Stochastic current switching in bistable resonant tunneling systems

O. A. Tretiakov[1] and K. A. Matveev[2,1]

[1]*Department of Physics, Duke University, Durham, North Carolina 27708, USA*
[2]*Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
(Received 2 November 2004; published 26 April 2005)

Current-voltage characteristics of resonant-tunneling structures often exhibit intrinsic bistabilities. In the bistable region of the *I-V* curve one of the two current states is metastable. The system switches from the metastable state to the stable one at a random moment in time. The mean switching time τ depends exponentially on the bias measured from the boundary of the bistable region V_{th}. We find full expressions for τ (including prefactors) as functions of bias, sample geometry, and in-plane conductivity. Our results take universal form upon appropriate renormalization of the threshold voltage V_{th}. We also show that in large samples the switching initiates inside, at the edge, or at a corner of the sample depending on the parameters of the system.

I. INTRODUCTION

Recent advances in experimental techniques have made possible the study of fast stochastic processes such as dynamic current switching in resonant tunneling structures. The electron transport in these devices has attracted a lot of attention since the pioneering work of Tsu and Esaki.[1] The interest was further stimulated by the discovery of the phenomenon of intrinsic bistability[2–6] in double-barrier resonant tunneling structures (DBRTS). Other resonant tunneling structures, such as superlattices, are also known to show bistable behavior.[7–10] Recent experiments[8–10] established that in the bistable region one of the current states is metastable, and the switching to the stable state was studied. Both the mean switching time and its distribution function were measured.[10]

The existence of intrinsic bistability is well understood theoretically.[11–14] It was shown[14] that in a certain range of bias, $\widetilde{V}_{th} < V < V_{th}$, for every value of V the current can take two different values, see Fig. 1. If one increases the bias starting from any value below $\widetilde{V}_{th}$, the current follows the upper branch of the *I-V* curve shown in Fig. 1 until V reaches V_{th}, where the current switches to the lower branch. On the other hand, if one decreases the bias from the values greater than V_{th}, the current follows the lower branch and then switches to the upper branch at $\widetilde{V}_{th}$.

The bistability can be understood by considering the potential profile of the DBRTS schematically shown in Fig. 2. If the level E_0 in the quantum well is below the bottom of the conduction band of the left lead, tunneling into the well is not possible, and the current through the heterostructure is zero. In this case the charge in the well $Q=0$. However, if a nonzero charge Q is added to the well, the level E_0 rises due to the charging effects and may become higher than the bottom of the conduction band of the left lead. Then, another steady state of current is possible. In this state the current into the well from the left lead is compensated by the current out of the well through the right barrier. Thus, it is possible to have two different current states at the same bias. (See, e.g., points **M** and **S** on the *I-V* curve, Fig. 1.)

The electric current in the device fluctuates, because the electrons tunnel in and out of the well at random moments in time. The resulting shot noise of current through the heterostructure gives rise to the metastability of some current states. The two solid lines in Fig. 1 correspond to the most probable values of current at a given bias. These two branches are stable, i.e., any *small* fluctuation of current near a solid line will decay with time, and the current will return to its value at the solid line. The dashed line between $\widetilde{V}_{th}$ and V_{th} corresponds to the unstable state. Here any deviation from the dashed line which raises or lowers the current will switch the system to the upper or lower stable current state, respectively.

Qualitative understanding of the metastability can be achieved by considering the system at a bias near V_{th}, e.g., point **M** on the upper branch of the *I-V* curve, Fig. 1. Then, as one can see from Fig. 1, a relatively small fluctuation can shift the current below the dashed line corresponding to the unstable state. If that happens, the system switches to the lower branch. The opposite process is much less probable, since the distance from the lower branch to the dashed line is much larger than that from the upper branch. Therefore, the lower branch is stable, and the system remains in that state.

The dependence of the mean switching time τ on the bias was addressed theoretically in Ref. 15. It was shown that near the threshold voltage V_{th} the logarithm of τ behaves as
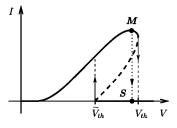

FIG. 1. The *I-V* curve of the DBRTS. The bistable region is present in the range of bias between $\widetilde{V}_{th}$ and V_{th}. The bold dashed line corresponds to the unstable current state.
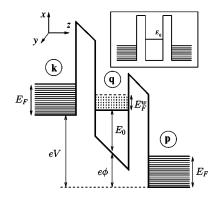
FIG. 2. Schematic potential profile of the double-barrier resonant tunneling structure. The structure consists of a quantum well separated from two leads by tunneling barriers. The electrons with three-dimensional wave vectors **k** and **p** fill all the states up to the Fermi energies E_F in the left and right leads, respectively. In the quantum well the motion of electrons in the z direction is quantized, and the electrons with two-dimensional transverse wave vectors **q** occupy all the states up to the Fermi energy E_F^w. The inset shows the potential profile at zero bias.

$$\ln \tau \propto \begin{cases} (V_{th} - V)^{3/2}, & L \ll r_0, \\ V_{th} - V, & L \gg r_0. \end{cases} \qquad (1)$$

Here L is the size of the sample, $r_0 \propto \sigma^{1/2}(V_{th}-V)^{-1/4}$ is a characteristic length scale, and σ is the in-plane conductivity. In small samples, $L \ll r_0$, the current switches from the metastable state to the stable one simultaneously over the entire area of the device. On the other hand, in large samples, $L \gg r_0$, the switching is initiated in a small critical region of radius r_0. After the switching has occurred in that region, it extends rapidly to the rest of the sample.

In this paper we show that if the sample is large, the switching can initiate not only inside, but also at the edge of the device. The latter process tends to be more efficient, since the exponential in the respective expression for τ is smaller than in the case of switching far from the edges of the sample (Sec. V). On the other hand, the switching at the edge can be initiated anywhere along the boundary of the device, and thus the prefactor of the switching rate $1/\tau$ due to these processes is proportional to the perimeter $\sim L$. Similarly, the prefactor of the rate of switching inside the device is proportional to the area $\sim L^2$, which makes these processes more efficient in larger samples.

We obtain analytically the full expressions for τ, including the pre-exponential factors. Apart from the dependence on sample dimensions, the calculation of the prefactors reveals the nontrivial dependence of the threshold voltage V_{th} on the degree of disorder of the sample. Formal evaluation of the prefactors in the case of nonuniform electron density in the well results in ultraviolet divergences. Similar divergences appear in quantum field theory, where they are eliminated with the use of a renormalization procedure.[16,17] The application of a similar technique to our problem leads to the renormalization of the threshold voltage which depends strongly on the conductivity of the quantum well (Secs. IV A 2 and V). Upon this renormalization $\ln \tau$ in large samples acquires logarithmic corrections to its linear voltage dependence.

The paper is organized as follows. In Sec. II we obtain the Fokker-Planck equation for tunneling in DBRTS which completely describes the electron transport in small samples. This equation enables us to find a simple result for the mean switching time in these samples. In Sec. III we derive the Fokker-Planck equation for the case of large samples which describes the dynamics of electron density in the well due to both the diffusion in the plane of the well and tunneling between the well and the leads. We use it to investigate the effect of weak density fluctuations on the decay of metastable current state in small samples (Sec. IV) and to study the switching in large samples (Sec. V). The application of our theory to the existing and future experiments is discussed in Sec. VI.

II. FOKKER-PLANCK EQUATION FOR TUNNELING IN DBRTS

The bistable current-voltage characteristic of DBRTS was studied theoretically in Refs. 11–14. The I-V curve shows the dependence of the *average* current on voltage applied to the device. In addition, shot noise was studied in the regime of small fluctuations.[12–14] On the other hand, the switching between the branches of the I-V curve is caused by large fluctuations of current. In this section we use the model of Ref. 14 to derive the Fokker-Planck equation for tunneling in DBRTS, which accounts for these large fluctuations, and thus describes the switching.

The model is illustrated in Fig. 2. The well is extended in the x-y plane. The motion in the z direction in the well is quantized, and the well is assumed to have only one resonant level of energy E_0. The two-dimensional wave vectors in the well are denoted by **q**. The left and right leads are three-dimensional; the wave vectors of electrons are denoted by **k** and **p**, respectively. The conduction bands in the leads are occupied up to the Fermi energy E_F. In typical devices E_0 is of the order of E_F; for definiteness we assume $E_0 > E_F$. The temperature T is assumed to be small compared to E_F and eV. The well is separated from the leads by two tunneling barriers with the transmission coefficients much smaller than unity.

In Ref. 14 the tunneling through the double barrier was described quantum mechanically using the Breit-Wigner formula. The level widths with respect to the decay to the right and left leads Γ_L, Γ_R were eventually taken to be much smaller than all other relevant energy scales. We make this assumption from the beginning, and describe the electron transport through the barriers using the sequential tunneling approach. This method is an alternative to the use of the Breit-Wigner formula. Unlike the latter, it cannot account for the coherent tunneling through the well at large $\Gamma_{L,R}$. On the other hand, at small $\Gamma_{L,R}$ it enables us to discuss both the I-V characteristic and the large fluctuations of current.[18]

In order to have a steady state of nonzero current in the device, the electrochemical potential in the well should lie between those in the left and right leads, i.e.,

$$eV + E_F > E_0 + e\phi + E_F^w > E_F. \qquad (2)$$

Here $E_F^w = \hbar^2 q_F^2/2m$ is the Fermi energy in the well with m being the effective mass. Then, in the limit of low temperature the inequalities (2) dictate that the tunneling is possible only in one direction, namely, from left to right, Fig. 2. The probability to tunnel through a barrier is given by the Fermi golden rule. The rates of electron tunneling into the well J_L and out of the well J_R take the form

$$J_L = \frac{4\pi}{\hbar} \sum_{\mathbf{qk}} |t_{k_z}|^2 \delta_{\mathbf{qk_\parallel}} f_{\mathbf{k}}(1-f_{\mathbf{q}}) \delta(eV+E(k_z)-E_0-e\phi),$$

$$\qquad (3)$$

$$J_R = \frac{4\pi}{\hbar} \sum_{\mathbf{qp}} |t_{p_z}|^2 \delta_{\mathbf{qp_\parallel}} f_{\mathbf{q}} \delta(E_0+e\phi-E(p_z)). \qquad (4)$$

Here $E(k) = \hbar^2 k^2/2m$; $f_{\mathbf{k}}$ and $f_{\mathbf{q}}$ are the Fermi occupation numbers in the left lead and the quantum well, respectively. In Eq. (4) we used the fact that the Fermi occupation numbers in the right lead $f_{\mathbf{p}} = 0$ at energies above $E_0 + e\phi$, because $E_0 > E_F$. Expressions (3) and (4) include an additional factor of 2, which accounts for electron spins. The matrix elements t_{p_z} (t_{k_z}) describe the transitions between the resonant level in the well and the state with z component of the wave vector p_z (k_z) in the right (left) lead. The conservation of the transverse momentum is taken into account by Kronecker deltas.

To simplify the expression for the tunneling rate (4) we use $\delta_{\mathbf{qp_\parallel}}$ to remove the sum over $\mathbf{p_\parallel}$. The remaining sum over $\mathbf{q}$ of Fermi function $f_{\mathbf{q}}$ gives exactly the number of electrons in the well with a given spin $N/2$. Then Eq. (4) reduces to

$$J_R = \frac{\Gamma_R}{\hbar} N. \qquad (5)$$

Here Γ_R is the level width with respect to tunneling into the right lead. We define the level widths for the two possible tunneling processes as

$$\Gamma_L = 2\pi \sum_{k_z} |t_{k_z}|^2 \delta(eV+E(k_z)-E_0-e\phi), \qquad (6a)$$

$$\Gamma_R = 2\pi \sum_{p_z} |t_{p_z}|^2 \delta(E_0+e\phi-E(p_z)). \qquad (6b)$$

To find J_L we use the Kronecker delta to remove the sum over $\mathbf{k_\parallel}$ in Eq. (3), while the value of $k_z^2 = (2m/\hbar^2)(E_0+e\phi-eV)$ is fixed by the delta function. At $\mathbf{k}=(\mathbf{q},k_z)$ and $T \to 0$ the sum over $\mathbf{q}$ of $f_{\mathbf{k}}(1-f_{\mathbf{q}})$ can be easily evaluated, and gives $(S/4\pi)(k_F^2-k_z^2-q_F^2)$ under the condition (2), where S is the area of the sample. Then, the expression (3) can be simplified as follows:

$$J_L = \frac{\Gamma_L}{\hbar} \left(\frac{Sm}{\pi\hbar^2}(E_F+eV-E_0-e\phi) - N \right). \qquad (7)$$

Here we used the expression $N = Sq_F^2/2\pi$ for the total number of electrons in the well. Note that at $eV > e\phi + E_0$ the level width (6a) vanishes, and thus $J_L = 0$.

In the sequential tunneling approximation the average number of electrons in the well can be determined from the condition $J_L = J_R$,

$$N = \frac{Sm}{\pi\hbar^2} \frac{\Gamma_L}{\Gamma_L+\Gamma_R}(E_F+eV-E_0-e\phi). \qquad (8)$$

One cannot directly obtain N from Eq. (8), since the potential ϕ depends on the number of electrons in the well. Considering the barriers as two capacitors, one finds from electrostatics the following expression for the electric potential of the well (Fig. 2):

$$\phi = \frac{V}{2} + \frac{eN}{2C}. \qquad (9)$$

Here we assumed for simplicity that the capacitances of the left and right barriers are equal to each other, and denoted the capacitance of each barrier as C.

One can obtain the current-voltage characteristic of the DBRTS by repeating the following steps of Ref. 14. First, one notices that the level widths are energy dependent,

$$\Gamma_L = g_L \sqrt{E_0(E_0-eV+e\phi)}, \qquad (10a)$$

$$\Gamma_R = g_R \sqrt{E_0(E_0+e\phi)}, \qquad (10b)$$

where $g_{L,R}$ are dimensionless constants. Since Γ_L and Γ_R depend on ϕ, they are also functions of N. Therefore to find N one must solve the pair of Eqs. (8) and (9). The latter leads to an equation on N, which has three solutions in the bistable region. One of the solutions corresponds to the average number of electrons on the unstable branch, while the other two correspond to N on the lower ($N=0$) and upper stable branches. Upon substitution of N into Eq. (5) one finds the dependence of the average current on bias, i.e., the bistable I-V curve,[14] which is schematically shown in Fig. 1.

To account for the noise, we go one step further and write the master equation for the time evolution of the distribution function $P(N,t)$ of the number of electrons in the well N. In terms of the tunneling rates (3) and (4), the master equation for $P(N,t)$ takes the form

$$\frac{\partial}{\partial t}P(N,t) = P(N-1,t)J_L(N-1) + P(N+1,t)J_R(N+1)$$

$$- P(N,t)[J_L(N)+J_R(N)]. \qquad (11)$$

The first two terms on the right-hand side of Eq. (11) account for the processes which increase the probability to have N electrons in the well, while the last term corresponds to the opposite processes.

In this section we consider the samples of large in-plane conductivity where the density in the well is uniform. Therefore, in the steady state of nonzero current the total number of particles in the well is proportional to the area of the sample. The linear dimensions of the sample are assumed to be large compared to the Bohr radius in the semiconductor. Thus the total number of electrons in the well is large, $N \gg 1$, and one can expand Eq. (11) in $1/N$. Keeping the terms up to the second order, the master equation reduces[19] to

$$\frac{\partial}{\partial t}P(N,t) = -\frac{\partial}{\partial N}[A(N)P(N,t)] + \frac{1}{2}\frac{\partial^2}{\partial N^2}[B(N)P(N,t)].$$

$$(12)$$

Here $A(N) = J_L(N) - J_R(N)$ and $B(N) = J_L(N) + J_R(N)$. Equation (12) is known as the Fokker-Planck equation, and is widely used for the description of various stochastic processes, see, e.g., Refs. 20,21.

The stationary solution of Eq. (12) can be easily obtained,

$$P_0(N) = \frac{\text{const}}{B(N)}e^{-U(N)}, \quad U(N) = -2\int_0^N \frac{A(N')}{B(N')}dN', \quad (13)$$

Ref. 20. The extrema of function $U(N)$ are determined by the condition $A \equiv J_L - J_R = 0$, which we used above to find the average current through the device. Therefore each extremum of $U(N)$ corresponds to one of the branches of the I-V curve. Outside the bistable region the current $I(V)$ is uniquely defined, and $U(N)$ has a single minimum. In the bistable region $U(N)$ has two minima and a maximum, which correspond to the locally stable current branches and the unstable branch, respectively (Fig. 1).

From the definitions of A and B, it is clear that their ratio is independent of the area of the sample S. Using the expression (13) and the fact that $N \propto S$, one can see that U is linearly proportional to S. Thus $U(N)$ is an extensive quantity, and its dependence on N and S has the general form $U(N,S) = Su(n)$, where $n = N/S$ is the electron density. Since the area of the sample is large, we have $U \gg 1$. Therefore, the distribution function P_0 is peaked sharply near the global minimum of $U(N)$.

The experiments[8–10] studying the switching between the branches of the I-V curve are set up as follows. One starts at $V < \tilde{V}_{th}$, Fig. 1, where only one value of current is possible. In this case $U(N)$ has only one minimum, as shown schematically by the dashed-dotted line in Fig. 3(a). If we increase the bias up to some value slightly above $\tilde{V}_{th}$, the function $U(N)$ will acquire a new minimum to the left of the old one, see the dashed line in Fig. 3(a). This corresponds to the appearance of the lower current branch of the I-V curve. The new minimum is a local one, and the main peak of the distribution function is still centered at the old minimum. Thus, the system remains on the upper branch of the I-V curve. Further increasing V, we transform $U(N)$ to the shape shown schematically by the solid line in Fig. 3(a). Here the right minimum of $U(N)$ is a local one, and if we leave the system in this state for a sufficiently long time, it will eventually switch to the left minimum. To switch from the local minimum to the global one, the system must overcome the barrier of height U_b, Fig. 3(a). From the form of the distribution function (13) it is clear that this process takes a long time $\tau \propto \exp(U_b)$. To perform the measurement of the switching time from the upper to the lower current branch, one increases the bias to the chosen value over an interval of time short compared to τ, and then waits until the system switches to the lower branch.
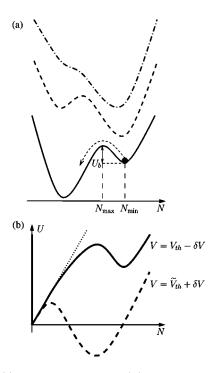

FIG. 3. (a) Generic behavior of $U(N)$ at different values of bias. Outside the bistable region $U(N)$ has one minimum (top curve). Inside the bistable region the function $U(N)$ has two minima and a maximum, which correspond to the locally stable current branches and the unstable branch, respectively (middle and bottom curves). (b) The sketch of $U(N)$ for the model of Fig. 1. Solid line corresponds to a bias slightly below V_{th}, whereas dashed line depicts $U(N)$ for the bias slightly above $\tilde{V}_{th}$.

As the voltage approaches its threshold value, the maximum at N_{max} and the local minimum of $U(N)$ at N_{min} (Fig. 3) move closer to each other, and at the threshold they coincide. At this point one can define a threshold electron density $n_{th} \equiv N_{max}/S = N_{min}/S$. In the vicinity of $n = n_{th}$ and $V = V_{th}$ the function $u(n)$ can be approximated by a cubic polynomial,

$$u(n) \approx -\alpha(n - n_{th}) + \frac{\gamma}{3}(n - n_{th})^3 + u_{th}, \quad \alpha = a(V_{th} - V).$$

$$(14)$$

Here the constant u_{th} is the value of u at $n = n_{th}$ and $V = V_{th}$.

To derive Eq. (14) microscopically, one has to consider $A(N)$ and $B(N)$ on the upper branch of the I-V curve in the vicinity of the threshold V_{th}. An analytical calculation of $I(V)$ is possible[14] if the dimensionless parameter

$$\lambda = \frac{me^2}{2\pi\hbar^2 c} \qquad (15)$$

is small, $\lambda \ll 1$. Here $c = C/S$ is the capacitance per unit area. In Appendix A we extend this approach to find $A(N), B(N)$, and the coefficients of expansion (14) at $\lambda \ll 1$.

The expansion (14) can be justified for any λ in the spirit of the Landau theory of second-order phase transitions.[22] The potential u is expected to be an analytic function of n and V. Thus, u can be expanded in Taylor series near the threshold, with $n - n_{th}$ playing the role of the order parameter. Since the

local minimum and the maximum of u coincide at the threshold, both the first and second derivatives of $u(n)$ vanish at $V=V_{th}$ and $n=n_{th}$. Therefore, the expansion starts with the third-order term. The sign of γ is not important; we choose $\gamma > 0$, which corresponds to the behavior of U near the right minimum as shown in Fig. 3(a). At $V \neq V_{th}$, the linear $\alpha(n-n_{th})$ and quadratic $\frac{1}{2}\beta(n-n_{th})^2$ terms are also present. Since $\alpha=\beta=0$ at $V=V_{th}$, we expect $\alpha \propto (V_{th}-V)$, $\beta \propto (V_{th}-V)$. We keep only the linear term in the expansion, because the second-order term is quadratic in small parameter $n-n_{th}$, and therefore is small compared to the linear one. In order for u to have a local minimum at $V < V_{th}$, the coefficient a should be positive.

Near the threshold the function $B(N)$ can be approximated by a constant,

$$B(N_{th}) = 2J_R = 2\Gamma_R N_{th}/\hbar. \tag{16}$$

In the case of constant B the Fokker-Planck equation (12) has been studied in detail. In particular, the exact expression for the mean switching time can be obtained including the prefactor (Ref. 20, Sec. XIII.2). In our notations it reads

$$\tau = \frac{4\pi}{B\sqrt{U''(N_{\min})|U''(N_{\max})|}} \exp(U_b). \tag{17}$$

For the potential (14) one can easily find the barrier height, $U_b = 4S\alpha^{3/2}/3\gamma^{1/2}$. The prefactor of Eq. (17) can be also straightforwardly evaluated, and one obtains the following expression for the mean switching time:

$$\tau = \frac{2\pi}{b\sqrt{\alpha\gamma}} \exp\left(\frac{4}{3}\frac{S[a(V_{th}-V)]^{3/2}}{\gamma^{1/2}}\right), \tag{18}$$

where $b=B/S$ is independent of the area of the sample. This result obviously agrees with Eq. (1) for small samples ($L \ll r_0$).

Expansion (14) is quite generic, and similar theoretical results were found in many different areas of physics.[23–29] In particular, Eq. (12) is also used to describe the motion of a Brownian particle in external potential, where N plays the role of the coordinate of the particle. Therefore, the logarithm of the mean escape time of the Brownian particle from a local minimum of potential is also expected to obey the 3/2-power law. Recently this behavior of the escape time was confirmed experimentally for the optically trapped Brownian particle.[27]

The lower branch of the I-V curve corresponds to the situation where the level in the well $E_0+e\phi$ is below the bottom eV of the conduction band in the left lead. In this case $J_L \equiv 0$ and $B=-A=J_R$. Consequently, as $N \to 0$ we have $U(N)=2N$, see Eq. (13). Since N cannot be negative, U reaches its minimum at the boundary $N=0$ of the range of allowed values of N, where the derivative $U'(N) \neq 0$, Fig. 3(b). The nonanalyticity of $U(N)$ near the left minimum does not affect the calculation of the time of switching from upper to the lower branch. Indeed, at a bias slightly below V_{th}, the function $U(N)$ is analytic near its maximum and the local minimum [solid line in Fig. 3(b)], and the description of the switching from the upper to the lower branch in terms of

Eqs. (14) and (17) is correct. However, the situation is different for the switching from the lower to the upper branch of the I-V curve. To study this process, we decrease the bias to the value slightly above $\tilde{V}_{th}$. The function $U(N)$ for this case is depicted schematically by the dashed line in Fig. 3(b). Here $U(N)$ is nonanalytic at its local minimum, and therefore we cannot use expressions (14) and (17) for the switching time. The nonanalytic behavior of $U(N)$ is a consequence of crudeness of our model, in which the current on the lower branch is exactly zero. On the other hand, the experimentally measured I-V curves show nonzero current on the lower branch. Thus, in a more detailed model which accounts for this nonzero current, the minimum corresponding to the lower branch of the I-V curve will be reached at nonzero N. The discussion based on Eqs. (14) and (17) will then be valid.

III. FOKKER-PLANCK EQUATION FOR TRANSPORT IN DBRTS OF LARGE AREA

In Sec. II we studied the decay of a metastable state in DBRTS under the assumption that the electron density in the quantum well is uniform. Then the switching time τ given by Eq. (18) grows exponentially with the area of the sample. Since electrons can tunnel at any point of the quantum well, the tunneling process creates a nonuniform electron density. On the other hand, the diffusion of particles in the well leads to spreading of the charge across the sample. In small samples the spreading is fast, and the density becomes uniform. In samples of large area the electron density may change significantly before the charge spreads over the entire well. In this case the switching between the two branches of the I-V curve is initiated in a small part of the sample, and the switching time is not exponential in the area S.

In this section we generalize the Fokker-Planck equation (12) to the case of nonuniform density $n(\mathbf{r})$, where $\mathbf{r}=(x,y)$ is a position in the well. In subsequent sections this equation will be used to study the decay of a metastable state in DBRTS of large area.

A. Equation for distribution function of electron density in an isolated quantum well

We begin by considering the simplest case of a quantum well not coupled to the leads. At finite temperature the electron density in the well fluctuates and can be described by a distribution function $P\{n(\mathbf{r}),t\}$. Here we derive the Fokker-Planck equation for the distribution function of electron density due to the in-plane diffusion of electrons in the well. In Sec. III B we add the tunneling through the barriers and obtain the Fokker-Planck equation for DBRTS of large area.

We consider density fluctuations at length scales much greater than the inelastic mean free path. These density fluctuations are slow in comparison with the energy relaxation time in the well. Therefore the system is in a local equilibrium, and the distribution of electrons at any point in the well is given by a Fermi function. Note that the chemical potential in this Fermi function is determined by the electron density,

and therefore varies from point to point following the dependence $n(\mathbf{r})$.

Let us choose a time interval Δt much smaller than the relaxation time for $P\{n(\mathbf{r}),t\}$ and large in comparison with the collision time, so that the motion of electrons can be treated as diffusive. Then one can write the following equation for the evolution of the distribution function:

$$P\{n(\mathbf{r}),t+\Delta t\} - P\{n(\mathbf{r}),t\}$$
$$= \int \int d\mathbf{r}_1 d\mathbf{r}_2 [P\{n(\mathbf{r})+\delta n_{12}(\mathbf{r}),t\} W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n(\mathbf{r})$$
$$+ \delta n_{12}(\mathbf{r})) - P\{n(\mathbf{r}),t\} W_{\Delta t}(\mathbf{r}_2,\mathbf{r}_1;n(\mathbf{r}))]. \quad (19)$$

Here $\delta n_{12}(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}_1) - \delta(\mathbf{r}-\mathbf{r}_2)$ is the correction to the density $n(\mathbf{r})$ due to the displacement of one electron from point $\mathbf{r}_2$ to $\mathbf{r}_1$; the probability density $W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n(\mathbf{r}))$ describes diffusion of an electron from a point $\mathbf{r}_1$ in the quantum well to point $\mathbf{r}_2$ during the time interval Δt. Since the diffusion rate may depend on the electron density, $W_{\Delta t}$ is a functional of $n(\mathbf{r})$.

Expanding the first term on the right-hand side of Eq. (19) up to the second order in $\delta n_{12}(\mathbf{r})$, one obtains the following equation:

$$\Delta P\{n,t\} = \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \left[\left(\frac{\delta}{\delta n(\mathbf{r}_1)} - \frac{\delta}{\delta n(\mathbf{r}_2)} \right) \right.$$
$$\times [W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n) - W_{\Delta t}(\mathbf{r}_2,\mathbf{r}_1;n)] + \frac{1}{2} \left(\frac{\delta}{\delta n(\mathbf{r}_1)} \right.$$
$$\left. \left. - \frac{\delta}{\delta n(\mathbf{r}_2)} \right)^2 [W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n) + W_{\Delta t}(\mathbf{r}_2,\mathbf{r}_1;n)] \right] P\{n,t\}.$$
$$(20)$$

The probability densities $W_{\Delta t}$ to diffuse from $\mathbf{r}_1$ to $\mathbf{r}_2$ and back are not independent,

$$W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n)e^{-\mu_1/T} = W_{\Delta t}(\mathbf{r}_2,\mathbf{r}_1;n)e^{-\mu_2/T}. \quad (21)$$

Here μ_1 and μ_2 are the electrochemical potentials at points $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. For the case of elastic scattering by impurities considered in Ref. 15 expression (21) directly follows from Eq. (8) of Ref. 15. Generalization of Eq. (21) to arbitrary scattering mechanism is discussed in Appendix B.

In order to find μ we need to account for the interactions between electrons. We limit ourselves to the charging energy approximation; the electron exchange and correlation effects are neglected. Then at low temperatures $T \ll E_F$, the values of the electrochemical potential are found by adding the electrostatic potential $e^2 n/c$ to the Fermi energy,

$$\mu_{1,2} = \frac{e^2}{\tilde{c}} n(\mathbf{r}_{1,2}). \quad (22)$$

Here the effective capacitance per unit area $\tilde{c}$ is defined by $e^2/\tilde{c} = e^2/c + 1/\nu$, and ν is the density of states in the well.

In short time Δt an electron can only diffuse over a short distance, so that $|\mu_1-\mu_2| \ll T$. Therefore using Eq. (21), one can expand the expression in the curly brackets on the right-hand side of Eq. (20) to the leading order in $(\mu_1-\mu_2)/T$, and with the help of Eq. (22) obtain

$$\Delta P\{n,t\} = \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \left[\frac{e^2}{\tilde{c}T} \left(\frac{\delta}{\delta n(\mathbf{r}_1)} - \frac{\delta}{\delta n(\mathbf{r}_2)} \right) \right.$$
$$\times [n(\mathbf{r}_1) - n(\mathbf{r}_2)] + \left(\frac{\delta}{\delta n(\mathbf{r}_1)} - \frac{\delta}{\delta n(\mathbf{r}_2)} \right)^2 \right]$$
$$\times W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n) P\{n,t\}. \quad (23)$$

To proceed further we need an expression for the transition probability density $W_{\Delta t}$. This quantity is affected by all the relevant processes of electron scattering, such as elastic scattering of electrons by impurities, electron-phonon and electron-electron scattering. Instead of accounting for all these processes explicitly, we express $W_{\Delta t}$ in terms of in-plane conductivity σ, which can in principle be measured experimentally. Assuming that electron motion is diffusive, we conclude that the average square of the distance traveled by an electron during a short time interval is proportional to Δt, i.e.,

$$\int W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n)|\mathbf{r}_1-\mathbf{r}_2|^2 d\mathbf{r}_2 = G\Delta t. \quad (24)$$

Here the constant G is proportional to the conductivity, $G = 4T\sigma/e^2$, see Appendix C.

At small Δt the transition probability density $W_{\Delta t}$ can be expanded as

$$W_{\Delta t}(\mathbf{r}_1,\mathbf{r}_2;n) = \delta(\mathbf{r}_1-\mathbf{r}_2) + \frac{T\sigma\Delta t}{e^2} \nabla^2 \delta(\mathbf{r}_1-\mathbf{r}_2) + \cdots . \quad (25)$$

The physical meaning of the first term in this expansion is that electron remains at its initial position $\mathbf{r}_1$ at $\Delta t=0$. Thus the second term is needed to account for the electron diffusion. The coefficient in the second term is found by applying the expansion (25) to Eq. (24).

Equation (23) can be simplified significantly using expansion (25), and eventually takes the form

$$\frac{\partial}{\partial t} P\{n,t\} = -\frac{\sigma}{e^2} \int d\mathbf{r} \frac{\delta}{\delta n} \left(\frac{e^2}{\tilde{c}} \nabla^2 n + T\nabla^2 \frac{\delta}{\delta n} \right) P\{n,t\}. \quad (26)$$

This is the Fokker-Planck equation for the evolution of the distribution function of electron density. The first term in Eq. (26) describes the spreading of the charge in the well, whereas the second term accounts for the thermal noise.

It is instructive to substitute into Eq. (26) the equilibrium distribution function $P_0\{n\}$. The latter has the Gibbs form $e^{-E/T}$, namely,

$$P_0\{n\} = \exp\left(-\frac{1}{T}\int \frac{e^2 n^2(\mathbf{r})}{2\tilde{c}} d\mathbf{r}\right).$$

Here the energy per unit area $\varepsilon = e^2 n^2/2\tilde{c}$ is chosen in a way that reproduces the electrochemical potential $\mu = \partial \varepsilon / \partial n$ in the form (22). It is easy to check that $P_0\{n\}$ satisfies the Fokker-Planck equation (26).

B. Combined Fokker-Planck equation for tunneling and diffusion

In this section we obtain the combined Fokker-Planck equation which incorporates both the tunneling through the barriers and diffusion inside the well. We begin by generalizing the tunneling Fokker-Planck equation (12) to the case of nonuniform electron density. This is accomplished by dividing the plane of the well into small pieces, so that the density is uniform within each piece. In the absence of in-plane diffusion, the distribution function of electron density in the entire plane is given by the product of distribution functions of its pieces, $P = \Pi_j P_j\{N_j\}$. Applying Eq. (12) to each piece we obtain the following Fokker-Planck equation for the distribution function of the entire quantum well,

$$\frac{\partial}{\partial t} P = \sum_j \frac{\partial}{\partial N_j}\left(-A(N_j) + \frac{1}{2}\frac{\partial}{\partial N_j}B(N_j)\right)P.$$

The functions $A(N_j)$ and $B(N_j)$ are extensive quantities, and it is convenient to rewrite them as $A(N_j) = \Delta S a(n)$ and $B(N_j) = \Delta S b(n)$, where ΔS is the area of each piece. Replacing the sum with the integral over the area of the sample and $\partial/\partial N_j$ with the functional derivative $\delta/\delta n(\mathbf{r}_j)$, we find the continuous form of this equation,

$$\frac{\partial}{\partial t} P\{n,t\} = \int d\mathbf{r}\, \frac{\delta}{\delta n}\left(-a(n(\mathbf{r})) + \frac{1}{2}\frac{\delta}{\delta n}b(n(\mathbf{r}))\right)P\{n,t\}. \tag{27}$$

Let us now take into account the in-plane diffusion of electrons, which was discussed in Sec. III A. Because the tunneling and diffusion are independent processes, we can add the right-hand sides of Eqs. (26) and (27) and obtain the combined Fokker-Planck equation for DBRTS of large area,

$$\frac{\partial}{\partial t} P\{n,t\} = \int d\mathbf{r}\, \frac{\delta}{\delta n}\left(-a(n) + \frac{1}{2}\frac{\delta}{\delta n}b(n) - \frac{\sigma}{\tilde{c}}\nabla^2 n\right.$$
$$\left. - T\frac{\sigma}{e^2}\nabla^2 \frac{\delta}{\delta n}\right)P\{n,t\}. \tag{28}$$

This equation generalizes Eq. (26) to the case of a quantum well coupled to the leads.

In the vicinity of the threshold V_{th} the function $b(n)$ can be approximated by a constant $b = b(n_{th})$. In addition, one can substitute $2a/b = 2A/B = -u'(n)$, cf. Eq. (13). At bias near V_{th} the function $u(n)$ is given by the approximate expression (14), and Eq. (28) can be rewritten as

$$\frac{\partial}{\partial t} P\{n,t\} = \frac{b}{2}\int d\mathbf{r}\, \frac{\delta}{\delta n}\left(-\alpha + \gamma(n - n_{th})^2\right.$$
$$\left. - 2\eta\nabla^2 n + \frac{\delta}{\delta n}\right)P\{n,t\}, \tag{29}$$

where we defined $\eta = \sigma/\tilde{c}b$. In Eq. (29) we omitted the term proportional to the temperature, since it is negligible at low T. (The exact criterion is discussed in Appendix D.) Thus from now on we study only the effect of the shot noise due to the tunneling of electrons at high bias $eV \gg T$, whereas the thermal noise is neglected.

The stationary solution of Eq. (29) is found by setting the left-hand side to zero,

$$P_0\{n\} = e^{-F\{n\}},$$

$$F\{n\} = \int d\mathbf{r}\left(-\alpha(n - n_{th}) + \frac{\gamma}{3}(n - n_{th})^3 + \eta(\nabla n)^2\right). \tag{30}$$

The functional $F\{n\}$ has two contributions: the first two terms account for the tunneling, and the remaining term is due to the in-plane diffusion.

C. Dimensionless Fokker-Planck equation

For the following discussion it is convenient to parametrize the electron density $n(\mathbf{r})$ in terms of a dimensionless function $z(\boldsymbol{\rho})$ that vanishes at the minimum of $u(n)$,

$$n(\mathbf{r}) = n_{\min} - 2\sqrt{\frac{\alpha}{\gamma}}z(\mathbf{r}/r_0), \tag{31a}$$

$$r_0 = \frac{\sqrt{\eta}}{(\alpha\gamma)^{1/4}} = \sqrt{\frac{\sigma}{\tilde{c}b\sqrt{\alpha\gamma}}}. \tag{31b}$$

Here the density at the minimum $n_{\min} = n_{th} + \sqrt{\alpha/\gamma}$ can be easily found from Eq. (14). The Fokker-Planck equation (29) in terms of $z(\boldsymbol{\rho})$ takes the form

$$\frac{\partial P\{z,t\}}{\partial t} = b\sqrt{\gamma\alpha}\int d\boldsymbol{\rho}\, \frac{\delta}{\delta z}\left(-\nabla_\rho^2 z + z - z^2 + \frac{1}{U_0}\frac{\delta}{\delta z}\right)P\{z,t\}, \tag{32}$$

where

$$U_0 = \frac{8\eta\alpha}{\gamma}. \tag{33}$$

The stationary solution P_0 of Eq. (32) is given by

$$P_0\{z\} = e^{-F}, \quad F = U_0\int d\boldsymbol{\rho}\left(\frac{(\nabla_\rho z)^2}{2} + \frac{z^2}{2} - \frac{z^3}{3}\right). \tag{34}$$

One can see that the characteristic value of the functional F is given by U_0, whereas the characteristic size r_0 plays the role of a typical length scale of stochastic fluctuations of electron density $n(\mathbf{r})$.

IV. DECAY OF THE METASTABLE STATE IN EXTENDED SAMPLES

In Sec. II we obtained the expression for the mean switching time in DBRTS under the assumption of uniform electron density in the well. This assumption is valid only if the dimensions of the sample are small compared to the length scale r_0 of the density fluctuations, Eq. (31b). If the sample is large, the fluctuations of electron density must be taken into account.

In Sec. III we obtained the Fokker-Plank equation (29) which describes the time evolution of the distribution function of electron density. Unlike Eq. (12) for the case of uniform density, this equation has an infinite number of variables, since the density is different at every point.

The most general form of the multidimensional Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x},t)}{\partial t} = \mathcal{L} P(\mathbf{x},t),$$

$$\mathcal{L} = -\sum_i \frac{\partial}{\partial x_j} K_i(\mathbf{x}) + \sum_{i,j} \frac{\partial^2}{\partial x_i \, \partial x_j} D_{ij}(\mathbf{x}). \quad (35)$$

Assuming that the system has a metastable state, one can consider its domain of attraction Ω. The domain boundary $\partial\Omega$ is a separatrix of the drift field $\mathbf{K}$. The mean time of the first passage out of the domain Ω has been found in Refs. 30,31. For the process described by Eq. (35) the mean switching time is obtained as doubled mean first-passage time[31] and takes the form

$$\tau = -\frac{2\int_\Omega d^d x P_0(\mathbf{x})}{\sum_i \int_{\partial\Omega} dS_i \sum_j D_{ij}(\mathbf{x}) P_0(\mathbf{x}) \frac{\partial f(\mathbf{x})}{\partial x_j}}. \quad (36)$$

Here P_0 is the stationary solution of Eq. (35). The form function $f(\mathbf{x})$ is a stationary solution of the adjoint equation,

$$\mathcal{L}^\dagger f(\mathbf{x},t) = \sum_j \left(K_j(\mathbf{x}) + \sum_i D_{ij}(\mathbf{x}) \frac{\partial}{\partial x_i} \right) \frac{\partial f(\mathbf{x})}{\partial x_j} = 0. \quad (37)$$

In addition, $f(\mathbf{x})$ is defined to vanish at the boundary $\partial\Omega$ and reach $f(\mathbf{x}) \simeq 1$ well inside Ω.

In subsequent sections we use the expression (36) to find the mean time of current switching in double-barrier structures.

A. Mean switching time in small samples

In samples with linear dimensions small compared with r_0 the density fluctuations are weak. In this section we study their effect on the mean switching time. We will show that even these weak fluctuations can result in significant change of τ.

1. Evaluation of the mean switching time

In order to bring the Fokker-Planck equation (32) to the form (35) we present $z(\boldsymbol{\rho})$ as an expansion

$$z(\boldsymbol{\rho}) = \sum_{i=0}^\infty x_i \phi_i(\boldsymbol{\rho}), \quad (38)$$

where $\phi_i(\boldsymbol{\rho})$ are the normalized eigenfunctions of the Laplace operator, $-\nabla_\rho^2 \phi_i(\boldsymbol{\rho}) = \epsilon_i \phi_i(\boldsymbol{\rho})$. In particular $\phi_0(\boldsymbol{\rho}) = r_0/\sqrt{S}$ and $\epsilon_0 = 0$. Since there is no current flowing through the boundaries of the sample, the eigenfunctions must satisfy the boundary conditions $\hat{n} \cdot \nabla \phi_i(\boldsymbol{\rho})|_{\text{boundary}} = 0$, where $\hat{n}$ is a unit vector normal to the boundary. The x_0 coordinate corresponds to the average electron density in the well, whereas the other coordinates describe small fluctuations of the density. The eigenvalues ϵ_i are numbered in order of increasing magnitude, $\epsilon_1 \sim r_0^2/S \gg 1$.

To obtain the $\mathbf{x}$ representation of the Fokker-Planck equation we substitute the expansion (38) into Eq. (32) and find

$$\mathcal{L} = b\sqrt{\gamma\alpha} \left[\sum_{i=0}^\infty \frac{\partial}{\partial x_i} \left((\epsilon_i + 1 - 2\phi_0 x_0) x_i + \frac{1}{U_0} \frac{\partial}{\partial x_i} \right) \right.$$
$$\left. + \frac{\partial}{\partial x_0} \phi_0 x_0^2 + \sum_{i,j,k=1}^\infty \xi_{ijk} \frac{\partial}{\partial x_i} x_j x_k \right], \quad (39)$$

where

$$\xi_{ijk} = \int d\boldsymbol{\rho} \, \phi_i(\boldsymbol{\rho}) \phi_j(\boldsymbol{\rho}) \phi_k(\boldsymbol{\rho}).$$

The stationary solution $P_0 = e^{-F}$ in terms of x_i can be found by substituting expression (38) into Eq. (34). Then the functional F takes the form

$$F\{\mathbf{x}\} = U_0 \left(\frac{1}{2} \sum_{i=0}^\infty (\epsilon_i + 1 - 2\phi_0 x_0) x_i^2 + \frac{2\phi_0 x_0^3}{3} \right.$$
$$\left. - \frac{1}{3} \sum_{i,j,k=1}^\infty \xi_{ijk} x_i x_j x_k \right). \quad (40)$$

One can easily verify that $\exp(-F\{\mathbf{x}\})$ solves the Fokker-Planck equation with $\mathcal{L}$ given by Eq. (39).

The stationary probability density P_0 is sharply peaked at the minimum of the functional F, i.e., at $z(\boldsymbol{\rho}) = 0$ ($\mathbf{x} = 0$). Therefore, keeping terms up to the second order in x_i in expansion (40), we can evaluate the integral in the numerator of Eq. (36) in Gaussian approximation,

$$\int_{-\infty}^\infty \prod_{i=0}^\infty dx_i P\{\mathbf{x}\} = \prod_{i=0}^\infty \sqrt{\frac{2\pi}{U_0(\epsilon_i + 1)}}. \quad (41)$$

In a multidimensional case in order to switch from the metastable state the system must pass from the local minimum of F to its global minimum. The switching process is dominated by the paths which go through the vicinity of the lowest saddle point separating the domains of attraction of metastable and stable states. The boundary of the domain Ω lies exactly at the saddle point and is orthogonal to the direction of the steepest descent.

The integral in the denominator of Eq. (36) is dominated by the saddle point of F. The latter is found from the condition $\delta F/\delta z = 0$. This equation has an obvious solution $z_s(\boldsymbol{\rho})$

=1. In **x** representation it corresponds to $x_0 = 1/\phi_0$ and $x_i = 0$ for $i \geq 1$. Expanding expression (40) near this point up to the second order in x_i we approximate F near the saddle point by

$$F\{\mathbf{x}\} \simeq U_0 \left[\frac{1}{6\phi_0^2} - \frac{1}{2}\left(x_0 - \frac{1}{\phi_0}\right)^2 + \frac{1}{2}\sum_{i=1}^{\infty}(\epsilon_i - 1)x_i^2 \right].$$

(42)

In small samples $\epsilon_1 > 1$, and therefore F has only one unstable direction x_0, whereas all other directions are stable. One can see from Eq. (42) that in this approximation the boundary $\partial\Omega$ is the plane $x_0 = 1/\phi_0$.

Since the boundary $\partial\Omega$ is orthogonal to the x_0 direction, the sum over i in the denominator of Eq. (36) reduces to a single term with $i = 0$. Comparing Eqs. (35) and (39) one finds that $D_{ij} = (b\sqrt{\gamma\alpha}/U_0)\delta_{ij}$. Noting that D_{ij} is diagonal, the sum over j also reduces to the only term with $j = 0$.

To find $\partial f/\partial x_0$ one needs to solve Eq. (37). Noting that $\epsilon_0 = 0$ and using Eq. (39), we can write the adjoint equation (37) near the saddle point as

$$\left[U_0\left(x_0 - \frac{1}{\phi_0}\right) + \frac{\partial}{\partial x_0} \right]\frac{\partial f}{\partial x_0} = 0.$$

(43)

Solving this equation, we obtain

$$\frac{\partial f}{\partial x_0} = -\sqrt{\frac{2U_0}{\pi}}e^{-(U_0/2)[x_0 - (1/\phi_0)]^2}.$$

(44)

Here the prefactor was found using the fact that $f = 1$ inside the domain Ω (i.e., at $x_0 \to -\infty$) and $f = 0$ at the domain boundary $x_0 = 1/\phi_0$.

Using Eqs. (42) and (44) we can evaluate the integral in the denominator of Eq. (36) in Gaussian approximation. Then dividing the numerator (41) by this integral, we find the following expression for the mean switching time:

$$\tilde{\tau} = \tau \Upsilon_0.$$

(45)

Here τ is the switching time (18) obtained without the inclusion of density fluctuations. The latter gives rise to the renormalization factor

$$\Upsilon_0 = \prod_{i=1}^{\infty}\sqrt{\frac{\epsilon_i - 1}{\epsilon_i + 1}}.$$

(46)

To estimate the product Υ_0 we assume a rectangular geometry of the sample with length L and width w. Then the eigenvalues ϵ_i are given by

$$\epsilon_i = \epsilon_{nm} = \pi^2 r_0^2\left(\frac{m^2}{L^2} + \frac{n^2}{w^2}\right),$$

(47)

where n, m are non-negative integers.

In small samples $\epsilon_{nm} \gg 1$ and the expression for $\ln \Upsilon_0$ can be expanded as

$$\ln \Upsilon_0 \simeq -\sum_{n,m=0}'\frac{1}{\epsilon_{nm}} = -\frac{1}{\pi^2 r_0^2}\sum_{n,m=0}'\frac{1}{\frac{m^2}{L^2} + \frac{n^2}{w^2}},$$

(48)

where the prime in the sum means that the term with $n = m = 0$ is excluded.

The infinite sum in Eq. (48) is logarithmically divergent. However, since the diffusion picture is only valid at distances greater than the mean free path l, the wave vectors of the density fluctuations cannot[32] exceed l^{-1}. Therefore, we need to cut the sum off at $m \leq L/l$ and $n \leq w/l$.

At $L \sim w \sim \sqrt{S}$ the sum (48) can be approximated by a two-dimensional integral and yields

$$\ln \Upsilon_0 \simeq -\frac{S}{2\pi r_0^2}\ln\frac{\sqrt{S}}{l}.$$

(49)

Note, that although in small samples the area S is small compared to r_0^2, the effect of density fluctuations may become significant at $l \ll \sqrt{S}$.

In the case of strip geometry, $w \ll L$, we separate the sum into two parts, with $n = 0$ and $n > 0$. The first part gives the sum of $1/m^2$ which can be explicitly evaluated and results in a small contribution $L^2/6r_0^2 \ll 1$ to $\ln \Upsilon_0$. In the second part we approximate the sum over m by the integral with an infinite upper limit. Then neglecting terms $\sim(L/r_0)^2$, we obtain the sum of $1/n$. Cutting off this sum as discussed above, we find

$$\ln \Upsilon_0 \simeq -\frac{Lw}{2\pi r_0^2}\ln\frac{w}{l}.$$

(50)

For simplicity, from now on we will consider samples with $w \sim L \sim \sqrt{S}$.

2. Renormalization of threshold voltage

Using Eqs. (18), (45), and (49) we find the following expression for the mean switching time in small samples:

$$\tilde{\tau} = \frac{2\pi}{b\sqrt{\alpha\gamma}}\exp\left(\frac{4}{3}\frac{S[a(V_{th} - V)]^{3/2}}{\gamma^{1/2}} - \frac{S\sqrt{\gamma a}(V_{th} - V)^{1/2}}{2\pi\eta}\ln\frac{\sqrt{S}}{l}\right).$$

(51)

The second term in the exponential of Eq. (51) represents the correction (49) due to the density fluctuations.

Let us consider the regime when the magnitude of this term is larger than unity, but still small compared to the first term in the exponential of Eq. (51). Then this correction can be interpreted as a shift of the threshold voltage in formula (18). Indeed, substituting $V_{th} \to V_{th} + \delta V_{th}$, with the shift

$$\delta V_{th} = -\frac{1}{4\pi a\eta}\gamma\ln\frac{\sqrt{S}}{l},$$

(52)

into Eq. (18) and expanding it up to the first order in δV_{th} we reproduce the result (51). In experiments the threshold voltage V_{th} is not known *a priori*. If one treats it as a fitting parameter, Eqs. (18) and (51) are equivalent up to the first order in δV_{th}.

The last term in the exponential of Eq. (51) formally diverges at $l \to 0$. Similar divergences have been studied in

quantum field theory in the problem of the decay of the false vacuum.[16,17,33-37] According to Eq. (14), the shift (52) of the threshold voltage is equivalent to adding a linear term $-a\delta V_{th}(n-n_{th})$ to the integrand of the functional (30). This corresponds to the standard in quantum field theory method of renormalization of action.[16,17] Such renormalization procedure removes all the divergences.

The origin of the renormalization of the threshold voltage can be understood as follows. The "action" F describes the so-called ϕ^3 field theory in two dimensions, where $\phi \equiv (n-n_{th})$ is a scalar field. An alternative approach to the renormalization of this scalar field theory is to integrate out the fast modes ϕ_F corresponding to large wave vectors, while keeping only slow modes ϕ_S with small wave vectors in the action F. One can find that the averaging of ϕ_F^2 gives the sum of inverse eigenvalues of Laplace operator identical to (48), so that the term $(\gamma/3)\phi^3$ after the integration over the fluctuations of the fast modes gives rise to $\gamma\langle\phi_F^2\rangle\phi_S = -a\delta V_{th}\phi_S$. Physically this renormalization corresponds to the averaging of the switching rate over fluctuations of the electron density n in the well with characteristic scales between the mean free path and the sample size.

Due to the renormalization of the threshold voltage the parameter α is modified as $\alpha \to \alpha + a\delta V_{th}$. Therefore, the quantities which depend on α, such that τ and r_0, are also renormalized. More precise expression for τ is given by Eq. (18) upon substitution of the renormalized α into it. On the other hand, the small corrections to the prefactor of τ due to the renormalization are more challenging to observe experimentally, and for comparison with experiment they can be ignored.

V. MEAN SWITCHING TIME IN LARGE SAMPLES

So far we studied samples of small area $S \ll r_0^2$. We found that the switching occurs when the electron density at the saddle point is uniform, because the diffusion processes are fast and they smooth out all density variations. In large samples, $S \gg r_0^2$, the diffusion is slower, and the system can reach the critical density in a small part of the well. After the switching occurs in that part, the switching process spreads rapidly throughout the entire well. In this section we study the switching time due to these nucleation processes.

To find the expression for the mean switching time τ in large samples we need to obtain the minimum and the saddle points of the functional F in Eq. (34). They can be found using the condition $\delta F/\delta z = 0$, i.e.,

$$-\nabla^2 z + z - z^2 = 0. \tag{53}$$

The boundary conditions for Eq. (53) should account for the fact that there is no current flowing through the boundaries of the sample. Since the current is proportional to the density gradient ∇n, according to Eq. (31a) these boundary conditions take the form

$$\hat{\mathbf{n}} \cdot \nabla z\big|_{\text{boundary}} = 0, \tag{54}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the boundary. The trivial solution $z(\boldsymbol{\rho}) = 0$ gives the minimum of the functional $F\{z\}$,
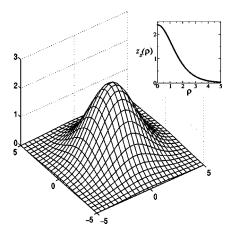

FIG. 4. The sketch of the density profile at the saddle point $z_s(\boldsymbol{\rho})$ corresponding to the solution of Eq. (53) with the boundary conditions (54). The precise radial dependence $z_s(\rho)$ obtained by solving Eq. (55) numerically is shown in the inset.

while the saddle points can be found as nontrivial solutions $z_s(\boldsymbol{\rho})$ of Eq. (53).

A. Nucleation processes in very large samples

Let us consider the switching in an infinite sample, $S \to \infty$. Due to the symmetry of the problem, the solutions of Eq. (53) should be azimuthally symmetric. Placing the origin of the coordinate system at the center of the switching region and writing Eq. (53) in polar coordinates, we find

$$z_s''(\rho) + \frac{1}{\rho}z_s'(\rho) - z_s(\rho) + z_s^2(\rho) = 0. \tag{55}$$

This equation should be solved with the boundary condition $z_s(\rho) = 0$ at $\rho \to \infty$, since otherwise $F\{z_s\} \propto S$, and the switching time $\tau \propto e^{F\{z_s\}}$ will be infinite at $S \to \infty$. One can show that this condition is consistent with Eq. (54), that is $z_s'(\infty) = 0$. Indeed, Eq. (53) can be interpreted as a Schrödinger equation for a particle in potential $-z_s$, i.e., $-(\nabla^2 + z_s)z_s = -z_s$. Therefore, $z_s(\rho)$ has the meaning of an eigenfunction of a bound state; its asymptotic behavior at large distances is $z_s \to e^{-\rho}/\sqrt{\rho}$, so that $z_s'(\infty) = 0$. The nontrivial solution of Eq. (55) with the boundary condition described earlier can be obtained numerically. The result is shown in the inset of Fig. 4.

The main exponential dependence of mean switching time τ_i in an infinite sample is given by $e^{F\{z_s\}}$. Substituting the numerical result for $z_s(\boldsymbol{\rho})$ into Eq. (34), one finds[15]

$$\tau_i = \tau_i^* \exp\left(8\zeta\frac{\eta a(V_{th} - V)}{\gamma}\right), \tag{56}$$

where the numerical constant

$$\zeta = \int d\boldsymbol{\rho}\left(\frac{(\nabla z_s)^2}{2} + \frac{z_s^2}{2} - \frac{z_s^3}{3}\right) \approx 7.751. \tag{57}$$

Equation (56) is the counterpart of the result (18) derived for small samples, $S \ll r_0^2$. Because of the dependence of r_0 on V, see Eqs. (14) and (31b), both types of behavior can be

observed in a single device by tuning the bias. At the cross-over, $r_0^2 \sim S$, the two results coincide.

The problem of stochastic current switching is similar to the problem of finding the probability of spontaneous decay of a metastable vacuum near a Peierls transition point in $(1+1)$ dimensional scalar field theory. The latter problem was solved in Ref. 37, and the exponential factor in the result for the decay time is analogous to Eq. (56). On the other hand, the prefactor of the decay time is essentially different from τ_i^*, since we study the shot noise described by the classical Fokker-Planck equation, while the false vacuum decay problem is inherently quantum mechanical.

1. Evaluation of the prefactor

In a finite sample the switching can occur anywhere in the well, hence the prefactor of the switching rate τ_i^{-1} must be proportional to the area S. Thus, while τ_i has a large exponential, Eq. (56), its prefactor τ_i^* is proportional to $1/S$ and can be small in large samples. Therefore, to fully understand the switching one needs to find τ_i^*.

The time evolution of distribution function $P\{z,t\}$ in large samples is given by the Fokker-Planck equation (32). To evaluate the prefactor of the mean switching time we again use the expression (36). The procedure is similar to the one for small samples described in Sec. IV A. However, the integration in Eq. (32) is now over a large sample, and therefore the density at the saddle point becomes nonuniform, Fig. 4. This significantly complicates the evaluation of the prefactor τ_i^*.

We evaluate both integrals in Eq. (36) in Gaussian approximation. As in Sec. IV A the integral in the numerator of Eq. (36) is dominated by the minimum of $F\{z\}$ and is given by the expression (41). The denominator of (36) is dominated by the saddle point. Presenting $z(\boldsymbol{\rho})$ near the saddle point as $z(\boldsymbol{\rho}) = z_s(\rho) + \tilde{v}(\boldsymbol{\rho})$, we obtain the expansion of $F\{z\}$ in the form

$$F\{z_s + \tilde{v}\} = F\{z_s\} + \frac{U_0}{2} \int d\boldsymbol{\rho} \, \tilde{v}(\boldsymbol{\rho})(-\nabla_\rho^2 - 2z_s + 1)\tilde{v}(\boldsymbol{\rho}).$$

$$(58)$$

It is convenient to evaluate the integral in Eq. (58) by expanding

$$\tilde{v}(\boldsymbol{\rho}) = \sum_{n,m} \tilde{x}_{nm} \tilde{\phi}_{nm}(\boldsymbol{\rho}), \tag{59}$$

where $\tilde{\phi}_{nm}(\boldsymbol{\rho})$ are the normalized solutions of the eigenvalue problem

$$[-\nabla_\rho^2 - 2z_s(\boldsymbol{\rho}) + 1]\tilde{\phi}_{nm}(\boldsymbol{\rho}) = \tilde{\lambda}_{nm}\tilde{\phi}_{nm}(\boldsymbol{\rho}). \tag{60}$$

The boundary conditions for this equation are given by Eq. (54).

Equation (60) can be interpreted as a Schrödinger equation for a particle in the attractive potential $-2z_s$ with energy $\tilde{\lambda}_{nm} - 1$. Since the potential is azimuthally symmetric, we can separate the variables as $\tilde{\phi}_{nm}(\boldsymbol{\rho}) = Q_{nm}(q\rho)\Psi_m(\varphi)$. The solutions for the azimuthal part $\Psi_m(\varphi)$ are given by $e^{\pm im\varphi}$. Below

it will be convenient to use their real combinations, $\cos m\varphi$ and $\sin m\varphi$, and introduce the following notations: $\Psi_0(\varphi) = 1/\sqrt{2\pi}$, $\Psi_m(\varphi) = (\cos m\varphi)/\sqrt{\pi}$ for $m = 1, 2, \ldots$, and $\Psi_m(\varphi) = (\sin m\varphi)/\sqrt{\pi}$ for $m = -1, -2, \ldots$.

Substituting the expansion (59) into Eq. (58) and using the orthonormality condition for the eigenfunctions, we find

$$F\{\tilde{\mathbf{x}}\} = F\{z_s\} + \frac{U_0}{2}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty} \tilde{\lambda}_{nm}\tilde{x}_{nm}^2. \tag{61}$$

The discussion leading to Eq. (61) did not rely on the assumption of large sample size. In the case of small samples Eq. (61) reproduces the expansion (42), if one identifies $\tilde{\lambda} = \epsilon - 1$. This relation is easily understood by noticing that in small samples the density at the saddle point is $z_s(\boldsymbol{\rho}) = 1$. Comparing the definition of ϵ given in the paragraph after Eq. (38) with Eq. (60), where we substitute $z_s = 1$, we reproduce $\tilde{\lambda} = \epsilon - 1$.

The form of Eq. (61) suggests that in the case of large samples it is more convenient to evaluate the integral in the denominator of Eq. (36) using variables $\tilde{x}_{nm}$ rather than x_{nm}. Since the eigenfunctions ϕ_{nm} and $\tilde{\phi}_{nm}$ are normalized, the expansion coefficients $\tilde{x}_{nm}$ are related to coefficients x_{nm} of expansion (38) via an orthogonal transformation. The Jacobian of this transformation equals unity, and therefore the integration over Πdx_{nm} in the denominator of Eq. (36) can be replaced by the integration over $\Pi d\tilde{x}_{nm}$.

In order to evaluate the integral in the denominator of Eq. (36) in the $\tilde{\mathbf{x}}$ representation, we need to find the eigenvalues $\tilde{\lambda}_{nm}$ of Eq. (60). All $\tilde{\lambda}_{nm}$ are positive with the exception of one negative eigenvalue, $\tilde{\lambda}_{00} < 0$, and two zero eigenvalues, $\tilde{\lambda}_{0,1} = \tilde{\lambda}_{0,-1} = 0$. Numerical solution of Eq. (60) yields $\tilde{\lambda}_{00} \approx -1.648$. This negative eigenvalue is associated with unstable deviation from z_s corresponding to the motion over the saddle point. In Eq. (36) the boundary $\partial\Omega$ of the domain of attraction of the metastable state is orthogonal to the $\tilde{x}_{00}$ direction, so that the integration in the denominator is performed only over the positive and zero modes. Since each positive $\tilde{\lambda}_{nm}$ corresponds to a Gaussian integral, the integration over them is straightforward. The integration over the zero modes is more challenging; to perform it we first need to understand their physical meaning.

The existence of two zero eigenvalues is due to the translational invariance of the functional $F\{z\}$ with respect to any shift of the center of the switching region in the plane of the quantum well. The two zero eigenvalues correspond to two orthogonal to each other directions in the plane along which such a shift can be performed. Indeed, a small shift $\Delta\boldsymbol{\rho}$ of the center of switching region results in the following small change in the saddle point density:

$$z_s(\boldsymbol{\rho} + \Delta\boldsymbol{\rho}) - z_s(\boldsymbol{\rho}) = \frac{\partial z_s}{\partial \rho_x}\Delta\rho_x + \frac{\partial z_s}{\partial \rho_y}\Delta\rho_y. \tag{62}$$

One can check by differentiating Eq. (53) with respect to $\rho_{x,y}$ that the derivatives $\partial z_s/\partial \rho_{x,y}$ are solutions of Eq. (60) with $\tilde{\lambda} = 0$. Furthermore, $\partial z_s/\partial \rho_x = z_s'(\rho)\cos\varphi$ and $\partial z_s/\partial \rho_y$

$=z_s'(\rho)\sin\varphi$, so the azimuthal quantum numbers corresponding to zero modes are $m=\pm1$ in our notations. Thus we conclude that $\partial z_s/\partial\rho_{x,y}=c_0\tilde{\phi}_{0,\pm1}(\boldsymbol{\rho})$, where c_0 is a constant.[38]

Substituting these expressions for $\partial z_s/\partial\rho_{x,y}$ into Eq. (62) and comparing it with the expansion (59), we find that the coefficients corresponding to zero modes are $\tilde{x}_{0,\pm1}=c_0\Delta\rho_{x,y}$. Thus the integral over the zero modes $\tilde{x}_{0,1}$ and $\tilde{x}_{0,-1}$ amounts to the integration over the possible positions of the center of the switching region in the sample,

$$\int d\tilde{x}_{0,1}\int d\tilde{x}_{0,-1}=c_0^2\int d(\Delta\rho_x)\int d(\Delta\rho_y)=\zeta\frac{S}{r_0^2}. \quad (63)$$

Here the constant c_0 was found using azimuthal symmetry of z_s and the fact that the eigenfunctions $\tilde{\phi}_{nm}(\boldsymbol{\rho})$ are normalized,

$$c_0^2=\int\left(\frac{\partial z_s}{\partial\rho_x}\right)^2 d\boldsymbol{\rho}=\frac{1}{2}\int[\nabla z_s(\boldsymbol{\rho})]^2 d\boldsymbol{\rho}=\zeta. \quad (64)$$

The relation between the last integral and the constant ζ defined by Eq. (57) is proven in Appendix E.

To find the denominator of Eq. (36) in the $\tilde{\mathbf{x}}$ representation we also need D_{ij} and $\partial f/\partial\tilde{x}_{00}$. They can be obtained from the $\tilde{\mathbf{x}}$ representation of the Fokker-Planck equation for large samples. Substituting $z(\boldsymbol{\rho})=z_s(\rho)+\tilde{v}(\boldsymbol{\rho})$ with $\tilde{v}$ in the form (59) into Eq. (32) and using the orthonormality condition for the eigenfunctions $\tilde{\phi}_{nm}(\boldsymbol{\rho})$, we obtain the Fokker-Planck equation $\dot{P}=\mathcal{L}P$ with

$$\mathcal{L}=b\sqrt{\gamma\alpha}\sum_{n,m}\left(\frac{\partial}{\partial\tilde{x}_{nm}}\tilde{\lambda}_{nm}\tilde{x}_{nm}+\frac{1}{U_0}\frac{\partial^2}{\partial\tilde{x}_{nm}^2}\right). \quad (65)$$

Here we neglected the terms of higher orders in $\tilde{x}_{nm}$. One can easily check that the solution of the stationary Fokker-Planck equation $\mathcal{L}P_0=0$ is $P_0=e^{-F}$ with F given by the Gaussian approximation (61).

Comparing Eqs. (35) and (65) we conclude that $D_{ij}=(b\sqrt{\gamma\alpha}/U_0)\delta_{ij}$. To find $\partial f/\partial\tilde{x}_{00}$ we need to solve Eq. (37) with $\mathcal{L}$ given by (65), that is

$$\left(-\tilde{\lambda}_{00}\tilde{x}_{00}+\frac{1}{U_0}\frac{\partial f}{\partial\tilde{x}_{00}}\right)\frac{\partial f}{\partial\tilde{x}_{00}}=0. \quad (66)$$

Solving it with the conditions $f=1$ inside the domain Ω (i.e., at $\tilde{x}_{00}\to-\infty$) and $f=0$ at the domain boundary $\tilde{x}_{00}=0$, we find that at the saddle point $\partial f/\partial\tilde{x}_{00}=-(2|\tilde{\lambda}_{00}|U_0/\pi)^{1/2}$.

Substituting Eq. (41) for the numerator of Eq. (36), and Eqs. (61) and (63) along with the expressions for D_{ij} and $\partial f/\partial\tilde{x}_{00}$ into the denominator of Eq. (36) we reproduce the result (56) with the prefactor given by

$$\tau_i^*\simeq\frac{\pi^2}{4\sqrt{|\tilde{\lambda}_{00}|}\zeta bS\alpha^2}\Upsilon''',\quad\Upsilon'''=\prod_{n,m}{}'''\sqrt{\frac{\tilde{\lambda}_{nm}}{\lambda_{nm}}}. \quad (67)$$

Here the product Υ''' excludes the factors corresponding to the three nonpositive eigenvalues $\tilde{\lambda}_{nm}$. The coefficients λ_{nm} denote the parameters $1+\epsilon_i$ used in Sec. IV A. They coincide with the eigenvalues of the Schrödinger equation (60) in the absence of the attractive potential $-2z_s(\rho)$.

To evaluate the infinite product Υ''' we need to find the continuous spectrum of Eq. (60). The radial part of $\tilde{\phi}_{nm}$ oscillates as a function of ρ with the wave vector q_n. The phase of these oscillations at $\rho\to\infty$ is shifted by $\delta_m(q_n)$ due to the scattering in the attractive potential $-2z_s$. The eigenvalues of the continuous spectrum are expressed in terms of these scattering phase shifts as follows:

$$\tilde{\lambda}_{nm}=1+q_n^2\simeq1+\left(\frac{\pi n}{R}\right)^2\left(1-\frac{\delta_m(\pi n/R)}{\pi n}\right)^2. \quad (68)$$

This result is derived for a round sample of dimensionless radius $R\gg1$; the derivation and the expression for the phase shifts δ_m are given in Appendix F. The expression for the eigenvalues λ_{nm} is given by Eq. (68) with $\delta_m=0$.

It is convenient to calculate the logarithm of Υ''', thereby transforming the product over n and m to a sum. Taking the large sample limit, $R\to\infty$, we replace the sum over n by an integral over $q=\pi n/R$. Then expanding the integrand in small parameter δ_m/n, we find

$$\ln\Upsilon'''\simeq-\frac{1}{\pi}\int_0^\infty\left(\sum_{m=-\infty}^\infty\delta_m(q)\right)\frac{q\,dq}{1+q^2}. \quad (69)$$

To investigate the convergence of the integral we need to evaluate the sum of the phase shifts at large q. This is accomplished with the help of the following "Friedel sum rule:"

$$\sum_{m=-\infty}^\infty\delta_m(q)\Bigg|_{q\to\infty}=2\zeta \quad (70)$$

proven in Appendix F. The asymptotic behavior (70) of the phase shifts implies that the integral in Eq. (69) diverges logarithmically at $q\to\infty$. This ultraviolet divergence signals that Υ''' is determined by a large wave vector cutoff or, equivalently, by some short distance scale. An analogous divergence appeared in the prefactor of the mean switching time in small samples, Sec. IV A. There we have shown that this short distance cutoff is of the order of the mean free path l. Following the same recipe, we cut off the integral in Eq. (69) at $q\sim r_0/l$, and with logarithmic accuracy find

$$\ln\Upsilon'''\simeq-\frac{2\zeta}{\pi}\ln\left(\frac{r_0}{l}\right). \quad (71)$$

Substituting this result into Eq. (67), we obtain the prefactor

$$\tau_i^*\sim\frac{1}{bS\alpha^2}\left(\frac{l}{r_0}\right)^{2\zeta/\pi}. \quad (72)$$

This expression completely describes the parametric dependence of the prefactor of the mean switching time in large samples. On the other hand, because of the ultraviolet divergence of Υ''', the numerical coefficient in τ_i^* cannot be determined without detailed treatment of charge transport at short length scales.[39]

2. Renormalization of threshold voltage in large samples

Expression (72) for the prefactor τ_i^* implies that in large samples the switching rate τ_i^{-1} diverges at $l\to0$. In Sec.

IV A 2 we encountered the same problem while considering small samples. There it was shown that the dependence of τ on the mean free path can be absorbed into the definition of the threshold voltage V_{th}. Following the same renormalization technique, one can shift the threshold voltage V_{th} by the amount

$$\delta V_{th} = -\frac{1}{4\pi a \eta}\gamma \ln\frac{r_0}{l}, \tag{73}$$

chosen in such a way that the resulting correction to the exponential in Eq. (56) cancels Y''' in the prefactor τ_i^*, see Eqs. (67) and (71). The renormalized result for the mean switching time then takes the form

$$\tau_i \sim \frac{1}{bS\alpha_R^2}\exp\left(\frac{8\zeta\eta\alpha_R}{\gamma}\right), \quad \alpha_R = \alpha + a\delta V_{th}. \tag{74}$$

This expression is equivalent to Eqs. (56) and (72) up to the correction in the prefactor τ_i due to the substitution $\alpha \to \alpha_R$.

The characteristic length scale r_0 is sensitive to the position of the threshold voltage, so its value must be renormalized. Since the size of the critical nucleus and δV_{th} are coupled to each other, Eq. (73), they should be evaluated self-consistently,

$$r_R = \frac{\sqrt{\eta}}{[(\alpha + a\delta V_{th})\gamma]^{1/4}}, \tag{75a}$$

$$\delta V_{th} = -\frac{1}{4\pi a\eta}\gamma\ln\frac{r_R}{l}. \tag{75b}$$

To find δV_{th} one can solve the system of equations (75) iteratively starting with $r_R = r_0$. The result (73) then should be understood as the first iteration of Eq. (75b).

Upon the substitution of the shift (75b) into Eq. (74), the logarithm of the switching time τ_i acquires an additional logarithmic dependence on voltage due to the bias-dependent renormalization of V_{th}. This dependence is physically meaningful and can, in principle, be tested experimentally. However, these corrections to the voltage dependence (56) of τ_i are small, and to the leading order $\ln\tau_i$ is still linear in voltage.

B. Nucleation near sample boundaries

In Sec. V A we studied the nucleation processes in very large samples assuming that the switching initiates far from the boundaries (i.e., at distances significantly greater than r_0). In this section we show that the switching can be more effective when it is initiated near the boundaries of the sample and evaluate the mean switching time for such processes.

1. Nucleation at a smooth edge

To study the nucleation near an edge which is smooth on the scale r_0, we model the sample by a half-plane and set up the coordinate system so that ρ_x is the coordinate along the boundary and ρ_y is positive inside the half-plane. Then the boundary condition (54) takes the form $z'_{\rho_y}(\rho_x,+0)=0$. If we place the center of the saddle-point solution z_s shown in Fig.

4 on the edge of the sample, the resulting function

$$z_e(\rho_x,\rho_y) = z_s([(\rho_x - \rho_x^{(0)})^2 + \rho_y^2]^{1/2}) \tag{76}$$

automatically satisfies not only Eq. (53) but also the boundary condition. Therefore, the expression (76) gives the saddle-point density for the half-plane.

One can argue that there are no other saddle-point solutions for edge switching. Indeed, suppose that we have a solution $\tilde{z}(\rho_x,\rho_y)$ of Eq. (53) for a half-plane. Then we can define the function $z(\rho_x,\rho_y)$ in the entire plane, so that $z=\tilde{z}(\rho_x,\rho_y)$ for $\rho_y>0$, and $z=\tilde{z}(\rho_x,-\rho_y)$ for $\rho_y<0$. By construction $z(\rho_x,\rho_y)$ satisfies Eq. (53) at $\rho_y\neq 0$. However, this procedure does not guarantee that the derivative z'_{ρ_y} is continuous at $\rho_y=0$; as a result $\partial^2 z/\partial\rho_y^2$ may have a delta-function contribution. More specifically, $z(\rho_x,\rho_y)$ satisfies the equation

$$-\nabla^2 z + z - z^2 = -2\tilde{z}'_{\rho_y}(\rho_x,+0)\delta(\rho_y). \tag{77}$$

If in addition $\tilde{z}(\rho_x,\rho_y)$ satisfies the boundary condition (54), i.e., $\tilde{z}'_{\rho_y}(\rho_x,+0)=0$, Eq. (77) coincides with Eq. (53) everywhere in the plane. Then by construction $z(\rho_x,\rho_y)=z_s(\rho)$, and therefore $\tilde{z}(\rho_x,\rho_y)$ is given by a half of the saddle-point solution z_s shown in Fig. 4 with its center on the boundary of the half-plane. Thus, there are no saddle-point solutions for edge switching except (76).

The main exponential dependence of the mean switching time τ is given by $e^{F\{z_s\}}$. In the definition (34) of $F\{z\}$ the integral is taken over the area of the sample. In the case of switching far from the boundaries it is over an entire plane, while for the edge switching this integral is over a half-plane. Therefore F is reduced by a factor of 2 compared to the case of switching far from the boundaries. Thus, instead of Eq. (56), the expression for τ at the edge takes the form

$$\tau_e = \tau_e^*\exp\left(4\zeta\frac{\eta a(V_{th}-V)}{\gamma}\right). \tag{78}$$

The evaluation of the prefactor τ_e^* is similar to the one for the switching in the middle of a large sample, Sec. V A 1. In that case we found two types of modes for the azimuthal part of the eigenfunctions $\tilde{\phi}_{nm}(\boldsymbol{\rho})$ of Eq. (60), namely, $\sin m\varphi$ and $\cos m\varphi$. At the edge only the eigenfunctions proportional to $\cos m\varphi$ are consistent with the boundary condition $z'_{\rho_y}(\rho_x,+0)=0$ on the dimensionless density z. In the notations of Sec. V A 1 these modes correspond to $m=0, 1, 2, \ldots$.

The functional $F\{z\}$ is invariant with respect to the shifts of density $z(\rho_x,\rho_y)$ along the edge of the sample. Thus $F\{z\}$ has a single zero mode $\tilde{x}_{01}$; it corresponds to the eigenfunction with the azimuthal part $\cos\varphi$. Integration over the zero mode, in analogy with Eq. (63), is performed as

$$\int d\tilde{x}_{01} = \sqrt{\frac{\zeta}{2}}\frac{\mathcal{P}}{r_0}, \tag{79}$$

where $\mathcal{P}$ is the perimeter of the sample.

To evaluate the prefactor we again use formula (36). Expression (41) for the numerator and the formulas for D_{ij} and $\partial f/\partial\tilde{x}_{00}$ in large samples are still applicable, as they were

obtained in a way independent of the exact form of the saddle-point density. Following the procedure of Sec. V A 1, we find the prefactor τ_e^* in the form

$$\tau_e^* = \frac{\pi^{3/2} Y''}{\sqrt{2\zeta|\tilde{\lambda}_{00}|} b \gamma^{1/4} \alpha^{5/4} \mathcal{P}}, \quad Y'' = \prod_{n,m \geq 0}'' \sqrt{\frac{\tilde{\lambda}_{nm}}{\lambda_{nm}}}, \quad (80)$$

cf. Eq. (67). The definition of Y'' assumes that the factors corresponding to the two lowest eigenvalues, $\tilde{\lambda}_{00}$ and $\tilde{\lambda}_{01}$, are excluded.

In the product Y'' the quantum number m changes from 0 to ∞, while in Y''' the same product is from $-\infty$ to ∞. Thus using the fact that $\tilde{\lambda}_{nm}$ and λ_{nm} are even functions of m, we obtain

$$\ln Y'' \simeq \frac{1}{2} \ln Y''' \simeq -\frac{\zeta}{\pi} \ln\left(\frac{r_0}{l}\right), \quad (81)$$

see Eq. (71).

Similarly to Eq. (72) we find the prefactor of τ for the edge switching

$$\tau_e^* \sim \frac{1}{b\mathcal{P}\gamma^{1/4}\alpha^{5/4}} \left(\frac{l}{r_0}\right)^{\zeta/\pi}. \quad (82)$$

Note that due to the ultraviolet divergence of Y'' we can evaluate τ_e^* only up to an undetermined constant.

Performing the same renormalization (73) of the threshold voltage as in Sec. V A 2, one can eliminate the explicit dependence $l^{\zeta/\pi}$ of the prefactor on the mean free path and obtain the following expression for the mean switching time at the edge:

$$\tau_e \sim \frac{1}{b\mathcal{P}\gamma^{1/4}\alpha_R^{5/4}} \exp\left(\frac{4\zeta\eta\alpha_R}{\gamma}\right). \quad (83)$$

The exponent in Eq. (83) is a factor of 2 smaller than the exponent of τ for the switching far from the boundaries, Eq. (74). Far from the threshold the exponential factor is dominant, and therefore edge switching is more efficient. To determine which switching mechanism is more efficient near the threshold, one needs to take into account the dependences of the prefactors in Eqs. (74) and (83) on the dimensions of the device.

2. Nucleation in a corner

In Sec. V B 1 we considered the processes of switching initiated near a smooth edge of the sample. In samples with pronounced corners, such as the devices of square or triangular shape, there is also a possibility of nucleation in a corner. As we will show, such processes may be more efficient than the nucleation in the interior and at the edges of the sample.

We consider a corner of angle $\theta < \pi$. Similarly to the discussion in the beginning of Sec. V B 1, one can show that the saddle-point solution $z_s(\boldsymbol{\rho})$ centered at the corner both solves the equation (53) and satisfies the boundary condition (54).

The subsequent consideration is similar to the one for the switching at a smooth edge of a large sample, Sec. V B 1. At

$\theta < \pi$ the functional $F\{z\}$ does not possess translational symmetry with respect to the shifts of $z(\boldsymbol{\rho})$, and therefore there are no zero modes. Due to the boundary condition (54) the allowed modes of the azimuthal part of the eigenfunction $\tilde{\phi}_{nm}(\boldsymbol{\rho})$ of Eq. (60) are $\cos(\pi m\varphi/\theta)$. Then instead of Eqs. (78) and (80), we obtain

$$\tau_c = \frac{2\pi Y'}{\sqrt{|\tilde{\lambda}_{00}|} b \sqrt{\gamma\alpha}} \exp\left(\frac{4\theta\zeta\eta\alpha}{\pi\gamma}\right), \quad Y' = \prod_{n,m \geq 0}' \sqrt{\frac{\tilde{\lambda}_{nm}}{\lambda_{nm}}}. \quad (84)$$

Here $\tilde{\lambda}_{nm}$ are the eigenvalues of Eq. (60) with the boundary conditions (54), which take the form $(\partial\tilde{\phi}_{nm}/\partial\varphi)|_{\varphi=0,\theta}=0$ for the corner switching. Unlike in Secs. V A 1 and V B 1, here at $n \geq 1$ the eigenvalues $\tilde{\lambda}_{nm}$ are given by Eqs. (68) and (F4) with m replaced by $\pi m/\theta$. The product Y' excludes the factor corresponding to the negative eigenvalue $\tilde{\lambda}_{00}$.

Following closely the calculations of Secs. V A 1 and V B 1, one can find the prefactor of τ_c, and the expression for the mean switching time takes the form

$$\tau_c \sim \frac{1}{b\sqrt{\gamma\alpha}} \left(\frac{l}{r_0}\right)^{\theta\zeta/\pi^2} \exp\left(\frac{4\theta\zeta\eta\alpha}{\pi\gamma}\right). \quad (85)$$

One might expect that at $\theta \to \pi$ this result should coincide with Eqs. (78) and (82) describing the edge switching. On the other hand, the prefactors for the edge and corner switching are qualitatively different, since the latter does not depend on the perimeter $\mathcal{P}$. This is due to the fact that at $\theta < \pi$ there is no zero mode, i.e., all $\tilde{\lambda}_{nm}$ except $\tilde{\lambda}_{00}$ are positive. At $\theta \to \pi$ the eigenvalue $\tilde{\lambda}_{01} \to 0$, which corresponds to the appearance of a zero mode. In this case one needs to apply the same procedure as in Sec. V B 1, which will lead to the result (82) for the prefactor.

Performing the same renormalization (73) of V_{th} as in Secs. V A 2 and V B 1, we find the expression for the mean switching time at a corner of angle θ,

$$\tau_c \sim \frac{1}{b\sqrt{\gamma\alpha_R}} \exp\left(\frac{4\theta\zeta\eta\alpha_R}{\pi\gamma}\right). \quad (86)$$

Note that at $\theta < \pi$ the exponent of τ for the corner switching is smaller than that for both interior and edge switching. This makes corner switching more efficient far from the threshold.

VI. DISCUSSION

In the preceding sections we studied the mean time τ of switching from the metastable to the stable current state in double-barrier resonant-tunneling structures. We calculated both the exponentials and prefactors of τ for switching in the small sample regime [Eq. (18)] and for the interior, edge, and corner switching in the large sample regime [Eqs. (74), (83), and (86), respectively]. In this section we discuss the dependence of the mean switching time on voltage for different structural parameters of DBRTS.

We concentrate on the case of round samples, such as the ones used in the recent experiments.[9,10] As we have shown,

when the voltage V is tuned close to the threshold, the size of critical nucleus r_0 is large compared to the radius of the sample L, and the device is in the small sample regime. If the voltage is far from V_{th}, the device is in the large sample regime, $L \gg r_0$. In a typical experiment τ is measured in a single device for different values of bias. We will therefore assume that all structural parameters and the size of the sample are fixed, and discuss the switching time as a function of voltage. For comparison with experiment we will not distinguish between r_R and r_0 in this section, since the logarithmic in voltage corrections due to the renormalization of the threshold voltage V_{th} are more challenging to observe.

Our approach is valid as long as the exponents in the expressions for the switching time, Eqs. (18), (56), and (78) are much greater than unity. To check when these conditions are satisfied, it is convenient to write the exponent in Eq. (56) as

$$\frac{8\zeta\eta\alpha}{\gamma} = \left(\frac{d}{r_0}\right)^4. \tag{87}$$

Here we introduced a new characteristic length scale

$$d \equiv \left(\frac{8\zeta\eta^3}{\gamma^2}\right)^{1/4} \tag{88}$$

and applied the definition of r_0 given by Eq. (31b). Note that the length scale d depends on structural parameters of the device, but not on the sample size or bias.

Similarly, the exponent of the switching time (18) in a small sample can be expressed in terms of d and r_0 as

$$\frac{4S\alpha^{3/2}}{3\gamma^{1/2}} = \frac{\pi}{6\zeta}\frac{L^2}{r_0^2}\left(\frac{d}{r_0}\right)^4, \tag{89}$$

where we used the fact that in round samples $S = \pi L^2$. This exponent is much greater than unity at $r_0 \ll (Ld^2)^{1/3}$. On the other hand, the regime of small sample is defined by the condition $r_0 \gg L$. Therefore, it exists only in sufficiently small samples, $L \ll d$. In this case close to the threshold there is a region of 3/2-power law behavior (18). As voltage tuned further away from V_{th} (i.e., at $r_0 \ll L$), it crosses over to the region of linear voltage dependence of $\ln \tau$ for the regime of large sample, see solid line in Fig. 5(a).

In large round samples the mean switching time τ is given by $\tau^{-1} = \tau_i^{-1} + \tau_e^{-1}$. Therefore, to find the slope of linear segment of the curve in Fig. 5(a), one must compare the rates of switching in the interior and at the edge. Using Eqs. (74) and (83), the ratio of the rates can be expressed as

$$\frac{\tau_e^{-1}}{\tau_i^{-1}} \sim \frac{d}{L} \exp\left[\frac{1}{2}\left(\frac{d}{r_0}\right)^4 - 3\ln\frac{d}{r_0}\right]. \tag{90}$$

At $L \ll d$ this result shows that the switching always occurs at the edge rather than in the interior of the sample.

To summarize, we found that in samples of radius $L \ll d$ starting at voltage difference $(V_{th} - V)$ corresponding to $r_0 \sim (Ld^2)^{1/3}$, one first observes the region of 3/2-power law dependence (18) of $\ln \tau$. Then, as $(V_{th} - V)$ increases, follows
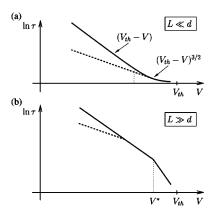

FIG. 5. Schematic dependence of the logarithm of the mean switching time τ on voltage. (a) In round samples at $L \ll d$, close to the threshold there is a region of 3/2-power law dependence of $\ln \tau$, followed by the region of linear dependence corresponding to the switching at the edge in the regime of the large sample. In samples with pronounced corners the region of linear dependence corresponds to the switching at the sharpest corner (dashed line). (b) At $L \gg d$, two regions of different linear behavior corresponding to the switching in the interior and at the edge of the large round sample are present. In samples with pronounced corners these two regions are followed by an additional region of linear voltage dependence corresponding to the switching at the corner of smallest angle θ shown by dashed line. The slope of this linear dependence is smaller by a factor of π/θ than that of edge switching.

the region of linear dependence (83) corresponding to the switching at the edge in the regime of large sample, see solid line in Fig. 5(a).

At $L \gg d$ the system is never in the small sample regime. In this case the dependence of $\ln \tau$ on voltage is linear, but it may be due to either interior or edge switching. According to Eq. (90), at $r_0 \lesssim d$ and very large L interior switching dominates. At very small r_0 the exponential in Eq. (90) becomes very large, and therefore the switching takes place at the edge. The crossover voltage V^* between these two regions of linear dependence can be determined from the condition $\tau_e^{-1} = \tau_i^{-1}$ applied to Eq. (90),

$$V^* \approx V_{th} - \frac{2\eta^2}{\gamma a d^4}\ln\frac{L}{d}. \tag{91}$$

Thus, in these large samples the interior switching (74) dominates between V^* and V_{th}, whereas at voltage below V^* the edge switching (83) prevails. The dependence of $\ln \tau$ on voltage for $L \gg d$ is shown schematically in Fig. 5(b) by solid line.

If the sample size is of order d, the dependence of $\ln \tau$ on voltage can be obtained from the dependences shown in Fig. 5(a) and Fig. 5(b). At $L \sim d$ the region of 3/2-power law dependence in Fig. 5(a) and the interior switching region in Fig. 5(b) disappear. Thus, at $L \sim d$ one can only observe the region of linear voltage dependence corresponding to the edge switching.

In samples with pronounced corners the dependence of $\ln \tau$ on voltage is different due to the possibility of corner switching. The mean switching time τ in these samples is

given by $\tau^{-1} = \tau_i^{-1} + \tau_e^{-1} + \tau_c^{-1}$. At $L \ll d$ an analysis similar to that for round samples shows that the region of linear dependence corresponds to the switching at the sharpest corner, Eq. (86). This dependence is illustrated by dashed line in Fig. 5(a). At $L \gg d$, the two regions of interior and edge switching are followed by an additional region corresponding to the switching at the sharpest corner as $(V_{th} - V)$ becomes large, see dashed line in Fig. 5(b).

Depending on the ratio of L and d two qualitatively different voltage dependences of $\ln \tau$ are expected. To see whether it is possible to observe them experimentally, we make a crude estimate of the parameter d. Substituting $\eta = \sigma/\tilde{c}b$ into Eq. (88) and using the estimates of γ and B found in Appendix A, we get

$$d \sim \frac{1}{\sqrt{n}} \left(\frac{\hbar\sigma}{e^2 T_R} \right)^{3/4} \left(\frac{T_L}{T_R} \right)^{1/2}. \tag{92}$$

To obtain this expression the capacitance of the device per unit area was estimated as $c \sim e^2 n/E_F$, and the energy of the level in the well E_0 was assumed to be of the order of E_F. The electron density n in the well is typically of the order of $2 \times 10^{11} \text{cm}^{-2}$. The transmission coefficients $T_{L,R}$ of the left and right barriers can be varied in the range from 1 to 10^{-4}, whereas the conductivity measured in units of $e^2/\hbar$ varies from 1 to 100. Assuming $T_L \sim T_R$, the low bound $d \sim 20$ nm is achieved at $\sigma \sim e^2/\hbar$ and $T_{L,R} \sim 1$. The upper bound $d \sim 600$ μm is achieved by substituting the maximum value of the conductivity and the minimum value of the transmission coefficient. These estimates show that both the cases of $L \ll d$ and $L \gg d$ are experimentally achievable in modern DBRTS, as the sample sizes range from 1 to 10^3 μm.

The available experimental data[10] confirm that the dependence of the mean switching time on voltage is indeed exponential. Based on Eq. (92) we estimate $d \sim 10$ μm, which is somewhat smaller than the radius of the sample $L = 60$ μm. Thus, one should expect the logarithm of the mean switching time to behave as shown in Fig. 5(b). (The switching time τ is referred to as the *relocation* time in Ref. 10.) On the other hand, it was observed in Ref. 10 that $\ln \tau$ bends upwards, which suggests that $L \ll d$, see Fig. 5(a). One of the possible explanations can be that this experiment was performed in superlattices, rather than in DBRTS studied in this paper, which makes our estimate of d unreliable. To test our theory in more detail, it would be interesting to carry out similar measurements of τ in several samples of different size but with the same structural parameters. This will ensure that both dependences depicted schematically in Fig. 5(a) ($L \ll d$) and Fig. 5(b) ($L \gg d$) will be observed. In addition, the exponential dependence in Ref. 10 is not very pronounced, since τ varies by only one order of magnitude. This suggests that τ was measured rather close to the threshold, and therefore the data captures only the initial part of either linear dependence for interior switching [Fig. 5(b)] or 3/2-power law dependence, Fig. 5(a). To observe the entire bias dependence shown in Fig. 5(a) or Fig. 5(b), a measurement of τ in a wider range of voltage is needed.

APPENDIX A: CALCULATION OF COEFFICIENTS *A* AND *B* IN EQ. (12)

In this appendix we find the functions $A(N)$ and $B(N)$ in Eq. (12) in the vicinity of the threshold. We will assume that the parameter (15) is small, $\lambda \ll 1$. It will be convenient here to consider A and B as functions of the electron density n rather than $N = nS$.

Let us write the expression for $A = J_L - J_R$ near the threshold. On the upper branch of the I-V curve, the level in the well lies within the conduction band in the left lead; from Eq. (9) we obtain $eV/2 + e^2N/2C + E_0 > eV$. On the lower branch, the level is below the bottom of the conduction band in the left lead $eV/2 + E_0 < eV$, so that no current can flow through the well and $N = 0$. Therefore, in the bistable region

$$eV - \frac{e^2N}{2C} < \frac{eV}{2} + E_0 < eV. \tag{A1}$$

At $\lambda \ll 1$, it follows from Eq. (8) that $e^2N/2C$ is small in comparison with E_F and E_0. One can then see from Eq. (A1) that $E_0 \simeq eV/2$, and Eq. (9) results in $e\phi \simeq eV/2$. Then from Eq. (10b) we find $\Gamma_R = \sqrt{2g_R E_0}$. The expression (7) for the rate J_L can also be simplified. For $\lambda \ll 1$ and $g_L \sim g_R$, to first order in λ the expression in the square brackets of Eq. (7) is $(Sm/\pi\hbar^2)E_F$. Using Eqs. (5), (7), and (10a) with all the above simplifications, close to the threshold $A = J_L - J_R$ can be approximated as

$$A(n,V) = \frac{g_L}{\hbar} \sqrt{E_0 \left(E_0 - \frac{eV}{2} + \frac{e^2n}{2c} \right)} \frac{Sm}{\pi\hbar^2} E_F - \frac{\sqrt{2g_R E_0}}{\hbar} nS. \tag{A2}$$

The density n on the metastable and unstable branches of the I-V curve is found by solving the equation $A(n,V) = 0$, which reduces to the quadratic equation,

$$\left(\frac{e^2n}{2c} \right)^2 - \left(\lambda \frac{g_L}{g_R} \right)^2 \frac{E_F^2}{2E_0} \left(\frac{e^2n}{2c} \right) + \left(\lambda \frac{g_L}{g_R} \right)^2 \frac{E_F^2}{2E_0} \left(\frac{eV}{2} - E_0 \right) = 0. \tag{A3}$$

At the threshold the two solutions for n coincide. This condition enables us to find the threshold voltage and density

$$V_{th} = \frac{2E_0}{e} \left(1 + \frac{\lambda^2}{8} \frac{g_L^2}{g_R^2} \frac{E_F^2}{E_0^2} \right), \tag{A4}$$

$$n_{th} = \frac{\lambda^2}{2} \left(\frac{g_L}{g_R} \right)^2 \frac{c}{e^2} \frac{E_F^2}{E_0}. \tag{A5}$$

Using Eqs. (16) and (A5), and the fact that near the threshold $\Gamma_R = \sqrt{2} g_R E_0$, we find the value of $B(n)$ at the threshold

$$B(n_{th}) = \sqrt{2} \lambda^2 \frac{1}{\hbar} \frac{g_L^2}{g_R} \frac{C}{e^2} E_F^2. \qquad (A6)$$

Using Eqs. (A4) and (A5) we expand $A(n,V)$ given by Eq. (A2) in Taylor series near (n_{th}, V_{th}) up to the first nonvanishing terms in $n-n_{th}$ and $V-V_{th}$, respectively. At the threshold $A(n)=0$ has only one solution, i.e., the first derivative with respect to n equals to zero, and therefore we need to expand up to the second order in n. The result can be presented as

$$A(n) = -\frac{B(n_{th})}{2}[-\alpha + \gamma(n-n_{th})^2], \qquad (A7a)$$

$$\alpha = \frac{2}{\lambda^2}\left(\frac{g_R}{g_L}\right)^2 \frac{E_0}{E_F^2} e(V_{th} - V), \qquad (A7b)$$

$$\gamma = \frac{2}{\lambda^4}\left(\frac{g_R}{g_L}\right)^4 \left(\frac{e^2}{c}\right)^2 \frac{E_0^2}{E_F^4}. \qquad (A7c)$$

Since $dU/dN = -2A/B$, the coefficients α and γ coincide with those used in Eq. (14).

Assuming rectangular potential profile in the well, parameters $g_{L,R}$ can be estimated in terms of the transmission coefficients of the barriers as $g_{L,R} = T_{L,R}/\pi$.

APPENDIX B: DERIVATION OF EQ. (21) FROM THE DETAILED BALANCE PRINCIPLE

Let us consider two very close to each other points $\mathbf{r}_1$ and $\mathbf{r}_2$ in the well. The system is in a local equilibrium, and the electron distributions are given by Fermi functions. We assume electrons to be sufficiently well coupled to the lattice, so that the temperature T is the same everywhere in the quantum well. Then the probabilities of diffusion between these two points are given by

$$W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n) = \sum_{ij} W_{ij} f_i (1-f_j) = \sum_{ij} W_{ij} f_i f_j e^{(\epsilon_j - \mu_2)/T},$$

$$W_{\Delta t}(\mathbf{r}_2, \mathbf{r}_1; n) = \sum_{ij} W_{ji} f_j (1-f_i) = \sum_{ij} W_{ji} f_i f_j e^{(\epsilon_i - \mu_1)/T}.$$
$$(B1)$$

Here i and j label the energy levels at positions $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively; $f_{i(j)}$ are the Fermi functions, and W_{ij} is the probability of transition from occupied level i to unoccupied level j.

In equilibrium the transition rates satisfy the detailed balance condition

$$W_{ij} e^{-\epsilon_i/T} = W_{ji} e^{-\epsilon_j/T}. \qquad (B2)$$

Our system is away from equilibrium, since the electrochemical potential $\mu(\mathbf{r})$ varies with the electron density $n(\mathbf{r})$. However, expression (B2) is still applicable for the relevant

electron scattering processes. For example, in the case of elastic scattering by impurities $\epsilon_i = \epsilon_j$, and $W_{ij} = W_{ji}$ due to time reversal symmetry, so that Eq. (B2) holds. Furthermore, one can easily check that for electron-phonon scattering expression (B2) is also valid, because the phonons are not sensitive to the change in electrochemical potential.

Strictly speaking in the presence of electron-electron scattering expression (B2) is incorrect. If electron during the transition from state i to j scatters off an electron at position $\mathbf{r}_1'$, the latter moves to position $\mathbf{r}_2'$. Then one finds an additional factor of $\exp[(\mu_1' - \mu_2')/T]$ on the right-hand side of Eq. (B2). However, because the electron-electron interaction is screened, the distance $\mathbf{r}_2' - \mathbf{r}_1'$ is of the order of the screening length in the well. The change of μ at such short distances is small compared to the temperature, and thus Eq. (B2) is still approximately correct.

Applying expression (B2) to Eqs. (B1) we obtain Eq. (21). Since during a short time interval Δt an electron can only diffuse over a short distance, the above proof is sufficient for the purposes of Sec. III A.

As an additional remark, let us show that the expression (21) also holds at larger distances. We consider the probability density $W_t(\mathbf{r}_i, \mathbf{r}_f; n)$ of diffusion from point $\mathbf{r}_i$ to a relatively distant point $\mathbf{r}_f$. Let us divide the time interval t into N small intervals $\Delta t = t/N$. Then W_t can be represented in terms of $W_{\Delta t}$ in the following way:

$$W_t(\mathbf{r}_i, \mathbf{r}_f; n) = \prod_{k=1}^{N} \int W_{\Delta t}(\mathbf{r}_k, \mathbf{r}_{k+1}; n) d\mathbf{r}_{k+1},$$

where $\mathbf{r}_1 = \mathbf{r}_i$ and $\mathbf{r}_{N+1} = \mathbf{r}_f$. The distances between the points $\mathbf{r}_k$ and $\mathbf{r}_{k+1}$ are small, so that the expression (21) is applicable. Since at small distances $\Delta\mu \ll T$, we can expand Eq. (21) up to the linear terms in $\Delta\mu/T$. Using this expansion we can rewrite each integrand in the above expression in terms of $W_{\Delta t}(\mathbf{r}_{k+1}, \mathbf{r}_k; n)$. Then evaluating the product over k we obtain Eq. (21). This completes the prove.

APPENDIX C: CALCULATION OF CONSTANT G IN EQ. (24)

In this appendix we find the constant G in Eq. (24) for an arbitrary scattering mechanism. This is accomplished by expressing G in terms of conductivity σ.

If a small electrochemical potential gradient is applied in the x direction, it gives rise to an electric current,

$$J = -L_y \sigma \frac{\partial}{\partial x}\left(\frac{\mu}{e}\right), \qquad (C1)$$

where L_y is the width of the sample.

Let us find the expression for the current along x axis at $x=0$ in terms of the transition probability density $W_{\Delta t}$. It is given by the difference in the number of electrons crossing the line $x=0$ from left to right and in the opposite direction in unit time, namely,

$$J = \frac{e}{\Delta t} \int_0^{L_y} dy_1 \int_0^{L_y} dy_2 \int_{-\infty}^0 dx_1 \int_0^\infty dx_2$$

$$\times [W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n) - W_{\Delta t}(\mathbf{r}_2, \mathbf{r}_1; n)]. \quad \text{(C2)}$$

In equilibrium, i.e., at $\partial \mu / \partial x = 0$, the expression for the difference of probability densities in the second line of Eq. (C2) vanishes. Away from equilibrium it can be found by using the "detailed balance" expression (21),

$$W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n) - W_{\Delta t}(\mathbf{r}_2, \mathbf{r}_1; n) \simeq \frac{\mu_1 - \mu_2}{T} W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n).$$

Expanding $\mu_1 - \mu_2 \simeq (x_1 - x_2) \partial \mu / \partial x$, one can see that the linearized form of Eq. (C2) reproduces Eq. (C1) with the conductivity given by

$$\sigma = \frac{e^2}{L_y T \Delta t} \int_0^{L_y} dy_1 \int_0^{L_y} dy_2 \int_{-\infty}^0 dx_1 \int_0^\infty dx_2$$

$$\times (x_2 - x_1) W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n). \quad \text{(C3)}$$

It is important to note that this expression is taken in the limit $\partial \mu / \partial x = 0$, so that $W_{\Delta t}$ in Eq. (C3) is an equilibrium quantity. Therefore $W_{\Delta t}$ depends only on the distance between $\mathbf{r}_1$ and $\mathbf{r}_2$, i.e., $W_{\Delta t}(\mathbf{r}_1, \mathbf{r}_2; n) = W_{\Delta t}(|\mathbf{r}_1 - \mathbf{r}_2|; n)$. Then substituting new variables $x = x_2 - x_1$ and $u = (x_1 + x_2)/2$ into the integral in Eq. (C3), and integrating over u, we find

$$\sigma = \frac{e^2}{2 L_y T \Delta t} \int_0^{L_y} dy_1 \int_0^{L_y} dy_2 \int_{-\infty}^\infty dx$$

$$\times x^2 W_{\Delta t}(\sqrt{x^2 + (y_1 - y_2)^2}; n).$$

Changing the variables to $y = y_1 - y_2$ and $v = (y_1 + y_2)/2$, and using the fact that $\int dv = L_y$, we obtain

$$\sigma = \frac{e^2}{4 T \Delta t} \int d\mathbf{r} \, r^2 W_{\Delta t}(|\mathbf{r}|; n). \quad \text{(C4)}$$

Finally, comparing Eqs. (24) and (C4) we find $G = 4 T \sigma / e^2$.

APPENDIX D: STATIONARY SOLUTION OF EQ. (28) NEAR THE THRESHOLD

In this appendix we discuss the stationary solution $P_s\{n\}$ of the Fokker-Planck equation (28) near the threshold. At bias near V_{th} function $b(n)$ can be approximated by a constant $b = b(n_{th})$. Then the equation for $P_s\{n\}$ takes the form

$$\left(u'(n) + \frac{\delta}{\delta n} - 2 \eta \nabla^2 n - T \frac{2\sigma}{e^2 b} \nabla^2 \frac{\delta}{\delta n} \right) P_s\{n\} = 0. \quad \text{(D1)}$$

Here $\eta = \sigma / \tilde{c} b$ and $2a/b = 2A/B = -u'(n)$, cf. Eq. (13).

It is convenient to present $P_s\{n\}$ in terms of a functional $F_s\{n\}$, such that $P_s\{n\} = \exp(-F_s\{n\})$. Then Eq. (D1) takes the following simple form:

$$-\frac{2\sigma T}{e^2 b} \nabla^2 y(\mathbf{r}) + y(\mathbf{r}) = f(\mathbf{r}), \quad \text{(D2)}$$

where we introduced $y(\mathbf{r}) = \delta F_s / \delta n$ and $f(\mathbf{r}) = u'(n) - 2 \eta \nabla^2 n$.

Solution of this equation is given by

$$y(\mathbf{r}) = \int d\mathbf{r}' f(\mathbf{r}') \mathcal{G}(\mathbf{r} - \mathbf{r}'), \quad \text{(D3)}$$

where $\mathcal{G}$ is presented in terms of the modified Bessel function K_0 as

$$\mathcal{G}(\mathbf{r}) = \frac{1}{2\pi r_T^2} K_0(r/r_T), \quad r_T = \sqrt{\frac{2\sigma T}{e^2 b}}. \quad \text{(D4)}$$

At low T the characteristic size r_T of the Green's function $\mathcal{G}$ is very small, so that $\mathcal{G}$ can be approximated by a δ-function. Then Eq. (D3) greatly simplifies,

$$\frac{\delta}{\delta n} F_s\{n\} = u'(n(\mathbf{r})) - 2 \eta \nabla^2 n(\mathbf{r}). \quad \text{(D5)}$$

The solution of this equation reproduces Eq. (30).

The exact criterion for validity of Eq. (D5) is given by the condition $r_T \ll r_0$, where r_0 is the characteristic size of the function $-u'(n) + 2 \eta \nabla^2 n$, see Eq. (31b). After substitution of the parameters of the problem from Eqs. (31b) and (A7) this criterion takes the form

$$T \ll \lambda^2 (1 + \lambda) \left(\frac{g_L}{g_R} \right)^3 \frac{E_F^3}{E_0^{3/2} \sqrt{e(V_{th} - V)}}. \quad \text{(D6)}$$

To estimate the right-hand side of (D6) we take the parameters $\lambda \sim 1$, $E_F \sim E_0$, $e(V_{th} - V) \lesssim E_F$ and $g_L \sim g_R$. Then the criterion (D6) reduces to $T \ll E_F$. Therefore, one can neglect the temperature term in Eq. (28) unless the structure is strongly asymmetrical, so that $g_L \ll g_R$.

APPENDIX E: PROPERTIES OF THE SADDLE-POINT SOLUTION $z_s(\rho)$

In this appendix we derive several relations between integrals involving $z_s(\rho)$. Our goal is to express the integrals $\int (\nabla z_s)^2 d\rho$ and $\int z_s d\rho$ in terms of ζ defined by Eq. (57).

Integrating Eq. (53) over the infinite plane and using the fact that $z_s(\rho)$ decays rapidly at large ρ, we find

$$\int z_s d\rho = \int z_s^2 d\rho. \quad \text{(E1)}$$

To express the integral in Eq. (64) in terms of the constant ζ, we transform it as

$$\int (\nabla z_s)^2 d\rho = -\int z_s \nabla^2 z_s d\rho = -\int z_s^2 d\rho + \int z_s^3 d\rho, \quad \text{(E2)}$$

where we used Eq. (53) to obtain the second line of (E2).

To express one of the integrals in the second line of Eq. (E2) in terms of the other, we take advantage of the azimuthal symmetry of the saddle-point solution. Multiplying Eq. (55) by $\rho^2 z_s'$ and integrating over ρ, we find

$$\frac{\rho^2 z_s'^2}{2}\bigg|_0^\infty + \int_0^\infty \rho^2 \frac{d}{d\rho}\left(\frac{z_s^3}{3} - \frac{z_s^2}{2}\right)d\rho = 0.$$

The first term in this equation vanishes, since $z_s \propto e^{-\rho}$ at $\rho \to \infty$. The second term can be simplified by integration by parts, resulting in

$$\int z_s^3 d\boldsymbol{\rho} = \frac{3}{2}\int z_s^2 d\boldsymbol{\rho}. \tag{E3}$$

Using Eqs. (57) and (E1)–(E3) we find the following expressions for the integrals in Eqs. (64) and (F6):

$$\int (\nabla z_s)^2 d\boldsymbol{\rho} = 2\zeta, \qquad \int z_s d\boldsymbol{\rho} = 4\zeta. \tag{E4}$$

APPENDIX F: SOLUTIONS OF EQ. (60)

In this appendix we find the eigenvalues of continuous spectrum of the Schrödinger equation (60). We consider a round sample of dimensionless radius $R = \sqrt{S/\pi r_0^2}$ with the critical fluctuation situated in the center. Note that since we are interested in the case of large samples, the size of the critical fluctuation is small compared to the sample size, i.e., $R \gg 1$.

The potential $-2z_s(\rho)$ is azimuthally symmetric, so it is convenient to solve equation (60) in polar coordinates. Separating the variables in as $\tilde{\phi}_{nm}(\boldsymbol{\rho}) = Q_{nm}(\rho)\Psi_m(\varphi)$, we can write the equation for the radial part as follows:

$$\left(-\frac{d^2}{d\rho^2} - \frac{1}{\rho}\frac{d}{d\rho} + \frac{m^2}{\rho^2} - 2z_s(\rho)\right)Q_{nm}(\rho) = q_{nm}^2 Q_{nm}(\rho), \tag{F1}$$

where $q_{nm}^2 \equiv \tilde{\lambda}_{nm} - 1$. This equation is subject to two boundary conditions: $Q_{nm}(\rho)$ is finite at the origin and $Q_{nm}'(R) = 0$.

Let us first consider an infinite sample. In the absence of the attractive potential $-2z_s$, the finite at the origin solutions to Eq. (F1) are the Bessel functions of the first kind $J_m(q\rho)$. Their asymptotic behavior at $\rho \to \infty$ is

$$J_m(q\rho) \simeq \sqrt{\frac{2}{\pi q\rho}}\cos\left[q\rho - \frac{\pi}{2}\left(m + \frac{1}{2}\right)\right]. \tag{F2}$$

In the presence of the attractive potential the asymptotic form of the radial part of the eigenfunction modifies as follows:

$$Q_m(\rho) \simeq \sqrt{\frac{2}{\pi q\rho}}\cos\left[q\rho - \frac{\pi}{2}\left(m + \frac{1}{2}\right) + \delta_m(q)\right]. \tag{F3}$$

Here $\delta_m(q)$ is the scattering phase shift due to the attractive potential.

For our purposes we only need the expression for δ_m at large wave vectors q. At $q \gg 1$ the phase shifts $\delta_m \ll 1$, and can thus be found in Born approximation,

$$\delta_m(q) = \pi \int_0^\infty z_s(\rho)J_m^2(q\rho)\rho \, d\rho, \tag{F4}$$

see also Eq. (14) in Ref. 40. Note that δ_m is indeed small at $q \gg 1$, because $J_m^2 \propto 1/q$.

In a finite sample the wave vectors q_{nm} are quantized. Using the asymptotic form (F3) and the boundary condition $Q_{nm}'(R) = 0$, we find

$$q_{nm} = \frac{\pi}{R}\left(\tilde{n} - \frac{\delta_m(q_{nm})}{\pi}\right), \tag{F5}$$

where $\tilde{n}$ is given by $n + 1/4$ if m is even, and by $n + 3/4$ if m is odd, with n being a non-negative integer. Then the eigenvalues $\tilde{\lambda}_{nm}$ are given by $1 + q_{nm}^2$. We use this result in Sec. V A 1 to calculate $\Pi_n\tilde{\lambda}_{nm}$. This product is dominated by the factors with large q_{nm}. Therefore in Eq. (68) we approximate $\tilde{n}$ by the radial quantum number n and the argument of $\delta_m(q_{nm})$ by $\pi n/R$.

In addition, in Sec. V A 1 we need an expression for the sum of the phase shifts (F4) over the azimuthal quantum numbers m. On the right-hand side of Eq. (F4) only the Bessel functions $J_m(q\rho)$ depend on m. Since the sum of $J_m^2(q\rho)$ over m equals unity,[41] we find

$$\sum_{m=-\infty}^{\infty} \delta_m = \pi \int_0^\infty z_s(\rho)\rho \, d\rho = 2\zeta. \tag{F6}$$

We used Eq. (E4) to express the above integral in terms of the constant ζ.

This result can also be derived by means of the Friedel sum rule which states that the sum of the phase shifts on the left-hand side of Eq. (F6) is given by πN, where N is the average number of levels in the attractive potential $U = -2z_s(\rho)$. Since the two-dimensional density of states $\nu_2 = 1/4\pi$, we find

$$N = \frac{1}{4\pi}\int 2z_s(\rho)d\boldsymbol{\rho}. \tag{F7}$$

Combining this expression with the Friedel sum rule we reproduce the result (F6).

[1] R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

[2] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. **58**, 1256 (1987).

[3] E. S. Alves *et al.*, Electron. Lett. **24**, 1190 (1988).

[4] A. Zaslavsky, V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Appl. Phys. Lett. **53**, 1408 (1988).

[5] R. K. Hayden, L. Eaves, M. Henini, D. K. Maude, and J. C. Portal, Phys. Rev. B **49**, 10 745 (1994).

[6] J. L. Jimenez, E. E. Mendez, X. Li, and W. I. Wang, Phys. Rev. B **52**, R5495 (1995).

[7] J. Kastrup, H. T. Grahn, K. Ploog, F. Prengel, A. Wacker, and E. Schöll, Appl. Phys. Lett. **65**, 1808 (1994).

[8] J. Kastrup, F. Prengel, H. T. Grahn, K. Ploog, and E. Schöll, Phys. Rev. B **53**, 1502 (1996).

[9] K. J. Luo, H. T. Grahn, and K. H. Ploog, Phys. Rev. B **57**, R6838 (1998).

[10] M. Rogozia, S. W. Teitsworth, H. T. Grahn, and K. H. Ploog, Phys. Rev. B **64**, 041308(R) (2001).

[11] F. W. Sheard and G. A. Toombs, Appl. Phys. Lett. **52**, 1228 (1988).

[12] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).

[13] A. Levy Yeyati, F. Flores, and E. V. Anda, Phys. Rev. B **47**, 10 543 (1993).

[14] Ya. M. Blanter and M. Büttiker, Phys. Rev. B **59**, 10 217 (1999).

[15] O. A. Tretiakov, T. Gramespacher, and K. A. Matveev, Phys. Rev. B **67**, 073303 (2003).

[16] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Oxford University Press, New York, 2001), Chap. 37.

[17] S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1985), Chap. 7, and references therein.

[18] According to our estimates, in experiments (Refs. 8–10) $\Gamma_{L,R} \ll eV \sim E_F$, and therefore our assumption of the sequential tunneling is justified.

[19] R. Landauer, J. Appl. Phys. **33**, 2209 (1962).

[20] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (Elsevier, Amsterdam, 1997).

[21] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer-Verlag, Berlin, New York, 1985).

[22] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1*, 3rd ed. (Butterworth-Heinemann, Oxford, 1980), Chap. XIV.

[23] J. Kurkijärvi, Phys. Rev. B **6**, 832 (1972).

[24] R. H. Victora, Phys. Rev. Lett. **63**, 457 (1989).

[25] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, J. Chem. Phys. **100**, 5735 (1994).

[26] M. I. Dykman and M. A. Krivoglaz, Physica A **104**, 480 (1980).

[27] M. I. Dykman, B. Golding, J. R. Kruse, L. I. McCann, and D. Ryvkine, cond-mat/0204621 (unpublished).

[28] M. I. Dykman, B. Golding, and D. Ryvkine, Phys. Rev. Lett. **92**, 080602 (2004).

[29] D. Ryvkine, M. I. Dykman, and B. Golding, Phys. Rev. E **69**, 061102 (2004).

[30] B. J. Matkowsky and Z. Schuss, SIAM J. Appl. Math. **33**, 365 (1977).

[31] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990), see Eq. (7.46).

[32] At wave vectors larger than $1/l$ the motion of electrons becomes ballistic, and therefore the conductivity $\sigma \to \infty$. Then it follows from the expression (31b) that $r_0 \to \infty$, and thus the sum (48) is cut off at wave vectors of the order of $1/l$.

[33] S. Coleman, Phys. Rev. D **15**, 2929 (1977); **16**, 1248(E) (1977).

[34] C. G. Callan and S. Coleman, Phys. Rev. D **16**, 1762 (1977).

[35] M. B. Voloshin, in *Vacuum and Vacua: the Physics of Nothing*, edited by Z. Zichichi (World Scientific, Singapore, 1996).

[36] M. B. Voloshin, I. Yu. Kobzarev, and L. B. Okun', Yad. Fiz. **20**, 1229 (1974) [Sov. J. Nucl. Phys. **20**, 644 (1975)].

[37] K. G. Selivanov, Zh. Eksp. Teor. Fiz. **94**, 57 (1988) [Sov. Phys. JETP **67**, 1548 (1988)].

[38] For the zero modes the quantum number $n=0$, because the radial part of the wave function $z_s'(\rho)$ has no zeros at finite ρ, and thus corresponds to the ground state at $m = \pm 1$.

[39] Apart from the continuous spectrum, Eq. (60) has one discrete positive eigenvalue $\tilde{\lambda}_{10} \approx 0.771$. It gives rise to a factor $\sqrt{\tilde{\lambda}_{10}}$ in Y''' that was not accounted for in Eq. (71). We neglect this factor along with the unknown numerical coefficient in Eq. (72).

[40] K. A. Matveev and A. I. Larkin, Phys. Rev. B **46**, 15 337 (1992).

[41] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).