec{q}, \omega), \quad (109)$$

implies that poles in the holon spectrum correspond to divergences in the Kondo interaction. As magnetism drives the Kondo interaction back to weak coupling, we expect  $J^*(\vec{q}, \omega)$  to pass through infinity, so the holons should become gapless at some point in this process.

We can, in fact, use the Luttinger–Ward approach to gain insight into the low-energy thermodynamics of the Kondo lattice.<sup>3</sup> Provided the replacement of frequency sums by continuous integrals is valid at low-temperatures, Luttinger argued that the low-energy thermodynamics is determined by the leading logarithm in the free energy functional (4), evaluated with the zero-temperature Green’s function. In our case, this leads to the relation

$$F(T) \approx \sum T \text{Str}[\ln(-\mathcal{G}_{(0)}^{-1})] \quad (T \rightarrow 0),$$

where  $\mathcal{G}_{(0)}$  is the zero temperature Green’s function. Carrying out the frequency integrals, this implies that the low-temperature entropy is given by

$$\begin{aligned} S(T) = & \int \frac{d^D k}{(2\pi)^D} \int \frac{d\omega}{\pi} \overbrace{\left( \frac{dn}{dT} \text{Im Tr} \ln -\mathcal{G}_b^{-1}(\vec{k}, \omega) + \frac{df}{dT} \text{Im Tr} \ln -\mathcal{G}_\chi^{-1}(\vec{k}, \omega) \right)}^{\text{spinon and holon entropy}} \\ & + \int \frac{d^D k}{(2\pi)^D} \int \frac{d\omega}{\pi} \underbrace{\left( \frac{df}{dT} \text{Im Tr} \ln -\mathcal{G}_c^{-1}(\vec{k}, \omega) \right)}_{\text{Fermi liquid}} \end{aligned} \quad (110)$$

where the zero-temperature propagators are to be used. The last term in this expression provides the  $T$ -linear entropy of the Fermi liquid; the first two terms are the “spinon” and “holon” contributions. For a Landau Fermi liquid to form, these last terms must clearly be gapped at low temperatures. This feature is observed in the large- $N$  solutions.<sup>26</sup> What it also makes clear, however, is that the holons cannot become gapless without a thermodynamic departure from Fermi liquid behavior, which implies some kind of quantum phase transition.

However, even though the holons are gapped in the Fermi liquid, they can exist as low-lying spinless, charged excitations. It is particularly interesting to speculate that the gap to holon formation closes at a quantum critical point. This is a possibility that can be studied in the large- $N$  limit, but the general arguments should hold at finite  $N$ , as long as the

Luttinger–Ward approach is valid. One way to go beyond the large- $N$  limit is to carry out a fully self-consistent treatment of the leading Luttinger–Ward free energy functional, updating the conduction electron self-energy and feeding the full conduction electron Green’s functions into the calculation of the self-energy for the  $\chi$  and Schwinger boson field. Such an approximation goes selectively beyond the large- $N$  limit, but it will satisfy the Ward identities. Moreover, assuming that the “filled shell” stability of the fully screened ground state is not an artifact of the large- $N$  limit, we also expect the gap in the Schwinger boson and the  $\chi$  fermion particle spectrum to be preserved at finite  $N$ .

Another way to check the picture emerging from this approach would be to examine low-dimensional systems. Various authors have examined the possibility of a transition from a small to large Fermi surface in the one-dimensional

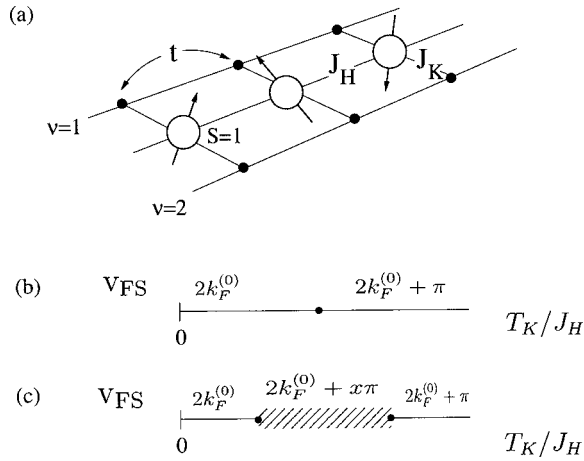


FIG. 4. (a) The two-channel, spin  $S=1$  Kondo chain. (b) In the simplest scenario, the transition from a Haldane chain with a small Fermi surface, to the heavy electron phase with a large Fermi surface will occur via a single phase transition. (c) The sum rules allow for an intermediate phase with a Fermi surface that lies between these two extremes.

spin-1/2 Kondo lattice.<sup>36,37</sup> The difficulty with this model is that the gaplessness of the Heisenberg spin-1/2 chain tends to make the Kondo coupling relevant, no matter how small the Kondo temperature. A model that avoids these difficulties is the  $S=1$ , two-channel Kondo lattice. This model involves

a Haldane spin-1 chain coupled via Kondo interactions to two, one-dimensional, noninteracting Hubbard models, as shown in Fig. 4.

When the Kondo interaction is weak, the Haldane chain is gapped, and the “Fermi surface” of the Kondo lattice is small, giving rise to Friedel oscillations at wave vector  $Q = 2k_F^{(0)}$ . When the Kondo interaction is strong, a Kondo lattice with a large Fermi surface is expected to form where  $Q = 2k_F^{(0)} + \pi$ . The way these two phases are linked is particularly interesting. If they are linked by a single quantum critical point, then we expect  $Q$  to simply jump at this point [Fig. 4(b)]. However, if the holons form a gapless phase, then over an intermediate range of parameters, we would expect the Friedel oscillations to exhibit an intermediate wave vector [Fig. 4(c)],

$$Q = 2k_F^{(0)} + x\pi. \quad (111)$$

## ACKNOWLEDGMENTS

We are grateful to Mathew Fisher, Kevin Ingersent, Andreas Ludwig, Catherine Pépin, and Gergely Zarand for discussions related to this work. This research was supported by the National Science Foundation Grants No. NSF DMR 0312495 and No. PHY99-07949. The authors are thankful for the hospitality of the KITP, where most of this research was carried out.

- <sup>1</sup>J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).
- <sup>2</sup>J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).
- <sup>3</sup>J. M. Luttinger, Phys. Rev. **121**, 1251 (1961).
- <sup>4</sup>J. Friedel, Philos. Mag. **43**, 153 (1952).
- <sup>5</sup>J. S. Langer and V. Ambegaokar, Phys. Rev. **121**, 1090 (1961).
- <sup>6</sup>D. C. Langreth, Phys. Rev. **150**, 516 (1966).
- <sup>7</sup>A. A. Abrikosov, Physics (Long Island City, N.Y.) **2**, 5 (1964).
- <sup>8</sup>H. Suhl, Phys. Rev. **138**, A515 (1965).
- <sup>9</sup>G. Gruner and A. Zawadowski, in *Progress in Low Temperature Physics*, edited by D. F. Brewer (North-Holland, Amsterdam, 1978), Vol. VIIB, p. 591.
- <sup>10</sup>R. M. Martin, Phys. Rev. Lett. **48**, 362 (1982).
- <sup>11</sup>B. I. Shraiman and E. D. Siggia, Phys. Rev. Lett. **62**, 1564 (1989).
- <sup>12</sup>B. L. Altshuler, A. V. Chubukov, A. Dashevskii, A. A. M. Finkel'stein, and D. K. Morr, Europhys. Lett. **41**, 401 (1998).
- <sup>13</sup>P. Coleman, C. Pépin, Q. Si, and R. Ramazashvili, J. Phys.: Condens. Matter **13**, R723 (2001).
- <sup>14</sup>T. Senthil, M. Vojta, and S. Sachdev, Phys. Rev. B **69**, 035111 (2004).
- <sup>15</sup>C. Pépin, Phys. Rev. Lett. **94**, 066402 (2005).
- <sup>16</sup>S. Kawarazaki *et al.*, Phys. Rev. B **61**, 4167 (2000).
- <sup>17</sup>S. Paschen, T. Lühmann, S. Wirth, P. Gegenwart, O. Trovarelli, Ch. Geibel, F. Steglich, P. Coleman, and Q. Si, Nature (London) **432**, 881 (2004).
- <sup>18</sup>M. Oshikawa, M. Yamanaka, and I. Affleck, Phys. Rev. Lett. **78**, 1984 (1997).
- <sup>19</sup>M. Oshikawa, Phys. Rev. Lett. **84**, 3370 (2000).
- <sup>20</sup>G. M. Eliashberg, Zh. Eksp. Teor. Fiz. **43**, 1005 (1962) [Sov. Phys. JETP **16**, 780 (1963)].
- <sup>21</sup>J.-P. Blaizot, E. Iancu, and A. Rebhan, Phys. Rev. D **63**, 065003 (2001).
- <sup>22</sup>G. M. Carneiro and C. J. Pethick, Phys. Rev. B **11**, 1106 (1975).
- <sup>23</sup>G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961); G. Baym, *ibid.* **127**, 1391 (1962).
- <sup>24</sup>O. Parcollet and A. Georges, Phys. Rev. Lett. **79**, 4665 (1997).
- <sup>25</sup>S. Powell, S. Sachdev, and H. P. Buchler, cond-mat/0502299 (unpublished).
- <sup>26</sup>J. Rech, P. Coleman, O. Parcollet, and G. Zarand, cond-mat/0507001 (unpublished).
- <sup>27</sup>P. Coleman and C. Pépin, Phys. Rev. B **68**, 220405 (2003).
- <sup>28</sup>P. Coleman and I. Paul, Phys. Rev. B **70**, 035111 (2004).
- <sup>29</sup>P. W. Anderson and Z. Zou, Phys. Rev. Lett. **60**, 132 (1988).
- <sup>30</sup>To avoid the ambiguities of unit cells, it is important here that the  $\vec{Q}$  vector does not have certain values. For a helimagnet, for example, we must avoid the commensurate case where  $\vec{Q} = (\pi/N, \pi/N, \pi/N)$ , where  $N=2$  is the parameter for an SP (2N) generalization.
- <sup>31</sup>M. Potthoff, cond-mat/0406671 (unpublished).
- <sup>32</sup>J. Rammer and H. Smith, Rev. Mod. Phys. **58**, 323 (1986).
- <sup>33</sup>A. Kamenev, cond-mat/0412296 (unpublished).
- <sup>34</sup>A. M. Clogston and P. W. Anderson, Bull. Am. Phys. Soc. **6**, 124 (1961).
- <sup>35</sup>For a detailed discussion of this point, see Appendix A in P. Coleman and I. Paul, Phys. Rev. B **70**, 035111 (2004).
- <sup>36</sup>J. C. Xavier, E. Novais, and E. Miranda, Phys. Rev. B **65**, 214406 (2002).
- <sup>37</sup>E. Pivovarov and Q. Si, cond-mat/0304129 (unpublished).