Theory of Coulomb-blockade oscillations in the conductance of a quantum dot

C. W. J. Beenakker

Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands

(Received 28 November 1990)

A linear-response theory is developed for resonant tunneling through a quantum dot of small capacitance, in the regime of thermally broadened resonances. The theory extends the classical theory of Coulomb-blockade oscillations by Kulik and Shekhter to the resonant-tunneling regime. Both the cases of negligible and strong inelastic scattering in the quantum dot are considered. Effects from the *non*-Fermi-Dirac distribution of electrons among the energy levels (occurring when kT is comparable to the level separation) are fully included. Explicit analytic results are obtained for the periodicity, amplitude, line shape, and activation energy of the conductance oscillations.

I. INTRODUCTION

The discreteness of the electron charge manifests itself in the conductance as a result of the Coulomb repulsion of individual electrons. The transfer by tunneling of one electron between two initially neutral regions, of mutual capacitance C, increases the electrostatic energy of the system by an amount of $e^2/2C$. At low temperatures and small applied voltages, conduction is suppressed because of the charging energy. This phenomenon (first reported by Gorter[1] in 1951) is known as the *Coulomb blockade* of single-electron tunneling.[2]

The Coulomb blockade can be removed by capacitive charging (by means of a gate electrode) of the region between two tunnel barriers.[2-4] The series conductance of the tunnel junctions shows oscillations as a function of the gate voltage, due to the periodic modulation of the charging energy. The theory of these *Coulomb-blockade oscillations* was developed by Kulik and Shekhter.[5,6] Theirs is a classical theory, in which the discreteness of the energy spectrum in the region confined by the tunnel barriers is ignored. That is an excellent approximation in metals, where the energy-level separation is in general much smaller than both the charging and the thermal energy.[6]

The situation is different in a semiconductor. In the two-dimensional electron gas of an inversion layer or heterostructure, the Fermi wavelength can be as large as 50 nm. That is two orders of magnitude larger than in a metal, and within reach of today's microfabrication techniques. Resonant tunneling studies have demonstrated energy-level separations $\Delta E \gtrsim 0.1$ meV in submicrometer-size regions in a two-dimensional electron gas, confined electrostatically by means of gate electrodes on top of a GaAs–(Al,Ga)As heterostructure.[7-9] For typical capacitances $C \lesssim 10^{-15}$ F, and at millidegrees Kelvin temperatures, one then has $e^2/C \simeq \Delta E \gg kT$. In this regime the classical theory of the Coulomb-blockade oscillations has to be replaced by a theory which includes the effects of the discreteness of the energy spectrum. That is the problem addressed in the present paper.

Our analysis is a linear-response theory, which yields the conductance of the quantum dot in the limit of vanishingly small source-drain voltage. That is the appropriate limit for the Coulomb-blockade oscillations. The charging energy manifests itself in a different way in the nonlinear current-voltage characteristics, in the form of a stepwise increase known as the *Coulomb staircase*.[2] Averin, Korotkov, and Likharev have recently investigated the effect of a discrete energy spectrum on the Coulomb staircase,[10,11] and the present work proceeds in a similar way.

The experimental motivation for this theoretical work came from the observations of conductance oscillations periodic in the density of a two-dimensional electron gas which is confined to a narrow channel.[12-16] The effect was discovered by Scott-Thomas *et al.*,[12] who interpreted it in terms of the formation of a charge-density wave or "Wigner crystal." In Ref. 17, van Houten and the author proposed the alternative explanation of Coulomb-blockade oscillations, where the charging energy is associated with a region of the narrow channel delimited by two dominant scattering centers. The issue of Coulomb blockade versus Wigner crystal has led to a lively debate,[18] which has not yet been settled.[14-16] We hope that the theory presented here will contribute towards a resolution.

The outline of this paper is as follows. In Sec. II we formulate the problem of the influence of the charging energy on resonant tunneling through a quantum dot, which is weakly coupled to two electron reservoirs. Our main assumption is that the thermal energy exceeds the width of the transmission resonance, so that the conductance peaks are thermally broadened. We specialize to the linear-response regime in Sec. III, and obtain an expression for the conductance [Eq. (3.14)] which can be evaluated straightforwardly, given the energy spectrum and tunnel rates. The present theory takes fully into account that the distribution of electrons among the energy levels in the quantum dot is *different* from the Fermi-Dirac distribution [cf. Eq. (4.14)], whenever the thermal energy is comparable to the level separation— a fact which has so far not been generally appreciated in this field. Limiting forms of the conductance formula

(3.14) in the classical and resonant tunneling regime are derived in Sec. IV. Up to that section we consider the case of no inelastic scattering in the quantum dot (but only in the reservoirs). In Sec. V we turn to the opposite case of strong inelastic scattering in the quantum dot. The results are applied to the Coulomb-blockade oscillations in Sec. VI, where simple analytical expressions are obtained for their periodicity, amplitude, line shape, and activation energy. We consider in that section only the conductance oscillations as a function of electron density (corresponding to oscillations as a function of gate voltage in the experiments mentioned above). The influence of the charging energy on the conductance oscillations as a function of magnetic field (i.e., on the Aharonov-Bohm effect), has been analyzed in Ref. 19.

II. FORMULATION OF THE PROBLEM

We consider a confined region which is weakly coupled via tunnel barriers to two electron reservoirs. The confined region, or "quantum dot," has single-electron energy levels at E_p ($p = 1, 2, \ldots$), labeled in ascending order and measured relative to the bottom of the potential well. Each level contains either one or zero electrons. Spin degeneracy can be included by counting each level twice, and other degeneracies can be included similarly.

Each reservoir is taken to be in thermal equilibrium at temperature T and chemical potential E_F. A continuum of states is assumed in the reservoirs, occupied according to the Fermi-Dirac distribution

$$f(E - E_F) = \left[1 + \exp\left(\frac{E - E_F}{kT}\right)\right]^{-1}. \qquad (2.1)$$

In Fig. 1 we show schematically a cross section of the geometry, and the profile of the electrostatic potential energy along a line through the tunnel barriers.

Because the number of electrons N localized in the quantum dot can take on only integer values, a charge imbalance, and hence an electrostatic potential difference $\phi(Q)$ can arise between the dot and the reservoirs in equilibrium ($Q = -Ne$ is the charge on the dot). We adopt the simple approximation usually made in studies of the Coulomb blockade,[2] of expressing ϕ in terms of an effective capacitance C between dot and reservoirs,

$$\phi(Q) = Q/C + \phi_{ext}, \qquad (2.2)$$

including also a contribution ϕ_{ext} from external charges. The electrostatic energy $U(N) \equiv \int_0^{-Ne} \phi(Q')dQ'$ then takes the form

$$U(N) = (Ne)^2/2C - Ne\phi_{ext}. \qquad (2.3)$$

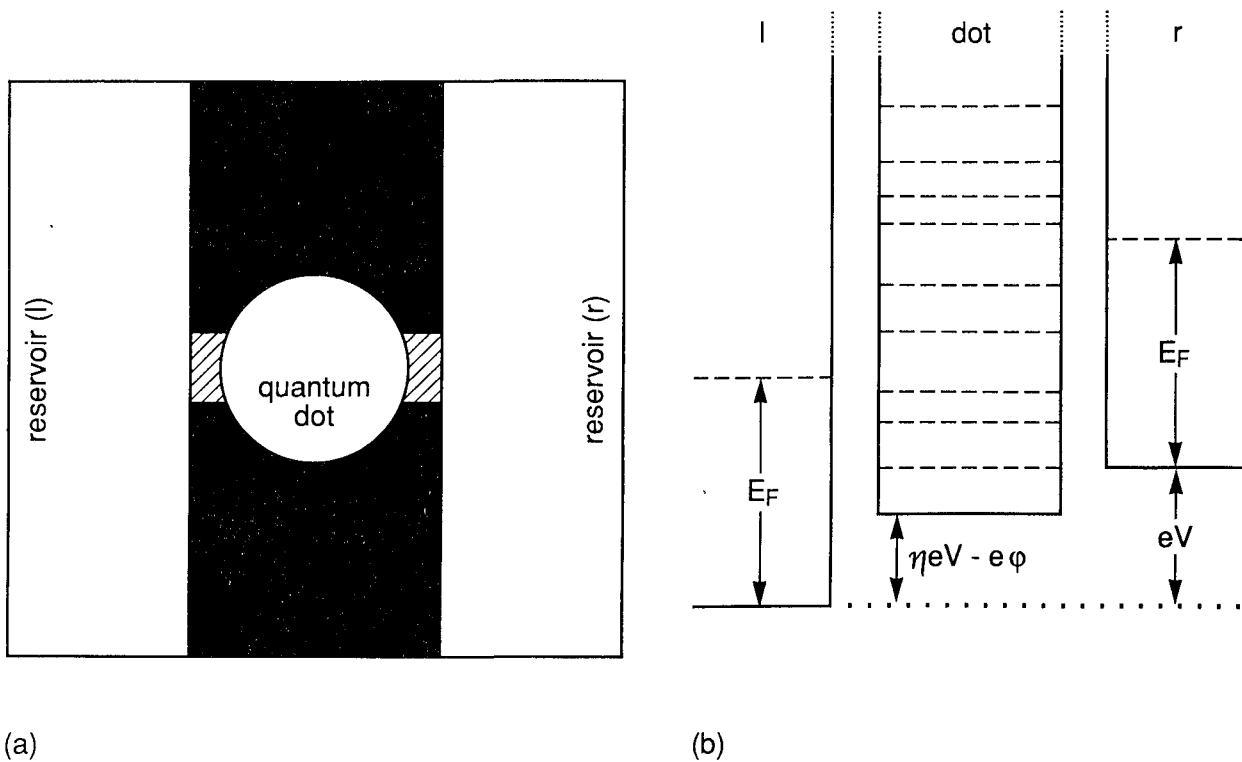(a)   (b)

FIG. 1.   (a) Schematic cross section of the geometry studied in this paper, consisting of a confined region ("quantum dot") weakly coupled to two electron reservoirs via tunnel barriers (hatched). (b) Profile of the electrostatic potential energy (solid curve) along a line through the tunnel barriers. The Fermi levels in the left and right reservoirs, and the discrete energy levels in the quantum dot, are indicated (dashed lines).

In a two-dimensional electron gas, the external charges are supplied by ionized donors and by a gate electrode (with an electrostatic voltage V_{gate} between gate and reservoir). One has $\phi_{\text{ext}} = \phi_{\text{donors}} + \alpha V_{\text{gate}}$, where α (as well as C) is a rational function of the capacitance matrix elements of the system. The quantity $Q_{\text{ext}} \equiv C\phi_{\text{ext}}$ plays the role of an "externally induced charge" on the dot, which can be varied continuously by means of V_{gate} (in contrast to Q which is restricted to integer multiples of e). In terms of Q_{ext} one can write

$$U(N) = (Ne - Q_{\text{ext}})^2/2C - Q_{\text{ext}}^2/2C, \qquad (2.4)$$

which is equivalent to Eq. (2.3). We emphasize that Q_{ext} is an externally controlled variable, via V_{gate}, regardless of the relative magnitude of the various capacitances in the system.

A current I can be passed through the dot by applying a potential difference V between the two reservoirs. The tunnel rate from level p to the left and right reservoirs in Fig. 1 is denoted by Γ_p^l and Γ_p^r, respectively. We assume that both kT and ΔE are $\gg h(\Gamma^l + \Gamma^r)$ (for all levels participating in the conduction), so that the finite width $h\Gamma = h(\Gamma^l + \Gamma^r)$ of the transmission resonance through the quantum dot can be disregarded. This assumption allows us to characterize the state of the quantum dot by a set of occupation numbers, one for each energy level. (As we will see, the restriction kT, $\Delta E \gg h\Gamma$ results in the conductance being much smaller than the quantum e^2/h, which is a necessary condition for the occurrence of the Coulomb blockade.[2]) We also assume conservation of energy in the tunnel process, thus neglecting contributions of higher order in Γ from tunneling via a *virtual* intermediate state in the quantum dot.[20,21] In this section, and in Secs. III and IV, we assume that inelastic scattering takes place exclusively in the reservoirs—not in the quantum dot. The effect of inelastic scattering in the quantum dot is considered in Sec. V.

Energy conservation upon tunneling from an initial state p in the quantum dot (containing N electrons) to a final state in the left reservoir at energy $E^{f,l}$ (in excess of the local electrostatic potential energy) requires that

$$E^{f,l}(N) = E_p + U(N) - U(N - 1) + \eta eV. \qquad (2.5)$$

Here η is the fraction of the applied voltage V which drops over the left barrier. (As we will see in Sec. III, this parameter η drops out of the final expression for the conductance.) The energy conservation condition for tunneling from an initial state $E^{i,l}$ in the left reservoir to a final state p in the quantum dot is

$$E^{i,l}(N) = E_p + U(N + 1) - U(N) + \eta eV, \qquad (2.6)$$

where [as in Eq. (2.5)] N is the number of electrons in the dot *before* the tunneling event. Similarly, for tunneling between the quantum dot and the right reservoir one has the conditions

$$E^{f,r}(N) = E_p + U(N) - U(N - 1) - (1 - \eta)eV, \qquad (2.7)$$

$$E^{i,r}(N) = E_p + U(N + 1) - U(N) - (1 - \eta)eV, \qquad (2.8)$$

where $E^{i,r}$ and $E^{f,r}$ are the energies of the initial and final states in the right reservoir.

The stationary current through the left barrier equals that through the right barrier, and is given by

$$\begin{aligned}
I = -e \sum_{p=1}^{\infty} \sum_{\{n_i\}} &\Gamma_p^l P(\{n_i\}) \\
&\times \{\delta_{n_p,0} f(E^{i,l}(N) - E_F) \\
&- \delta_{n_p,1}[1 - f(E^{f,l}(N) - E_F)]\}.
\end{aligned} \qquad (2.9)$$

The second summation is over all realizations of occupation numbers $\{n_1, n_2, \ldots\} \equiv \{n_i\}$ of the energy levels in the quantum dot, each with stationary probability $P(\{n_i\})$. (The numbers n_i can take on only the values 0 and 1.) In equilibrium, this probability distribution is the Gibbs distribution in the grand canonical ensemble:

$$P_{\text{eq}}(\{n_i\}) = Z^{-1} \exp\left[-\frac{1}{kT}\left(\sum_{i=1}^{\infty} E_i n_i + U(N) - NE_F\right)\right], \qquad (2.10)$$

where $N \equiv \sum_i n_i$, and Z is the partition function,

$$Z = \sum_{\{n_i\}} \exp\left[-\frac{1}{kT}\left(\sum_{i=1}^{\infty} E_i n_i + U(N) - NE_F\right)\right]. \qquad (2.11)$$

The nonequilibrium probability distribution P is a stationary solution of the kinetic equation

$$\begin{aligned}
\frac{\partial}{\partial t} P(\{n_i\}) = 0 = &-\sum_p P(\{n_i\})\delta_{n_p,0}\left[\Gamma_p^l f(E^{i,l}(N) - E_F) + \Gamma_p^r f(E^{i,r}(N) - E_F)\right] \\
&- \sum_p P(\{n_i\})\delta_{n_p,1}\left\{\Gamma_p^l\left[1 - f(E^{f,l}(N) - E_F)\right] + \Gamma_p^r\left[1 - f(E^{f,r}(N) - E_F)\right]\right\} \\
&+ \sum_p P(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots)\delta_{n_p,0}
\end{aligned}$$

$$\times \left\{ \Gamma_p^l \left[1 - f(E^{j,l}(N+1) - E_F) \right] + \Gamma_p^r \left[1 - f(E^{j,r}(N+1) - E_F) \right] \right\}$$

$$+ \sum_p P(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots) \delta_{n_p, 1}$$

$$\times \left[\Gamma_p^l f(E^{i,l}(N-1) - E_F) + \Gamma_p^r f(E^{i,r}(N-1) - E_F) \right]. \qquad (2.12)$$

The kinetic equation (2.12) for the stationary distribution function is equivalent to the set of detailed balance equations (one for each $p = 1, 2, \ldots$)

$$P(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots)\{\Gamma_p^l[1 - f(E^{j,l}(\tilde{N}+1) - E_F)] + \Gamma_p^r[1 - f(E^{j,r}(\tilde{N}+1) - E_F)]\}$$

$$= P(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots)[\Gamma_p^l f(E^{i,l}(\tilde{N}) - E_F) + \Gamma_p^r f(E^{i,r}(\tilde{N}) - E_F)], \quad (2.13)$$

with the notation $\tilde{N} \equiv \sum_{i \neq p} n_i$.

A similar set of equations formed the basis for the work of Averin, Korotkov, and Likharev on the Coulomb staircase in the nonlinear I-V characteristic of a quantum dot.[10] To simplify the solution of the kinetic equation, they assumed that the charging energy e^2/C is much greater than the average level spacing ΔE. In the present paper we restrict ourselves to the regime of linear response, appropriate for the Coulomb-blockade oscillations. Then the conductance can be calculated exactly and analytically.

III. LINEAR RESPONSE

The (two-terminal) linear-response conductance G of the quantum dot is defined as $G = I/V$ in the limit $V \to 0$. To solve the linear-response problem we substitute

$$P(\{n_i\}) \equiv P_{\text{eq}}(\{n_i\}) \left(1 + \frac{eV}{kT} \Psi(\{n_i\}) \right) \qquad (3.1)$$

into the detailed balance equation (2.13), and linearize with respect to V. One finds

$$P_{\text{eq}}(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots) \left\{ \Psi(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots)(\Gamma_p^l + \Gamma_p^r)[1 - f(\varepsilon)] - [\Gamma_p^l \eta - \Gamma_p^r(1 - \eta)]kTf'(\varepsilon) \right\}$$

$$= P_{\text{eq}}(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots) \left\{ \Psi(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots)(\Gamma_p^l + \Gamma_p^r)f(\varepsilon) + [\Gamma_p^l \eta - \Gamma_p^r(1 - \eta)]kTf'(\varepsilon) \right\}, \quad (3.2)$$

where $f'(\varepsilon) \equiv df(\varepsilon)/d\varepsilon$, and we have abbreviated $\varepsilon \equiv E_p + U(\tilde{N}+1) - U(\tilde{N}) - E_F$.

Equation (3.2) can be simplified by making subsequently the substitutions

$$1 - f(\varepsilon) = f(\varepsilon)e^{\varepsilon/kT}, \qquad (3.3)$$

$$P_{\text{eq}}(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots) = P_{\text{eq}}(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots)e^{-\varepsilon/kT}, \qquad (3.4)$$

$$kTf'(\varepsilon)(1 + e^{-\varepsilon/kT}) = -f(\varepsilon). \qquad (3.5)$$

The factors P_{eq} and f cancel, and one is left with the simple equation

$$\Psi(n_1, \ldots, n_{p-1}, 1, n_{p+1}, \ldots) = \Psi(n_1, \ldots, n_{p-1}, 0, n_{p+1}, \ldots) + \frac{\Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} - \eta. \qquad (3.6)$$

The solution is

$$\Psi(\{n_i\}) = \text{const} + \sum_{i=1}^{\infty} n_i \left(\frac{\Gamma_i^r}{\Gamma_i^l + \Gamma_i^r} - \eta \right). \qquad (3.7)$$

The constant first term in Eq. (3.7) takes care of the normalization of P to first order in V, and need not be determined explicitly. Notice that the first-order nonequilibrium correction Ψ to P_{eq} is *zero* if $\eta = \Gamma_i^r/(\Gamma_i^l + \Gamma_i^r)$ for all i. This will happen, in particular, for two identical tunnel barriers (when $\eta = \frac{1}{2}$, $\Gamma_i^l = \Gamma_i^r$). Because of the symmetry of the system, the distribution function then contains only terms of *even* order in V.

Now we are ready to calculate the current I through the quantum dot to first order in V. Linearization of Eq. (2.9), after substitution of Eq. (3.1) for P, gives

$$I = -e\frac{eV}{kT}\sum_p \sum_{\{n_i\}} \Gamma_p^l P_{eq}(\{n_i\})(\delta_{n_p,0}\eta kT f'(\varepsilon) + \delta_{n_p,1}\eta kT f'(\varepsilon) + \Psi(\{n_i\})\delta_{n_p,0}f(\varepsilon) - \Psi(\{n_i\})\delta_{n_p,1}[1 - f(\varepsilon)])$$

$$= \frac{e^2 V}{kT}\sum_p \sum_{\{n_i\}} \Gamma_p^l P_{eq}(\{n_i\})\delta_{n_p,0}f(E_p + U(N+1) - U(N) - E_F)$$

$$\times [\eta + \Psi(n_1,\ldots,n_{p-1},1,n_{p+1},\ldots) - \Psi(n_1,\ldots,n_{p-1},0,n_{p+1},\ldots)]$$

$$= \frac{e^2 V}{kT}\sum_p \sum_{\{n_i\}} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{eq}(\{n_i\})\delta_{n_p,0}f(E_p + U(N+1) - U(N) - E_F). \tag{3.8}$$

In the second equality we have again made use of the identities (3.3)–(3.5), and in the third equality we have substituted Eq. (3.6). Notice that the parameter η has dropped out of the final expression for I.

We define the equilibrium probability distributions

$$P_{eq}(N) = \sum_{\{n_i\}} P_{eq}(\{n_i\})\delta_{N,\sum_i n_i} = \frac{\exp[-\Omega(N)/kT]}{\sum_N \exp[-\Omega(N)/kT]}, \tag{3.9}$$

$$F_{eq}(E_p \mid N) = \frac{1}{P_{eq}(N)} \sum_{\{n_i\}} P_{eq}(\{n_i\})\delta_{n_p,1}\delta_{N,\sum_i n_i}$$

$$= \exp\left(\frac{\mathcal{F}(N)}{kT}\right) \sum_{\{n_i\}} \exp\left(-\frac{1}{kT}\sum_{i=1}^\infty E_i n_i\right) \delta_{n_p,1}\delta_{N,\sum_i n_i}. \tag{3.10}$$

Here $\Omega(N)$ is the thermodynamic potential of the quantum dot, and $\mathcal{F}(N)$ is the free energy of the internal degrees of freedom:

$$\Omega(N) = \mathcal{F}(N) + U(N) - NE_F, \tag{3.11}$$

$$\mathcal{F}(N) = -kT \ln\left[\sum_{\{n_i\}} \exp\left(-\frac{1}{kT}\sum_{i=1}^\infty E_i n_i\right) \delta_{N,\sum_i n_i}\right]. \tag{3.12}$$

The function $P_{eq}(N)$ is the probability that the quantum dot contains N electrons in equilibrium; the function $F_{eq}(E_p \mid N)$ is the conditional probability in equilibrium that level p is occupied given that the quantum dot contains N electrons. In terms of these distribution functions, the conductance $G = I/V$ resulting from Eq. (3.8) equals

$$G = \frac{e^2}{kT}\sum_{p=1}^\infty \sum_{N=0}^\infty \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{eq}(N)[1 - F_{eq}(E_p \mid N)]f(E_p + U(N+1) - U(N) - E_F). \tag{3.13}$$

In view of Eqs. (3.3) and (3.4), Eq. (3.13) can equivalently be written in the form

$$G = \frac{e^2}{kT}\sum_{p=1}^\infty \sum_{N=1}^\infty \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{eq}(N)F_{eq}(E_p \mid N)[1 - f(E_p + U(N) - U(N-1) - E_F)]. \tag{3.14}$$

This equation is the central result of the present paper.

IV. LIMITING FORMS OF THE CONDUCTANCE FORMULA (3.14)

Equation (3.14) reduces to the result of Kulik and Shekhter[6] in the limit $kT \gg \Delta E$, i.e., when the discrete energy spectrum may be treated as a continuum. In that classical limit one may approximate $F_{eq}(E_p \mid N)$ by the Fermi-Dirac distribution

$$F_{eq}(E_p \mid N) = f(E_p - \mu(N)) \quad \text{if } kT \gg \Delta E, \tag{4.1}$$

where the chemical potential $\mu(N)$ is to be determined from the equation

$$\sum_{p=1}^\infty f(E_p - \mu(N)) = N. \tag{4.2}$$

The distribution function $P_{eq}(N)$ takes its classical form

$$P_{class}(N) = \frac{\exp\{-[U(N) + N(\bar\mu - E_F)]/kT\}}{\sum_N \exp\{-[U(N) + N(\bar\mu - E_F)]/kT\}}, \tag{4.3}$$

where $\bar\mu$ is the chemical potential of the dot in equilibrium. The summation over p in Eq. (3.14) may be re-

placed by an integration over E, multiplied by the density of states $\rho(E)$ in the quantum dot. If $kT \ll \bar{\mu}$, E_F, one may in general disregard the energy dependence of the density of states and of the tunnel rates. One can then carry out the integration by means of the formula

$$\int_{-\infty}^{\infty} dy\, f(y)[1 - f(y + x)] = x(1 - e^{-x/kT})^{-1} \equiv g(x).$$

(4.4)

The conductance becomes

$$G = \frac{e^2 \rho}{kT} \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r}$$

$$\times \sum_{N=1}^{\infty} P_{\text{class}}(N) g(U(N) - U(N-1) + \bar{\mu} - E_F)$$

$$\text{if } \Delta E \ll kT \ll \bar{\mu}, E_F, \quad (4.5)$$

where Γ and ρ are evaluated at energy $\bar{\mu}$, and we have used that $\mu(N) \approx \text{const} \equiv \bar{\mu}$ for all N for which $P_{\text{class}}(N)$ differs appreciably from zero. Equation (4.5) is the result of Kulik and Shekhter.[6]

If, in addition to $kT \gg \Delta E$, also $kT \gg e^2/C$ (while still $kT \ll \bar{\mu}$, E_F), then the effect of the charging energy may be ignored as well. In that limit one has $g(x) = kT$, so that Eq. (4 5) reduces to

$$G = e^2 \rho \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r} \equiv G_\infty \quad \text{if } \Delta E, e^2/C \ll kT \ll \bar{\mu}, E_F.$$

(4.6)

The high-temperature resistance $1/G_\infty$ is the sum of the tunnel resistances $1/e^2 \rho \Gamma^l$ and $1/e^2 \rho \Gamma^r$ of the left and right barriers.

In the low-temperature regime $kT \ll \Delta E$ (while still $kT \gg \hbar\Gamma$), Eq. (3.14) can also be written in a simplified form. In that regime the term with $p = N = N_{\text{min}}$ gives the dominant contribution to the sum over p and N. The integer N_{min} minimizes the absolute value of

$$\Delta(N) = E_N + U(N) - U(N-1) - E_F. \quad (4.7)$$

We denote $\Delta_{\text{min}} \equiv \Delta(N_{\text{min}})$. By definition, N_{min} is such

that $P_{\text{eq}}(N)$ is negligibly small for N unequal to either N_{min} or $N_{\text{min}} - 1$, so that

$$P_{\text{eq}}(N_{\text{min}})$$

$$= \frac{\exp[-\Omega(N_{\text{min}})/kT]}{\exp[-\Omega(N_{\text{min}})/kT] + \exp[-\Omega(N_{\text{min}} - 1)/kT]}$$

$$= f(\Delta_{\text{min}}) \quad \text{if } kT \ll \Delta E. \quad (4.8)$$

In the second equality we have used that $\Omega(N) = \sum_{i=1}^{N} E_i + U(N) - N E_F$ in the low-temperature limit. Since, moreover, $F_{\text{eq}}(E_N \mid N) = 1$ in this limit, Eq. (3 14) reduces to

$$G = -e^2 \frac{\Gamma^l_{N_{\text{min}}} \Gamma^r_{N_{\text{min}}}}{\Gamma^l_{N_{\text{min}}} + \Gamma^r_{N_{\text{min}}}} f'(\Delta_{\text{min}})$$

$$\text{if } \hbar\Gamma \ll kT \ll \Delta E, \quad (4.9)$$

where we have used the identity

$$f(x)[1 - f(x)] = -kT f'(x). \quad (4.10)$$

Equation (4.9) can be seen as the usual resonant tunneling formula for a thermally broadened resonance, generalized to include the effect of the charging energy on the resonance condition.

Finally, we consider the limiting form of Eq. (3.14) in the regime $kT \ll e^2/C$ of large charging energy, but with comparable thermal energy and level spacing ($kT \simeq \Delta E$). Then the sum over N reduces to the single term $N = N_{\text{min}}$, but the sum over p has to be retained. Moreover, instead of Eq. (4.8) one has

$$P_{\text{eq}}(N_{\text{min}}) = f(\Omega(N_{\text{min}}) - \Omega(N_{\text{min}} - 1))$$

$$= f(\Delta_{\text{min}} - TS(N_{\text{min}}) + TS(N_{\text{min}} - 1))$$

$$\text{if } kT \ll e^2/C, \quad (4.11)$$

where the entropy $S(N)$ of the quantum dot is obtained from the free energy in the usual way

$$\mathcal{F}(N) = \sum_{i=1}^{N} E_i - TS(N). \quad (4.12)$$

Equation (3.14) now takes the form

$$G = \frac{e^2}{kT} \sum_{p=1}^{\infty} \frac{\Gamma^l_p \Gamma^r_p}{\Gamma^l_p + \Gamma^r_p} F_{\text{eq}}(E_p \mid N_{\text{min}}) [1 - f(\Delta_{\text{min}} - E_{N_{\text{min}}} + E_p)]$$

$$\times f(\Delta_{\text{min}} - TS(N_{\text{min}}) + TS(N_{\text{min}} - 1)) \quad \text{if } kT \ll e^2/C. \quad (4.13)$$

The sum over p in Eq. (4.13) cannot be simplified further if $kT \simeq \Delta E$, but can be evaluated numerically in a straightforward manner (once the energy levels and tunnel rates are given).

It is worth emphasizing that, in this regime $kT \simeq \Delta E$ of comparable thermal energy and level spacing, the distribution $F_{\text{eq}}(E_p \mid N)$ of N electrons among the levels in the quantum dot *differs* appreciably from the Fermi-Dirac distribution (4.1). For example, in the case of a

two-level system $\{E_1, E_2\}$ with $N = 1$, one has from the Gibbs distribution the result

$$F_{\text{eq}}(E_p \mid 1) = \frac{e^{-E_p/kT}}{e^{-E_1/kT} + e^{-E_2/kT}}$$

$$= \left[1 + \exp\left(\frac{E_p - \mu}{kT^*}\right)\right]^{-1}$$

with $\mu \equiv \tfrac{1}{2}(E_1 + E_2)$ $T^* \equiv \tfrac{1}{2}T$. (4.14)

The distribution function can, in this case, be written in the Fermi-Dirac form, but with a fictitious temperature T^* which is one-half the true temperature T. The difference between the true distribution and the Fermi-Dirac distribution (when $kT \simeq \Delta E$) was properly accounted for in some of the previous work[22]—but not in several more recent publications.[10,23]

V. EFFECTS OF INELASTIC SCATTERING

Only elastic tunneling events contribute to $\partial P/\partial t$ in the kinetic equation (2.12). Inelastic scattering is assumed to take place exclusively in the reservoirs, not in the quantum dot. In the present section we relax this assumption. One effect of inelastic scattering is to increase the width $h(\Gamma^l + \Gamma^r) \equiv h\Gamma_{el}$ of the transmission resonance by an amount $h\Gamma_{in}$.[24,25] We continue to make the assumption $kT \gg h\Gamma \equiv h(\Gamma_{el} + \Gamma_{in})$ that the thermal

energy is much greater than the resonance width, so that this effect of inelastic scattering does not play a role. A second effect of inelastic scattering is to thermalize the distribution of electrons among the levels in the quantum dot. This thermalization occurs on the time scale of the energy relaxation time τ_ε. Generally, $\tau_\varepsilon \gtrsim 1/\Gamma_{in}$. We consider here, for comparison with the previous sections, the case $\tau_\varepsilon \ll 1/\Gamma_{el}$ of full thermalization. The analysis given below thus applies to the regime $h\Gamma_{el} \ll h\Gamma_{in} \ll kT$.

Full thermalization means that the conditional probability distribution function $F(E_p \mid N)$ (which is the probability of finding level p occupied given that the quantum dot contains N electrons) retains its equilibrium form (3.10) also for a nonzero applied voltage. Only the probability $P(N)$ of finding N electrons in the quantum dot may differ from the equilibrium distribution (3.9). Instead of the set of detailed balance equations (2.13), one now has the single equation

$$P(N+1)\sum_{p=1}^{\infty} F_{eq}(E_p \mid N+1)\{\Gamma_p^l[1 - f(E^{f,l}(N+1) - E_F)] + \Gamma_p^r[1 - f(E^{f,r}(N+1) - E_F)]\}$$

$$= P(N)\sum_{p=1}^{\infty}[1 - F_{eq}(E_p \mid N)][\Gamma_p^l f(E^{i,l}(N) - E_F) + \Gamma_p^r f(E^{i,r}(N) - E_F)]. \quad (5.1)$$

We substitute $P(N) = P_{eq}(N)[1 + (eV/kT)\Psi(N)]$, and linearize with respect to V. A similar calculation as described in Sec. III leads to the recursion relation

$$\Psi(N+1) = \Psi(N) + \frac{\langle\Gamma^r\rangle_N}{\langle\Gamma^l + \Gamma^r\rangle_N} - \eta, \quad (5.2)$$

with the notation

$$\langle\Gamma\rangle_N \equiv \sum_{p=1}^{\infty} \Gamma_p[1 - F_{eq}(E_p \mid N)]$$

$$\times f(E_p + U(N+1) - U(N) - E_F). \quad (5.3)$$

There is no need to solve Eq. (5.2), since only the difference $\Psi(N+1) - \Psi(N)$ appears in the expression for the current.

The resulting conductance G_{therm} in the case of rapid thermalization may be written in the two equivalent forms

$$G_{therm} = \frac{e^2}{kT}\sum_{N=0}^{\infty} P_{eq}(N)\frac{\langle\Gamma^l\rangle_N\langle\Gamma^r\rangle_N}{\langle\Gamma^l + \Gamma^r\rangle_N}$$

$$= \frac{e^2}{kT}\sum_{N=1}^{\infty} P_{eq}(N)\frac{\langle\langle\Gamma^l\rangle\rangle_N\langle\langle\Gamma^r\rangle\rangle_N}{\langle\langle\Gamma^l + \Gamma^r\rangle\rangle_N}, \quad (5.4)$$

where the double brackets denote

$$\langle\langle\Gamma\rangle\rangle_N \equiv \sum_{p=1}^{\infty} \Gamma_p F_{eq}(E_p \mid N)$$

$$\times[1 - f(E_p + U(N) - U(N-1) - E_F)]. \quad (5.5)$$

The two expressions for the conductance in Eq. (5.4) are equivalent because of the identity

$$P_{eq}(N)\langle\Gamma^{l,r}\rangle_N = P_{eq}(N+1)\langle\langle\Gamma^{l,r}\rangle\rangle_{N+1}. \quad (5.6)$$

The conductance G in the case of no inelastic scattering, obtained in Sec. III [Eqs. (3.13) and (3.14)], may be written in the present notation as

$$G = \frac{e^2}{kT}\sum_{N=0}^{\infty} P_{eq}(N)\left\langle\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}\right\rangle_N$$

$$= \frac{e^2}{kT}\sum_{N=1}^{\infty} P_{eq}(N)\left\langle\left\langle\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}\right\rangle\right\rangle_N. \quad (5.7)$$

Equations (5.4) and (5.7) become identical if either the tunnel rates for the two barriers Γ_p^l and Γ_p^r are different but independent of the level index p—or if they are the same. (In particular, one has $G_{therm} = G$ in the regime $h\Gamma \ll kT \ll \Delta E$ where only a single thermally broadened resonance contributes to the conductance.) The equivalence of G and G_{therm} under these conditions is special for the linear-response conductance. The nonlinear current-voltage characteristic depends on the rate of inelastic scattering even for level-independent tunnel rates.[10]

The regime $kT \lesssim h\Gamma$ cannot be treated by the method used in this paper. For noninteracting electrons, the influence of inelastic scattering in this regime was studied by Stone and Lee[24] and by Büttiker.[25] Their result (for $kT \ll h\Gamma \ll \Delta E$) is that the conductance has the Breit-Wigner form:

$$G_{BW} = \mathcal{G}\frac{e^2}{h}\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}\frac{\Gamma}{(\varepsilon/\hbar)^2 + (\Gamma/2)^2}. \quad (5.8)$$

Here $\mathcal{G}$ is the degeneracy of the resonant level, and ε is the energy separation of that level from the Fermi level in the

reservoirs. Inelastic scattering has the effect of reducing the conductance on resonance by a factor $\Gamma_{el}/(\Gamma_{el}+\Gamma_{in})$. This is to be contrasted with the regime $\hbar\Gamma \ll kT \ll \Delta E$, where inelastic scattering has no effect on the conductance. The reason for the equivalence of G and G_{therm} in the latter regime is that the thermally averaged conductance $-\int G_{BW}f'(\varepsilon)\,d\varepsilon \approx \int G_{BW}\,d\varepsilon/kT$ is independent of Γ_{in}.[24]

A few words on terminology,[25] to make contact with the resonant tunneling literature. Tunneling in the regime $\Gamma_{el} \gg \Gamma_{in}$ of the previous sections is referred to as "coherent resonant tunneling"; In the regime $\Gamma_{el} \ll \Gamma_{in}$ of the present section it is known as "coherent sequential tunneling." Phase coherence plays a role in both these regimes, by establishing the discrete energy spectrum in the quantum dot. The classical, or incoherent, regime is entered when kT or $\hbar\Gamma_{in}$ become greater than ΔE. The discreteness of the energy spectrum can then be ignored.

VI. APPLICATION TO THE COULOMB-BLOCKADE OSCILLATIONS

A. Periodicity

The periodicity of the Coulomb-blockade oscillations can be obtained from the low-temperature expression (4.9) for the conductance of the quantum dot. That equation describes a series of peaks centered at $\Delta_{min} = 0$. In view of Eqs. (2.3) and (4.7), the resulting condition for a conductance peak is that

$$E_F = E_N + U(N) - U(N-1)$$
$$= E_N + (N-\tfrac{1}{2})\frac{e^2}{C} - e\phi_{ext}, \qquad (6.1)$$

for some integer N (which then by definition equals N_{min}). Equation (6.1) equates the equilibrium electrochemical potential of the quantum dot to the Fermi energy of the reservoirs. For an elementary derivation of Eq. (6.1), involving only equilibrium considerations, see Ref. 19.

The conductance of the quantum dot oscillates as a function of the Fermi energy (or electron density) of the reservoirs. Each period the number of electrons in the quantum dot changes by 1. The periodicity ΔE_F follows from Eq. (6.1). If E_F is increased at constant ϕ_{ext}, one has simply

$$\Delta E_F = \Delta E + \frac{e^2}{C} \equiv \Delta E^*. \qquad (6.2)$$

The periodicity of the conductance oscillations is governed by the "renormalized" level spacing ΔE^*. In the absence of charging effects, ΔE_F is determined by the irregular spacing ΔE of the single-electron levels in the quantum dot. The charging energy e^2/C regulates the spacing, once $e^2/C \gtrsim \Delta E$. The spin degeneracy of the levels is lifted by the charging energy. In a plot of G versus E_F this leads to a doublet structure of the oscillations, with a spacing alternating between e^2/C and $\Delta E + e^2/C$.

Experimentally, *both* E_F and ϕ_{ext} are varied by chang-

ing the voltage on the gate electrode which defines a confined region in a two-dimensional electron gas. A change in gate voltage may also affect the shape of the confining potential, and hence the single-electron levels E_p. The determination of the gate-voltage periodicity of the Coulomb-blockade oscillations is for these reasons a rather complicated electrostatic problem, which we will not address in this paper. Note that such a calculation will also have to take into account the fact that the *electrochemical* potential μ_{gate} between gate and reservoirs is the experimentally adjustable variable, rather than the *electrostatic* potential V_{gate}.[26]

B. Amplitude

Observation of the Coulomb-blockade oscillations requires sufficiently low temperatures, such that $kT \lesssim \max(\Delta E, e^2/C)$. Concerning the temperature dependence of the amplitude of the oscillations, we distinguish the two asymptotic regimes $kT \ll \Delta E$ and $\Delta E \ll kT \ll e^2/C$.

If $kT \ll \Delta E$, only a single energy level in the quantum dot participates in the conduction. This is the level labeled by N_{min} in Eq. (4.9). The peak height G_{max}, according to that equation, is given by

$$G_{max} = \frac{e^2}{4kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l+\Gamma^r} \quad \text{if } \hbar\Gamma \ll kT \ll \Delta E, \qquad (6.3)$$

where the tunnel rates refer to level N_{min}. Note that Eq. (6.3) holds regardless of the relative magnitude of ΔE and e^2/C. The peak height increases monotonically as $kT/\Delta E \to 0$, as long as kT is greater than the resonance width $\hbar\Gamma$. The Breit-Wigner formula (5.8) implies for $kT \lesssim \hbar\Gamma$ a saturation of the peak height at a value which is at most $\mathcal{G}e^2/h$.

In the case $\Delta E \ll kT \ll e^2/C$, a continuum of energy levels in the quantum dot participates in the conduction. This is the classical regime studied by Kulik and Shekhter.[6] We include a discussion of this regime for completeness and for comparison with the resonant tunneling regime $kT \lesssim \Delta E$. If $\Delta E \ll kT \ll e^2/C$, only the term $N = N_{min}$ contributes to the sum in Eq. (4.5), where N_{min} minimizes the absolute value of $\Delta(N) = U(N)-U(N-1)+\bar\mu-E_F$ [being the classical correspondence to Eq. (4.7)]. We define $\Delta_{min} \equiv \Delta(N_{min})$. Equation (4.5) reduces to

$$G = \frac{e^2\rho}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l+\Gamma^r}P_{class}(N_{min})g(\Delta_{min})$$

$$= \frac{e^2\rho}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l+\Gamma^r}\frac{\Delta_{min}}{\exp(\Delta_{min}/kT)-\exp(-\Delta_{min}/kT)}$$

$$\text{if } \Delta E \ll kT \ll e^2/C. \qquad (6.4)$$

In the second equality we have used Eq. (4.3) for the classical distribution function, together with the fact that $P_{class}(N) = 0$ if $N \neq N_{min}, N_{min}-1$. The peak height resulting from Eq. (6.4),

$$G_{max} = \frac{e^2\rho}{2}\frac{\Gamma^l\Gamma^r}{\Gamma^l+\Gamma^r} \quad \text{if } \Delta E \ll kT \ll e^2/C, \qquad (6.5)$$

is temperature independent. The reason is that the $1/T$ temperature dependence of G_{max} associated with tunneling through an individual energy level [Eq. (6.3)] is canceled by the T dependence of the number ρkT of levels participating in the conduction. (Note that this cancellation holds only if the tunnel rates are energy independent within the interval kT.)

In the regime $kT \simeq \Delta E \ll e^2/C$ of comparable thermal energy and level spacing, one cannot use the asymptotic formulas (6.3) and (6.5). We have studied this intermediate regime by direct evaluation of Eq. (3.14). Results for G_{max} are plotted in Fig. 2, for the case of equidistant levels with level-independent tunnel rates. Note that the value $\frac{1}{2}G_\infty$ at which G_{max} saturates when $\Delta E \ll kT \ll e^2/C$ lies below the high-temperature limit (4.6). This implies that the value of the conductance on a maximum of the oscillations rises again when kT exceeds e^2/C, as a result of the overlap of adjacent peaks.

C. Line shape

To compare the line shape of the Coulomb-blockade oscillations in the resonant tunneling regime with the classical result of Kulik and Shekhter,[6] we write Eqs. (4.9) and (6.4) in the form

$$G/G_{max} = \cosh^{-2}\left(\frac{\Delta_{min}}{2kT}\right)$$
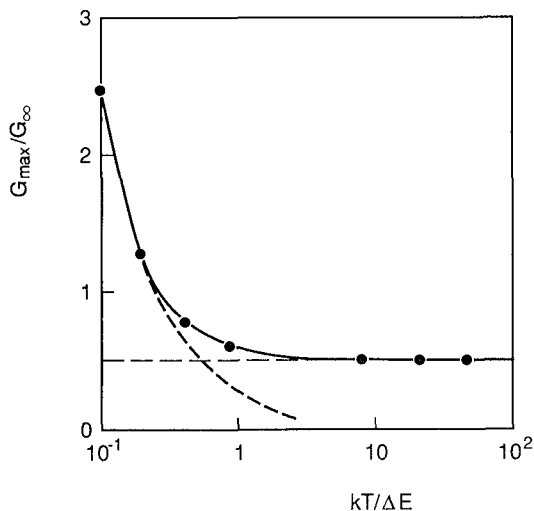
$$\text{if } h\Gamma \ll kT \ll \Delta E, \quad (6.6)$$

FIG. 2. Temperature dependence of the amplitude G_{max} of the Coulomb-blockade oscillations, in the regime $h\Gamma \ll kT \ll e^2/C$. The dashed curves are the asymptotic results (6.3) and (6.5). The dots follow from a numerical evaluation of Eq. (3.14) (the solid curve through the dots is a guide to the eye). The calculation was performed for the case of equidistant nondegenerate energy levels (at separation ΔE), all with the same tunnel rates Γ^l and Γ^r. The conductance is normalized by the classical value G_∞ in the absence of the Coulomb blockade [defined in Eq. (4.6), with $\rho \equiv 1/\Delta E$].
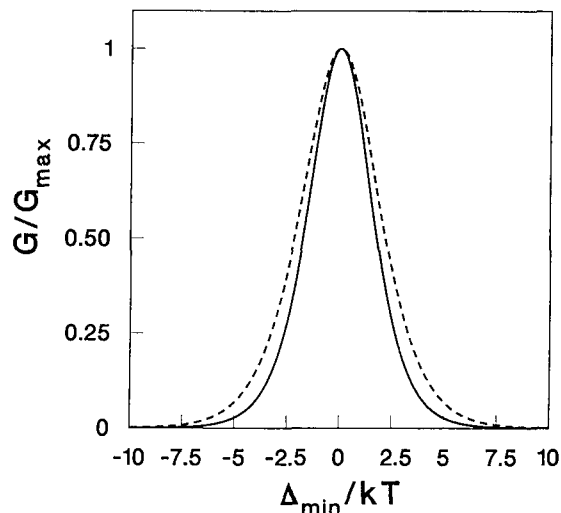
FIG. 3. Comparison of the thermally broadened conductance peak in the resonant tunneling regime $h\Gamma \ll kT \ll \Delta E$ (solid curve) and in the classical regime $h\Gamma \ll \Delta E \ll kT \ll e^2/C$ (dashed curve). The conductance is normalized by the peak height G_{max}, given by Eqs. (6.3) and (6.5) in the two regimes. The energy Δ_{min} is proportional to the Fermi energy in the reservoirs, cf. Eq. (4.7).

$$G/G_{max} = \frac{\Delta_{min}/kT}{\sinh(\Delta_{min}/kT)} \approx \cosh^{-2}\left(\frac{\Delta_{min}}{2.5\,kT}\right)$$

$$\text{if } \Delta E \ll kT \ll e^2/C. \quad (6.7)$$

The second equality in Eq. (6.7) is approximate, but holds to better than 1%. Although the line shapes are different in the two regimes, they are practically indistinguishable if the temperature is used as a fit parameter [as in the experiment of Meirav, Kastner, and Wind,[15] where an excellent fit is obtained to the line shape (6.6)]. In Fig. 3 we compare the two line shapes at the *same* value of the temperature. In that case the difference is clearly noticeable.

The above results imply that a measured temperature dependence of the peak height and width in a quantum dot with well-separated energy scales $h\Gamma$, ΔE, and e^2/C, contains in principle all the information one needs to extract the values of these characteristic energies. The signature of the classical regime $kT \gtrsim \Delta E$ is a peak which becomes narrower with decreasing temperature—while maintaining the same height. In the resonant tunneling regime $kT \lesssim \Delta E$, in contrast, the peak becomes both narrower and higher on lowering T, as long as $kT \gtrsim h\Gamma$.

D. Degenerate energy levels

Degenerate energy levels contribute a single peak to the conductance in the case of noninteracting electrons. The charging energy removes this degeneracy. The influence of Coulomb repulsion on resonant tunneling through a single twofold spin-degenerate state (for $h\Gamma \ll kT \ll e^2/C$) has been studied by Glazman and Matveev,[27] in
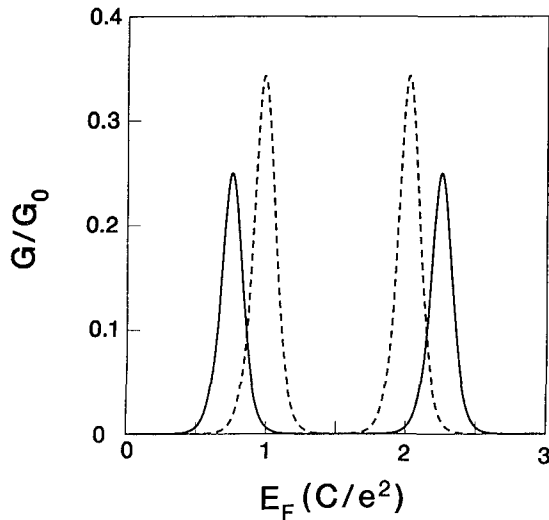
FIG. 4. Conductance vs Fermi energy for a two-level system, consisting either of two nondegenerate levels (solid curve, for $E_1 = 0.25e^2/C$, $E_2 = 0.75e^2/C$), or of a single twofold degenerate level (dashed curve, for $E_1 = E_2 = 0.5e^2/C$). The plot is calculated from Eq. (3.14), for $kT = 0.05e^2/C$, $\phi_{ext} = 0$, and for level-independent tunnel rates. The conductance is normalized by $G_0 \equiv (e^2/kT)\Gamma^l\Gamma^r(\Gamma^l + \Gamma^r)^{-1}$.

connection with experiments on tunneling through metal-oxide-semiconductor structures.[28] To illustrate the generality of the present theory, we show how their special case follows directly from Eq. (4.13).

We apply Eq. (4.13) to a degenerate two-level system ($E_1 = E_2 \equiv E$, $\Gamma_1^{l,r} = \Gamma_2^{l,r} \equiv \Gamma^{l,r}$). The first of the two conductance peaks corresponds to $N_{min} = 1$, $\Delta_{min} = E + U(1) - U(0) - E_F \equiv \varepsilon$. Each of the two levels is occupied with equal probability, $F_{eq}(E \mid 1) = \frac{1}{2}$, and the entropies

are given by $S(1) = k\ln 2$, $S(0) = 0$. Consequently, Eq. (4.13) reduces to

$$G = \frac{e^2}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}[1 - f(\varepsilon)]f(\varepsilon - kT\ln 2)$$

$$= \frac{e^2}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}(\tfrac{3}{2} + e^{-\varepsilon/kT} + \tfrac{1}{2}e^{\varepsilon/kT})^{-1}. \quad (6.8)$$

The second peak of the doublet is the mirror image of the first, and is given by Eq. (6.8) on redefining $\varepsilon \equiv -[E + U(2) - U(1) - E_F]$.

The conductance doublet for a twofold-degenerate energy level is plotted in Fig. 4, as a dashed curve. Each peak is slightly asymmetrical, falling off more rapidly on the side facing the other peak of the doublet. The peak height is $(6 - 4\sqrt{2})(e^2/kT)\Gamma^l\Gamma^r(\Gamma^l + \Gamma^r)^{-1}$, in agreement with Glazman and Matveev.[27] The solid curve shows the effect of removal of the degeneracy (e.g., by the Zeeman energy in the case of spin degeneracy). Once the level splitting $\Delta E \gg kT$, each of the two conductance peaks is given by Eq. (4.9). The peaks have become symmetrical, and are about 25% smaller than in the case of degenerate levels.[27]

E. Activation energy

The renormalized level spacing $\Delta E^* \equiv \Delta E + e^2/C$, which according to Eq. (6.2) determines the periodicity of the Coulomb-blockade oscillations, equals twice the activation energy of the conductance minima. To see this, we consider first a two-level system $\{E_1, E_2 \equiv E_1 + \Delta E\}$, for a Fermi energy $E_F = E_1 + \frac{1}{2}\Delta E + \frac{1}{2}[U(2) - U(0)]$ halfway between the two conductance peaks. If both levels have the same tunnel rates, this point is by symmetry the minimum of the conductance doublet. Starting from Eq. (3.14), one finds after some algebra that at this value of E_F the conductance minimum G_{min} equals

$$G_{min} = 2\frac{e^2}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}f(\varepsilon^+)[1 - f(\varepsilon^-)]\frac{f(\varepsilon^-) + 1 - f(\varepsilon^+)}{f(\varepsilon^+) + 1 - f(\varepsilon^-)}$$

with $\varepsilon^+ \equiv \frac{1}{2}(\Delta E + e^2/C) \equiv \frac{1}{2}\Delta E^*$, $\varepsilon^- \equiv \frac{1}{2}(\Delta E - e^2/C)$. (6.9)

If $e^2/C \ll \Delta E$, one has $\varepsilon^+ \approx \varepsilon^-$, so that Eq. (6.9) is just twice the expression (4.9) for a single thermally broadened resonance, evaluated at $\frac{1}{2}\Delta E$ from the maximum. For a non-negligible charging energy, one cannot simply construct the doublet as a superposition of two individual resonances, and the more complicated expression (6.9) is needed. Equation (6.9) can be simplified if the separation of the peaks is much larger than their width, which implies $kT \ll \Delta E + e^2/C$. Then $f(\varepsilon^+)$ can be approximated by zero in the quotient of Fermi-Dirac distributions appearing in Eq. (6.9). The result is

$$G_{min} \approx 2\frac{e^2}{kT}\frac{\Gamma^l\Gamma^r}{\Gamma^l + \Gamma^r}[1 + f(\varepsilon^-)]\exp(-\varepsilon^+/kT). \quad (6.10)$$

It follows that G_{min} depends exponentially on the tem-

perature, $G_{min} \propto \exp(-E_{act}/kT)$, with activation energy

$$E_{act} = \frac{1}{2}(\Delta E + e^2/C) = \frac{1}{2}\Delta E^*. \quad (6.11)$$

The exponential decay of the conductance at the minima of the Coulomb-blockade oscillations results from the suppression of tunneling processes which conserve energy in the intermediate state in the quantum dot (cf. Sec. II). Tunneling via a *virtual* intermediate state is not suppressed at low temperatures, and contributes a small temperature-independent residual conductance.[20,21]

The result (6.11) for the activation energy, derived for a two-level system, holds more generally when only two levels compete in the conduction. This occurs in the resonant tunneling regime $kT \ll \Delta E$. In the classical

regime $kT \gg \Delta E$ a continuum of energy levels has a non-negligible tunnel probability, and the analysis has to be modified. Equation (4.5) is then the appropriate starting point. The resulting activation energy turns out to be $e^2/2C$, still consistent with Eq. (6.11).[29]

[1] C. J. Gorter, Physica 17, 777 (1951).

[2] For reviews of the field of single-electron tunneling in metals, we refer to K. K. Likharev, IBM J. Res. Dev. 32, 144 (1988); D. V. Averin and K. K. Likharev, in Mesoscopic Phenomena in Solids, edited by B. L. Al'tshuler, P. A. Lee, and R. A. Webb (Elsevier, Amsterdam, 1991); a recent review of single-electron tunneling in semiconductors is H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in Single Charge Tunneling, NATO Advanced Study Institute, Series B: Physics, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1991).

[3] M. Amman, K. Mullen, and E. Ben-Jacob, J. Appl. Phys. 65, 339 (1989).

[4] L. I. Glazman and R. I. Shekhter, J. Phys. Condens. Matter 1, 5811 (1989).

[5] R. I. Shekhter, Zh. Eksp. Teor. Fiz. 63, 1410 (1972) [Sov. Phys.—JETP 36, 747 (1973)].

[6] I. O. Kulik and R. I. Shekhter, Zh. Eksp. Teor. Fiz. 68, 623 (1975) [Sov. Phys.—JETP 41, 308 (1975)].

[7] B. J. van Wees, L. P. Kouwenhoven, C. J. P. M. Harmans, J. G. Williamson, C. E. Timmering, M. E. I. Broekaart, C. T. Foxon, and J. J. Harris, Phys. Rev. Lett. 62, 2523 (1989).

[8] C. G. Smith, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C 21, L893 (1988).

[9] For a review of the field of quantum transport in semiconductor nanostructures, see C. W. J. Beenakker and H. van Houten, Solid State Phys. 44, 1 (1991).

[10] D. V. Averin and A. N. Korotkov, Zh. Eksp. Teor. Fiz. 97, 1661 (1990) [Sov. Phys.—JETP 70, 937 (1990)]; A. N. Korotkov, D. V. Averin, and K. K. Likharev, Physica B 165 & 166, 927 (1990).

[11] The Coulomb staircase problem in a quantum dot has also been addressed by A. Groshev [Phys. Rev. B 42, 5895 (1990)], to interpret a resonant tunneling experiment performed by M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, T. M. Moore, and A. E. Wetsel, Phys. Rev. Lett. 60, 535 (1988).

[12] J. H. F. Scott-Thomas, S. B. Field, M. A. Kastner, H. I. Smith, and D. A. Antoniadis, Phys. Rev. Lett. 62, 583 (1989).

[13] U. Meirav, M. A. Kastner, M. Heiblum, and S. J. Wind, Phys. Rev. B 40, 5871 (1989).

[14] S. B. Field, M. A. Kastner, U. Meirav, J. H. F. Scott-Thomas, D. A. Antoniadis, H. I. Smith, and S. J. Wind, Phys. Rev. B 42, 3523 (1990).

[15] U. Meirav, M. A. Kastner, and S. J. Wind, Phys. Rev. Lett. 65, 771 (1990).

[16] A. A. M. Staring, H. van Houten, C. W. J. Beenakker, and C. T. Foxon, in High Magnetic Fields in Semiconductor Physics III, edited by G. Landwehr (Springer, Berlin, 1991).

[17] H. van Houten and C. W. J. Beenakker, Phys. Rev. Lett. 63, 1893 (1989).

[18] M. A. Kastner, S. B. Field, U. Meirav, J. H. F. Scott-Thomas, D. A. Antoniadis, and H. I. Smith, Phys. Rev. Lett. 63, 1894 (1989).

[19] C. W. J. Beenakker, H. van Houten, and A. A. M. Staring, following paper, Phys. Rev. B 44, 1657 (1991); also in Granular Nanoelectronics, NATO Advanced Study Institute, Series B: Physics, edited by D. K. Ferry, J. Barker, and C. Jacoboni (Plenum, New York, in press).

[20] L. I. Glazman and K. A. Matveev, Pis'ma Zh. Eksp. Teor. Fiz. 51, 425 (1990) [JETP Lett. 51, 484 (1990)].

[21] D. V. Averin and Yu. V. Nazarov, Phys. Rev. Lett. 65, 2446 (1990).

[22] R. Landauer and M. Büttiker, Phys. Rev. Lett. 54, 2049 (1985); M. Büttiker, Phys. Rev. B 36, 3548 (1987).

[23] A. Kumar, S. E. Laux, and F. Stern, Phys. Rev. B 42, 5166 (1990).

[24] A. D. Stone and P. A. Lee, Phys. Rev. Lett. 54, 1196 (1985).

[25] M. Büttiker, Phys. Rev. B 33, 3020 (1986); IBM J. Res. Dev. 32, 63 (1988).

[26] S. Colak (private communication).

[27] L. I. Glazman and K. A. Matveev, Pis'ma Zh. Eksp. Teor. Fiz. 48, 403 (1988) [JETP Lett. 48, 445 (1988)].

[28] A. B. Fowler, G. L. Timp, J. J. Wainer, and R. A. Webb, Phys. Rev. Lett. 57, 138 (1986).

[29] Meirav, Kastner, and Wind (Ref. 15) have reported an activation energy which is a factor 3.5 smaller than the energy $e^2/2C_g$ associated with the capacitance C_g between the confined region and the gate substrate. In interpreting this result, it should be kept in mind that the dot-reservoir capacitance C, which determines the activation energy, may well be larger than C_g in their structure, due to the presence of an additional gate on top of the heterostructure.