Transport in Multilevel Quantum Dots: From the Kondo Effect to the Coulomb Blockade Regime

A. Levy Yeyati, F. Flores, and A. Martín-Rodero

Departamento de Física Teórica de la Materia Condensada C-V, Universidad Autónoma de Madrid, E-28049 Madrid, Spain
(Received 23 October 1998)

A new theoretical method is introduced to study coherent electron transport in an interacting multilevel quantum dot. The method yields the correct behavior in the limit of both weak and strong coupling to the leads, giving a unified description of the Coulomb blockade and the Kondo effect. Results for the density of states and the temperature-dependent conductance for a two-level dot are presented. The relevance of these results in connection to recent experiments on the Kondo effect in semiconducting quantum dots is discussed.

Coherent electron transport in the presence of strong electron-electron interactions has been one of the central issues in the field of mesoscopic systems [1]. Semiconducting quantum dots (QDs) provide an almost ideal system where the predictions of the theory can be tested. For instance, recent experiments have demonstrated the possibility of exploring the Kondo effect, a prototypical correlation effect, using this technology [2,3].

From a theoretical point of view, correlation effects have been mainly analyzed in this type of systems by means of the single-level Anderson model [4,5]. Within this model the Kondo effect arises due to fluctuations in the spin of an unpaired electron [6]. So far, there have been few attempts to include a multilevel spectrum for describing either single QDs with quasidegenerate levels or coupled QDs beyond a rate equation approach [7]. The actual situation in semiconducting QDs should require the inclusion of a multilevel spectrum whenever many single-particle states with a small level separation are involved in the transport process. This could be the case in experiments measuring the phase of the transmission amplitude through a QD in the Coulomb blockade (CB) regime [8] and also in recent experiments on the Kondo effect [2,3].

The aim of this letter is to introduce a new theoretical approach for describing correlation effects in multilevel QDs. The approach is constructed to yield the correct behavior in the limit of both infinite and vanishing charging energy. This is achieved by introducing an interpolative self-energy for the one-electron Green functions, an approach which has been successfully applied to several interacting systems including the equilibrium and non-equilibrium Anderson models [5,9]. This type of approach has been recently rediscovered and applied to analyze the Mott transition in the Hubbard model for arbitrary band filling [10,11].

For describing the multilevel QD we consider a model Hamiltonian which is a generalization of the single-level Anderson model, $H = H_{\text{dot}} + H_{\text{leads}} + H_T$, where $H_{\text{dot}} = \sum_m \epsilon_m \hat{d}_m^\dagger \hat{d}_m + U \sum_{l>m} \hat{n}_m \hat{n}_l$ corresponds to the uncoupled QD with M levels ($\hat{n}_m = \hat{d}_m^\dagger \hat{d}_m$);

$H_{\text{leads}} = \sum_{k \in L,R} \epsilon_k \hat{c}_k^\dagger \hat{c}_k$ to the uncoupled leads, and $H_T = \sum_{m,k \in L,R} t_{m,k} \hat{d}_m^\dagger \hat{c}_k + \text{H.c.}$ describes the coupling between the dot and the leads. The labels m and l ($1 \leq m, l \leq M$) in H denote the different dot levels including spin quantum numbers.

Our main objective is the determination of the dot retarded Green functions from which the different level charges and the linear conductance can be obtained. In a frequency representation they can be written as $G_m(\omega) = [\omega - \epsilon_m^{\text{HF}} - \Sigma_m(\omega) - i\Gamma_{m,L}(\omega) - i\Gamma_{m,R}(\omega)]^{-1}$, where $\epsilon_m^{\text{HF}} = \epsilon_m + U \sum_{l \neq m} n_l$ is the Hartree-Fock level (we adopt the notation $n_l = \langle \hat{n}_l \rangle$), $i\Gamma_{m,L(R)}(\omega) = \sum_{k \in L(R)} t_{m,k}^2/(\omega - \epsilon_k + i0^+)$ and $\Sigma_m(\omega)$ is a self-energy that takes into account electron-electron interactions beyond the Hartree-Fock approximation. We shall neglect the indirect coupling between dot states [12] and adopt the usual approximation of $\Gamma_{m,L}$, $\Gamma_{m,R}$ being independent of energy. In our approach [5,9], we look for an interpolative self-energy yielding the correct $U/\Gamma_m \to 0, \infty$ limits.

In the $U/\Gamma_m \to \infty$ or "atomic" limit, $G_m(\omega)$ can be obtained using the equation of motion method [13], which yields

$$G_m^{(at)}(\omega) = \frac{\langle \prod_{l \neq m}(1 - \hat{n}_l) \rangle}{\omega - \epsilon_m + i0^+}$$
$$+ \sum_{l \neq m} \frac{\langle \hat{n}_l \prod_{(s \neq l) \neq m}(1 - \hat{n}_s) \rangle}{\omega - \epsilon_m - U + i0^+} + \cdots$$
$$+ \frac{\langle \prod_{l \neq m} \hat{n}_l \rangle}{\omega - \epsilon_m - (M-1)U + i0^+}, \quad (1)$$

In this expression all possible charge states of the dot give a contribution. Their evaluation requires the knowledge of many particle correlations $\langle \hat{n}_1 \hat{n}_2 \rangle, \langle \hat{n}_1 \hat{n}_2 \hat{n}_3 \rangle, \ldots$, etc. However, for sufficiently large U, fluctuations in the dot charge with respect to the mean charge $\mathcal{N}$ by more than one electron become negligible and $G_m^{(at)}(\omega)$ is accurately given by the three poles expression

$$G_m^{(at)}(\omega) \simeq \frac{A_{N-1}^m}{\omega - \epsilon_m - U(N-1) + i0^+} + \frac{A_N^m}{\omega - \epsilon_m - UN + i0^+} + \frac{A_{N+1}^m}{\omega - \epsilon_m - U(N+1) + i0^+}, \quad (2)$$

where $N = \text{Int}[\mathcal{N}]$. In order to yield the exact first three momenta of Eq. (1) the weight factors A_N^m should satisfy the following sum rules:

$$A_{N-1}^m + A_N^m + A_{N+1}^m = 1,$$

$$(N-1)A_{N-1}^m + NA_N^m + (N+1)A_{N+1}^m = \sum_{l \neq m} n_l, \quad (3)$$

$$(N-1)^2 A_{N-1}^m + N^2 A_N^m + (N+1)^2 A_{N+1}^m = \sum_{l \neq m} n_l + \langle \hat{n}\hat{n}\rangle_m,$$

where $\langle \hat{n}\hat{n}\rangle_m = \sum_{(l \neq k) \neq m} \langle \hat{n}_l \hat{n}_k \rangle$. The special case $N = 0$ ($N = M - 1$) has to be treated as $N = 1$ ($N = M - 2$). Notice that this expression for $G_m^{(at)}(\omega)$ is fully determined by the average charges n_l and the two-body correlations $\langle \hat{n}_l \hat{n}_k \rangle$. From $G_m^{(at)}(\omega)$, one can define an atomic self-energy, $\Sigma_m^{(at)} = \omega - \epsilon_m^{\text{HF}} - [G^{(at)}(\omega)]_m^{-1}$. Using Eqs. (2) and (3), $\Sigma_m^{(at)}$ can be written as

$$\Sigma_m^{(at)} = \frac{a_m U^2 (\omega - \epsilon_m + i0^+) + b_m U^3}{(\omega - \epsilon_m + i0^+)^2 + c_m U(\omega - \epsilon_m + i0^+) + d_m U^2}, \quad (4)$$

where $a_m = (\mathcal{N} - n_m)[1 - (\mathcal{N} - n_m)] + \langle \hat{n}\hat{n}\rangle_m$, $c_m = \mathcal{N} - n_m - 3N$, $d_m = \langle \hat{n}\hat{n}\rangle_m + 3N^2 - 1 - (3N - 1)(\mathcal{N} - n_m)$, and $b_m = N^2(1 - N) - (\mathcal{N} - n_m)d_m$. On the other hand, in the $U/\Gamma_m \to 0$ limit, Σ_m is given by the second order expression [9]

$$\Sigma_m^{(2)}(\omega) = U^2 \sum_{l \neq m} \int_{-\infty}^{\infty} d\epsilon_1 d\epsilon_2 d\epsilon_3 \frac{\tilde{\rho}_m(\epsilon_1)\tilde{\rho}_l(\epsilon_2)\tilde{\rho}_l(\epsilon_3)}{\omega + \epsilon_2 - \epsilon_1 - \epsilon_3 + i0^+}[f_1 f_3(1 - f_2) + (1 - f_1)(1 - f_3)f_2], \quad (5)$$

where $f_i = f(\epsilon_i)$ denotes the Fermi distribution function at the leads, and $\tilde{\rho}_m(\omega)$ are effective densities of states given by $\pi\tilde{\rho}_m(\omega) = \Gamma_m/[(\omega - \tilde{\epsilon}_m)^2 + \Gamma_m^2]$. The effective levels $\tilde{\epsilon}_m$ are introduced in order to fulfill the Fermi-liquid relations associated with charge conservation [Friedel sum rule (FSR) [14]] as discussed below.

In order to determine an interpolative scheme between the two limits, let us first notice that $\Sigma_m^{(2)} \to U^2 \alpha_m/(\omega - \tilde{\epsilon}_m)$, where $\alpha_m = \sum_{l \neq m} \tilde{n}_l(1 - \tilde{n}_l)$, when $\Gamma_m/\omega \to 0$. On the other hand, $a_m \to \alpha_m$ in the small U limit and thus $\Sigma_m^{(at)} \to U^2 \alpha_m/(\omega - \epsilon_m)$ in this case. These properties allow us to define the interpolative self-energy, replacing $\omega - \epsilon_m$ by $\alpha_m/\Sigma_m^{(2)} + \Delta\epsilon_m$, where $\Delta\epsilon_m = \tilde{\epsilon}_m - \epsilon_m$, in Eq. (4) for $\Sigma_m^{(at)}$, which yields

$$\Sigma_m = \left(\frac{\Sigma_m^{(2)}}{\alpha_m}\right) \frac{a_m + (a_m \Delta\epsilon_m/U + b_m/\alpha_m)(\Sigma_m^{(2)}/U)}{1 + (2\Delta\epsilon_m/U + c_m)(\Sigma_m^{(2)}/\alpha_m U) + (c_m \Delta\epsilon_m/U + d_m)(\Sigma_m^{(2)}/\alpha_m U)^2}. \quad (6)$$

This expression provides the generalization to the multi-level case of the interpolative self-energy first introduced in [9] for the single-level Anderson model. The single-level case ($M = 2$) is readily obtained from Eq. (6) when $N = 1$ taking $\langle \hat{n}\hat{n}\rangle_m = 0$. It is evident by construction that $\Sigma_m \to \Sigma_m^{(at)}$ when $U/\Gamma_m \to \infty$ and $\Sigma_m \to \Sigma_m^{(2)}$ in the small U limit.

The final step in this approach is to determine the level charges n_m, the correlation functions $\langle \hat{n}_l \hat{n}_k \rangle$ and the effective levels $\tilde{\epsilon}_m$ self-consistently. The charges and the correlations functions are determined through the relations $n_m = -\int_{-\infty}^{\infty} f(\omega) \text{Im} G_m(\omega) d\omega/\pi$ and $\sum_{l \neq m} \langle \hat{n}_l \hat{n}_m \rangle = -\int_{-\infty}^{\infty} f(\omega) \text{Im}[\Sigma_m(\omega) G_m(\omega)] d\omega/(\pi U)$. This last relation follows from the equation of motion of the retarded Green functions. It is important to stress that the self-consistent determination of the two-body correlations is *essential* to get the correct values of the charge for large U. The effective level is determined by impos-

ing the condition $\int_{-\infty}^{\infty} f(\omega) \text{Im}[G_m(\omega)\partial\Sigma_m/\partial\omega] d\omega = 0$, which at zero temperature reduces to the Luttinger theorem [15] ensuring the fulfillment of the FSR [14] $n_m = -\text{Im}[\ln G_m(E_F)]/\pi$. In Ref. [5] we showed that the condition of consistency between the effective and the final charges is nearly equivalent to imposing the fulfillment of the FSR for the simple Anderson model. Similar self-consistency conditions have been proposed in Refs. [10,11].

We have first applied this formalism to the case of a doubly degenerate level ($M = 4$ when including spin) which is a simple generalization of the single-level Anderson model. The inset in Fig. 1 shows the charge per level n_m as a function of the leads Fermi energy E_F. For the case shown in Fig. 1, corresponding to $\epsilon_m = 0$, $\Gamma_m/U = 0.075$ (same for all levels), and zero temperature, one can observe a modulation in the charge resembling the typical Coulomb staircase of the $\Gamma_m/U \to 0$ limit.
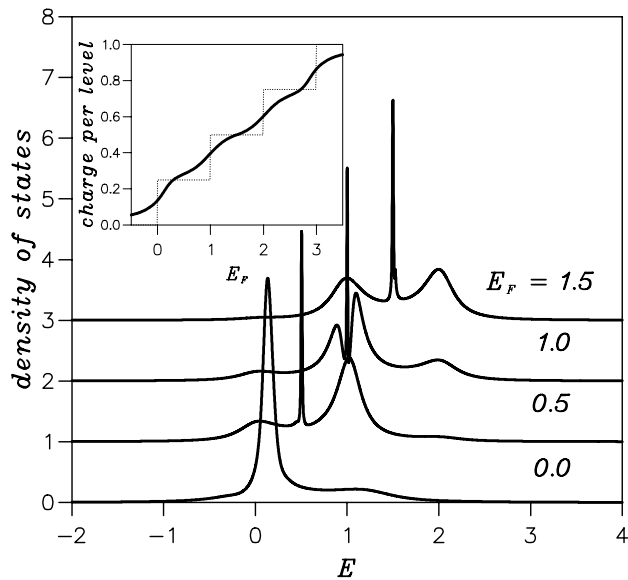
FIG. 1. Density of states for a fully degenerate two-level dot with $\Gamma_m/U = 0.075$ and $\epsilon_m = 0$ for different values of the leads Fermi energy E_F. The inset shows the charge per level and spin as a function of E_F (solid line) as compared to the $\Gamma_m/U \to 0$ case (dotted line). Energies are measured in units of U.

FIG. 2. Total conductance for the same case of Fig. 1 as a function of E_F for different temperature values T/U: (a) 0.0005, (b) 0.0025, (c) 0.005, (d) 0.01, and (e) 0.03.

Figure 1 also shows the density of states (DOS) associated with the dot levels for different values of E_F/U corresponding to the steps and the center of the plateaus in the charging curve. For $E_F/U = 0$, $n_m \simeq 0.125$, the system is in the so-called mixed-valence regime, the spectrum being similar to the one found using large N expansions in the $U \to \infty$ limit [16] with a resonance at a renormalized level just above E_F. For $E_F/U = 0.5$ and 1.5 there are approximately one and two electrons, respectively, inside the dot. For these cases the system is in the Kondo regime, the DOS exhibiting a sharp peak around E_F halfway between two broader resonances (of width $\sim\Gamma$) approximately separated by U. The weight of these resonances is the same for the $E_F/U = 1.5$ case due to the electron-hole symmetry in the half-filled dot, while there is a pronounced asymmetry for $E_F/U = 0.5$. For a half-integer occupation of the dot such as in the $E_F/U = 1.0$ case the DOS exhibits a more complex structure with three broad resonances around $E = 0$, U, and $2U$, and a narrow Kondo peak still present at E_F. The overall shape is reminiscent of an average between the cases with one and two electrons in the dot.

The evolution of the Kondo peak as a function of E_F and temperature should be reflected in the dot linear conductance G, which can be readily obtained from the Green functions [17]. This is shown for different temperatures in Fig. 2. When approaching zero temperature the conductance behaves like $G = 4(e^2/h)\sin^2[\pi n_m]$ as expected from the FSR. The conductance decreases very rapidly with temperature in the region $0.5 < E_F < 2.5$, where the DOS is controlled by the Kondo peak. Outside this region one can notice a slight increase of conductance with
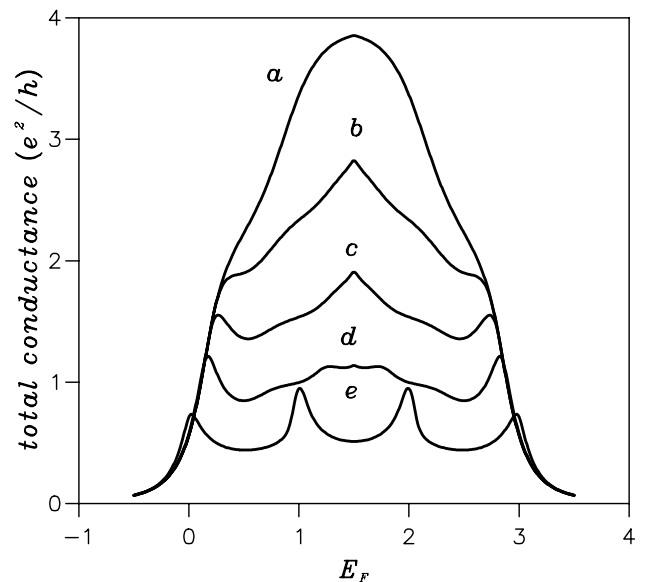
temperature. At temperatures large enough to be above the Kondo temperature T_K, which can be estimated by the condition $G(T_K) \simeq G(0)/2$ [2], the conductance tends to exhibit the usual CB peaks at the charge degeneracy points. We should state that our approach does not provide an accurate estimate of T_K as a function of the model parameters as the exponential decrease of the Kondo peak weight for very large U [18] is not strictly recovered. This limitation does not affect, however, the qualitative behavior of the conductance except for a rescaling of the temperature values.

So far we have analyzed the case of a completely degenerate level. In an actual QD, geometrical asymmetries could result in an effective splitting ΔE of the dot levels. When $\Delta E \gg \Gamma$ the physical situation could be described by a single-level Anderson model. It is interesting to analyze the evolution from this situation to the quasidegenerate case $\Delta E \to 0$ previously discussed. Figure 3 illustrates the case $\Delta E \sim \Gamma$. As can be observed in the charging curves shown as an inset, the splitting tends to block the charging of the upper levels. This blocking effect becomes nearly complete when $\Delta E > \Gamma$, which is the actual case in the experiments of Ref. [2]. Figure 4 shows the linear conductance for a set of parameters which yields a similar behavior to the experimental data in Figs. 2(a) and 2(b) in the first paper in Ref. [2]. We have taken $\Delta E/U = 0.4$, $\Gamma_1/U = 0.075$, and $\Gamma_2/U = 0.1$, i.e., a case where the upper level is somewhat more strongly coupled to the leads (in Ref. [2], it was suggested that $U \simeq 1-2$ meV, $\Delta E \sim U$, and $\Gamma \simeq 0.2$ meV). The selected temperature range in Fig. 4, $T/U = 0.02-0.05$ approximately corresponds to the experimental values $T \simeq 400-800$ mK). One can notice that the two conductance peaks associated with the lower level are strongly asymmetric due to the
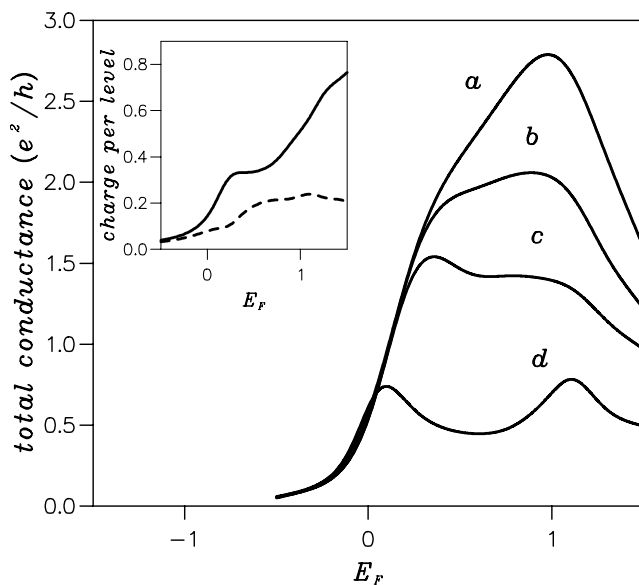
FIG. 3. Total conductance for a two-level dot with splitting $\Delta E/U = 0.1$ as a function of E_F for different temperature values T/U: (a) 0.0005, (b) 0.0025, (c) 0.005, and (d) 0.03. The inset shows the charge per spin on the two levels as a function of E_F. Only the first half of the curve is shown as the total conductance is symmetrical with respect to $(3U + \Delta)/2$.

presence of the second level, in qualitative agreement with the data of Ref. [2]. The other two peaks, associated with the second level, are somewhat less resolved as found experimentally. The overall effect of temperature is again to reduce the total conductance in the range for E_F where the Kondo effect is present, while an increase is observed outside this range. Both effects, the increase of conductance with temperature between each pair of peaks as well as the asymmetry, can be considered as manifestations of the multilevel structure of an actual QD.

The most remarkable consequences of this multilevel structure would be observed in a situation corresponding to a smaller splitting, where "collective" Kondo-like fea-
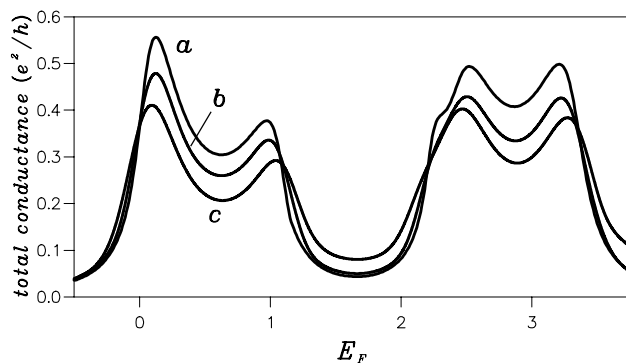

FIG. 4. Total conductance for a two-level dot with $\Delta E/U = 0.4$, $\Gamma_1/U = 0.075$, and $\Gamma_2/U = 0.1$ as a function of E_F for different temperature values T/U: (a) 0.02, (b) 0.03, and (c) 0.05. Parameters are chosen in order to qualitatively represent the experimental data in Figs. 2(a) and 2(b) of Ref. [2].

tures such as the ones depicted in Figs. 2 and 3 should appear. Vertical dots with cylindrical symmetry constitute an almost ideal realization of the twofold degenerate case. Some caution should be used, however, regarding the effects of Hund's rule (not included in the present approach) which might introduce some deviation with respect to the behavior depicted in Fig. 2 when the exchange interaction between the degenerate levels is not much smaller than U.

[1] H. van Houten, C. W. J. Beenakker, and A. A. M. Staring in *Single Electron Tunneling,* edited by H. Grabert and M. H. Devoret (Plenum Press, New York, 1991).

[2] D. Goldhaber-Gordon *et al.,* Nature (London) **391**, 156 (1998); D. Goldhaber-Gordon *et al.,* Phys. Rev. Lett. **81**, 5225 (1998).

[3] S. M. Cronenwett *et al.,* Science **281**, 540 (1998).

[4] T. K. Ng and P. A. Lee, Phys. Rev. Lett. **61**, 1768 (1988); L. I. Glazman and M. E. Raikh, JETP Lett. **47**, 452 (1988); S. Hershfield, J. H. Davies and J. W. Wilkins, Phys. Rev. Lett. **67**, 3720 (1991); Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **70**, 2601 (1993).

[5] A. Levy Yeyati, A. Martín-Rodero, and F. Flores, Phys. Rev. Lett. **71**, 2991 (1993).

[6] See also L. I. Glazman and K. A. Matveev, Sov. Phys. JETP **71**, 1031 (1991).

[7] T. Inoshita *et al.,* Phys. Rev. B **48**, 14 725 (1993); C. A. Stafford and S. Das Sarma, Phys. Rev. Lett. **72**, 3590 (1994); T. Pohjola *et al.,* Europhys. Lett. **40**, 189 (1997); W. Izumida, O. Sakai, and Y. Shimizu, J. Phys. Soc. Jpn. **67**, 2444 (1998).

[8] A. Yacoby *et al.,* Phys. Rev. Lett. **74**, 4047 (1995); R. Schuster *et al.,* Nature (London) **385**, 417 (1997).

[9] A. Martín-Rodero *et al.,* Solid State Commun. **44**, 911 (1982); A. Martín-Rodero *et al.,* Phys. Rev. B **33**, 1814 (1986).

[10] H. Kajueter and G. Kotliar, Phys. Rev. Lett. **77**, 131 (1996).

[11] M. Potthoff, T. Wegner, and W. Nolting, Phys. Rev. B **55**, 16 132 (1997).

[12] This approximation is justified for vertical quantum dots where orbital and spin quantum numbers are conserved in tunneling. Generalization of the present approach for including indirect coupling between dot states is in progress.

[13] See, for instance, J. J. Palacios, L. Liu, and D. Yoshioka, Phys. Rev. B **55**, 15 735 (1997).

[14] D. C. Langreth, Phys. Rev. **150**, 516 (1966).

[15] J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).

[16] N. E. Bickers, Rev. Mod. Phys. **59**, 845 (1987).

[17] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).

[18] A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge, England, 1993).