Polarization charge relaxation and the Coulomb staircase in ultrasmall double-barrier tunnel junctions

C. Schönenberger and H. van Houten
Philips Research Laboratories, Eindhoven, The Netherlands

C.W.J. Beenakker
Instituut-Lorentz, University of Leiden, Leiden, The Netherlands

Experimental results are reported on the Coulomb staircase in a double-barrier tunnel junction formed by the tip of a cryogenic scanning tunneling microscope, an ultrasmall Au particle (4 nm in diameter), a ZrO$_2$ tunnel oxide-barrier, and a Au covered substrate. Two discrepancies with the orthodox model (global rule) are frequently found: an enhanced asymptotic separation of the current-voltage characteristic, and an anomalous suppression of the first current steps in the region around zero voltage. These observations are tentatively attributed to the effect of slow dielectric relaxation of polarization charge induced in the tunnel oxide. This notion is supported, although some discrepancies remain, by a calculation of Coulomb staircases for large relaxation times according to the local rule.

1. Introduction

The quantized nature of the electron charge greatly influences the electric transport in networks of small tunnel junctions, through the dependence of the tunneling probability on the change in charging energy brought about by the tunneling of a single electron [1]. A widely studied structure is composed of a small island, coupled over two tunnel junctions to an external voltage source. In order to observe single-electron tunneling (SET) effects, two conditions have to be fulfilled: the total capacitance $C$ of the island to its environment should be so small that the single-electron charging energy $e^2/2C$ exceeds the thermal energy $kT$, and the resistances of the tunnel junctions should be larger than the resistance quantum $h/e^2$. The generic SET-effect is the Coulomb blockade: the suppression of the tunneling current at low voltages $|U| \ll e/2C$. An additional manifestation (present, if the double-barrier tunnel junction is strongly asymmetric) is the Coulomb staircase: a sequence of equidistantly spaced steps in the current-voltage characteristic. Each step corresponds to the addition of a single excess electron to the island.

The continuing progress in nanofabrication technology has allowed one to explore the physics of single-electron tunneling in metal [2] and semiconductor [3] based tunnel junctions with dimensions down to $\sim 50$ nm. Here, SET effects are observed at $T \approx 4$ K. A much smaller size regime can be explored by using small metal particles (2–5 nm), sandwiched between a metallic substrate with an oxide tunnel barrier and the metallic tip of a scanning-tunneling microscope (STM) [4]. The increased charging energy $e^2/C$ in such a system has recently allowed two of us
to demonstrate SET effects conclusively at room temperature [5]. The higher temperature scale is obviously of importance for device applications of SET. In addition, the new size regime may give rise to deviations from the conventional theory (the so-called orthodox model [1]).

In this paper we discuss low-temperature (4.2 K) current–voltage (I–U) characteristics obtained on small metal particles using STM. These characteristics display significant deviations from the Coulomb staircase predicted by the orthodox model. In an attempt to understand the discrepancies with the experimental data, we calculated the influence of a finite relaxation time $\tau_r$ of induced charge in the oxide tunnel barrier on the I–U characteristic employing both the local and global rule [7]. Our motivation was that the small capacitance might invalidate the assumption $\tau_r < \hbar C/e^2$ of the orthodox model. As we will discuss, some aspects of the experiment can indeed be accounted for, but in particular, we hope that this paper will stimulate further work that will completely elucidate the measured deviations.

2. Experimental results

The current–voltage characteristics are measured in a low-temperature STM operating in a helium gas atmosphere at 4.2 K, while the tip is held fixed over one metal particle. The substrate–tunnel-barrier–particle system (see inset of fig. 1) consists of a stack of a 100 nm thick Au film (conducting substrate), followed by a 0.5–1 nm thick layer of ZrO$_2$ (oxide tunnel barrier of junction 2), covered by Au particles $\approx$4.5 nm in diameter. The capacitance $C_2$ of the particle–substrate junction is larger than the capacitance $C_1$ of the particle-STM-tip junction due to the large dielectric constant $\varepsilon_2 \approx 10$ of the ZrO$_2$. We estimate $C_2 \approx 10^{-19}$ F, corresponding to $e^2/C \approx 150$ meV. Experimental and preparative details can be found in ref. [6].

The experiments reported here are in the regime $C_2 \gg C_1$, $R_2 \gg R_1$. In fig. 1 a measured I–U characteristic (thick solid curve) is compared with a fit (thin solid curve) to the orthodox model, i.e. neglecting relaxation effects. The fitting parameters (capacitance $C_i$, resistance $R_i$ of the tunnel junctions $i = 1, 2$, and the offset charge $q_0$) are obtained as follows [8]: The period $U_p = e/C_2$ of the step pattern, i.e. the voltage interval $U_p$ between neighbouring steps, yields $C_2$. The shift of the staircase with respect to $U = 0$ determines $q_0$. The slope on the plateaux (in between steps) is given by $K_L = C_1/C$ with $C = C_1 + C_2$. The total resistance $R = R_1 + R_2$ is obtained from the asymptotic slope of the I–U characteristic for large voltages. The resistance ratio $r_2 = R_2/R$ is estimated from the voltage range over which the staircase is visible. The parameters for the curve in fig. 1 are: $q_0/e = -0.13$, $r_2 = 0.96$ ($R_2 = 24 R_1$), $\kappa_1 = 0.05$ ($C_2 = 19 C_1$), $U_p = 0.15$ V, and $R = 14$ GΩ.

There are two clear discrepancies between the theory and the experiment: first, the horizontal displacement $U_a$ between the linear asymptotes for large positive and small negative applied voltages, shown in fig. 1 by thin solid lines, is considerably larger in the experimental curve. For $R_2 \gg R_1$ and $C_2 \gg C_1$, which is the case here, the orthodox model predicts $U_a = e/C_2$ which is equal to $U_p = 0.15$ V. From the experiment, however, $U_a = 0.28$ V exceeds $U_p$ by a factor of...
2. Second, the first step on either side of $U = 0$ is suppressed by approximately a factor 3. The rest of the $I-U$ characteristic, i.e. the staircase at larger voltages, is in good agreement with the theory.

Fig. 2 shows another example. Here, the step pattern decays rapidly with increasing voltage due to the small resistance ratio $R_2/R_1 \approx 1.5$. At first sight, one notices that the region around $U = 0$ appears as a large gap, larger by a factor of two than $U_p = e/C_2 \approx 0.18$ V. This time, however, the magnified view of the central part of the curve (inset of fig. 2) reveals that the current is not fully suppressed to zero in this region, but that the $I-U$ characteristic consists of two linear segments with a slope of $\approx 8 \times 10^{-3}$ (in units of $R^{-1}$). To fit this curve by the orthodox model requires $q_0 = e/2$ (thin solid curve in Fig. 2). The reproduced segment slope is then given by $C_2/C_1 \approx C_2/C_2$. The fit suggests $C_2$ to be larger than $C_1$ by approximately a factor of 100. This is inconsistent with other measured $I-U$ characteristic that yield $C_2/C_1 \approx 5-20$, as expected from the dielectric constant $\varepsilon_2 \approx 10$ of the ZrO$_2$.

In summary, we have demonstrated the following two anomalies in our experimental $I-U$ characteristic that are not accounted for by the orthodox model: First, the current around $U = 0$ is suppressed by a factor of 3-10, and second, the offset $U_a$ between the two asymptotes at large voltages is about twice the terrace width $U_p$. The described deviations were found in many of our measurements, but not in all. Occasionally, we have obtained $I-U$ characteristics that conform more closely to the orthodox model (see for example fig. 3 in ref. [6]).

The prediction of the orthodox model that both $U_a$ and $U_p$ equal $e/C_2$ (for $R_2 \gg R_1$ and $C_2 \gg C_1$), might be reconciled with the experimental finding $U_a > U_p$ if somehow two different $C_2$'s would govern $U_a$ and $U_p$. We asked ourselves the question whether such a situation could originate from the strong frequency dependence of the dielectric constant $\varepsilon_2$ of the dielectric (oxide tunnel barrier). While for ZrO$_2$ $\varepsilon_2 \approx 10$ for low frequencies $\omega$, it is reduced to $\approx 4$ for $\omega > \omega_0$, where $\omega_0$ is a typical frequency of the optical phonon modes. The high frequency reduction in $\varepsilon_2$ occurs because the ionic part of the dielectric polarizability does not contribute to $\varepsilon_2$ for $\omega > \omega_0$. To test these qualitative ideas we studied theoretically how a finite relaxation time $\tau_r$ of the ionic contribution to $\varepsilon_2$ changes the Coulomb staircase.

3. Effect of relaxation of polarization charge on the Coulomb staircase

We have investigated whether the deviation of the measured Coulomb staircase from the prediction of the orthodox model can be explained by the finite relaxation time of polarization charges in the oxide layer. In the orthodox
model these charges are assumed to be induced instantaneously upon a tunnel event. This model applies if the relaxation time \( \tau_r \) of the induced charges is small compared to the time scale \( \tau_c = \frac{\hbar}{e^2/C} \) on which an energy loss of the order of the charging energy \( e^2/C \) can be tolerated quantum-mechanically. In this case energy differences before and after the tunnel event have to be calculated according to the global rule, i.e. after all charges have relaxed to their asymptotic value [7]. In the opposite regime, if \( \tau_r \gg \tau_c \), the local rule applies: Energy differences are calculated with the induced charges held fixed at their value before the tunnel event [7]. A comparison of global versus local rule for a double-junction geometry has been carried out by Odintsov et al. [9], and by Ingold et al. [10]. They concentrated on the case of a symmetric junction, which does not show a Coulomb staircase. There is a third time scale in the problem, which is the mean time \( \tau_t = eI/\hbar \) between tunnel events. It is usually assumed that \( \tau_t \gg \tau_r \), so that the charges relax completely between successive tunnel events. Tsukada et al. [11] have investigated the case that \( \tau_t \) is not much bigger than \( \tau_r \), so that the induced charge lags behind its asymptotic value. They did not make contact with the notion of global versus local rule, and it is not clear to us how they calculated the energy differences. Note that if \( \tau_t \gg \tau_r \), we are automatically in the regime \( \tau_t \gg \tau_c \) of the local rule \( (R_1, R_2 \gg \hbar/e^2) \).

The equivalent circuit which we have studied is shown in fig. 3. The two metal-particle junctions are modeled by capacitors \( C_1 \) and \( C_2^\infty \), which are assumed to charge and discharge instantaneously. The induced charges in the oxide layer contribute a capacitance \( C_0 \) to junction 2. \( C_2^\infty \) corresponds to the high frequency part of the capacitor \( C_2 \), i.e. \( C_2^\infty = C_2(\omega \to \infty) \), while the static capacitance \( C_2(\omega = 0) \) is given by \( C_0 + C_2^\infty \). \( C_0 \) is charged via a resistance \( R_0 \), and therefore has a finite relaxation time, given by \( R_0/\tau_r = 1/C_0 + 1/(C_1 + C_2^\infty) \). Between tunnel events, the charge \( Q_0 \) on capacitor 0 relaxes exponentially, with time constant \( \tau_r \), to its asymptotic value

\[
Q_0^\infty(Q) = \frac{(C_0/C)(C_1U + Q)}{C + C_1 + C_2^\infty} \tag{1}
\]
determined by the total charge \( Q = Q_1 + Q_2 + Q_0 \) on the particle and by the bias voltage \( U \). We define \( C = C_0 + C_1 + C_2^\infty \). The voltage drops over junctions 1 and 2 are given by

\[
\Delta U_1(Q, Q_0) = \frac{C_2^\infty U + Q - Q_0}{C_1 + C_2^\infty}, \tag{2}
\]

\[
\Delta U_2(Q, Q_0) = \frac{C_2^\infty U + Q - Q_0}{C_1 + C_2^\infty}. \tag{3}
\]

A tunnel event is assumed to change \( Q \) instantaneously by \( \pm e \). At zero temperature, the tunnel rate through junction 1 is given by

\[
r_1^\pm(Q) = \begin{cases} (e^2 R_1)^{-1} \Delta E_1^\pm & \text{if } \Delta E_1^\pm > 0, \\ 0 & \text{if } \Delta E_1^\pm < 0, \end{cases} \tag{4}
\]

for the process \( Q \to Q \pm e \), where \( R_1 \) is the tunnel resistance. Similarly, for junction 2 the process \( Q \to Q \pm e \) has rate \( r_2^\pm(Q) \) given by eq. (4) with the subscript 1 replaced by 2. The energy differences \( \Delta E_1^\pm \) and \( \Delta E_2^\pm \) are determined by the voltages \( \Delta U_1 \) and \( \Delta U_2 \) over the two junctions, by expressions of the form

\[
\Delta E_1^\pm = \int_{Q_0}^{Q + e} \Delta U_1 \, dQ', \tag{5}
\]

\[
\Delta E_2^\pm = - \int_{Q_0}^{Q - e} \Delta U_2 \, dQ'. \tag{6}
\]

The integrals can be evaluated according to the two different rules mentioned above. The global rule prescribes

\[
\Delta E_1^\pm = \int_{Q_0}^{Q + e} \Delta U_1(Q', Q_0(Q')) \, dQ', \tag{7}
\]

i.e. \( Q_0 \) is treated as an integration variable. The local rule, in contrast, holds \( Q_0 \) fixed at its value \( Q_0(t_0) \) just before the tunnel event (occurring at time \( t_0 \)),

\[
\Delta E_1^\pm = \int_{Q_0}^{Q + e} \Delta U_1(Q', Q_0(t_0)) \, dQ'. \tag{8}
\]
The expressions for the energy differences can be written in a unified way by defining the capacitance $C_\beta = C_1 + C_2 + \beta C_0$, where $\beta = 1$ for the global rule and $\beta = 0$ for the local rule. The result is

$$\Delta E_1 = \pm e \Delta U_1 - \frac{e^2}{2\epsilon_0},$$

$$\Delta E_2 = \pm e \Delta U_2 - \frac{e^2}{2\epsilon_0}.$$  

Here the voltage differences $\Delta U_1$ and $\Delta U_2$ are to be evaluated just before the tunnel event.

An analytical treatment is possible in the limit that the relaxation time $\tau_r$ of the induced charge is either much longer or much shorter than the time $\tau_t$ between tunnel events. The crossover regime $\tau_r = \tau_t$ can not be treated analytically, and would require a numerical simulation of the rate equations (as in ref. [11]). For $\tau_r < \tau_t$, we can assume that the value of $Q_0$ just before the tunnel event $Q \to Q \pm e$ has the asymptotic value $Q_0^\omega(Q)$ given by eq. (1). The probability $P(Q)$ to find a charge $Q$ on the particle can then be obtained straightforwardly from the detailed balance equation [7]

$$[r_1^-(Q + e) + r_2^-(Q + e)]P(Q + e) = [r_1^+(Q) + r_2^+(Q)]P(Q).$$  

From the probability distribution of the charge one obtains the average current $I$ as a function of $U$. A calculated $I-U$ characteristic for a strongly asymmetric junction ($R_2 = 24R_1, C_2 = C_0 = 9C_1$) is shown in fig. 4. The solid curve is the result from the orthodox model, using the local rule. The dotted curve follows from the local rule. The difference appears in the first step of the staircase, which is reduced in magnitude and shifted to higher voltages. As a result the Coulomb blockade interval of zero current is broadened, in agreement with the results of refs. [9, 10] for a symmetric junction. The broadening is by a factor $(C_0 + C_1 + C_2)/(C_0 + C_1)$.

For $\tau_r \gg \tau_t$ we should apply the local rule, but in addition would have to take into account that the induced charge $Q_0(t)$ just before a tunnel event might lag behind its asymptotic value $Q_0^\omega(Q)$. We then have to consider the point probability distribution $P(Q, Q_0) = P(Q|Q_0)P(Q)$, where $P(Q|Q_0)$ is the conditional probability for charge $Q$ on the particle for a given induced charge $Q_0$, and $P(Q_0)$ is the a priori probability of the induced charge. Since $P(Q|Q_0)$ relaxes on the time scale $\tau_t$, which is assumed to be much less than the time scale $\tau_r$, over which $Q_0$ changes, we may determine the conditional probability by the detailed balance eq. (11). We can then calculate for each value of $Q_0$ the average $\bar{Q}_0^\omega(Q_0) = \Sigma Q_0 P(Q_0)P(Q|Q_0)$. The stationary distribution of induced charge is $P(Q_0) = \delta(Q_0 - Q_0^\omega)$, where $Q_0^\omega$ is the solution of the equation $Q_0^\omega - Q_0^\omega = 0$. To make this self-consisted calculation tractable, we need a simple analytical solution to the detailed balance eq. (11). We used, the two-state approximation of Wan et al. [12], valid for a strongly asymmetry junction ($R_2 > R_1$). This solution is based on the fact that the conditional distribution of charge $P(Q|Q_0)$ is strongly peaked at one particular value $Q_0^\omega$ of $Q$. The transition from $Q_{\max}$ to $Q_{\max} + 1$ occurs in very small voltage intervals around the steps of the Coulomb staircase. Outside of these voltage intervals we have $\bar{Q}_0^\omega = Q_0^\omega = Q_0^\omega$, i.e. the stationary value of the induced charge equals the asymptotic value associated with a charge $Q_{\max}$.
on the particle. We conclude that in the case of two very asymmetric junctions (the relevant case for the Coulomb staircase), the induced charge does not lag behind its asymptotic value, the reason being basically that the distribution of charge on the particle is sharply peaked at one particular value, just as in equilibrium.

In summary, relaxation of induced charge modifies the predictions of the orthodox model for the Coulomb staircase, if $\tau > \tau_C$. The relative magnitude of $\tau$ and $\tau_C$ is not essential. The height of the first step is reduced, while its width increases. Subsequent steps are modified only by an offset, their height and width remain the same.

4. Conclusion

To compare the theory with the experiments, we first estimate the model parameters that apply to the measurements. Since $\varepsilon_z$ is reduced by approximately a factor of 2 between low and high frequencies, we assume $C_2^\infty \approx C_0 \gg C_1$. We have $\tau_C = \hbar C/e^2 \approx 7 \times 10^{-15}$ s. In the experiment, the relaxation time $\tau_r$ is due to the finite response time of the ionic contribution to the dielectric constant of the oxidic layer. The response time is given by $C_0 R_o$ and approximated by $\omega_0^{-1}$. For a typical optical phonon energy of $\hbar \omega_0 \approx 10$ meV, we estimate $\tau_r = \omega_0^{-1} (C_2(\omega = \infty)/C_2(\omega = 0)) \approx 1/2 \omega_0 \approx 30 \times 10^{-15}$ s. We are therefore in the regime $\tau > \tau_C$ for which the local rule applies.

The theory predicts that the gap $U_a$ derived from the high-voltage asymptotes is broadened by a factor $(C_0 + C_1 + C_2^\infty)/(C_1 + C_2^\infty) = C_2(\omega = 0)/C_2(\omega = \infty) \approx 2$ in good agreement with the experiment. In addition, the reduction in magnitude of the first step and the fact that subsequent steps are not changed in height and width agrees with the experiment. However, what is not correctly reproduced is the broadening of the Coulomb blockade of zero current and the associated shift in the position of the steps. Fig. 1 shows a Coulomb gap of size equal to the width of the subsequent steps. We are at present not certain if this difference can be resolved within the present theoretical model. However, we would like to point out that the experiment is close to the intermediate regime $\tau \approx \tau_C$, while the model calculation was for $\tau \gg \tau_C$. It would be desirable to calculate Coulomb staircases in this intermediate regime using the general formulas of refs. [9,10]. From the experimental side, support of the proposed interpretation may be found in measurements of $I-U$ characteristics for dielectrics that differ in the ratio of the high versus low frequency dielectric constant. We are planning to perform such an investigation.

Acknowledgements

We are grateful to H.C. Donkersloot and J.M. Kerkhof for growing the samples. Research at Leiden University is supported by the Dutch Science Foundation NWO/FOM.

References


