

Theory of the thermopower of a quantum dot

C. W. J. Beenakker

Instituut-Lorentz, University of Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands

A. A. M. Staring

Philips Research Laboratories, P.O. Box 80000, 5600 JA Eindhoven, The Netherlands

(Received 15 April 1992)

A linear-response theory is presented for the thermopower of a quantum dot of small capacitance. In the *classical* regime (thermal energy kT much greater than the level spacing ΔE), the thermopower oscillates around zero in a sawtooth fashion as a function of Fermi energy (as long as kT is small compared to the charging energy e^2/C). The periodicity of the oscillations is the same as that of the previously studied Coulomb-blockade oscillations in the conductance, and is determined by the difference in ground-state energies on addition of a single electron to the quantum dot. In the *quantum* regime of resonant tunneling ($kT \ll \Delta E$), a fine structure is predicted to develop on the oscillations. Unlike the Coulomb-blockade oscillations, the periodicity of the fine structure is determined by the excitation spectrum at a *constant* number of electrons on the quantum dot.

I. INTRODUCTION

Most of the research on transport properties of mesoscopic systems has focused on electrical conduction.¹ The thermal and thermoelectric transport coefficients are more difficult to measure on small length scales than the conductance, and have therefore attracted less attention. Recent experiments²⁻⁴ have shown that it is, in fact, possible to create an appreciable temperature difference in a nanostructure and to measure the voltage V induced by this temperature difference ΔT (the Seebeck effect). The ratio $S = -V/\Delta T$ under conditions of zero electrical current is the thermopower, one of the two thermoelectric (or "off-diagonal") transport coefficients. (The other, the coefficient of the Peltier effect, is related to S by an Onsager relation⁵). The thermopower depends sensitively on the energy dependence of the transmission probabilities of the system,^{6,7} and, in general, contains different information than the two diagonal transport coefficients (the electrical and thermal conductances).

Two mesoscopic thermoelectric phenomena are by now reasonably well understood. These are the quantum-size effect in the thermopower of a quantum point contact^{4,8-10} (consisting of peaks in S which line up with the plateaus of quantized conductance), and the quantum interference effect in the thermopower of a disordered conductor^{2,3,11-14} (the analog of universal conductance fluctuations). These two phenomena belong, respectively, to the ballistic and to the diffusive transport regime. Electron-electron interactions do not play a dominant role in these two regimes. Very recently, thermoelectric experiments¹⁵ have been performed in a transport regime, where the interactions are of crucial importance. This is the regime of single-electron tunneling through a quantum dot of small capacitance. The measured thermopower¹⁵ shows pronounced oscillations as a

function of gate voltage (i.e., essentially as a function of Fermi energy). The thermopower oscillations have the same periodicity as the oscillations in the conductance of the same structure. The latter so-called "Coulomb-blockade oscillations" are known to result from a periodic modulation of the charging energy associated with the tunneling of a single electron into the quantum dot.¹⁶ The theory of the Coulomb-blockade oscillations in the conductance has been developed both for the classical case of a continuous energy spectrum^{17,18} and for the case of resonant tunneling through discrete energy levels in a quantum dot.¹⁹⁻²¹

Theoretical studies of the effect of Coulomb repulsion on thermoelectric transport have thus far been carried out for the case of a single tunnel junction,²² but not for the double-junction geometry relevant to a quantum dot. The purpose of the present paper is to provide such a theory. We do this by extending the theory by one of us²⁰ for single-electron tunneling in response to a voltage difference, to single-electron tunneling in response to both voltage and temperature differences. The requirement of zero net tunnel-current yields a relation between V and ΔT , which in the regime of linear response contains the thermopower as proportionality coefficient ($V = -S\Delta T$). As in previous work on the conductance oscillations,¹⁷⁻²¹ the Coulomb repulsion is treated within the framework of the "orthodox model" of single-electron tunneling.²³ Two major simplifications of this model are that virtual tunnel processes are neglected and that the electrostatic energy is described by the classical charging energy $(Ne)^2/2C$ (with N the number of electrons on the dot and C its capacitance to the surroundings). The first simplification requires that the conductance of the quantum dot is much smaller than the conductance quantum e^2/h . The second simplification requires that the screening length is much smaller than the size of the

dot. Both conditions are well fulfilled in typical metal structures, but not so well in a semiconducting two-dimensional electron gas. Much of the recent theoretical work²⁴ on the Coulomb blockade of the conductance in a double-junction geometry deals with modifications of the orthodox model which are required when either or both of the above conditions are not fulfilled. In the present study we will stay within the orthodox model and see what effects this model predicts for the thermopower of a quantum dot.

As we will show, the oscillations in the thermopower resulting from the Coulomb blockade are qualitatively different from those in the conductance. First of all, in the *classical* regime $kT \gg \Delta E$ (with T the temperature and ΔE the level spacing in the dot), the thermopower oscillations are skewed (positive slope smaller than negative slope in absolute value), while the conductance oscillations are symmetric (equal positive and negative slopes). If $kT \ll e^2/C$, while still $kT \gg \Delta E$, the thermopower oscillates between $\pm e/4CT$ in the form of a sawtooth. The periodicity $\Delta E_F = \Delta E + e^2/C$ of the oscillations as a function of Fermi energy equals the energy required to add a single electron to the quantum dot in the ground state. In the *quantum* regime $kT \ll \Delta E$ the sawtooth thermopower oscillations develop a fine structure, provided $\Delta E < e^2/C$. The periodicity $\delta E_F = \Delta E$ of the fine structure equals the energy between thermally excited states of the quantum dot at a constant number of electrons on the dot. The magnitude of the fine structure is a fraction of order $\Delta E/(e^2/C)$ of the main oscillations. A corresponding fine structure exists on the conductance oscillations, but in that case the relative amplitude is exponentially small, of order $\exp(-\Delta E/kT)$ (which is why it was not noticed before).

The thermopower is thus a different “spectroscopic” tool than the conductance²⁵ and capacitance,²⁶ which measure the *addition* spectrum (differences in ground state energies on increasing N) and not the *excitation* spectrum (energy differences at constant N). As demonstrated recently,^{21,27} the fine structure on the nonlinear current-voltage characteristic (Coulomb staircase) can also provide information on the excitation spectrum. A fundamental difference is that the thermopower oscillations are a linear response phenomenon, which (in contrast to the Coulomb staircase) involves only an infinitesimally small perturbation of the system from equilibrium.

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 at different voltages and different temperatures. We specialize to the linear-response regime in Sec. III, and obtain an expression for the thermopower [Eq. (3.13)] which can be evaluated straightforwardly, given the energy spectrum and tunnel rates. This formula corresponds for the thermopower to the conductance formula derived in Ref. 20. The derivation in Secs. II and III follows that paper closely. Limiting forms of the thermopower formula (3.13) in the classical and quantum regimes are derived in Secs. IV and V, respectively. Simple analytical expressions are obtained for the periodicity, amplitude, and line

shape of the thermopower oscillations in the two regimes, and compared with plots which are calculated directly from Eq. (3.13). We have also included a comparison with the conductance oscillations, to illustrate the similarities and differences between the two phenomena.

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 ($p = 1, 2, \dots$), calculated by treating the electron-electron interaction in a mean-field (Hartree) approximation. The levels are 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. In principle, the position of the levels may depend on the number of electrons in the quantum dot, but for simplicity we will ignore such dependence in what follows. Each reservoir is taken to be in thermal equilibrium, but between the reservoirs there is a temperature difference $\Delta T = T_l - T_r$ as well as a voltage difference $V = -\Delta\mu/e$ ($\Delta\mu = \mu_l - \mu_r$ is the difference in electrochemical potential and $-e$ is the electron charge). The states in the left (l) and right (r) reservoirs are occupied according to the Fermi-Dirac distributions

$$f_l(E - E_F) = \left[1 + \exp\left(\frac{E - E_F}{kT_l}\right) \right]^{-1}, \quad (2.1)$$

$$f_r(E - E_F) = \left[1 + \exp\left(\frac{E - E_F}{kT_r}\right) \right]^{-1},$$

where the Fermi energy E_F is measured relative to the local conduction-band bottom in the reservoirs. The thermopower S of the quantum dot is defined as minus the ratio of the voltage and temperature differences under the condition that the current I between the two reservoirs is zero,

$$S \equiv - \lim_{\Delta T \rightarrow 0} \frac{V}{\Delta T} \Big|_{I=0}. \quad (2.2)$$

The limit $\Delta T \rightarrow 0$ ensures that we remain in the regime of linear response. 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 even if $V = 0$ ($Q = -Ne$ is the charge on the dot). As discussed in Sec. I, we adopt the “orthodox model” of the Coulomb blockade,²³ in which ϕ is expressed in terms of an effective (N -independent) capacitance C between dot and reservoirs,

$$\phi(Q) = Q/C + \phi_{\text{ext}}, \quad (2.3)$$

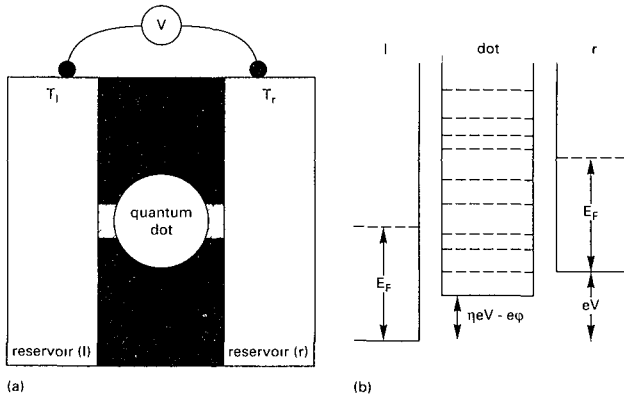


FIG. 1. (a) Schematic cross section of the geometry studied in this paper, consisting of a confined region (“quantum dot”) weakly coupled to two electron reservoirs via tunnel barriers (hatched). (b) Profile of the electrostatic potential energy (solid curve) along a line through the tunnel barriers.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

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

$$U(N) = (Ne)^2/2C - Ne\phi_{\text{ext}}. \quad (2.4)$$

The tunnel rate from level p to the left and right reser-

voirs is denoted by Γ_p^l and Γ_p^r , respectively. A possible dependence of the tunnel rates on N is ignored. We assume that both the thermal energy kT and the level separation ΔE are much greater than $\hbar(\Gamma^l + \Gamma^r)$, so that virtual tunnel processes (and the resulting finite width of the transmission resonance through the quantum dot) can be disregarded. This assumption (which also implies that the conductance of the dot is much smaller than e^2/h) allows us to characterize the state of the quantum dot by a set occupation numbers, one for each energy level. The transport through the dot can then be described by rate equations.²³ We also assume that inelastic scattering takes place exclusively in the reservoirs — not in the quantum dot.

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 conduction-band bottom), requires that

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

Here η is the fraction of the voltage V which drops over the left barrier. 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

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

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

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

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

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

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

The second summation is over all realizations of occupation numbers $\{n_1, n_2, \dots\} \equiv \{n_i\}$ 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:

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

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

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

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

$$\begin{aligned}
 \frac{\partial}{\partial t}P(\{n_i\}) = 0 = & - \sum_p P(\{n_i\})\delta_{n_p,0}[\Gamma_p^l f_l(E^{i,l}(N) - E_F) + \Gamma_p^r f_r(E^{i,r}(N) - E_F)] \\
 & - \sum_p P(\{n_i\})\delta_{n_p,1}(\Gamma_p^l \{1 - f_l[E^{f,l}(N) - E_F]\} + \Gamma_p^r \{1 - f_r[E^{f,r}(N) - E_F]\}) \\
 & + \sum_p P(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots)\delta_{n_p,0}\{\Gamma_p^l [1 - f_l(E^{f,l}(N+1) - E_F)] \\
 & \qquad \qquad \qquad + \Gamma_p^r [1 - f_r(E^{f,r}(N+1) - E_F)]\} \\
 & + \sum_p P(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)\delta_{n_p,1}\{\Gamma_p^l f_l[E^{i,l}(N-1) - E_F] + \Gamma_p^r f_r(E^{i,r}(N-1) - E_F)\}.
 \end{aligned}
 \tag{2.12}$$

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

$$\begin{aligned}
 P(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots)\{\Gamma_p^l [1 - f_l(E^{f,l}(\tilde{N} + 1) - E_F)] + \Gamma_p^r [1 - f_r(E^{f,r}(\tilde{N} + 1) - E_F)]\} \\
 = P(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)[\Gamma_p^l f_l(E^{i,l}(\tilde{N}) - E_F) + \Gamma_p^r f_r(E^{i,r}(\tilde{N}) - E_F)],
 \end{aligned}
 \tag{2.13}$$

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

III. LINEAR RESPONSE

To solve the linear-response problem we substitute

$$P(\{n_i\}) \equiv P_{\text{eq}}(\{n_i\}) [1 + \Psi(\{n_i\})]
 \tag{3.1}$$

into the detailed balance equation (2.13), and expand to first order in ΔT and V . We define $T_r \equiv T$, $T_l \equiv T + \Delta T$, and

$$f(\varepsilon) \equiv [1 + \exp(\varepsilon/kT)]^{-1},
 \tag{3.2}$$

so that we can write $f_r(\varepsilon) = f(\varepsilon)$, $f_l(\varepsilon) = f(\varepsilon) - (\varepsilon\Delta T/T)f'(\varepsilon) + \mathcal{O}(\Delta T)^2$. The result of the linearization of Eq. (2.13) is

$$\begin{aligned}
 P_{\text{eq}}(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots)\{\Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots)(\Gamma_p^l + \Gamma_p^r)[1 - f(\varepsilon)] \\
 - eV[\Gamma_p^l \eta - \Gamma_p^r(1 - \eta)]f'(\varepsilon) + (\varepsilon\Delta T/T)\Gamma_p^l f'(\varepsilon)\} \\
 = P_{\text{eq}}(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)\{\Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)(\Gamma_p^l + \Gamma_p^r)f(\varepsilon) \\
 + eV[\Gamma_p^l \eta - \Gamma_p^r(1 - \eta)]f'(\varepsilon) - (\varepsilon\Delta T/T)\Gamma_p^l f'(\varepsilon)\},
 \end{aligned}
 \tag{3.3}$$

where we have abbreviated $\varepsilon \equiv E_p + U(\tilde{N} + 1) - U(\tilde{N}) - E_F$.

Equation (3.3) can be simplified by making subsequently the substitutions

$$1 - f(\varepsilon) = f(\varepsilon)e^{\varepsilon/kT},
 \tag{3.4}$$

$$P_{\text{eq}}(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) = P_{\text{eq}}(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)e^{-\varepsilon/kT},
 \tag{3.5}$$

$$kT f'(\varepsilon)(1 + e^{-\varepsilon/kT}) = -f(\varepsilon).
 \tag{3.6}$$

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

$$\Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) = \Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) + \frac{eV}{kT} \left(\frac{\Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} - \eta \right) + \frac{\varepsilon\Delta T}{kT^2} \frac{\Gamma_p^l}{\Gamma_p^l + \Gamma_p^r}.
 \tag{3.7}$$

This equation cannot be solved explicitly for $\Delta T \neq 0$, because ε (defined above) depends on the set $\{n_i\}$ (through \tilde{N}). Fortunately, no explicit solution is needed to determine the current.

Linearization of the expression (2.9) for the current gives, after substitution of Eq. (3.1) for P ,

$$\begin{aligned}
I &= -e \sum_p \sum_{\{n_i\}} \Gamma_p^l P_{\text{eq}}(\{n_i\}) \{eV \delta_{n_p,0} \eta f'(\varepsilon) + eV \delta_{n_p,1} \eta f'(\varepsilon) - (\varepsilon \Delta T / T) \delta_{n_p,0} f'(\varepsilon) - (\varepsilon \Delta T / T) \delta_{n_p,1} f'(\varepsilon) \\
&\quad + \Psi(\{n_i\}) \delta_{n_p,0} f(\varepsilon) - \Psi(\{n_i\}) \delta_{n_p,1} [1 - f(\varepsilon)]\} \\
&= e \sum_p \sum_{\{n_i\}} \Gamma_p^l P_{\text{eq}}(\{n_i\}) \delta_{n_p,0} f(\varepsilon) \left(\eta \frac{eV}{kT} - \frac{\Delta T}{kT^2} \varepsilon + \Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) \right. \\
&\quad \left. - \Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) \right) \\
&= \frac{e}{kT} \sum_{p=1}^{\infty} \sum_{N=0}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(N) [1 - F_{\text{eq}}(E_p | N)] f[E_p + U(N+1) - U(N) - E_F] \\
&\quad \times \left(eV - \frac{\Delta T}{T} [E_p + U(N+1) - U(N) - E_F] \right). \tag{3.8}
\end{aligned}$$

In the second equality we have again made use of the identities (3.4)–(3.6), and in the third equality we have substituted Eq. (3.7). We have defined the equilibrium probability distributions

$$P_{\text{eq}}(N) = \sum_{\{n_i\}} P_{\text{eq}}(\{n_i\}) \delta_{N, \sum_i n_i}, \tag{3.9}$$

$$F_{\text{eq}}(E_p | N) = \frac{1}{P_{\text{eq}}(N)} \sum_{\{n_i\}} P_{\text{eq}}(\{n_i\}) \delta_{n_p,1} \delta_{N, \sum_i n_i}. \tag{3.10}$$

The function $P_{\text{eq}}(N)$ is the probability that the quantum dot contains N electrons in equilibrium; The function $F_{\text{eq}}(E_p | N)$ is the conditional probability in equilibrium that level p is occupied given that the quantum dot contains N electrons.

To obtain the thermopower we calculate $S = -V/\Delta T$ for $I = 0$. The result from Eq. (3.8) is

$$\begin{aligned}
S &= -\frac{e}{kT^2 G} \sum_{p=1}^{\infty} \sum_{N=0}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} [E_p + U(N+1) - U(N) - E_F] \\
&\quad \times P_{\text{eq}}(N) [1 - F_{\text{eq}}(E_p | N)] f[E_p + U(N+1) - U(N) - E_F], \tag{3.11}
\end{aligned}$$

where the conductance G of the quantum dot is given by

$$G = \frac{e^2}{kT} \sum_{p=1}^{\infty} \sum_{N=0}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(N) [1 - F_{\text{eq}}(E_p | N)] f[E_p + U(N+1) - U(N) - E_F]. \tag{3.12}$$

In view of Eqs. (3.4) and (3.5), Eqs. (3.11) and (3.12) can equivalently be written in the form

$$\begin{aligned}
S &= -\frac{e}{kT^2 G} \sum_{p=1}^{\infty} \sum_{N=1}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} [E_p + U(N) - U(N-1) - E_F] \\
&\quad \times P_{\text{eq}}(N) F_{\text{eq}}(E_p | N) \{1 - f[E_p + U(N) - U(N-1) - E_F]\}, \tag{3.13}
\end{aligned}$$

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

Equation (3.14) for the conductance was previously obtained in Ref. 20. Equation (3.13) for the thermopower is the central result of the present paper.

IV. CLASSICAL LIMIT

In the limit $kT \gg \Delta E$ the discrete energy spectrum may be treated as a continuum. In that classical limit one may approximate $F_{\text{eq}}(E_p | N)$ by the Fermi-Dirac distribution

$$F_{\text{eq}}(E_p | N) = f[E_p - \mu(N)] \quad \text{if } \Delta E \ll kT, \tag{4.1}$$

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 $P_{\text{eq}}(N)$ takes its classical form

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

where $\bar{\mu}$ is the chemical potential of the dot in equilibrium. The summations over p in Eqs. (3.13) and (3.14) may be replaced by integrations over E , multiplied by the density of states ρ in the quantum dot. For simplicity, we disregard here the energy dependence of the density of states and of the tunnel rates. The expression for the thermopower then becomes

$$S = -\frac{1}{eT} \frac{\sum_{N=1}^{\infty} P_{\text{class}}(N) \int_{-\infty}^{\infty} d\varepsilon \varepsilon f(\varepsilon - \Delta(N)) [1 - f(\varepsilon)]}{\sum_{N=1}^{\infty} P_{\text{class}}(N) \int_{-\infty}^{\infty} d\varepsilon f(\varepsilon - \Delta(N)) [1 - f(\varepsilon)]} \quad \text{if } \Delta E \ll kT, \quad (4.4)$$

where $\Delta(N) \equiv U(N) - U(N-1) + \bar{\mu} - E_F$. [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.]

The integral in the denominator of Eq. (4.4) is elementary:

$$g(x) \equiv \int_{-\infty}^{\infty} dy f(y-x) [1 - f(y)] = x(1 - e^{-x/kT})^{-1}. \quad (4.5)$$

The integral in the numerator,

$$h(x) \equiv \int_{-\infty}^{\infty} dy y f(y-x) [1 - f(y)], \quad (4.6)$$

can be evaluated by noting that the expression

$$2h(x) - xg(x) = \int_{-\infty}^{\infty} dy y \{f(y-x)[1 - f(y)] + f(y)[1 - f(y+x)]\} \quad (4.7)$$

vanishes since it is an integral over an odd function. As a result,

$$h(x) = \frac{1}{2}xg(x). \quad (4.8)$$

Substituting the formulas (4.5) and (4.8) into Eq. (4.4), we obtain for the thermopower the result

$$S = -\frac{1}{2eT} \frac{\sum_{N=1}^{\infty} P_{\text{class}}(N) \Delta(N) g(\Delta(N))}{\sum_{N=1}^{\infty} P_{\text{class}}(N) g(\Delta(N))} \quad \text{if } \Delta E \ll kT. \quad (4.9)$$

If, in addition to $\Delta E \ll kT$, also $kT \ll e^2/C$, then only the term $N = N_{\text{min}}$ where N_{min} minimizes $|\Delta(N)|$ contributes to the sums in the numerator and denominator of Eq. (4.9). We define $\Delta_{\text{min}} \equiv \Delta(N_{\text{min}})$. Equation (4.9) reduces to

$$S = -\frac{1}{2eT} \Delta_{\text{min}} = -\frac{1}{2eT} \left((N_{\text{min}} - \frac{1}{2}) \frac{e^2}{C} + \bar{\mu} - e\phi_{\text{ext}} - E_F \right) \quad \text{if } \Delta E \ll kT \ll e^2/C. \quad (4.10)$$

According to Eq. (4.10), the thermopower oscillates around zero in a sawtooth manner as a function of the Fermi energy, jumping discontinuously between $\pm e/4CT$ each time N_{min} changes by one. The peak-to-peak amplitude of the oscillations equals $(e^2/2C)/kT$ in units of k/e , and is therefore a direct measure of the relative magnitude of charging and thermal energies. The (positive) slope of the sawtooth $dS/dE_F = 1/2eT$ depends only on the temperature, not on the capacitance. The periodicity of the thermopower oscillations is the same as that of the Coulomb-blockade oscillations in the conductance, which in the classical limit are given by¹⁷

$$G = G_{\text{max}}^{\text{class}} \frac{\Delta_{\text{min}}/kT}{\sinh(\Delta_{\text{min}}/kT)}, \quad (4.11)$$

$$G_{\text{max}}^{\text{class}} \equiv \frac{e^2 \rho}{2} \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r}. \quad (4.12)$$

The conductance peaks are located at the zeros of the thermopower, at values of the Fermi energy for which $\Delta(N) = 0$, $N = 1, 2, \dots$

Although the periodicity is the same, the amplitude and line shape of the thermopower oscillations (4.10) is entirely different from what would follow from a naive application of Mott's rule to Eq. (4.11) for the conductance:

$$S_{\text{Mott}} \equiv -\frac{\pi^2 k}{3 e} kT \frac{d \ln G}{d E_F} = -\frac{\pi^2 k}{3 e} \left[\text{coth} \left(\frac{\Delta_{\min}}{kT} \right) - \frac{kT}{\Delta_{\min}} \right]. \quad (4.13)$$

Equation (4.13) would predict an amplitude of order k/e of the thermopower oscillations which is parametrically smaller (by a factor of order e^2/CkT) than the correct result (4.10). Of course, Mott's rule is derived (e.g., in Ref. 28) for a noninteracting electron gas with a weakly energy-dependent conductance, and is therefore clearly not applicable to the Coulomb-blockade regime. Still, its breakdown even as "a rule of thumb" is noteworthy.

To illustrate the sawtooth thermopower oscillations, we have computed S and G from our basic Eqs. (3.13) and (3.14) for parameters in the classical regime. The results are plotted in Fig. 2.

V. QUANTUM LIMIT

We now turn to the low-temperature limit $kT \ll \Delta E$, when the discreteness of the energy spectrum of the confined region can no longer be ignored. This is the quantum regime, in which the Coulomb blockade and resonant tunneling interplay. We still assume $kT \gg h(\Gamma^l + \Gamma^r)$, so that the resonances are thermally broadened and Eq. (3.13) applies. In the limit $kT \ll \Delta E$ this general expression for the thermopower can be simplified as follows. First of all, we note that the term with $N = N_{\min}$ gives the dominant contribution to the sums over N in Eqs. (3.13) and (3.14). The integer N_{\min} minimizes $|\Delta(N)|$, where

$$\Delta(N) = E_N + U(N) - U(N-1) - E_F. \quad (5.1)$$

We define

$$\Delta_{\min} \equiv \Delta(N_{\min}) = E_{N_{\min}} + (N_{\min} - \frac{1}{2}) \frac{e^2}{C} - e\phi_{\text{ext}} - E_F. \quad (5.2)$$

The expression for the thermopower thus contains only sums over p ,

$$S = -\frac{1}{eT} \frac{\sum_{p=1}^{\infty} \gamma_p (\Delta_p + \Delta_{\min}) F_{\text{eq}}(E_p | N_{\min}) [1 - f(\Delta_p + \Delta_{\min})]}{\sum_{p=1}^{\infty} \gamma_p F_{\text{eq}}(E_p | N_{\min}) [1 - f(\Delta_p + \Delta_{\min})]} \quad \text{if } kT \ll \Delta E, \quad (5.3)$$

where we have defined $\gamma_p \equiv \Gamma_p^l \Gamma_p^r (\Gamma_p^l + \Gamma_p^r)^{-1}$ and $\Delta_p \equiv E_p - E_{N_{\min}}$. Furthermore we have, in the low-temperature limit

$$F_{\text{eq}}(E_p | N_{\min}) \approx \begin{cases} 1 & \text{for } 1 \leq p \leq N_{\min} \\ e^{-\Delta_p/kT} & \text{for } p > N_{\min}, \end{cases} \\ \equiv \min(1, e^{-\Delta_p/kT}), \quad (5.4)$$

$$1 - f(\Delta_p + \Delta_{\min}) \approx \min(1, e^{(\Delta_p + \Delta_{\min})/kT}). \quad (5.5)$$

Equation (5.5) does not hold if $|\Delta_p + \Delta_{\min}| \lesssim kT$ — but for $kT \ll \Delta E$ this is a vanishingly small interval in Fermi energy, which can be safely disregarded. Equations (5.4) and (5.5) combined give

$$F_{\text{eq}}(E_p | N_{\min}) [1 - f(\Delta_p + \Delta_{\min})] \approx \min(1, e^{\Delta_{\min}/kT}) \begin{cases} 1 & \text{if } -\Delta_{\min} < \Delta_p \leq 0 \\ 1 & \text{if } 0 \leq \Delta_p < -\Delta_{\min} \\ 0 & \text{otherwise,} \end{cases} \quad (5.6)$$

again with disregard of the threshold interval $|\Delta_p + \Delta_{\min}| \lesssim kT$. Substitution into Eq. (5.3) yields the result

$$S = -\frac{1}{eT} \frac{\sum_{p=N_{\min}}^{N_c} \gamma_p (\Delta_p + \Delta_{\min})}{\sum_{p=N_{\min}}^{N_c} \gamma_p} \quad \text{if } kT \ll \Delta E. \quad (5.7)$$

For $\Delta_{\min} < 0$ the sum over p extends over the integers $N_{\min}, N_{\min} + 1, \dots, N_c$, where N_c is the largest integer such that $\Delta_p + \Delta_{\min} < 0$. For $\Delta_{\min} > 0$ the sum over p extends over $N_{\min}, N_{\min} - 1, \dots, N_c$, with N_c the smallest integer such that $\Delta_p + \Delta_{\min} > 0$.

Since Δ_{\min} depends linearly on E_F , Eq. (5.7) tells us that the thermopower in the quantum limit is a piecewise linear function of the Fermi energy, with a derivative $dS/dE_F = 1/eT$ which is twice as large as the clas-

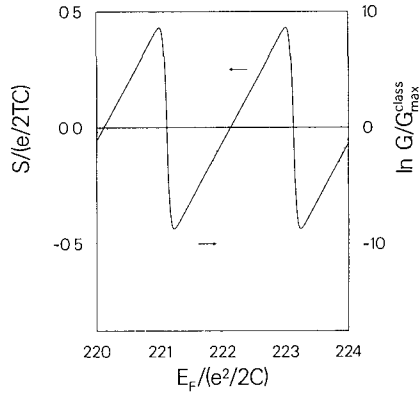


FIG. 2. Thermopower oscillations as a function of Fermi energy in the classical regime (solid curve). The Coulomb-blockade oscillations in the conductance are shown as well (dotted), on a logarithmic scale. The curves are computed from Eqs. (3.13) and (3.14), for a series of equidistant nondegenerate levels with $\Delta E = 0.01 e^2/2C$, $kT = 0.05 e^2/2C$, and taking level-independent tunnel rates.

sical slope $1/2eT$ [Eq. (4.10)]. The thermopower jumps discontinuously when either of the integers N_{\min} or N_c changes by one. Jumps of the first kind (change in N_{\min}) occur when $\Delta(N_{\min}) = -\Delta(N_{\min} + 1)$, i.e., when

$$E_F = \frac{1}{2}(E_{N_{\min}} + E_{N_{\min}+1}) + N_{\min} \frac{e^2}{C} - e\phi_{\text{ext}}. \quad (5.8)$$

Jumps of the second kind (change in N_c) occur when $\Delta_p + \Delta_{\min} = 0$ for some integer $p \neq N_{\min}$, i.e., when

$$E_F = E_p + (N_{\min} - \frac{1}{2}) \frac{e^2}{C} - e\phi_{\text{ext}}, \quad p \neq N_{\min}. \quad (5.9)$$

At finite temperature these two kinds of discontinuities manifest themselves as oscillations in S as a function of E_F with two different periodicities. The *long-period oscillations* [Eq. (5.8)] are due to changes in the number of electrons on the dot in the ground state. Their periodicity $\Delta E_F = \Delta E + e^2/C$ is determined by differences in ground-state energies. If $\mathcal{E}(N)$ denotes the ground-state energy for N electrons on the dot, and $\mu(N) \equiv \mathcal{E}(N+1) - \mathcal{E}(N)$ the chemical potential associated with this ground state, then $\Delta E_F = \mu(N+1) - \mu(N)$. The *short-period oscillations* [Eq. (5.9)] have periodicity $\delta E_F = \Delta E$ determined by energy differences between ground state and excited states of the quantum dot at a *constant* number of electrons on the dot.

Equation (5.7) takes a particularly simple form in the special case of equidistant energy levels ($E_p = p\Delta E$) with level-independent tunnel rates:

$$\begin{aligned} S &= -\frac{1}{eT} \frac{\sum_{p=N_{\min}}^{N_c} [(p - N_{\min})\Delta E + \Delta_{\min}]}{|N_c - N_{\min} + 1|}, \\ &= -\frac{1}{eT} \left(\frac{\Delta E}{2} (N_c - N_{\min}) + \Delta_{\min} \right), \\ &= -\frac{1}{eT} \left[-\frac{\Delta E}{2} \text{Int} \left(\frac{\Delta_{\min}}{\Delta E} \right) + \Delta_{\min} \right]. \end{aligned} \quad (5.10)$$

Here $\text{Int } x$ is the integer part of x for $x > 0$, and minus the integer part of $|x|$ for $x < 0$. Since $|\Delta_{\min}| \leq \frac{1}{2}(\Delta E + e^2/C)$, Eq. (5.10) reduces to $S = -\Delta_{\min}/eT$ if $e^2/C \leq \Delta E$. The thermopower thus has only the long-period oscillations [consisting of a sawtooth with periodicity $\Delta E + e^2/C$ and amplitude $(\Delta E + e^2/C)/eT$], if the charging energy is less than the level spacing. The short-period oscillations appear as soon as $e^2/C > \Delta E$. For $e^2/C \gg \Delta E$, the short-period oscillations are a fine structure on the envelope $S_{\text{envelope}} = -\Delta_{\min}/2eT$, obtained from Eq. (5.10) by replacing $\text{Int } x$ by x . Note that S_{envelope} is nothing but the classical expression (4.10) of the thermopower, obtained in Sec. IV by ignoring the discreteness of the energy spectrum.

It is interesting to compare Eqs. (5.7) and (5.10) for the thermopower with the corresponding expressions for the conductance in the quantum limit $kT \ll \Delta E$,

$$G = \frac{e^2}{4kT} \cosh^{-2}(\Delta_{\min}/2kT) \sum_{p=N_{\min}}^{N_c} \gamma_p, \quad (5.11)$$

$$G = G_{\text{max}}^{\text{quantum}} \cosh^{-2}(\Delta_{\min}/2kT) \left[\text{Int} \left(\frac{|\Delta_{\min}|}{\Delta E} \right) + 1 \right],$$

$$G_{\text{max}}^{\text{quantum}} \equiv \frac{e^2}{4kT} \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r}. \quad (5.12)$$

Equation (5.12) is for the case of equidistant energy levels with level-independent tunnel rates. Equation (5.11) follows from Eq. (3.14) by a similar analysis as for the thermopower, using also the result²⁰ $P_{\text{eq}}(N_{\min}) = f(\Delta_{\min})$ (valid if $kT \ll \Delta E$). We find that the fine structure in the thermopower has a corresponding fine structure in the conductance, consisting of a steplike feature with periodicity ΔE . However, in contrast to the fine structure in the thermopower, the fine structure in the conductance is exponentially small on the scale of the conductance peak itself: The n th conductance step has magnitude $4G_{\text{max}}^{\text{quantum}} \exp(-n\Delta E/kT) \ll G_{\text{max}}^{\text{quantum}}$. (The exponentially small magnitude of the fine structure in the conductance is the reason that it was not noticed in Ref. 20.)

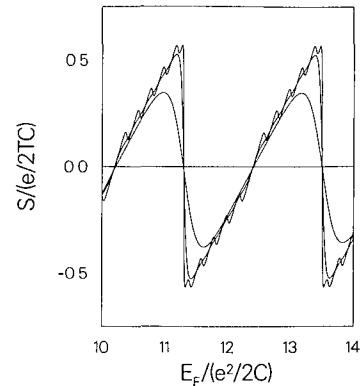


FIG. 3. Development of fine structure on the thermopower oscillations on lowering the temperature from $kT = (0.2 \text{ to } 0.05 \text{ to } 0.01) \times e^2/2C$. The curves are computed from Eq. (3.13), for a series of equidistant nondegenerate levels with $\Delta E = 0.2 e^2/2C$, and taking level-independent tunnel rates.

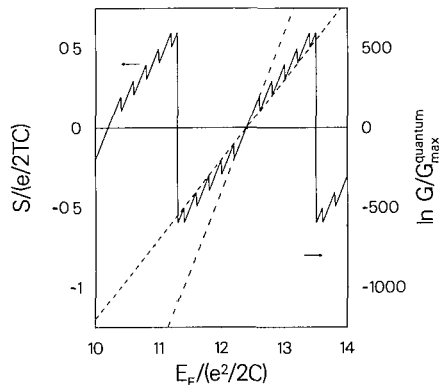


FIG. 4. Low-temperature limit of the thermopower oscillations of Fig. 3 (solid curve). The conductance oscillations are shown for comparison (dotted). The curves are computed from Eqs. (3.13) and (3.14), for a series of equidistant non-degenerate levels with $\Delta E = 0.2 e^2/2C$, $kT = 0.001 e^2/2C$, and taking level-independent tunnel rates. The dash-dotted line has slope $dS/dE_F = 1/eT$, the dashed line has a slope which is twice as small.

To illustrate the development of fine structure on the thermopower oscillations, we have computed S from our basic Eq. (3.13) for parameters between the classical and the quantum regime. The results are plotted in Fig. 3. The quantum limit is shown in Fig. 4. Notice that the slope $dS_{\text{envelope}}/dE_F = 1/2eT$ of the envelope (dashed) is twice as small as the slope $dS/dE_F = 1/eT$ of the piecewise linear segments (dash-dotted). The conductance [calculated from Eq. (3.14)] is plotted in the same figure, for comparison (dotted curve). The steplike fine structure on the conductance is not visible, because of its exponentially small magnitude. We emphasize that the spacing of the fine structure is determined by the level spacing, according to Eq. (5.9), and is only equally spaced if the levels themselves are equally spaced (as for the model calculations in the figures). In Fig. 5 we show the effect of spin degeneracy of the energy levels on the thermopower. The average density of states is the same in the top and bottom panel, but in the top panel the levels

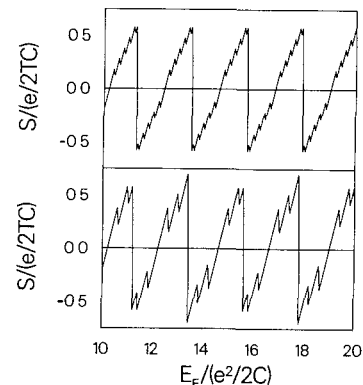


FIG. 5. Effect of spin degeneracy on the fine structure. The curves are computed from Eq. (3.13), at a temperature $kT = 0.001 e^2/2C$, for a series of equidistant levels with level-independent tunnel rates. In the top panel the levels are *nondegenerate*, with spacing $\Delta E = 0.2 e^2/2C$ (as in Fig. 4). In the bottom panel the levels are *twofold degenerate*, with spacing $\Delta E = 0.4 e^2/2C$ (so that the average density of states is the same).

are nondegenerate, while in the bottom panel each level is twofold degenerate. This change in the energy-level spectrum of the quantum dot has essentially no effect on the frequency of the long-period oscillations, while the frequency of the short-period oscillations is changed by a factor of 2. Notice also that the fine structure differs from one oscillation to the other in the case of spin-degenerate levels, reflecting the fact that the excitation spectrum is different if there are an even or an odd number of electrons on the dot.

ACKNOWLEDGMENTS

Valuable discussions with B. W. Alphenaar, H. van Houten, and L. W. Molenkamp are gratefully acknowledged. Research at the University of Leiden is supported financially by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek" (NWO) via the "Stichting voor Fundamenteel Onderzoek der Materie" (FOM).

¹Two general reviews of electrical conduction in mesoscopic systems are: *Mesoscopic Phenomena in Solids*, edited by B. L. Al'tshuler, P. A. Lee, and R. A. Webb (North-Holland, Amsterdam, 1991); C. W. J. Beenakker and H. van Houten, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1991), Vol. 44, p. 1.

²G. M. Gusev, Z. D. Kvon, and A. G. Pogosov, Pis'ma Zh. Eksp. Teor. Fiz. **51**, 151 (1990) [JETP Lett. **51**, 171 (1990)].

³B. L. Gallagher, T. Galloway, P. Beton, J. P. Oxley, S. P. Beaumont, S. Thoms, and C. D. W. Wilkinson, Phys. Rev. Lett. **64**, 2058 (1990).

⁴L. W. Molenkamp, H. van Houten, C. W. J. Beenakker, R. Eppenga, and C. T. Foxon, Phys. Rev. Lett. **65**, 1052 (1990); L. W. Molenkamp, Th. Gravier, H. van Houten, O.

J. A. Buijk, M. A. A. Mabeoone, and C. T. Foxon, *ibid.* **68**, 3765 (1992).

⁵S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (Dover, New York, 1984).

⁶U. Sivan and Y. Imry, Phys. Rev. B **33**, 551 (1986).

⁷P. N. Butcher, J. Phys. Condens. Matter **2**, 4869 (1990).

⁸P. Streda, J. Phys. Condens. Matter **1**, 1025 (1989).

⁹C. R. Proetto, Phys. Rev. B **44**, 9096 (1991).

¹⁰Y. Okuyama, T. Sakuma, and N. Tokuda, Surf. Sci. **263**, 258 (1992).

¹¹A. V. Anisovich, B. L. Al'tshuler, A. G. Aronov, and A. Yu. Zyuzin, Pis'ma Zh. Eksp. Teor. Fiz. **45**, 237 (1987) [JETP Lett. **45**, 295 (1987)].

¹²R. A. Serota, M. Ma, and B. Goodman, Phys. Rev. B **37**, 6540 (1988).

- ¹³G. B. Lesovik and D. E. Khmel'nitskiĭ, Zh. Eksp. Teor. Fiz. **94**, 164 (1988) [Sov. Phys. — JETP **67**, 957 (1988)].
- ¹⁴D. P. DiVincenzo (unpublished).
- ¹⁵A. A. M. Staring, L. W. Molenkamp, B. W. Alphenaar, H. van Houten, C. W. J. Beenakker, and C. T. Foxon (unpublished).
- ¹⁶A review of single-electron tunneling in semiconductor nanostructures is: H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in *Single Charge Tunneling*, edited by H. Grabert and M. H. Devoret, Vol. 294 of *NATO Advanced Study Institute, Series B: Physics* (Plenum, New York, 1992), p. 167.
- ¹⁷I. O. Kulik and R. I. Shekhter, Zh. Eksp. Teor. Fiz. **68**, 623 (1975) [Sov. Phys. JETP **41**, 308 (1975)].
- ¹⁸L. I. Glazman and R. I. Shekhter, J. Phys. Condens. Matter **1**, 5811 (1989).
- ¹⁹Y. Meir, N. Wingreen, and P. A. Lee, Phys. Rev. Lett. **66**, 3048 (1991).
- ²⁰C. W. J. Beenakker, Phys. Rev. B **44**, 1646 (1991).
- ²¹D. V. Averin, A. N. Korotkov, and K. K. Likharev, Phys. Rev. B **44**, 6199 (1991).
- ²²M. Amman, E. Ben-Jacob, and J. Cohn, Z. Phys. B **85**, 405 (1991).
- ²³The "orthodox model" of single-electron tunneling in metals is reviewed by 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 (North-Holland, Amsterdam, 1991).
- ²⁴L. I. Glazman and K. A. Matveev, Zh. Eksp. Teor. Fiz. **98**, 1834 (1990) [Sov. Phys. — JETP **71**, 1031 (1990)]; D. V. Averin and Yu. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990); A. A. Odintsov, G. Falci, and G. Schön, Phys. Rev. B **44**, 13089 (1991); G.-L. Ingold, P. Wyrowski, and H. Grabert, Z. Phys. B **85**, 443 (1991); W. Häusler, B. Kramer, and J. Mašek, *ibid.* **85**, 435 (1991); A. Groshev, T. Ivanov, and V. Valtchinov, Phys. Rev. Lett. **66**, 1082 (1991); A. Nakano, R. K. Kalia, and P. Vashishta, Phys. Rev. B **44**, 8121 (1991); A. N. Korotkov and Yu. V. Nazarov, Physica B **173**, 217 (1991); N. F. Johnson and M. C. Payne, Phys. Rev. B **45**, 3819 (1992); P. L. McEuen, E. B. Foxman, J. Kinaret, U. Meirav, M. A. Kastner, N. S. Wingreen, and S. J. Wind, *ibid.* **45**, 11419 (1992).
- ²⁵P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, Phys. Rev. Lett. **66**, 1926 (1991).
- ²⁶R. C. Ashoori, H. L. Stormer, J. S. Weiner, L. N. Pfeiffer, S. J. Pearton, K. W. Baldwin, and K. W. West, Phys. Rev. Lett. **68**, 3088 (1992).
- ²⁷B. Su, V. J. Goldman, and J. E. Cunningham, Science **255**, 313 (1992); P. Guéret, N. Blanc, R. Germann, and H. Rothuizen, Phys. Rev. Lett. **68**, 1896 (1992); A. T. Johnson, L. P. Kouwenhoven, W. de Jong, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon (unpublished).
- ²⁸J. M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, London, 1972).