The Kondo effect, the Kondo lattice, and the heavy Fermi liquid

Subir Sachdev

Department of Physics, Harvard University,
Cambridge, Massachusetts, 02138, USA

School of Natural Sciences,
Institute for Advanced Studies,
Princeton, NJ 08540, USA

Topological Aspects of Strong Correlations and Gauge Theories

International Centre for Theoretical Sciences
Tata Institute of Fundamental Research
Bengaluru, Sep 6, 2021

The Kondo model was introduced to describe the behavior of impurities of transition metal ions (*e.g.* Mn) in simple metals (*e.g.* Cu). It was observed that the impurity ion acquires a local magnetic moment which interacts non-trivially with the host conduction electrons. More recently, similar models have also been used to describe small quantum dots coupled to mobile electrons in leads. The importance of the Kondo model in condensed matter physics rests on its central role in the development of our understanding of the consequences of strong interactions in metals.

I. RESONANT LEVEL MODEL

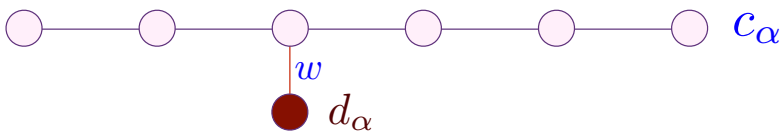

FIG. 1. Resonant level model of free electrons: conduction electrons c_α hybridize with a localized d_α state with amplitude w. The c_α move on a $d > 1$ dimensional lattice, although only one dimension is shown.

We begin by a simple model of non-interacting electrons on a lattice with a single impurity (see Fig. 1)

$$H_{\rm RLM} = \sum_{\mathbf{k},\alpha} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + \sum_\alpha \left[\varepsilon_d \, d^\dagger_\alpha d_\alpha - w \left(d^\dagger_\alpha c_{0\alpha} + c^\dagger_{0\alpha} d_\alpha \right) \right] \tag{1}$$

The $c_{\mathbf{k}\alpha}$ are the conduction electrons with dispersion $\varepsilon_{\mathbf{k}}$ in a perfect lattice. They scatter off an additional impurity atom at the origin of spatial co-ordinates, and the electron on the impurity atom is represented by d_α. The scattering is represented by the tunneling matrix element w between the impurity atom and the lattice site at the origin, with

$$c_{0\alpha} \equiv \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} c_{\mathbf{k}\alpha} \, . \tag{2}$$

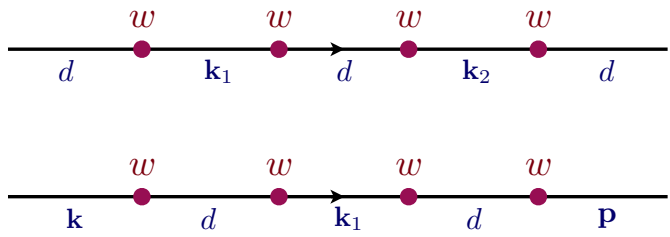

FIG. 2. Feynman diagrams for G_{dd} and the conduction electrons.

This is a model of non-interacting electrons, and despite the lack of translational invariance, it is not difficult to solve it exactly. We can sum all the Feynman diagrams in powers of w, as shown in Fig. 2, and obtain the Green's function of the d electron

$$\begin{aligned}[G_{dd}(i\omega_n)]^{-1} &= i\omega_n - \varepsilon_d - \frac{1}{V}\sum_{\mathbf{k}} \frac{|w|^2}{i\omega_n - \varepsilon_{\mathbf{k}}} \\ &= i\omega_n - \varepsilon_d - \int d\varepsilon \frac{d(\varepsilon)|w|^2}{i\omega_n - \varepsilon}\,, \end{aligned} \tag{3}$$

where $d(\varepsilon)$ is the single spin density of states of the conduction electrons. We approximate the density of states by its value of the Fermi level, $d(0)$, and absorb the real part of the integral over ε into a renormalization of ε_d. Then we obtain the final answer

$$G_{dd}(i\omega_n) = \frac{1}{i\omega_n - \varepsilon_d + i\Gamma \mathrm{sgn}(\omega_n)} \tag{4}$$

where

$$\Gamma = \pi w^2 d(0)\,. \tag{5}$$

The density of electronic states on the d site is now seen to be Lorentzian of width Γ

$$\begin{aligned}\rho_d(\omega) &= -\frac{1}{\pi}\mathrm{Im}\, G_{dd}(\omega + i\eta) \\ &= \frac{1}{\pi}\frac{\Gamma}{(\omega - \varepsilon_d)^2 + \Gamma^2} \end{aligned} \tag{6}$$

This is the resonant level: the d electron is mostly on the d site with energy ε_d, but it has a lifetime of $1/\Gamma$, as it can 'decay' by *coherent* tunneling into the conduction band. The quotes around decay, indicate that this is not an incoherent decay involving exchange of energy with other electrons, and the single

particle eigenstates of H_{RLM} are infinitely long lived, albeit on plane waves.

We can also obtain the conduction electron Green's function, which is now not diagonal in $\mathbf{k}$:

$$G(\mathbf{k},\mathbf{p},i\omega_n) = \frac{\delta_{\mathbf{k},\mathbf{p}}}{i\omega_n - \varepsilon_{\mathbf{k}}} + \frac{w^2}{V}\frac{G_{dd}(i\omega_n)}{(i\omega_n - \varepsilon_{\mathbf{k}})(i\omega_n - \varepsilon_{\mathbf{p}})} \tag{7}$$

With these expressions, we can establish a nice result for the change in the electron density due to the presence of the d level, which is reminiscent of the Luttinger relation for a Fermi liquid. In the present context, this relation is often called the Friedel sum rule, when expressed in terms of the phase shift of the scattering of the conduction electrons from the impurity; the connection to the Luttinger relation will be further explored in Section V B. The change in the electron density (with a factor of 2 for spin) is

$$\begin{aligned}
\delta\rho_e &= 2\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[G_{dd}(i\omega) + \sum_{\mathbf{k}}\left(G(\mathbf{k},\mathbf{k},i\omega) - \frac{1}{i\omega - \varepsilon_{\mathbf{k}}}\right)\right]e^{i\omega 0^+} \\
&= 2\int_{-\infty}^{\infty}\frac{d\omega}{2\pi} i\frac{d}{d\omega}\ln\left[G_{dd}(i\omega)\right]e^{i\omega 0^+} \\
&= -2i\int_{-\infty}^{0}\frac{dz}{2\pi}\frac{\partial}{\partial z}\ln\left[\frac{G_{dd}(z+i0^+)}{G_{dd}(z+i0^-)}\right] \\
&= \frac{-i}{\pi}\ln\left[\frac{G_{dd}(i0^+)}{G_{dd}(i0^-)}\right] \\
&= 1 + \frac{2}{\pi}\tan^{-1}\left[\frac{\mathrm{Re}\, G_{dd}^{-1}(i0^+)}{\mathrm{Im}\, G_{dd}^{-1}(i0^+)}\right] .
\end{aligned} \tag{8}$$

The expression (8) is exact, and does not rely on the 'flat density of states' approximations used to obtain (4). In terms of (4) we have

$$\delta\rho_e = 1 - \frac{2}{\pi}\tan^{-1}\left[\frac{\varepsilon_d}{\Gamma}\right] \tag{9}$$

The right hand side of (9) is $1/\pi$ times the scattering phase shift of the conduction electrons at the Fermi level [1]. As expected, the density varies from $\delta\rho_e = 2$, when the d level is far below the Fermi level, to $\delta\rho_e = 0$, when the d level is far above the Fermi level.

Unlike the case of the Fermi liquid, notice now that $\mathrm{Im}\, G_{dd}^{-1}(z)$ does not vanish as $z \to 0$, but equals the non-zero constant Γ.

This is a consequence of the lack of translational invariance, not the breakdown of the quasiparticle concept. The width of the resonant level, Γ, signifies a coherent mixing of the localized d level state with the continuum of conduction electron states. It is not a measure of the scattering of quasiparticles which exchange energy. So in disordered systems, we have to use more complicated correlators to deduce the lifetime of the true quasiparticles, which are not momentum eigenstates.

For subsequent considerations when we include the effect of interactions, it is useful to characterize the system by its response to a uniform applied Zeeman field coupling as $-h\sum_i S_{zi}$, where $\vec{S}_i$ is the electron spin operator on the site i. We characterize the response by the local spin susceptibility $\chi_i = \langle S_{zi}\rangle/h$. For a system without an impurity (or far from the impurity, when present), this spin susceptibility can be computed just like the compressibility, and we obtain the Pauli susceptibility of a metal, χ_P (which is the spin susceptibility per unit volume)

$$\chi_P = \frac{d(0)}{2}\,. \tag{10}$$

With an impurity, we are interested in the behavior of χ_i in the vicinity of the impurity. In the 'flat density of states' approximation, which was made between (3) and (4), the response on the d site is dominated entirely by the response to the local field: then we can then easily compute the impurity susceptibility on the d site as

$$\begin{aligned}\chi_{\rm imp} &= -\frac{1}{2}\sum_{\omega_n}\left[G_{dd}(i\omega_n)\right]^2 \\ &= \frac{\Gamma}{2\pi(\varepsilon_d^2+\Gamma^2)}\,.\end{aligned} \tag{11}$$

At resonance, *i.e.* for a d level which is at the Fermi level ($\varepsilon_d = 0$) and Γ small, the susceptibility $\chi_{\rm imp} \sim 1/\Gamma$ is greatly enhanced over other sites.

II. ADDING INTERACTIONS

Now we add interactions to $H_{\rm RLM}$. It is not difficult to argue that interactions in the conduction band will not do much, and

merely renormalize the bare conducting electrons to electron-like quasiparticles with a modified dispersion. However, interactions on the impurity site can have a strong effect, and we will only include those. In the applications to quantum dots, this is the analog of including the 'Coulomb blockade' on the quantum dot. In this manner, we obtain the Hamiltonian of the Anderson model

$$H_A = H_{\rm LRM} + U_d\, d^\dagger_\uparrow d_\uparrow d^\dagger_\downarrow d_\downarrow \,. \tag{12}$$

In the context of the perturbation theory in U, it is easy to see from diagrammatic perturbation expansion in Fig. 2 that the relationship between the conduction electron and d electron Green's function in (7) still applies—we simply have to replace the d Green's function by a renormalized Green's function including all interactions. These interactions modify the form of the d electron Green's function from (3) to

$$[G_{dd}(i\omega_n)]^{-1} = i\omega_n - \varepsilon_d - \Sigma_{dd}(i\omega_n) - \frac{1}{V}\sum_{\mathbf{k}} \frac{|w|^2}{i\omega_n - \varepsilon_{\mathbf{k}}} \,. \tag{13}$$

Here $\Sigma_{dd}(i\omega_n)$ is the only change from the presence of the interaction U_d. We can also apply the analysis from Fermi liquid theory to conclude that this change does not modify the Friedel sum rule in (8). This follows here from the relation

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} G_{dd}(i\omega) \frac{d}{d\omega} \Sigma_{dd}(i\omega) = 0 \,, \tag{14}$$

which follows because of the continued existence of the Luttinger-Ward functional.

The task for the remainder of this chapter is to understand the behavior of Σ_{dd} as a function of temperature and frequency for large U_d, and deduce consequences for other physical observables. A first guess would be to simply compute $\Sigma_{dd}(\omega)$ in a perturbative expansion in powers of U_d, and hope that the result applies also for large U_d, as it did for Fermi liquid theory, albeit with possibly large renormalizations of various parameters. In fact, this hope is realized. However, it took some time for the condensed matter community to appreciate this, and much new physics emerged from understanding the

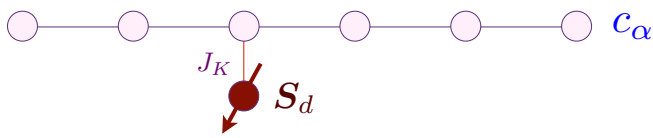

FIG. 3. Kondo impurity model: conduction electrons c_α with a spin $\boldsymbol{S}_d$ with exchange interactions J_K.

intricate structure of the crossover to large U_d in this analysis. An important part of the new physics is the emergence of a new energy scale, the Kondo temperature T_K, which can be much smaller than all other energy scales, when U_d is large. For $T \ll T_K$, we do indeed the realize Fermi liquid behavior of a non-interacting resonant level model, but with strong renormalizations. And for $T_K \ll T \ll E_F$, we have 'local moment' physics that we will describe shortly.

The analysis for large U_d proceeds most naturally by performing a canonical transformation to an effective Hamiltonian acting on the low energy subspace. This will turn out to be the Kondo Hamiltonian. Let us examine the spectrum on the d level, in the limit that $|\varepsilon_d|$ and U_d are much larger than w. Then there are 4 possible states on the d level, with energies

$$\begin{aligned} |0\rangle \quad &\Rightarrow \quad E = 0 \\ d^\dagger_\alpha |0\rangle \quad &\Rightarrow \quad E = \varepsilon_d \\ d^\dagger_\alpha |0\rangle \quad &\Rightarrow \quad E = 2\varepsilon_d + U_d \,. \end{aligned} \tag{15}$$

The non-trivial situation arises when $\varepsilon_d \ll 0, 2\varepsilon_d + U_d$. In this case, the d level is doubly degenerate, and has either a spin up or a spin down electron, and other states on the d site have a much larger energy. So we can replace the d site by a $S = 1/2$ spin, and by a 'Schrieffer-Wolff transformation', we obtain the celebrated Kondo impurity model, sketched in Fig. 3

$$H_K = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + J_K \boldsymbol{S}_d \cdot c^\dagger_{0\alpha} \frac{\boldsymbol{\sigma}_{\alpha\beta}}{2} c_{0\beta} \tag{16}$$

Here $\vec{S}_d$ is a $S = 1/2$ spin operator acting on the 2 states $d^\dagger_\uparrow |0\rangle$, $d^\dagger_\downarrow |0\rangle$ on the d site, and recall (2) for the conduction electron

operator on the impurity site $c_{0\alpha}$. So H_K describes the conduction electrons interacting with a $S = 1/2$ spin with an exchange interaction J_K, which is antiferromagnetic, as in the Hubbard model, the Kondo exchange interaction is antiferromagnetic. Its value is

$$J_K = 2w^2 \left(\frac{1}{-\varepsilon_d} + \frac{1}{\varepsilon_d + U_d} \right) . \tag{17}$$

Note that for $\varepsilon_d \ll 0, 2\varepsilon_d + U_d$, both energy denominators are large and positive. If we take the limit of 2 sites that we considered for the Hubbard model, this reduces to the familiar expression $J = 4t^2/U$. The Schrieffer-Wolff transformation also generates an additional potential scattering term $\sim w^2/U_d$ for the conduction electrons which we have dropped in (16).

III. RENORMALIZATION THEORY

As a first step towards understanding the large U_d limit, we can perform a perturbation expansion of H_K in powers of J_K. As we will see below, such an expansion leads to correlators which have a logarithmic dependence upon external frequency (at $T = 0$). This naturally suggests an application of the renormalization group, which allows us to resum the logarithmically singular terms. From such an analysis we find that system becomes strongly coupled below an energy scale T_K, which is non-zero even for very small J_K; an estimate of T_K will appear below. The perturbative expansion in J_K fails for temperatures $T < T_K$, and this strong coupling regime will be addressed systematically in Section IV by another method.

Generating a perturbation expansion in powers of J_K is not entirely straightforward because the d spins are now no longer free fermions, and there is no Wick's theorem for the correlations of the spin operators. However, it is nevertheless possible to use diagrammatic methods by using the Schwinger fermion decomposition of the spin operator along with a unit fermion constraint:

$$\boldsymbol{S}_d = \frac{1}{2} f_\alpha^\dagger \boldsymbol{\sigma}_{\alpha\beta} f_\beta \quad , \quad f_\alpha^\dagger f_\alpha = 1. \tag{18}$$

Remarkably, for the case of single spin, it is possible to impose the constraint in (18) exactly, diagram-by-diagram, using the

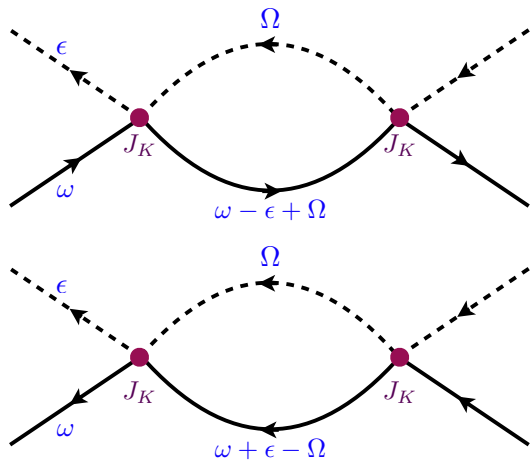

FIG. 4. Renormalization of the Kondo exchange coupling J_K. The full line is the spinon f_α, while the dashed line is the conduction electron.

Abrikosov method. This method involves imposing a chemical potential $-\lambda$ on the Schwinger fermions, and taking the $\lambda \to \infty$ limit to impose the single fermion constraint. So we consider the Hamiltonian, generalizing (16),

$$H_K = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + \lambda f^\dagger_\alpha f_\alpha + J_K \left(f^\dagger_\gamma \frac{\boldsymbol{\sigma}_{\gamma\delta}}{2} f_\delta \right) \cdot \left(c^\dagger_{0\alpha} \frac{\boldsymbol{\sigma}_{\alpha\beta}}{2} c_{0\beta} \right) \tag{19}$$

The constraint in (18) is implemented by computing $\left\langle \mathcal{O} f^\dagger_\alpha f_\alpha \right\rangle_\lambda / \left\langle f^\dagger_\alpha f_\alpha \right\rangle_\lambda$, where $\mathcal{O}$ is any observable, and taking the limit $\lambda \to \infty$. In practice, this is straightforward to implement, and usually involves omitting graphs in which the f_α flow both forward and backward in time. Note also that we are now representing the impurity spin by a single fermionic spinon f_α. This is not the same as the original d fermion because it obeys the constraint in (18).

The key physics of the Kondo model becomes evident upon considering the renormalization of the J_K coupling to second order in J_K^2. This is given by the 2 graphs in Fig. 4. The first

graph in Fig. 4 evaluates to

$$\frac{J_K^2}{2}\frac{1}{V}\sum_{\mathbf{k}}\int\frac{d\Omega}{2\pi}\frac{1}{(i\Omega-\varepsilon_{\mathbf{k}})(i\omega-i\epsilon+i\Omega-\lambda)} = \frac{J_K^2}{2}\frac{1}{V}\sum_{\mathbf{k}}\frac{\theta(-\varepsilon_{\mathbf{k}})}{i\omega-i\epsilon+\varepsilon_{\mathbf{k}}-\lambda}\,, \tag{20}$$

while the second evaluates to

$$\frac{J_K^2}{2}\frac{1}{V}\sum_{\mathbf{k}}\int\frac{d\Omega}{2\pi}\frac{1}{(i\Omega-\varepsilon_{\mathbf{k}})(i\omega+i\epsilon-i\Omega-\lambda)} = \frac{J_K^2}{2}\frac{1}{V}\sum_{\mathbf{k}}\frac{-\theta(-\varepsilon_{\mathbf{k}})}{i\omega+i\epsilon-\varepsilon_{\mathbf{k}}-\lambda}\,, \tag{21}$$

We should take the frequency of the incoming spinon to be just above threshhold *i.e.* with $i\omega-\lambda$ small. We can also set the frequency of the external conduction electron to zero, ϵ. Then the integral over $\mathbf{k}$ leads to a logarithmic dependence upon the external frequency of the spinon. We can treat this using the renormalization group, by only integrating out high energy conduction electrons. In the simplest model, we assume a flat band density of states that extends for $|\varepsilon_{\mathbf{k}}<D$, where $2D$ is the bandwith. In the RG computation, we only integrate out the highest energy electrons with $D-\delta D<|\varepsilon_{\mathbf{k}}|<D$. Then the sum of (20) and (21) evaluates to a renormalization of the exchange coupling

$$J_K \to J_K + \frac{J_K^2}{V}\sum_{D-\delta D<|\varepsilon_{\mathbf{k}}|<D}\frac{1}{\varepsilon_{\mathbf{k}}} \tag{22}$$

Performing the $\mathbf{k}$ integral, and writing $\delta D = D\delta\ell$, we obtain the 'poor man's' RG flow

$$\frac{dJ_K}{d\ell} = d(0)J_K^2 + \mathcal{O}(J_K^3)\,. \tag{23}$$

This flow is sketched in Fig. 5. For the ferromagnetic Kondo problem, $J_K<0$, the RG flow is towards $J_K=0$: in this case the impurity spin is essentially decoupled from the conduction electrons, and its coupling to the conduction electrons can be treated perturbatively. However, our interest here is the anti-ferromagnetic case with $J_K>0$, in which case (23) informs us

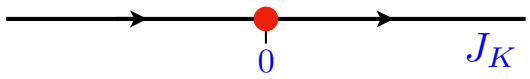

FIG. 5. Flow of the RG equation (23). For ferromagnetic exchange, $J_K < 0$, the flow is to the fixed point at $J_K = 0$. For antiferromagnetic exchange, $J_K > 0$, the flow is towards strong coupling, $J_K \to \infty$.

that J_K increases without bound under the RG flow, no matter how small the initial positive value of J_K. Specifically, the integral of (23) is

$$J_K(\ell) = \frac{1}{1/J_K(0) - d(0)\ell} \tag{24}$$

If we now start with a very small positive bare value $J_K(\ell = 0)$, we see from (24) that the renormalized exchange is of order unity at $\ell = \ell^* \sim 1/(d(0)J_K)$; equivalently, when $De^{-\ell^*} \sim T_K$, where T_K is the Kondo temperature

$$T_K \sim D \exp\left(-\frac{1}{J_K d(0)}\right) . \tag{25}$$

From the point of view of the Kondo Hamiltonian, the expression for T_K is non-perturbative, given its singular dependence upon J_K. However, one should note that this expression is ultimately non-singular at small U_d as $U_d \sim 1/J_K$: this is a hint that the low energy physics is actually adiabatically connected to the free resonant level model, albeit with strong renormalizations, as we will now discuss. But first, it should be noted that the flow of J_K to infinity predicted by (23) is not a reliable prediction of the present analysis because we cannot trust (23) when J_K becomes large—it was obtained in a perturbative expansion in J_K. Computations by Wilson using a numerical renormalization group scheme showed that the flow is indeed to $J_K \to \infty$, and the predictions of (23) are qualitatively correct.

Given this flow to large J_K, we can understand the qualitative fate of the model by examining the ground state of H_K in the limit of large J_K. In this limit, the energy is minimized if the impurity spin $\boldsymbol{S}_d$ locks into a spin singlet with a *single*

conduction electron at the site 0. No other electron can occupy this site, and therefore we can describe the remaining electrons by the renormalized free electron Hamiltonian

$$H_R = \sum_{\mathbf{k},\alpha} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + V_0 c^\dagger_{0\alpha} c_{0\alpha} \tag{26}$$

and take the limit $V_0 \to \infty$ to prevent any other electrons from occupying the impurity site. As H_R is free electron like, it is not difficult to take this limit using scattering theory, and we obtain an effective Fermi liquid description of the scattering states. Indeed, the remarkable conclusion is that in the strong coupling limit at $T \ll T_K$, the Kondo model reduces to a model of non-interacting electrons qualitatively similar to the resonant level model of Section I.

We can use this interpretation to deduce some important features of the T dependence of the impurity spin susceptibility $\chi_{\rm imp}$, introduced towards the end of Section I. At temperatures $T \gg T_K$, the perturbation theory in J_K is reliable, and so at leading order, the impurity susceptibility is given by the Curie susceptibility of an isolated spin 1/2

$$\chi_{\rm imp} = \frac{1}{4T} \quad , \quad T_K \ll T \ll U_d \, . \tag{27}$$

For $T \ll T_K$, we expect a mapping to the resonant level model, which has a finite impurity susceptibility $\sim 1/\Gamma$, as obtained in (11). This $\chi_{\rm imp}$ is determined by the width of the resonant level Γ, and a natural guess is that the width of the Kondo 'resonance' (as it is known) should be T_K. So we have

$$\chi_{\rm imp} \sim \frac{1}{T_K} \quad , \quad T \ll T_K \, , \tag{28}$$

where the co-efficient depends upon the precise definition of T_K. We can combine (27) and (28) into a crossover function between the two limiting regimes

$$\chi_{\rm imp} = \frac{1}{4T} \, \Phi(T/T_K) \, . \tag{29}$$

An important implication of the Kondo RG flow is that the crossover function $\Phi(\bar{T})$ is a *universal* function for $J_K \ll D$, determined by the renormalization group flow of the Kondo

model from $J_K = 0$ to $J_K = \infty$. At $\bar{T} \gg 1$, we have $\Phi \to 1$ so that we have the susceptibility of a free moment. For $\bar{T} \ll 1$ we have $\Phi(\bar{T}) \sim \bar{T}$ so that $\chi_{\rm imp}$ is finite.

Similar universal crossovers apply to other observables of a Kondo impurity in a metal.

IV. LARGE N THEORY

This section describes a method which can yield explicit results for crossover functions such as those in (29). This is obtained by generalizing the SU(2) spin symmetry of the Kondo model to SU(N), and examining the large N limit. It yields the correct qualitative behavior both at low and high T, and the crossover between these limits. And as will see below in Section V, the large N is also a powerful tool in examining the Kondo lattice model.

First, we realize the spin $\boldsymbol{S}_d$ by the spinon f_α in (18). In this section, we will implement the constraint in (18) by a Lagrange multiplier in the path integral, in contrast to the Abrikosov method in Section III.

To enable the generalization to SU(N), we first write the SU(2) model in a manner which does not involve the Pauli matrices. We use the identity

$$\boldsymbol{\sigma}_{\alpha\beta} \cdot \boldsymbol{\sigma}_{\gamma\delta} = 2\delta_{\alpha\delta}\delta_{\beta\gamma} - \delta_{\alpha\beta}\delta_{\gamma\delta} \tag{30}$$

to write the Kondo interaction as

$$\frac{J_K}{2}\boldsymbol{S}_d \cdot c^\dagger_{0\gamma}\boldsymbol{\sigma}_{\gamma\delta}c_{0\delta} = -\frac{J_K}{2}\left(f^\dagger_\alpha c_{0\alpha}\right)\left(c^\dagger_{0\beta}f_\beta\right) - \frac{J_K}{4}\left(f^\dagger_\alpha f_\alpha\right)\left(c^\dagger_{0\beta}c_{0\beta}\right) \tag{31}$$

After using the constraint in (18), the second term in (31) is just a shift in the local chemical potential, and we will ignore it from now on. The generalization to SU(N) is now straightforward: the indices $\alpha, \beta = 1 \ldots N$, and the constraint is

$$f^\dagger_\alpha f_\alpha = \frac{N}{2}\,. \tag{32}$$

As we will see below, to obtain a suitable large N saddle point, we also need to replace $J_K/2$ by J_K/N.

We can now write the path integral for the Kondo model

$$\mathcal{Z}_K = \int \mathcal{D}f_\alpha \mathcal{D}c_{\mathbf{k}\alpha} \mathcal{D}\lambda \exp\left(-\int_0^\beta d\tau \left[\mathcal{L}_0 + \mathcal{L}_1\right]\right)$$
$$\mathcal{L}_0 = \sum_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha}\left(\frac{\partial}{\partial\tau} + \varepsilon_{\mathbf{k}}\right) c_{\mathbf{k}\alpha} + f^\dagger_\alpha\left(\frac{\partial}{\partial\tau} + i\lambda\right) f_\alpha - i\lambda\frac{N}{2}$$
$$\mathcal{L}_1 = -\frac{J_K}{N}\left(f^\dagger_\alpha c_{0\alpha}\right)\left(c^\dagger_{0\beta} f_\beta\right) \tag{33}$$

The large N theory is obtained by decoupling the Kondo exchange term by a Hubbard-Stratonovich field $P(\tau)$. Then we obtain

$$\mathcal{Z}_K = \int \mathcal{D}\lambda \mathcal{D}P \mathcal{D}f_\alpha \mathcal{D}c_{\mathbf{k}\alpha} \exp\left(-\int_0^\beta d\tau \left[\mathcal{L}_0 + \mathcal{L}_Q\right]\right)$$
$$\mathcal{L}_Q = \frac{N|P|^2}{J_K} - P f^\dagger_\alpha c_{0\alpha} - P^* c^\dagger_{0\alpha} f_\alpha \tag{34}$$

Before proceeding, we notice an important property of $\mathcal{Z}_K$: it is invariant under an emergent U(1) gauge symmetry, under which

$$f_\alpha \to f_\alpha e^{i\phi(\tau)}$$
$$P \to P e^{i\phi(\tau)}$$
$$\lambda \to \lambda - \frac{\partial\phi}{\partial\tau}\,. \tag{35}$$

This gauge symmetry is not so crucial in the Kondo model, as it is always possible to work in a convenient fixed gauge, but it will play a crucial role when we consider the Kondo lattice in Section V.

We return to taking the large N limit of $\mathcal{Z}_K$. In the form (34), the action is quadratic in the fermions, and so we can integrate them out. As all the fermions have N components, this will yield an effective action for P and λ which has a N prefactor. Consequently, the large N limit is obtained by replacing P and λ by their saddle-point values. We go ahead and do this and replace P by $\overline{P}$, and $i\lambda$ by $\overline{\lambda}$ (we are anticipating here that the saddle-point value of λ is purely imaginary. Then the problem reduces to the following free fermion Hamiltonian

$$\overline{H}_K = \frac{N|\overline{P}|^2}{J_K} - \overline{P} f^\dagger_\alpha c_{0\alpha} - \overline{P}^* c^\dagger_{0\alpha} f_\alpha + \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha}$$
$$+ \overline{\lambda} f^\dagger_\alpha f_\alpha - \overline{\lambda}\frac{N}{2}\,. \tag{36}$$

Our remaining task is to find the ground state energy of $\overline{H}_K$, and demand that it is stationary with respect to variations in $\overline{P}$ and $\overline{\lambda}$. The latter task is simplified by the Feynman-Hellman theorem: in any eigenstate $|G\rangle$ of $\overline{H}_K$ we have

$$\frac{\partial}{\partial \overline{P}^*}\langle G|\overline{H}_K|G\rangle = \langle G|\frac{\partial \overline{H}_K}{\partial \overline{P}^*}|G\rangle \tag{37}$$

and so

$$\overline{P} = \frac{J_K}{N}\left\langle c^\dagger_{0\alpha} f_\alpha \right\rangle . \tag{38}$$

Similarly, the corresponding equation for $\overline{\lambda}$ is

$$\frac{1}{N}\left\langle f^\dagger_\alpha f_\alpha \right\rangle = \frac{1}{2} . \tag{39}$$

We can evaluate these expectation values from the Green's functions of $\overline{H}_K$, which are the same as those of the resonant level model, H_{RLM} in (1). As in (4), we have

$$G_{ff}(i\omega_n) = \frac{1}{i\omega_n - \overline{\lambda} + i\Gamma_P \mathrm{sgn}(\omega_n)} \tag{40}$$

where

$$\Gamma_P = \pi |\overline{P}|^2 d(0) , \tag{41}$$

and

$$G_{fc_0} = -G_{ff}(i\omega_n)\frac{\overline{P}}{V}\sum_{\mathbf{k}} \frac{1}{i\omega_n - \varepsilon_{\mathbf{k}}} . \tag{42}$$

For the frequency summations required for (38) and (39), we employ the spectral representation

$$G_{ff}(i\omega_n) = \int_{-\infty}^{\infty} \frac{d\Omega}{\pi}\frac{A_f(\Omega)}{i\omega_n - \Omega} , \tag{43}$$

where

$$A_f(\Omega) = \frac{\Gamma_P}{(\Omega - \overline{\lambda})^2 + \Gamma_P^2} \tag{44}$$

is the Lorentzian spectral density of the f level. Then (38) becomes

$$\begin{aligned} \overline{P} &= \frac{J_K \overline{P}}{V}\int_{-\infty}^{\infty}\frac{d\Omega}{\pi}A_f(\Omega)\sum_{\mathbf{k}} T\sum_{\omega_n}\frac{-1}{(i\omega_n - \varepsilon_{\mathbf{k}})(i\omega_n - \Omega)} \\ &= J_K \overline{P}\int_{-\infty}^{\infty}\frac{d\Omega}{\pi}A_f(\Omega)\int d\varepsilon\, d(\varepsilon)\frac{f(\varepsilon) - f(\Omega)}{\Omega - \varepsilon} . \end{aligned} \tag{45}$$

Similarly, (39) is

$$\int_{-\infty}^{\infty} \frac{d\Omega}{\pi} A_f(\Omega) f(\Omega) = \frac{1}{2} \tag{46}$$

We now have to determine the saddle point values of $\overline{P}$ and $\overline{\lambda}$ by solving (45) and (46). Fortunately, the solution of (46) is simple

$$\overline{\lambda} = 0\,, \tag{47}$$

because then we pick up exactly half of the Lorentzian spectral density. So the renormalized f level is exactly at the Fermi level *i.e.* it is exactly resonant (this is sometimes called the Abrikosov-Suhl or Kondo resonance). The value of $\overline{P}$ is determined by (45), which is the analog here of the BCS equation which determined the gap parameter. We first evaluate the integral of ε in the limit of a flat density of states at $T = 0$

$$\begin{aligned}
\int d\varepsilon\, d(\varepsilon) \frac{f(\varepsilon) - f(\Omega)}{\Omega - \varepsilon} &\approx d(0) \int_{-D}^{D} d\varepsilon \frac{\theta(-\varepsilon) - \theta(-\Omega)}{\Omega - \varepsilon} \\
&= d(0) \left[\ln \left| \frac{D + \Omega}{\Omega} \right| - \theta(-\Omega) \ln \left| \frac{D + \Omega}{D - \Omega} \right| \right] \\
&\approx d(0) \ln \left| \frac{D}{\Omega} \right| ,
\end{aligned} \tag{48}$$

when $|\Omega| \sim \Gamma_P \ll D$. So (45) yields

$$\begin{aligned}
\frac{1}{J_K d(0)} &= \int_{-\infty}^{\infty} \frac{d\Omega}{\pi} A_f(\Omega) \ln \left| \frac{D}{\Omega} \right| \\
&= \ln \left(\frac{D}{\Gamma_P} \right)
\end{aligned} \tag{49}$$

So, we have our main result: the value of $\overline{P}$ is determined by (41) from the width of the Kondo resonance, which is

$$\Gamma_P = D \exp \left(-\frac{1}{J_K d(0)} \right) , \tag{50}$$

consistent with the estimate of the Kondo temperature in (25). We can now compute other physical properties in the large N expansion, and it is natural that they will be determined by the exponentially low energy scale Γ_P in (50).

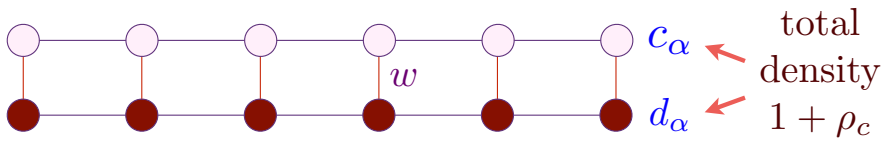

FIG. 6. Anderson lattice model: conduction band electrons c_α and d-band electrons d_α with band-mixing hybridization w. Two electrons on the same site of the d band repel with energy U_{dd}.

V. THE KONDO LATTICE

The Kondo lattice is the preferred model to describe the physics of a number of intermetallic compounds. These compounds contain a transition metal or a rare-earth metal with a localized orbital with a strong local Coulomb interactions which prefer a net magnetic moment on each site. This moment then interactions with the mobile conduction electrons arising from the lighter elements. The key difference from the previous chapter is that the moments are not isolated impurity sites, but arranged periodically in a perfect lattice. So the Bloch crystal momentum will be a good quantum number to describe the electronic states, including those associated with the moments on the electronic sites.

We begin by generalizing the interacting resonant level model, *i.e.* the Anderson model in (12), to the Anderson lattice model sketched in Fig. 6. The resonant d site is now replaced by a lattice of d sites, each of which mix with the conduction electrons c_α with the hybridization w, and there is an on-site repulsion U_d on every d site:

$$H_{AL} = \sum_{\mathbf{k}} \left[\varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + \varepsilon^d_{\mathbf{k}} d^\dagger_{\mathbf{k}\alpha} d_{\mathbf{k}\alpha} \right] + \sum_i \Big[-w \left(d^\dagger_{i\alpha} c_{i\alpha} + c^\dagger_{i\alpha} d_{i\alpha} \right) + U_d \, d^\dagger_{i\uparrow} d_{i\uparrow} d^\dagger_{i\downarrow} d_{i\downarrow} \Big] \tag{51}$$

Note that we neglect the weaker interactions on the c sites. As the two bands mix, only the total number of electrons is conserved, and we denote the total density per unit cell as $1+\rho_c$, with $0 < \rho_c < 1$.

We will be interested in the large U_d limit, with the chemical

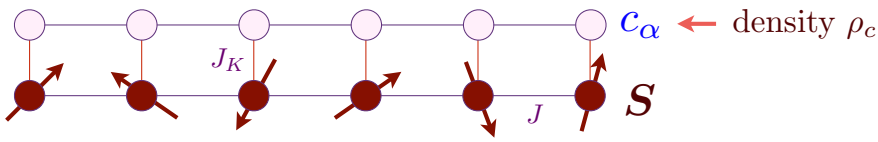

FIG. 7. Kondo lattice model: conduction electrons c_α coupled to $S = 1/2$ spins $\boldsymbol{S}$.

potential chosen so that there is exactly one electron in every d site. Then, we can perform the Schrieffer-Wolff transformation on (51), following the same procedure as Section II, and hence obtain the lattice generalization of the Kondo impurity model in (16). This Kondo lattice model is sketched in Fig. 7. The Kondo lattice Hamiltonian is expressed in terms of $S = 1/2$ spins $\boldsymbol{S}_i$ on each d site:

$$H_{KL} = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} + \frac{J_K}{2} \sum_i \boldsymbol{S}_i \cdot c^\dagger_{i\alpha} \boldsymbol{\sigma}_{\alpha\beta} c_{i\beta} \,. \tag{52}$$

The Schrieffer-Wolff transformation will also generate an exchange interaction between the d sites, but we defer consideration of the resulting Kondo-Heisenberg model. As the d sites have electron density which is exactly unity, the density of conduction electrons is now ρ_c, as shown in Fig. 7.

We will apply the large N method to the Kondo lattice model in (52), and find that the Kondo impurity model has a natural and simple generalization to the lattice. Kondo screening applies also to the lattice model, and we obtain a 'heavy Fermi liquid' (HFL) state involving both the conduction electrons and the local moments. The Friedel sum rule of the Kondo impurity model now maps to a Luttinger theorem for the HFL, and this shows that the volume enclosed by the Fermi surfaces counts *all* electrons: the conduction electrons and the local moments, for a total density of $1 + \rho_c$. The narrow Kondo resonance width translates, as we shall see, to a large renormalized mass at this large Fermi surface.

A. The heavy Fermi liquid

We proceed with an analysis of (52) using the large N approach of Section IV. The initial steps are exactly the same: we represent the spin by constrained fermionic spinons $f_{i\alpha}$, now with an additional site label. We impose the constraint by a Lagrange multiplier $\lambda_i(\tau)$ on each site, and decouple the Kondo interaction by a Hubbard-Stratonovich field $P_i(\tau)$ on each site. Finally, we reduce the theory to its large N saddle point, where the Lagrange multiplier is replaced by a site-independent value $i\lambda_i(\tau) \Rightarrow \overline{\lambda}$, and also for $P_i(\tau) \Rightarrow \overline{P}$. Then the large N saddle point is replaced by a saddle-point Kondo lattice Hamiltonian of free fermions generalizing (36) is (V now is the number of lattice sites)

$$\overline{H}_{KL} = \frac{NV|\overline{P}|^2}{J_K} + \sum_{\mathbf{k}} \left[-\overline{P} f^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} - \overline{P}^* c^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} + \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} \right] + \overline{\lambda} \sum_{\mathbf{k}} f^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} - \overline{\lambda} \frac{NV}{2} \,. \tag{53}$$

The most important difference from the impurity model is that the f spinons have now acquired a momentum label, and the Hamiltonian is diagonal in momentum. The saddle point equations determining the values of $\overline{\lambda}$ and $\overline{P}$ are now (replacing (38) and (39))

$$\overline{P} = \frac{J_K}{NV} \sum_{\mathbf{k}} \left\langle c^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} \right\rangle \tag{54}$$

$$\frac{1}{2} = \frac{1}{NV} \sum_{\mathbf{k}} \left\langle f^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} \right\rangle \,. \tag{55}$$

It is easy to compute the Green's functions of $\overline{H}_{KL}$ by summing diagrams order-by-order in $\overline{P}$, as shown in Fig. 8 (compare Fig. 2 for the resonant level model). We will find it convenient to write them in the following form

$$[G_{cc}(\mathbf{k}, i\omega_n)]^{-1} = i\omega_n - \varepsilon_{\mathbf{k}} - \frac{|\overline{P}|^2}{i\omega_n - \overline{\lambda}} \tag{56}$$

$$G_{fc}(\mathbf{k}, i\omega_n) = \frac{-\overline{P}}{(i\omega_n - \overline{\lambda})} G_{cc}(\mathbf{k}, i\omega_n) \tag{57}$$

$$G_{ff}(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n - \overline{\lambda}} + \frac{|\overline{P}|^2}{(i\omega_n - \overline{\lambda})^2} G_{cc}(\mathbf{k}, i\omega_n) \tag{58}$$

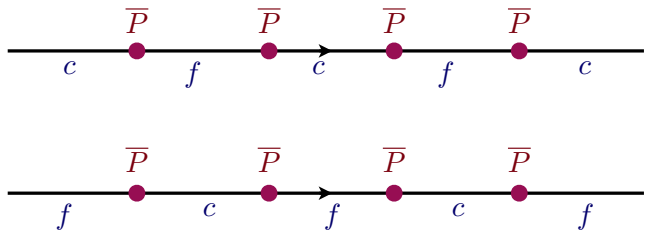

FIG. 8. Feynman diagrams for G_{cc} and G_{ff}. All lines carry the same momentum and frequency.

These equations correspond to (40) and (42) for the resonant level model. We now insert G_{fc} in the saddle-point equation (54), and obtain

$$\frac{\overline{P}}{J_K} = \frac{T}{V}\sum_{\mathbf{k},\omega_n} \frac{-\overline{P}}{(i\omega_n - \varepsilon_{\mathbf{k}})(i\omega_n - \overline{\lambda}) - |\overline{P}|^2}\,. \tag{59}$$

To evaluate the frequency summation, we notice that the denominator in (59) has poles at the energies $z = E^{\pm}_{\mathbf{k}}$ where

$$2E^{\pm}_{\mathbf{k}} = \varepsilon_{\mathbf{k}} + \overline{\lambda} \pm \left[(\varepsilon_{\mathbf{k}} - \overline{\lambda})^2 + 4|\overline{P}|^2\right]^{1/2} \tag{60}$$

These are, of course, the single particle eigenenergies of $\overline{H}_{KL}$. Evaluating the frequency summation in (59), we obtain (compare to (45))

$$\frac{\overline{P}}{J_K} = \frac{\overline{P}}{V}\sum_{\mathbf{k}} \frac{f(E^-_{\mathbf{k}}) - f(E^+_{\mathbf{k}})}{E^+_{\mathbf{k}} - E^-_{\mathbf{k}}}\,. \tag{61}$$

We will work under conditions in which the total density of the conduction electrons per site $\rho_c < 1$. We will see below in our consideration of the Luttinger theorem that under this condition we can have $E^+_{\mathbf{k}} > 0$ for all $\mathbf{k}$, while $E^-_{\mathbf{k}} < 0$ for some finite domain of $\mathbf{k}$: this will be the portion of the Brillouin zone inside the Fermi surface (see Fig. 9). Then at $T = 0$, (61) reduces to

$$\frac{\overline{P}}{J_K} = \frac{\overline{P}}{V}\sum_{\mathbf{k}} \frac{\theta(-E^-_{\mathbf{k}})}{\left[(\varepsilon_{\mathbf{k}} - \overline{\lambda})^2 + 4|\overline{P}|^2\right]^{1/2}} \tag{62}$$

Similarly, from (55), we obtain at $T = 0$ (compare to (46))

$$\frac{1}{2} = \frac{1}{V}\sum_{\mathbf{k}} \frac{1}{2}\left(1 + \frac{\varepsilon_{\mathbf{k}} - \overline{\lambda}}{\left[(\varepsilon_{\mathbf{k}} - \overline{\lambda})^2 + 4|\overline{P}|^2\right]^{1/2}}\right)\theta(-E^-_{\mathbf{k}}) \tag{63}$$

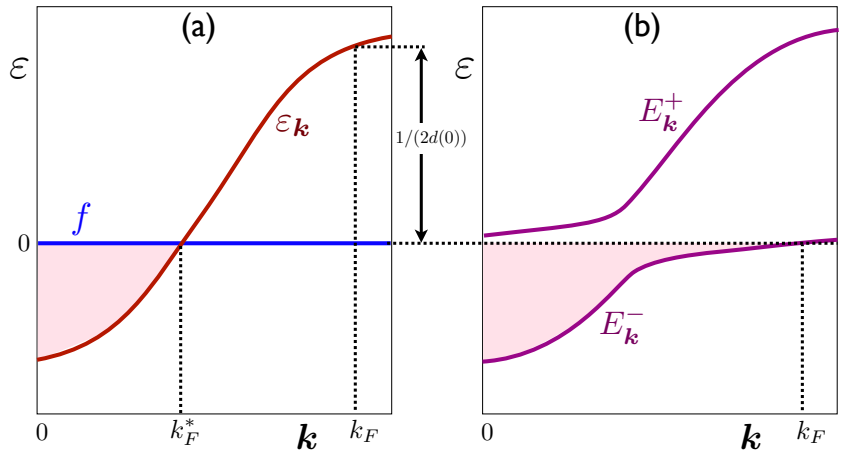

FIG. 9. Schematic band structures in the FL and FL* phases. In (a), we show the conduction electron band $\varepsilon_{\mathbf{k}}$, and the decoupled f band at zero energy; the conduction electron states with energy $\varepsilon_{\mathbf{k}} < 0$ are occupied and have density ρ_c, while the f band is half-filled. The HFL state is obtained when the mixing between the bands is non-zero (given by $\overline{P}$ in (69)), and the f band is shifted by $\overline{\lambda}$ in (67). This results in the bands shown in (b). Only the lower band in (b) is occupied, up to the wavevector k_F, for a total density of $1 + \rho_c$. The values of k_F in (a) and (b) are the same, with the value of ε_{k_F} specified by (70).

We have to solve (62) and (63) to obtain the values of the saddle-point parameters, $\overline{\lambda}$ and $\overline{P}$. We will examine the nature of the solution more carefully below after consideration of the Luttinger theorem. However, we can already notice an important point: the Kondo logarithmic divergence as $\overline{P} \to 0$, found in (48) for the Kondo impurity model, is also present in the Kondo lattice model. This is clear from (62), which has a logarithmic divergence at $\overline{P} = 0$ when the conduction band crosses the f level *i.e.* when $\varepsilon_{\mathbf{k}} = \overline{\lambda}$, provided the density of conduction electron states is finite at the Fermi level. This means that no matter how small we make J_K, we can have a solution with a non-zero $\overline{P}$; this is also the lowest energy solution, and so we obtain a heavy Fermi liquid. We will see in the next section that a minimum energy solution with $\overline{P} = 0$ becomes possible

once we allow the f moments to interact directly with each other.

B. Luttinger relation

In this initial application of the Luttinger relation at $N = \infty$, we follow the same route as in the Kondo impurity problem. From the expressions in (56), (58) it is easy to explicitly verify that

$$G_{cc}(\mathbf{k}, i\omega_n) + G_{ff}(\mathbf{k}, i\omega_n) = i\frac{\partial}{\partial\omega} \ln\left[(i\omega_n - \varepsilon_{\mathbf{k}})(i\omega_n - \overline{\lambda}) - |\overline{P}|^2\right] \tag{64}$$

Proceeding as earlier, we obtain the Luttinger constraint

$$\rho_c + 1 = \frac{2}{V}\sum_{\mathbf{k}} \theta(-E^-_{\mathbf{k}}) \tag{65}$$

The unity on the left hand side is the density of the f electrons, and recall that ρ_c is the density of conducting electrons. So clearly, the f electrons are included in the volume enclosed by the Fermi surface $E^-_{\mathbf{k}} = 0$, and we have a large Fermi surface.

C. Solution of saddle point equations

We now present an analytic solution of the saddle point equations (62), (63), (65) in the limit of small $\overline{P}$, for the case of a flat band density of states $d(\varepsilon) = (1/V)\sum_{\mathbf{k}} \delta(\varepsilon - \varepsilon_{\mathbf{k}}) \approx d(0)$; the structure of the solution was sketched in Fig. 9.

The Fermi surface is present at $E^-_{\mathbf{k}} = 0$, and from (59) this translates to $\varepsilon_{\mathbf{k}} = |\overline{P}|^2/\overline{\lambda}$. From (65), we therefore deduce that the limits of the summation over $\mathbf{k}$ in (62), (63), (65) are

$$-\frac{\rho_c + 1}{2d(0)} + \frac{|\overline{P}|^2}{\overline{\lambda}} < \varepsilon_{\mathbf{k}} < \frac{|\overline{P}|^2}{\overline{\lambda}} \tag{66}$$

We can now evaluate (63) in the limit $\overline{P} \to 0$, and obtain the value of $\overline{\lambda}$

$$\overline{\lambda} = 2d(0)|\overline{P}|^2 \tag{67}$$

(In contrast, recall that for the Kondo impurity model, we had $\overline{\lambda} = 0$.) We now see from the upper limit in (66) that the Fermi

surface is at $\varepsilon_{\mathbf{k}} = 1/(2d(0))$ when $\overline{P}$ is small, but non-zero, as shown in Fig. 9. This should be contrasted from the Fermi surface location $\varepsilon_{\mathbf{k}} = 0$ for a decoupled conduction electron band. This increase in $\varepsilon_{\mathbf{k}}$ is precisely that needed to accomodate the $1/2$ electron per site per spin component associated with the f fermions *i.e.* we have a large Fermi surface.

We can now obtain the value of $\overline{P}$ by evaluating (59) to logarithmic accuracy

$$\frac{1}{J_K} = 2d(0)\ln(D/|\overline{P}|)\,, \tag{68}$$

where D is an energy of order the bandwidth. So $\overline{P}$ is exponentially small,

$$|\overline{P}| \sim D\exp\left(-\frac{1}{2d(0)J_K}\right)\,, \tag{69}$$

the same as the estimate for the Kondo impurity model in the argument of the exponent.

Finally, let us examine the structure of the conduction electron Green's function near the Fermi level. We expand the expression for G_{cc} in (56) for small ω_n, and obtain

$$G_{cc}(\mathbf{k}, i\omega_n) \approx \frac{Z}{i\omega_n - Z\left[\varepsilon_{\mathbf{k}} - \dfrac{1}{2d(0)}\right]}\,, \tag{70}$$

where

$$Z = |2d(0)\overline{P}|^2\,. \tag{71}$$

So there is an exponentially small quasiparticle residue Z, and a quasiparticle effective mass $m^* = m/Z$ which is exponentially large. Notice also the shift in the Fermi energy in (70), illustrated in Fig. 9. These are characteristic properties of the heavy Fermi liquid (HFL).

Emergent gauge fields in metals and violations of the Luttinger theorem, in Kondo lattice and single band models.

Subir Sachdev

Department of Physics, Harvard University,
Cambridge, Massachusetts, 02138, USA

School of Natural Sciences,
Institute for Advanced Studies,
Princeton, NJ 08540, USA

Topological Aspects of Strong Correlations and Gauge Theories

International Centre for Theoretical Sciences
Tata Institute of Fundamental Research
Bengaluru, Sep 7, 2021

VI. FL* IN THE KONDO LATTICE

So far, a key feature of our discussions of the Kondo impurity model and the Kondo lattice model has been the singular nature of the limit $J_K \to 0$. Even for very small J_K, we have found that the $\overline{P} = 0$ state is unstable to the turning on of an exponentially small $\overline{P}$. In a renormalization group language, this is the statement that an infinitesimal J_K is a marginally relevant perturbation to the $J_K = 0$ fixed point. The $\overline{P} \neq 0$ state was then found to be a renormalized Fermi liquid, with well-defined quasiparticles obeying the Friedel sum rule for the Kondo impurity model; for the Kondo lattice model, the qusiparticles possess a Luttinger volume Fermi surface which counts the spins as electrons.

We will now turn on a direct interaction between the Kondo spins, and find a situation in which the $J_K = 0$ fixed point is stable, and we obtain a novel stable state with $\overline{P} = 0$, and no broken symmetry.

We will consider the Kondo-Heisenberg model with Hamiltonian

$$H_{KH} = H_{KL} + J \sum_{\langle ij \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j \,, \tag{72}$$

where H_{KL} was specified in (52), and we now have an antiferromagnet exchange interaction, J, between nearest-neighbor sites, as is generated from a Schrieffer-Wolff transformation of the Anderson lattice model in (51). We can proceed with a $1/N$ expansion for H_{KH} by combining the treatment above for the Kondo model H_{KL}, with that for the U(1) spin liquid. This implies that in addition to the decoupling field $P_i(\tau)$ used to obtain (53), we will have the analogs of the decoupling fields $Q_{ij}(\tau)$ between the d sites. This leads to the following new terms in the Lagrangian

$$\mathcal{L}_Q = \sum_{\langle ij \rangle} \left[\frac{N|Q_{ij}|^2}{J} - Q_{ij} f^\dagger_{j\alpha} f_{i\alpha} - Q^*_{ij} f^\dagger_{i\alpha} f_{j\alpha} \right] , \tag{73}$$

where Q_{ij} is the link Hubbard-Stratonovich field between the d sites. A crucial role will now be played by the emergent U(1) gauge symmetry, which combines the Kondo gauge symmetry

of (35) with that of the spin liquid gauge symmetry:

$$\begin{aligned} f_{i\alpha} &\to f_{i\alpha} e^{i\phi_i(\tau)} \\ P_i &\to P_i e^{i\phi_i(\tau)} \\ \lambda_i &\to \lambda_i - \partial_\tau \phi_i(\tau) \\ Q_{ij} &\to Q_{ij} e^{-i(\phi_i(\tau)-\phi_j(\tau))} \,. \end{aligned} \tag{74}$$

This transformation leaves the full Lagrangian invariant for a gauge transformation $\phi_i(\tau)$ which can have an arbitrary dependence on τ and lattice site i. Note that the phase of Q_{ij} transforms just like the vector potential of a U(1) gauge field. So if we write

$$Q_{ij} = |Q_{ij}| \exp(i a_{ij}) \,, \tag{75}$$

then

$$a_{ij} \to a_{ij} + \phi_j(\tau) - \phi_i(\tau) \,; \tag{76}$$

In the continuum limit the transformation of A is precisely the analog of the vector potential of the Maxwell theory, while that of λ is that of the scalar potential. So a_{ij} and λ_i together realize an emergent, lattice U(1) gauge field. For the a_{ij} to an independent propagating degree of freedom, we do need to expand about a saddle point in which the saddle point values of $|Q_{ij}|$ is non-zero: we will assume that is the case in the remaining discussion.

Given the identification of a_{ij} and λ_i with a U(1) gauge field, the other fields in (74) are easily seen to be *matter* fields which are charged under emergent the U(1) gauge symmetry. The $f_{i\alpha}$ are fermionic spinons that carry a unit gauge charge, and the P_i are bosons which also carry a unit gauge charge.

All the saddle points we have considered so far had

$$P_i = \overline{P} \neq 0 \quad , \quad \text{HFL} \,. \tag{77}$$

With our new-found identification of P as a gauge-charge boson, we see that any phase of matter satisfying (77) is a *Higgs phase* of the emergent U(1) gauge field, with $\overline{P}$ the Higgs condensate. The fluctuations of the gauge fields are quenched by this Higgs condensate, and they become overdamped modes in

the particle-hole continuum. This is why we did not have to seriously consider the U(1) gauge field in our analyses so far.

This section considers the possibility of a new saddle point in which

$$P_i = \overline{P} = 0 \quad , \quad \mathrm{FL}^* \, . \tag{78}$$

Now, in the language of gauge theories, the U(1) gauge symmetry is unbroken, and there is no Higgs condensate. It is essential to account for the fluctuations of the gauge field in a proper description of such a phase.

Let us examine the possibilities for the saddle point in the context of the mean-field Hamiltonian. Now the Hamiltonian is generalized to

$$\begin{aligned} \overline{H}_{KH} = \frac{NV|\overline{P}|^2}{J_K} + \sum_{\mathbf{k}} \left[-\overline{P} f^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} - \overline{P}^* c^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} + \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\alpha} c_{\mathbf{k}\alpha} \right] \\ + \frac{N\mathfrak{z}V|\overline{Q}|^2}{2J} - \overline{\lambda}\frac{NV}{2} + \sum_{\mathbf{k}} \varepsilon^f_{\mathbf{k}} \, f^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} \, , \end{aligned} \tag{79}$$

where $\mathfrak{z}$ is the number of nearest-neighbors of each lattice site, we have assumed a spatially uniform saddle point value of $|Q_{ij}|$ equal to $\overline{Q}$, and the dispersion of the f fermions is given by

$$\varepsilon^f_{\mathbf{k}} = \overline{\lambda} - \overline{Q} \sum_{a=1}^{\mathfrak{z}} e^{i\mathbf{k}\cdot\mathbf{e}_a} \, , \tag{80}$$

with $\mathbf{e}_a$ vectors connecting nearest neighbor sites. The saddle point equations (54), (55) are supplemented by an additional equation for $\overline{Q}$

$$\overline{Q} = \frac{2J}{N\mathfrak{z}V} \sum_{\mathbf{k}} \sum_{a=1}^{\mathfrak{z}} e^{i\mathbf{k}\cdot\mathbf{e}_a} \left\langle f^\dagger_{\mathbf{k}\alpha} f_{\mathbf{k}\alpha} \right\rangle \, . \tag{81}$$

The equations (54), (55), (81) can now be solved for the values of $\overline{\lambda}$, $\overline{Q}$, and $\overline{P}$. In the light of (77), (78), we pay particular attention to the equation of $\overline{P}$. This can be written in the form (61), with the quasiparticle dispersions now given by

$$2E^\pm_{\mathbf{k}} = \varepsilon_{\mathbf{k}} + \varepsilon^f_{\mathbf{k}} \pm \left[(\varepsilon_{\mathbf{k}} - \varepsilon^f_{\mathbf{k}})^2 + 4|\overline{P}|^2 \right]^{1/2} \, . \tag{82}$$

As we are interested in the possibility of a solution like (78), let us take the $\overline{P} \to 0$ limit of (61), which can be written as

$$\frac{\overline{P}}{J} = \frac{\overline{P}}{V} \sum_{\mathbf{k}} \frac{f(\varepsilon_{\mathbf{k}}) - f(\varepsilon^f_{\mathbf{k}})}{\varepsilon^f_{\mathbf{k}} - \varepsilon_{\mathbf{k}}} \, . \tag{83}$$

Now, the crucial point is the summation on $\mathbf{k}$ in (83) is *finite* for a generic dispersion $\epsilon^f_{\mathbf{k}}$ like that in (80). In the case $\overline{Q} = 0$ that we considered in Section V A, the f band is dispersionless, and then the summation on $\mathbf{k}$ in (83) is logarithmically divergent (see (61)). With $\overline{Q} \neq 0$, and the summation finite, the only solution possible for (83) for small J_K is $\overline{P} = 0$. Specifically, a fractionalized Fermi liquid (FL*) phase is obtained for $J_K < J_{Kc}$ with

$$\frac{1}{J_{Kc}} = \frac{1}{V} \sum_{\mathbf{k}} \frac{f(\varepsilon_{\mathbf{k}}) - f(\varepsilon^f_{\mathbf{k}})}{\varepsilon^f_{\mathbf{k}} - \varepsilon_{\mathbf{k}}} \, . \tag{84}$$

The reader should notice a similarity to the 'Stoner criterion' for spin density waves, with an inverse interaction strength on the left hand side, and a fermionic susceptibility on the right hand side.

We conclude with a brief statement of the physical properties of the FL* phase. With $\overline{P} = 0$, the conduction electrons are decoupled from the Kondo spins at mean field level. From (56), the conduction electron Green's function is

$$G_{cc}(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n - \varepsilon_{\mathbf{k}}} \, . \tag{85}$$

In comparison to (70), there is no shift of $1/(2d(0))$ in the Fermi energy to accommodate the f electrons, and so we have a *small* Fermi surface, as shown in Fig. 9(a). There is also no corresponding renormalization of the mass of the conduction electrons.

The f spinons are decoupled from the conduction electrons, and form an independent U(1) spin liquid at mean field level. Beyond mean-field, the spinons will interact with the emergent U(1) gauge field, and this has interesting consequences for the spin spectrum of the spin liquid. There will also be fluctuations of the hybridization boson P_i about $\overline{P} = 0$, and this will lead to some modification of the spinon spectrum from the presence of the conduction electrons. However, this coupling does not lead to a disappearance of the fractionalized spinon excitations, which are stabilized by their charges under the emergent U(1) gauge field.

A. Emergent gauge fields and generalized Luttinger relations

We saw in Section V B that the large N theory of the HFL phase yielded a 'large' Fermi surface corresponding to a density of $1+\rho_c$ electrons. Similarly in Section VI, the large N theory of the FL* phase yielded a 'small' Fermi surface of conduction electrons alone, corresponding to a density of ρ_c. In this section we will show that the results hold to all orders in $1/N$, and also describe the connection to the discussion of the Luttinger relation.

From the point of view of the Anderson lattice Hamiltonian in (51), the large Fermi surface appears to be the only possible answer for all values of U_d. In general, the Anderson lattice Hamiltonian will lead to a 2×2 matrix Green's function, at all orders in perturbation theory in U_d:

$$\boldsymbol{G}^{-1}(\boldsymbol{k}, i\omega_n) = \begin{pmatrix} i\omega_n - \varepsilon_{\mathbf{k}} & w \\ w & i\omega_n - \varepsilon^d_{\mathbf{k}} \end{pmatrix} - \boldsymbol{\Sigma}(\boldsymbol{k}, i\omega_n) \qquad (86)$$

where the self energy $\boldsymbol{\Sigma}$ is also a 2×2 matrix. There is only one global U(1) symmetry associated with electron number in (51), and so the Luttinger relation must compute the total density $1+\rho_c$. We can relate the total density to the size of the Fermi surface by the multiband identity:

$$\mathrm{Tr}\, \boldsymbol{G}(\mathbf{k}, i\omega) = i\frac{\partial}{\partial\omega} \ln\left[\det \boldsymbol{G}(\mathbf{k}, i\omega)\right] - i\mathrm{Tr}\left[\boldsymbol{G}(\mathbf{k}, i\omega)\frac{\partial}{\partial\omega}\boldsymbol{\Sigma}(\mathbf{k}, i\omega)\right] . \qquad (87)$$

Then an analysis analogous to that for the single band model leads to a constraint of $1+\rho_c$ on the total size of one or more Fermi surfaces.

This analysis of the Anderson lattice model appears to leave no room for the small Fermi surface of the FL* phase. However, it must be kept in mind that the above analysis is perturbative in U, even though it holds to all orders in U.

To understand the FL* phase, we have to turn to the Kondo lattice model (72), and understand its U(1) symmetries more carefully. The Kondo lattice model has a global U(1) symmetry of electron number of conservation which counts *only* the

Symmetry	f	c	P
U(1)	0	1	-1
$\mathrm{U}(1)_{\mathrm{gauge}}$	1	0	1
$\mathrm{U}(1)_{\mathrm{diag}}$	1	1	0

TABLE I. Symmetry charges

conduction electrons. However it has an additional $\mathrm{U}_{\mathrm{gauge}}(1)$ symmetry specified by (74), associated with fixed electron number on each d site. So the total symmetry of the Kondo lattice model is $\mathrm{U}(1)\times\mathrm{U}_{\mathrm{gauge}}(1)$. The fate of this enlarged symmetry is distinct in the two phases:
(i) HFL phase: the $\mathrm{U}(1)\times\mathrm{U}_{\mathrm{gauge}}(1)$ symmetry is broken to a diagonal $\mathrm{U}(1)_{\mathrm{diag}}$ symmetry by the condensation of the Higgs boson P. Recall that $P \sim c_\alpha^\dagger f_\alpha$, and so P carries charges of both U(1) and $\mathrm{U}(1)_{\mathrm{gauge}}$, associated with charges of $c^\dagger$ and f. However the condensation of P leaves $\mathrm{U}(1)_{\mathrm{diag}}$ unbroken, under which c and f have the same charge

$$\mathrm{U}_{\mathrm{diag}} : c_\alpha \to c_\alpha e^{i\phi_d}, \quad f_\alpha \to f_\alpha e^{i\phi_d} \tag{88}$$

We summarize the nature of these symmetries in Table I. We can only deduce a Luttinger relation for the unbroken $\mathrm{U}(1)_{\mathrm{diag}}$ symmetry, which counts the number of both c and f fermions. In this manner, we obtain the large Fermi surface of the HFL phase, as already obtained in the Anderson lattice model.
(ii) FL* phase: the $\mathrm{U}(1)\times\mathrm{U}_{\mathrm{gauge}}(1)$ remains unbroken. Now there should be 2 separate Luttinger relations associated with these 2 symmetries. Only the c fermions carry the global U(1) charge, and so the usual Luttinger arguments imply a small Fermi surface, as we discuss further below. The Luttinger relations also apply to $\mathrm{U}_{\mathrm{gauge}}(1)$, and lead to constraints on the spinon excitation structure of the spin liquid [2, 3].

Finally, a few further comments on the stability of the small Fermi surface, and the Luttinger (*i.e.* Luttinger*) relation that applies in the FL* phase. We can compute the conduction electron's Green's function in a $1/N$ expansion of the gauge theory about the $\overline{Q} = 0$ saddle point, and results can be written

as

$$G_{cc}(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n - \varepsilon_{\mathbf{k}} - \Sigma_{cc}(\mathbf{k}, \omega_n)} \,. \tag{89}$$

As long as we are expanding about the $\overline{Q} = 0$ saddle point, it is not difficulty to see that the self energy in (85) is obtained from a Luttinger-Ward functional, and so satisfies

$$\sum_{\mathbf{k}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} G_{cc}(\mathbf{k}, i\omega) \frac{\partial}{\partial \omega} \Sigma_{cc}(\mathbf{k}, i\omega) = 0 \,. \tag{90}$$

Then, proceeding in the usual route, we obtain the Luttinger* relation

$$\rho_c = \frac{1}{V} \sum_{\mathbf{k}} \theta(-\varepsilon^*_{\mathbf{k}}) \tag{91}$$

with $\varepsilon^*_{\mathbf{k}} = \varepsilon_{\mathbf{k}} + \Sigma_{cc}(\mathbf{k}, 0)$; contrast this to the usual Luttinger relation in (65) for the large Fermi surface.

It is also interesting to compare the conduction electron Green's function in (85) with that in (56). We can consider the f electron contribution in (56), equal to $|\overline{P}|^2/(i\omega_n - \overline{\lambda})$, as a conduction electron 'self energy': this self energy diverges at zero frequency for $\overline{\lambda} = 0$, leading to zeros of the Green's function which have been the focus of some attention in the literature[4–12]. We see from our treatment that such zeros are resolved [13] in two possible ways:
(i) In the HFL phase with $\overline{P} \neq 0$, we have $\overline{\lambda}$ non-zero, and given by $\overline{\lambda} = 2d(0)|\overline{P}|^2$ in (67). It is this non-zero $\overline{\lambda}$ which leads to the shift in the apparent Fermi surface $\varepsilon_{\mathbf{k}} = 0$ to the actual large Fermi surface $\varepsilon_{\mathbf{k}} = |\overline{P}|^2/\overline{\lambda}$, as in (70).
(ii) In the FL* phase with $\overline{P} = 0$ this apparently divergent contribution to the self energy is absent, and the actual self energy, which is Σ_{cc} in (89), can be obtained from a Luttinger-Ward functional. Now there is a stable small Fermi surface at $\varepsilon_{\mathbf{k}} = 0$ (with $1/N$ fluctuation corrections, at $\varepsilon^*_{\mathbf{k}} = 0$) obeying the Luttinger* relation in (91).

VII. FL* IN THE SINGLE-BAND HUBBARD MODEL

[1] A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, 1997).

[2] P. Bonderson, M. Cheng, K. Patel, and E. Plamadeala, "Topological Enrichment of Luttinger's Theorem," arXiv e-prints (2016), arXiv:1601.07902 [cond-mat.str-el].

[3] D. V. Else and T. Senthil, "Strange metals as ersatz Fermi liquids," (2020), arXiv:2010.10523 [cond-mat.str-el].

[4] B. L. Altshuler, A. V. Chubukov, A. Dashevskii, A. M. Finkel'stein, and D. K. Morr, "Luttinger theorem for a spin-density-wave state," Europhys. Lett. **41**, 401 (1998), cond-mat/9703120.

[5] F. H. Essler and A. M. Tsvelik, "Weakly coupled one-dimensional Mott insulators," Phys. Rev. B **65**, 115117 (2002), cond-mat/0108382.

[6] I. Dzyaloshinskii, "Some consequences of the Luttinger theorem: The Luttinger surfaces in non-Fermi liquids and Mott insulators," Phys. Rev. B **68**, 085113 (2003).

[7] R. M. Konik, T. M. Rice, and A. M. Tsvelik, "Doped Spin Liquid: Luttinger Sum Rule and Low Temperature Order," Phys. Rev. Lett. **96**, 086407 (2006), cond-mat/0511268.

[8] K.-Y. Yang, T. M. Rice, and F.-C. Zhang, "Phenomenological theory of the pseudogap state," Phys. Rev. B **73**, 174501 (2006), cond-mat/0602164.

[9] T. D. Stanescu and G. Kotliar, "Fermi arcs and hidden zeros of the Green function in the pseudogap state," Phys. Rev. B **74**, 125110 (2006).

[10] K. B. Dave, P. W. Phillips, and C. L. Kane, "Absence of Luttinger's Theorem due to Zeros in the Single-Particle Green Function," Phys. Rev. Lett. **110**, 090403 (2013), arXiv:1207.4201 [cond-mat.str-el].

[11] C. Berthod, T. Giamarchi, S. Biermann, and A. Georges, "Breakup of the Fermi surface near the Mott transition in

low-dimensional systems," Phys. Rev. Lett. **97**, 136401 (2006), arXiv:cond-mat/0602304.

[12] S. Sakai, Y. Motome, and M. Imada, "Evolution of Electronic Structure of Doped Mott Insulators: Reconstruction of Poles and Zeros of Green's Function," Phys. Rev. Lett. **102**, 056404 (2009), arXiv:0809.0950 [cond-mat.str-el].

[13] M. S. Scheurer, S. Chatterjee, W. Wu, M. Ferrero, A. Georges, and S. Sachdev, "Topological order in the pseudogap metal," Proc. Nat. Acad. Sci. **115**, E3665 (2018), arXiv:1711.09925 [cond-mat.str-el].