Shot noise in resonant-tunneling structures

Giuseppe Iannaccone
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa
Massimo Macucci
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa
Bruno Pellegrini
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa

G. Iannaccone, M. Macucci, B. Pellegrini, *Shot noise in resonant tunneling structures* , Physical Review B, **55**, pp.4539-4550 (1997).

Shot noise in resonant-tunneling structures

G. Iannaccone,* M. Macucci, and B. Pellegrini
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica e Telecomunicazioni, Università degli studi di Pisa, Via Diotisalvi 2, I-56126 Pisa, Italy
(Received 18 July 1996)

We propose a quantum-mechanical approach to noise in resonant-tunneling structures that can be applied in the whole range of transport regimes, from completely coherent to completely incoherent. In both limiting cases, well-known results which have appeared in the literature are recovered. Shot noise reduction due to both Pauli exclusion and Coulomb repulsion, and their combined effect, are studied as a function of the rate of incoherent processes in the well (which are taken into account by means of a phenomenological relaxation time), and of temperature. Our approach allows the study of noise in a variety of operating conditions (i.e., equilibrium, subpeak voltages, second-resonance voltages), and as a function of temperature, explaining experimental results and predicting interesting results, such as the dependence of noise on filled-emitter states and the prediction of both increasing and decreasing shot noise with increasing temperature, depending on the structure. It also allows the determination of the major contributions to shot-noise suppression by performing noise measurements at the second-resonance voltage. [S0163-1829(97)03207-4]

I. INTRODUCTION

In recent years, researchers have shown an increasing interest in noise in resonant-tunneling structures. In fact, from an experimental point of view, noise measurements can provide information about the structure and the transport properties of resonant-tunneling devices complementary to those given by dc characteristics and small signal ac responses. On the other hand, the correct prediction of noise properties is a good check for the validity of transport models for such devices.

Lesovik[1] predicted that, in the case of completely coherent transport, shot noise could be lower than its classical value for totally independent electron crossings through the structure, i.e., the so-called full shot noise;[2] Li and co-workers[3] showed the first experimental evidence of such a phenomenon in double barrier diodes.Since then, many theoretical studies appeared in the literature, based on both coherent[4–9] and semiclassical models,[10–17] while few experimental results are available.[3,13,16,18–22] If the time of observation T is much longer than the average time τ_T an electron takes to traverse the whole device, the noise spectral density at low frequencies ($\ll 1/\tau_T$) can be reduced only if consecutive current pulses are correlated, i.e., if the pulse distribution is sub-Poissonian.[2] Two are the mechanisms which have been considered responsible for introducing such a correlation in resonant-tunneling structures: Pauli exclusion,[9–12] and electrostatic repulsion,[13,14] which both tend to smooth fluctuations of the number of electrons in the well region.

In agreement to what seems to be confirmed by experimental measurements, most theoretical studies predict a maximum suppression of one half of the classical shot-noise value, which can be obtained if the transmission probabilities of the two barriers are equal.[23] However, it is most astonishing that such results have been obtained both with coherent[1,4–6,8,11,19] and semiclassical[10,11] models, and even if time correlations between consecutive traversals of the two barriers are discarded.[24]

In this paper, we propose an approach addressing noise properties of generic resonant-tunneling structures in the whole range of transport regimes, from completely coherent to completely incoherent, from a quantum-mechanical viewpoint. In the limit of coherent transport the result of Lesovik[1] is recovered, while, in the opposite limit of loss of coherence for all electrons traversing the well, the semiclassical results of Davies[11] are obtained. We also consider the combined effects of the Pauli exclusion and of the Coulomb repulsion, and show that a maximum shot-noise suppression of one half is to be expected independently of the coherence of transport, at least up to a given amount of collisions in the well.

We also study noise behavior of resonant-tunneling structures in various bias conditions, i.e., equilibrium, subpeak voltages, and second-resonance voltages. In particular, as a check for our model, we recover the Johnson-Nyquist[25] noise at equilibrium. Moreover, we study noise dependence on temperature, which has been measured experimentally,[20] but has received little attention from a theoretical point of view.

The outline of the paper is as follows: in Sec. II we discuss our model for transport in resonant-tunneling structures and for transitions through the barriers, and introduce the simplifications and approximations needed to address the problem analytically. In Sec. III we calculate the time-dependent currents and the current power spectral density, while in Sec. IV we focus on noise in typical operating conditions, i.e., various applied voltages and operating temperatures. The particular case of large well structures (where the characteristic time for fluctuations of the number of carriers in the well is larger than the time of observation) is addressed in Sec. IV. The Summary section ends the paper.

II. MODEL

In a recent paper,[26] it has been shown that the sequential tunneling approach can be adopted to describe the whole range of transport regimes in resonant-tunneling structures,

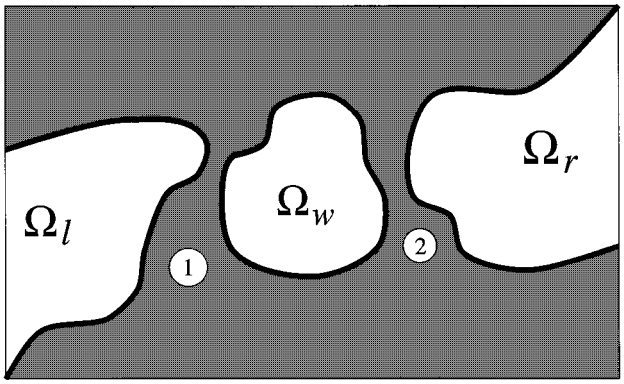

FIG. 1. A generic resonant-tunneling structure consists of three isolated regions Ω_l, Ω_w, Ω_r weakly coupled by tunneling barriers, here indicated with 1 and 2. Coupling between different regions has to be small enough to be treated with the first-order perturbation theory.

from purely coherent to completely incoherent. This approach is applicable when the structure can be seen as consisting of three isolated regions Ω_l, Ω_w, and Ω_r, i.e., the left reservoir, the well region, and the right reservoir, respectively, that are weakly coupled through the two tunneling barriers 1 and 2, as is sketched in Fig. 1, so that first-order perturbation theory is applicable.

Let each allowed state in Ω_l, Ω_w, Ω_r be characterized by a set of parameters α_l, α_w, α_r, respectively. The density of states and the occupation factor in region Ω_s $(s=l,w,r)$ are $\rho_s(\alpha_s)$ and $f_s(\alpha_s)$, respectively. Following Bardeen,[27] tunneling is treated as an electronic transition between levels in different regions. Given that first-order perturbation theory is applicable, tunneling probabilities per unit time are given by the Fermi ''golden rule.''[28]

We also take into account the effects of elastic and inelastic collisions in the well by means of a single phenomenological relaxation time $\tau_{\rm coll}$: an electron in the well has a probability $dt/\tau_{\rm coll}$ of experiencing a collision in the infinitesimal time interval dt, and electrons emerge from collisions with a thermal quasiequilibrium energy distribution and a completely random phase. Based on a similar model (in which a relaxation length was used instead of a relaxation time), a compact formula for the density of states in a quantum well has been obtained.[29]

It is worth noticing that in the relaxation-time approximation[30] all collisions are effective in randomizing phase and relaxing energy. For simplicity, we adopt this model, and we do not discuss the details of the energy dependence of the relaxation time. In fact, for the purpose of this paper, we just need to be aware of the fact that collisions affect the density of states by broadening and lowering the resonance peaks,[31–34] and affect also the occupation factor f_w in the well, which we divide into three components,[26]

$$f_w(\alpha_w)=f_{w0}(\alpha_w)+f_w^l(\alpha_w)+f_w^r(\alpha_w), \tag{1}$$

where f_{w0} is the Fermi-Dirac occupation probability associated to the quasi-Fermi level E_{fw} in the well, and f_w^l and f_w^r are the occupation factors for electrons which have come from the left and the right electrodes, respectively, and have not undergone a collision in the well.

Suppose that N electrons are in the well: the probability that in the time interval dt an electron enters the well through barrier m $(m=1,2)$ is $g_m(N)dt$. Following Davies *et al.*[11] we call $g_m(N)$ the ''partial generation rate'' for barrier m. The probability that in the time interval dt an electron escapes from the well through barrier m is $r_m(N)dt$, where $r_m(N)$ is the ''partial recombination rate'' for barrier m. Of course, we can also define the total generation rate $g(N)\equiv g_1(N)+g_2(N)$, and the total relaxation rate $r(N)\equiv r_1(N)+r_2(N)$.

A. Generation and recombination rates

Let $\hat{H}_m$ $(m=1,2)$ be the perturbation Hamiltonian due to barrier m. While in Bardeen's approach[27] tunneling is discussed from a many-particle point of view, we use here a single-particle description, and account for the effect of electron-electron interaction within a Hartree approximation. Of course, this means that our model is applicable provided that exchange and correlation contributions are negligible. We start by considering tunneling through the first barrier and calculate the matrix element $M_{1lw}\equiv\langle\alpha_w|H_1|\alpha_l\rangle$ for an electron transition from a state $|\alpha_l\rangle$ in Ω_l to a state $|\alpha_w\rangle$ in Ω_w. Of course, M_{1lw} depends on the charge density in the three regions, i.e., on the detailed occupation of the electron states.

The probability per unit time $\nu_{(l\to|\alpha_w\rangle)}$ that an electron in Ω_l jumps into an unoccupied state $|\alpha_w\rangle$ in Ω_w is given by the Fermi ''golden rule'' (to first order in $\hat{H}_1$):[28]

$$\nu_{(l\to|\alpha_w\rangle)}\equiv\frac{2\pi}{\hbar}\int|M_{1lw}|^2\rho_l f_l d\alpha_l. \tag{2}$$

Therefore, the transition rate g_1 from Ω_l to Ω_w is obtained by integrating $\nu_{(l\to|\alpha_w\rangle)}$ over all unoccupied states in Ω_w, i.e.,

$$\begin{aligned} g_1&=\int\nu_{(l\to|\alpha_w\rangle)}\rho_w(1-f_w)d\alpha_w \\ &=\frac{2\pi}{\hbar}\int\int|M_{1lw}|^2\rho_l\rho_w f_l(1-f_w)d\alpha_l d\alpha_w. \end{aligned} \tag{3}$$

Analogously, we can obtain the recombination rate r_1: the probability per unit time $\nu_{(|\alpha_w\rangle\to l)}$ that an electron in a state $|\alpha_w\rangle$ in Ω_w jumps in Ω_l is

$$\nu_{(|\alpha_w\rangle\to l)}\equiv\frac{2\pi}{\hbar}\int|M_{1lw}|^2\rho_l(1-f_l)d\alpha_l, \tag{4}$$

where we have used $M_{1lw}=M_{1wl}^*$; now r_1 is easily obtained by integrating $\nu_{(|\alpha_w\rangle\to l)}$ over occupied states in Ω_w, i.e.,

$$\begin{aligned} r_1&=\int\nu_{(|\alpha_w\rangle\to l)}\rho_w f_w d\alpha_w \\ &=\frac{2\pi}{\hbar}\int\int|M_{1lw}|^2\rho_l\rho_w f_w(1-f_l)d\alpha_l d\alpha_w. \end{aligned} \tag{5}$$

We wish to point out that both g_1 and r_1 are functionals of f_w, which appears explicitly in (3) and (5), and, through

the Hartree term, affects the total Hamiltonian of the system, and hence both the densities of states ρ_w, ρ_l, and the transition matrix elements M_{1lw}. Since we are interested in considering the effects on shot noise of Pauli exclusion and Coulomb repulsion associated to electrons occupying allowed states in the well, it should also be noticed that the effect of Pauli exclusion is accounted for in (3) through the term $(1-f_w)$, while Coulomb repulsion affects g_1 and r_1 through the term $|M_{1lw}|^2$, which depends on the charge density in the device. The passages from (2) to (5) and the above considerations can be repeated for transitions through barrier 2, by simply substituing pedices 1 with 2, and l with r.

B. Simplifying assumptions

In order to study the effects on shot noise of fluctuations in the distribution of the occupied states in the well, we have to make some simplifying, but justified, assumptions. First, we assume that the occupation factors f_l and f_r for states in the external regions do not fluctuate, which means that thermalization mechanisms in these regions are highly effective in establishing an equilibrium distribution. Then, we also assume that the effect of fluctations of f_w on the potential profile is weak enough that the densities of states in the three regions can be considered as constant. Therefore, fluctuating terms in g_1 and r_1 are f_w, of course, and $|M_{1lw}|^2$, which depends on f_w via the Poisson equation.

In realistic structures, the well region contains many states, which definitely makes the problem of considering g_1 and r_1 as functionals of the occupation factor of each level in the well not tractable. Hence, we need to make a further strong assumption: that generation and recombination rates depend on f_w only through the total number of electrons in the well N, defined as

$$N \equiv \int \rho_w f_w d\alpha_w. \tag{6}$$

In other words, we assume

$$\begin{aligned} g_m &= g_m(N) \quad \text{for } m=1,2, \\ r_m &= r_m(N) \quad \text{for } m=1,2. \end{aligned} \tag{7}$$

The total generation and recombination rates are

$$g(N) = g_1(N) + g_2(N), \tag{8}$$

$$r(N) = r_1(N) + r_2(N). \tag{9}$$

The eigenvalues of N are positive integers, and cannot be greater than the total number of states in the well $N_0 \equiv \int \rho_w d\alpha_w$. Though it is not necessary nor rigorous, we will often consider N as a number large enough to be treated as a continuous quantity, and, for example, will write derivatives of functions of N.

On one hand, it is very reasonable to state that the self-consistent energy profile, and therefore $|M_{1lw}|^2$, depend essentially on the total number of electrons in the well, and only to a second order on the detailed shape of the probability density distribution, i.e., on which particular states are actually occupied. On the other hand, we have to note that our assumption discards the effect of the detailed shape of the term f_w in both (3) and (5). We shall consider possible drawbacks of this last approximation later on, when discussing the results obtained. It is worth noticing that the expressions for generation and recombination rates used by Davies *et al.*, i.e., $g(N)=(N_0-N)/\tau_e$, and $r(N)=N/\tau_c$ [Eq. (3.10) of Ref. 11], are a particular case of (8) and (9).

C. Steady-state distribution of electrons in the well

The steady-state distribution $p_0(N)$ of electrons in the well can be obtained by using the conditions that $r(0)=0$ and N cannot be negative along with the detailed balance on the rates:[11]

$$r(N+1)p_0(N+1) = g(N)p_0(N), \tag{10}$$

which, by induction, yields

$$p_0(N) = p_0(0) \prod_{m=1}^{N} \frac{g(m-1)}{r(m)}. \tag{11}$$

One can then obtain $p_0(0)$ after imposing probability normalization, i.e.,

$$\sum_{N=0}^{N_0} p_0(N) = 1. \tag{12}$$

D. First-order approximation of generation and recombination rates

If the distribution of the total number of electrons in the well is narrow enough we can greatly improve the tractability of the problem by linearizing recombination and generation rates. Let $\tilde{N}$ be the number for which $g(\tilde{N})=r(\tilde{N})$ and let us define $\Delta N \equiv N-\tilde{N}$. We develop transition rates to the first order in ΔN, i.e.,

$$g(N) = \begin{cases} g(\tilde{N}) - \Delta N/\tau_g & \text{for } \Delta N < g(\tilde{N})\tau_g \\ 0 \quad \text{for } \Delta N \geq g(\tilde{N})\tau_g \end{cases}, \tag{13}$$

$$r(N) = \begin{cases} 0 \quad \text{for } \Delta N < -r(\tilde{N})\tau_r \\ r(\tilde{N}) + \Delta N/\tau_r \quad \text{for } \Delta N \geq -r(\tilde{N})\tau_r \end{cases}, \tag{14}$$

where the characteristic times τ_g and τ_r for generation and recombination are defined as

$$\frac{1}{\tau_g} \equiv -\left.\frac{dg}{dN}\right|_{N=\tilde{N}}, \tag{15}$$

$$\frac{1}{\tau_r} \equiv \left.\frac{dr}{dN}\right|_{N=\tilde{N}}. \tag{16}$$

In Fig. 2 a qualitative picture of the approximation made is shown. Now, by substituting (13) and (14) into (11), we obtain

$$\frac{p_0(N+1)}{p_0(N)} = \frac{g(\tilde{N}) - \Delta N/\tau_g}{r(\tilde{N}) + (\Delta N+1)/\tau_r} = \frac{\tau_r}{\tau_g}\frac{\tilde{N}_0 - L}{L+1} \tag{17}$$

if we define $L \equiv \Delta N + g(\tilde{N})\tau_r$ and $\tilde{N}_0 \equiv g(\tilde{N})(\tau_r+\tau_g)$. Since, according to (13) and (14) (Fig. 2), the number of electrons in the well cannot be lower than $\tilde{N}-r(\tilde{N})\tau_r$, for which the

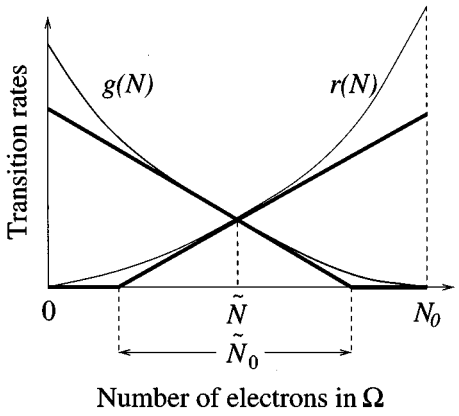

FIG. 2. A qualitative sketch of the total generation and recombination rates is shown (thin lines), along with the linearization described in Sec. II D. N_0 is the total number of states in the well, $\widetilde{N}$ is the number of electrons for which $g(\widetilde{N})=r(\widetilde{N})$, $\widetilde{N}_0$ is the maximum allowed excursion for N in the linearized approximation.

recombination rate is zero, nor greater than $\widetilde{N}+g(\widetilde{N})\tau_g$, for which the generation rate is zero, then $p_0(N)$ is nonzero only between these limits, and L runs from 0 to $\widetilde{N}_0$. From (17) we obtain for $p_0(N)$ the binomial distribution

$$p_0(N)=\binom{\widetilde{N}_0}{L}\frac{\tau^{\widetilde{N}_0}}{\tau_g^L\tau_r^{\widetilde{N}_0-L}}, \tag{18}$$

where $\tau^{-1}\equiv\tau_g^{-1}+\tau_r^{-1}$. As can be seen, (18) reduces to Eq. (3.13) of Ref. 11, if one assumes the transition rates given by Eq. (3.10) of the same reference, so that, as a consequence, $\widetilde{N}_0$ reduces to N_0, and L to N.

We now make the reasonable hypothesis that all the stochastic processes we are considering are ergodic, and indicate with $\langle a\rangle$ both the expectation value and the time average of any quantity a. Equation (18) allows us to calculate the average value of N and the variance:

$$\langle N\rangle\equiv\sum Np_0(N)=\widetilde{N}, \tag{19}$$

$$\mathrm{var}(N)\equiv\sum N^2p_0(N)-\langle N\rangle^2=\widetilde{N}_0\frac{\tau^2}{\tau_g\tau_r}=\langle g\rangle\tau. \tag{20}$$

It is worth noticing that, according to (20), the standard deviation $[\mathrm{var}(N)]^{1/2}$ is much lower than the allowed range $\widetilde{N}_0$ of variation for N, therefore we can reasonably assume the linearization of transition rates to be applicable. Moreover, given that $\langle N\rangle=\widetilde{N}$, we have

$$\langle r\rangle=\langle g\rangle=g(\widetilde{N})=g(\langle N\rangle). \tag{21}$$

It is useful to linearize partial recombination rates through each barrier, analogously to what we have done in (13) and (14) for $g(N)$ and $r(N)$. We have to define $1/\tau_{rm}\equiv(dr_m/dN)|_{N=\widetilde{N}}$ and $1/\tau_{gm}\equiv(-dg_m/dN)|_{N=\widetilde{N}}$, $m=1,2$. Therefore we have $\langle g_m\rangle=g_m(\widetilde{N})$ and $\langle r_m\rangle=r_m(\widetilde{N})$, for $m=1,2$. The steady-state current $\langle i\rangle$ at the device electrodes is given by the net transition rate through either barrier

$$\langle i\rangle=q\langle g_1-r_1\rangle=q\langle r_2-g_2\rangle. \tag{22}$$

We also define quantities that will be used in the following paragraphs:

$$\tau_1^{-1}\equiv\tau_{g1}^{-1}+\tau_{r1}^{-1},\quad \tau_2^{-1}\equiv\tau_{g2}^{-1}+\tau_{r2}^{-1}. \tag{23}$$

Of course we have

$$\tau_g^{-1}=\tau_{g2}^{-1}+\tau_{g1}^{-1},\quad \tau_r^{-1}=\tau_{r1}^{-1}+\tau_{r2}^{-1}, \tag{24}$$

$$\tau^{-1}=\tau_1^{-1}+\tau_2^{-1}=\tau_g^{-1}+\tau_r^{-1}. \tag{25}$$

E. Autocorrelation function

The autocorrelation function $c_{NN}(t)$ of $N(t)$ is defined as

$$c_{NN}(t)\equiv\langle\Delta N(0)\Delta N(t)\rangle. \tag{26}$$

By taking the time derivative we have

$$\frac{d}{dt}c_{NN}(t)=\left\langle\Delta N(0)\frac{d\Delta N(t)}{dt}\right\rangle=\langle\Delta N(0)[g(N)-r(N)]\rangle, \tag{27}$$

where we have used the rate equation[11,35] $dN(t)/dt=g(N)-r(N)$. From (13), (14), (25), and (26), we can write $dc_{NN}/dt=-c_{NN}/\tau$, from which we finally have

$$c_{NN}=\langle g\rangle\tau e^{-|t|/\tau}, \tag{28}$$

where we have used (20) and the fact that $c_{NN}(0)=\mathrm{var}(N)$. Let us point out that τ has the role of characteristic time of fluctuations in the number of electrons N.

III. NOISE

A. Time-dependent current

In this section, we are going to calculate the time-dependent current and its noise spectral density in the case of a constant voltage applied between the electrodes. According to the Ramo-Shockley theorem,[36] and to the electrokinematics theorem[37] that generalizes it to any system, to the electromagnetic field, and to quantum mechanics,[38] when an electron tunnels through one of the barriers, it generates a pulse in the current of the external circuit;[39] the time integral of the current pulse associated to the traversal of barrier m $(m=1,2)$ is $\lambda_m q$, where λ_m is equal to the ratio of the voltage drop across barrier m to the total applied voltage. Of course, we have $\lambda_1+\lambda_2=1$. In terms of the quasi-Fermi levels of the three regions, we have $\lambda_1=(E_{fl}-E_{fw})/(E_{fl}-E_{fr})$, and $\lambda_2=(E_{fw}-E_{fr})/(E_{fl}-E_{fr})$. Suppose that we observe the system in the interval $(0,T)$, in which the current $i(t)$ has the form

$$i(t)=\lambda_1 q\left[\sum_j f_j^{g1}(t-t_j^{g1})-\sum_j f_j^{r1}(t-t_j^{r1})\right]+\lambda_2 q\left[\sum_j f_j^{r2}(t-t_j^{r2})-\sum_j f_j^{g2}(t-t_j^{g2})\right], \tag{29}$$

where f_j^{g1} gives the shape of the current pulse due to a single generation via barrier 1 starting at time t_j^{g1}. Traversals of the barrier and current pulse shapes are not identical, therefore we have to associate a different function f_j^{g1} to each pulse.

What all functions f_j have in common is the normalization to unity, and the fact that their Fourier transform $F_j(\omega)$ is flat and equal to unity for frequencies much smaller than the inverse of the traversal time of each barrier. At such frequencies the Fourier transform $I(\omega)$ of the current is[40]

$$I(\omega)\equiv\int i(t)\exp(-i\omega t)dt$$
$$=\lambda_1 q\left[\sum_j \exp(-i\omega t_j^{g1})-\sum_j \exp(-i\omega t_j^{r1})\right]$$
$$+\lambda_2 q\left[\sum_j \exp(-i\omega t_j^{r2})-\sum_j \exp(-i\omega t_j^{g2})\right]. \tag{30}$$

The integral and the sums run over all the pulses occuring in the interval of observation $(0,T)$. We wish to point out that at low frequencies (here ''low'' means again much smaller than the inverse of the traversal time of the device), the power spectral density of current fluctuations $S(\omega)$ is not influenced by the pulse shape, therefore we do not expect $S(\omega)$ to be dependent on the particular values of λ_1 and λ_2 [as shown in Eq. (B3) of Appendix B].

B. Noise spectral density

The power spectral density $S(\omega)$ of the current is defined as

$$S(\omega)\equiv\frac{2}{T}\langle|I(\omega)|^2\rangle-4\pi\langle i\rangle^2\delta(\omega), \tag{31}$$

where, again, T is the time of observation. By substituting (30) in (31) we obtain 16 terms of the type

$$\left\langle\sum_k\sum_j \exp[-i\omega(t_j^\alpha-t_k^\beta)]\right\rangle, \tag{32}$$

where $\alpha,\beta=g_1,g_2,r_1,r_2$. We can analyze these terms following Ref. 11: if $\alpha=\beta$ the diagonal terms $j=k$ are equal to unity, and their sum gives the number of α hops occurring from time 0 to T, i.e., on average, $\langle\beta\rangle T$.

For $\alpha\neq\beta$, or $\alpha=\beta$ and $j\neq k$, we can define $h_{\alpha\beta}(t)$ as the probability per unit time that a α hop occurs at time t given that a β hop occurred at time 0. Therefore we can write (32) as

$$\left\langle\sum_k\int_0^T \exp[-i\omega(t-t_k^\beta)]h_{\alpha\beta}(t-t_k^\beta)dt\right\rangle. \tag{33}$$

As can be seen, the integral in (33) is independent of time and is simply the Fourier transform $H_{\alpha\beta}(\omega)$ of $h_{\alpha\beta}(t)$. The sum over k in (33) contains, on average, $\langle\beta\rangle T$ terms, therefore (32) becomes

$$\langle\beta\rangle T[\delta_{\alpha\beta}+H_{\alpha\beta}(\omega)]. \tag{34}$$

The detailed derivation of all the correlation functions $H_{\alpha,\beta}(\omega)$ is reported in Appendix A. Substitution of terms like (34) into $S(\omega)$ finally yields, for $\omega\ll 1/\tau$ [see Eq. (B3) Appendix B],

$$\frac{S(\omega)}{2q^2}=\tau^2\left(\frac{\langle g_1+r_1\rangle}{\tau_2^2}+\frac{\langle g_2+r_2\rangle}{\tau_1^2}\right). \tag{35}$$

As expected, at frequencies much smaller than the inverse of the transit time of electrons through the whole device, the relative sizes of the current pulses corresponding to the traversal of the two barriers are not relevant, therefore the dependence of $S(\omega)$ on λ_1 and λ_2 is canceled out.

We can arbitrarily decompose the net current into a component coming from the left and going towards the right I_l, given by the current $q\langle g_1\rangle$ entering the well through barrier 1 multiplied by the portion $\langle r_2\rangle/\langle r\rangle$ exiting through barrier 2, and a component coming from the right I_r, so that we have

$$I_l\equiv q\frac{\langle g_1\rangle\langle r_2\rangle}{\langle r\rangle},\quad I_r\equiv q\frac{\langle g_2\rangle\langle r_1\rangle}{\langle r\rangle}. \tag{36}$$

Substitution of (36) into (35) yields (see Appendix C)

$$\frac{S(\omega)}{2q}=I_l\left[1-\frac{2\tau^2}{\tau_1\tau_2}\left(1-\frac{\tau_1\langle r_1\rangle}{\tau_2\langle r_2\rangle}\right)\right]$$
$$+I_r\left[1-\frac{2\tau^2}{\tau_1\tau_2}\left(1-\frac{\tau_2\langle r_2\rangle}{\tau_1\langle r_1\rangle}\right)\right]. \tag{37}$$

In the next section we shall apply (35) and (37) to derive the expression of noise at several operating conditions.

IV. NOISE SUPPRESSION IN DIFFERENT OPERATING CONDITIONS

A. Thermal equilibrium

1. Conductance at equilibrium

We shall show that, at equilibrium, the noise spectral density $S(\omega)$ given by (37) yields the well-known Johnson-Nyquist[25] result. This is a good test for our model, and confirms the fact that shot noise and thermal-equilibrium noise are special forms of a more general noise formula,[8,34,41–44] i.e., for the system considered in this paper, Eq. (37). At equilibrium, the occupation factors in each of the three regions are given by Fermi-Dirac statistics, and, of course, zero net average currents flow through each barrier, i.e., $\langle g_1\rangle=\langle r_1\rangle$ and $\langle g_2\rangle=\langle r_2\rangle$.

It is reasonable to suppose that, if we apply an infinitesimal perturbation from the equilibrium condition, the electrons in the three regions still obey quasiequilibrium distributions, i.e., the occupation factor for each region has the Fermi-Dirac form, and is associated to a quasi-Fermi level which tends to that in equilibrium.

Let us calculate the conductance G at equilibrium. We need to apply a small voltage V between the right and the left electrode, as sketched in Fig. 3: we have $qV=E_{fl}-E_{fr}$. Let us define $\epsilon_1\equiv E_{fl}-E_{fw}$ and $\epsilon_2\equiv E_{fw}-E_{fr}$ so that we also have $qV=\epsilon_1+\epsilon_2$. The total current is $\langle i\rangle=I_l-I_r$. From (22) we have

$$G=\left.\frac{d\langle i\rangle}{dV}\right|_{V=0}=\left.\frac{q\langle r_2\rangle}{\langle r\rangle}\frac{d\langle g_1-r_1\rangle}{dV}\right|_{V=0}$$
$$+\left.\frac{q\langle r_1\rangle}{\langle r\rangle}\frac{d\langle r_2-g_2\rangle}{dV}\right|_{V=0}. \tag{38}$$

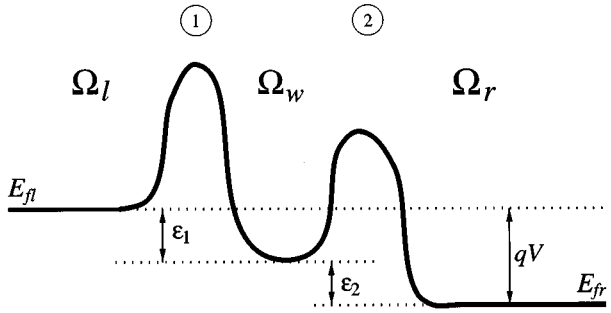

FIG. 3. The longitudinal section of a resonant-tunneling device is sketched; E_{fl}, E_{fw}, and E_{fr} are the quasi-Fermi levels of Ω_l, Ω_w, Ω_r, respectively. A small voltage V is applied between the electrodes, and ϵ_1, ϵ_2 are the partial potential energy drops across barriers 1 and 2, respectively.

The first derivative can be written in the form

$$\left.\frac{d\langle g_1 - r_1\rangle}{d\epsilon_1}\right|_{\epsilon_1=0} \left.\frac{d\epsilon_1}{dV}\right|_{V=0} = \frac{\langle g_1\rangle}{k_B\Theta}\left.\frac{d\epsilon_1}{dV}\right|_{V=0}, \tag{39}$$

where the equality comes from (B8), k_B is the Boltzmann constant, and Θ is the temperature. By using (39) and the corresponding equation for transitions through the second barrier, we write (38) as

$$G = \frac{q}{k_B\Theta\langle r\rangle}\left(\langle r_2\rangle\langle g_1\rangle \left.\frac{d\epsilon_1}{dV}\right|_{V=0} + \langle r_1\rangle\langle g_2\rangle \left.\frac{d\epsilon_2}{dV}\right|_{V=0}\right). \tag{40}$$

If we remember that at equilibrium $\langle r_2\rangle\langle g_1\rangle = \langle r_1\rangle\langle g_2\rangle$, that $\epsilon_1 + \epsilon_2 = qV$, and the definition (36) of I_l, we finally obtain G as

$$G = \frac{qI_l}{k_B\Theta}. \tag{41}$$

2. Thermal noise

In order to recover the Johnson-Nyquist noise, we just need to obtain a simple relation between τ_1 and τ_2 at equilibrium. From (15), (16), and (25), for the calculation of τ_1 we have to change the number of electrons in the well region by a small amount. At equilibrium, as we said above, the effect of this operation is to shift the quasi-Fermi level in the well with respect to those in the left and right regions. From (23) and (39) we have

$$\frac{1}{\tau_1} = \left.\frac{d(r_1 - g_1)}{d\epsilon_1}\right|_{\substack{\epsilon_1=0 \\ N=\tilde{N}}} \left.\frac{d\epsilon_1}{dN}\right|_{N=\tilde{N}}$$

$$= -\frac{\langle r_1\rangle}{k_B\Theta}\left.\frac{d\epsilon_1}{dN}\right|_{N=\tilde{N}}. \tag{42}$$

We know that $\epsilon_1 + \epsilon_2 = 0$, because at equilibrium the left and the right electrodes are at the same potential, therefore by comparing (42) and the corresponding equation for τ_2 we can finally write

$$\tau_1\langle r_1\rangle = \tau_2\langle r_2\rangle. \tag{43}$$

This is an interesting result: in fact, when we put it into (37) the factors responsible for noise suppression are canceled out, and we simply obtain

$$S(\omega) = 2q(I_l + I_r) = 4Gk_B\Theta, \tag{44}$$

i.e., the Johnson-Nyquist noise [the second equality comes from (41) and the fact that $I_l = I_r$ at equilibrium].

B. High bias

When the voltage V applied between the electrodes is large enough that most of the electrons injected in the device come from only one of the electrodes, we say the device is in a condition of high bias. Without loss of generality, we can assume that a positive voltage is applied to the right electrode, therefore high bias means $\langle g_2\rangle = 0$. The left-going current component I_r vanishes too, and the total current $\langle i\rangle = I_l$. From (37) we can obtain the shot-noise factor[2] as

$$\gamma \equiv \frac{S(\omega)}{2q\langle i\rangle} = 1 - \frac{2\tau^2}{\tau_1\tau_2}\left(1 - \frac{\tau_1\langle r_1\rangle}{\tau_2\langle r_2\rangle}\right). \tag{45}$$

It is worth noticing that, from (25), we have

$$\frac{2\tau^2}{\tau_1\tau_2} \leqslant \frac{1}{2}, \tag{46}$$

and the equal sign holds only if $\tau_1 = \tau_2$. Equation (46) implies that $\gamma \geqslant 1/2$. We also notice that, as expected, recombination through barrier one, corresponding to injected electrons not contributing to the net current, reduces the shot-noise suppression. Maximum suppression ($\gamma = 1/2$) is obtained only when $\tau_1 = \tau_2$ and $\langle r_1\rangle$ is zero, which means that V is high enough that Ω_l states in the resonant energy range of the well are fully occupied or that electrons in the well occupy states below the left electrode conduction band edge. We wish to point out that here τ_1 and τ_2 take into account both the Pauli exclusion principle and space charge effects; therefore, (45) means that the combined effects of both phenomena cannot push the shot-noise factor below 1/2.

The validity of this conclusion depends on the validity of the approximations that we have made throughout this paper. Let us recall the most relevant ones: we have used a first-order approximation for the generation and recombination rates as a function of N. Actually, this is not very limiting: in fact strong suppression requires strong correlation between electron transitions (incorrelated transitions give full shot noise); and the more transitions are correlated, the less the number of electrons in the well fluctuates around its average value, attributing validity to our first-order approximation.

The second strong assumption is that generation and recombination rates depend on the total number of electrons in the well, and not on the distribution of occupied levels. This simplification, as we said above, was necessary to make the

problem tractable while still capturing the physics of the Coulomb blockade and of the Pauli principle, but, on the other hand, prevents us from evaluating the possible effects on noise of the ''shape'' fluctuations in the distribution of occupied states in the well. Inclusion of these effects is possible if one addresses a very idealized situation, for example, a well with only two allowed levels, where simply two parameters determine the electron distribution, as Egues *et al.*[12] have done. In that case, a minimum shot-noise factor of approximately 0.45 is obtained, when the characteristic time for transitions between the two levels in the well is equal to the characteristic time for generations from the emitter, and recombinations towards the collector.

However, characteristic times for energy relaxation and phase randomization in the well in real devices are much lower than the characteristic times of escape from the well, as shown by the poor peak-to-valley ratio of experimental devices, compared to that predicted by completely coherent models. Therefore, it is very reasonable that at any moment, for a given number of electrons in the well, the electron distribution is practically the equilibrium one, in a sort of adiabatic approximation. In such a case, our assumption that g and r depend essentially on N is practically exact: the more inelastic collisions are efficient in establishing an equilibrium distribution, the more we can be confident of $g_m = g_m(N)$, and $r_m = r_m(N)$, $m = 1,2$.

On the basis of these considerations, we do not expect the effects that we have neglected to play a significant role, and the shot noise factor to drop below one half. Among available experimental studies, only the one by Brown[13] exhibits a value of γ smaller than one half, but the estimated accuracy of his results is not reported.

1. One-dimensional structures

Most experimental resonant-tunneling structures can be treated as one-dimensional devices: the problems are therefore simplified and many quantities of interest can be obtained analytically. In a one-dimensional structure, a state in any region can be decomposed in its longitudinal component $|E\rangle$, its transverse component $|\mathbf{k}_T\rangle$, and its spin component $|\sigma\rangle$, i.e., for $s = l,w,r$, $|\alpha_s\rangle = |E_s\rangle \otimes |\mathbf{k}_T\rangle \otimes |\sigma\rangle$. Electron transitions through either barrier conserve spin, longitudinal energy, and transverse wave vector, therefore the problem of calculating generation and recombination rates can be solved just in the longitudinal direction, and the results can be then integrated over transverse wave vectors and doubled to account for spin degeneracy (we discard, for simplicity, other spin effects). We can use the one-dimensional transition matrix elements derived in Appendix A of Ref. 26. For barrier 1 we have

$$|\langle E_l|H_1|E_w\rangle|^2 = |M_{1lw}(E_l,E_w)|^2 = \hbar^2 \nu_l(E_l)\,\nu_w(E_w)\,T_1(E_w)\,\delta(E_w - E_l), \tag{47}$$

where $\nu_l(E_l)$ and $\nu_w(E_w)$ are usually called attempt frequencies (because of their resemblance to the classical number of bounces on the barrier per second) for the states $|E_l\rangle$ and $|E_w\rangle$, respectively, and T_1 is the tunneling probability of barrier 1.

Let us define $\rho_s'(E_s)$ $(s = l,w,r)$ as the density of longitudinal states in Ω_s, and $\rho_T(\mathbf{k}_T)$ the density of transverse states. From (3) and (47), we have

$$g_1 = 4\pi\hbar \int dE \int d\mathbf{k}_T \nu_l(E)\,\nu_w(E)\,T_1(E)\,\rho_l'(E)\,\rho_w'(E) \times \rho_T(\mathbf{k}_T) f_l(E,\mathbf{k}_T)[1 - f_w(E,\mathbf{k}_T)]. \tag{48}$$

We assume, for simplicity, that no size effects are present in the left electrode, and that the longitudinal density of states ρ_l satisfy the condition $2\pi\hbar\,\nu_l(E)\rho_l'(E) = u(E - E_{\mathrm{cbl}})$, where u is the step function, and E_{cbl} is the conduction band edge of the left electrode. We can write

$$g_1 = 2\,\nu_w(E_R) T_1^g \int dE\; u(E - E_{\mathrm{cbl}})\,\rho_w' \int d\mathbf{k}_T \rho_T f_l (1 - f_w), \tag{49}$$

where E_R is an arbitrary resonant energy in the well, and T_1^g is defined as

$$T_1^g \equiv \frac{\int dE\,\nu_w T_1 u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_l (1 - f_w)}{\nu_w(E_R) \int dE\; u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_l (1 - f_w)}, \tag{50}$$

i.e., is practically an average of T_1 weighted on suitable couples of states for transitions from Ω_l to Ω_w. Analogously, for recombination we can write

$$r_1 = 2\,\nu_w(E_R) T_1^r \int dE u(E - E_{\mathrm{cbl}})\,\rho_w' \int d\mathbf{k}_T \rho_T f_w (1 - f_l), \tag{51}$$

where T_1^r is defined as

$$T_1^r \equiv \frac{\int dE\,\nu_w T_1 u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_w (1 - f_l)}{\nu_w(E_R) \int dE u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_w (1 - f_l)}. \tag{52}$$

It is worth noticing that if ρ_w' has a unique strong resonance for $E = E_R$, then we have $T_1^g = T_1^r = T_1(E_R)$.

Now, we need to define a few quantities of interest, namely, N_{l0}, $\widetilde{f}_l$, and η_l. We have

$$N_{l0} \equiv \int dE u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T f_l, \tag{53}$$

$$\widetilde{f}_l \equiv \frac{\int dE u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_l f_w}{\int dE u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_w}, \tag{54}$$

$$\eta_l \equiv \frac{\int dE u(E - E_{\mathrm{cbl}})\rho_w' \int d\mathbf{k}_T \rho_T f_w}{\int dE \rho_w' \int d\mathbf{k}_T \rho_T f_w}. \tag{55}$$

N_{l0} has a simple interpretation as the number of electrons with longitudinal energies greater than E_{cbl} that would be in the well if the occupation factor in the well was equal to that in the left electrode; $\widetilde{f}_l$ is the average of f_l over occupied states in the well above the conduction band edge of Ω_l, while η_l is the ratio of the number of electrons in the well with longitudinal energies greater than the conduction band edge of Ω_w to the total number N of electrons in Ω_l (therefore $0 \leqslant \eta_l \leqslant 1$). From (53)–(55), (49) and (51) can be written as

$$g_1 = 2\nu_w(E_R)T_1^g(N_{l0} - \widetilde{f}_l \eta_l N), \tag{56}$$

$$r_1 = 2\nu_w(E_R)T_1^r \eta_l N(1-\widetilde{f}_l). \tag{57}$$

To treat the problem analytically, we assume that $\widetilde{f}_l$ and η_l are only weakly dependent on the number of electrons in the well, so that the major dependence of g_1 and r_1 on N is the explicit one in Eqs. (56) and (57), and those through the tunneling probabilities T_1^g and T_1^r, which are due to electrostatic repulsion; the characteristic times for generation and recombination processes τ_{g_1} and τ_{r_1} are obtained as

$$\frac{1}{\tau_{g_1}} = -\langle g_1 \rangle \left. \frac{d\ln T_1^g}{dN} \right|_{N=\widetilde{N}} + 2\nu_w(E_R)T_1^g \widetilde{f}_l \eta_l, \tag{58}$$

$$\frac{1}{\tau_{r_1}} = \langle r_1 \rangle \left. \frac{d\ln T_1^r}{dN} \right|_{N=\widetilde{N}} + 2\nu_w(E_R)T_1^r \eta_l (1-\widetilde{f}_l). \tag{59}$$

If we discard space charge effects, the first right-hand terms of both (58) and (59) vanish, so that for τ_1 we have

$$\frac{1}{\tau_1} = \frac{1}{\tau_{g_1}} + \frac{1}{\tau_{r_1}} = 2\nu_w(E_R)\eta_l[T_1^g \widetilde{f}_l + T_1^r(1-\widetilde{f}_l)]. \tag{60}$$

The same passages can be repeated for the second barrier, once we simply substitute all pedices l with r and 1 with 2. Since we are considering the case of high bias, where $\widetilde{f}_r = 0$ and $\eta_r = 1$, we have $\langle r_2 \rangle = 2\nu_w(E_R)T_2^r N$ and $1/\tau_2 = 2\nu_w T_2^r$. By substituting these results, (57), and (60) into (45), we obtain

$$\gamma = 1 - \frac{2T_2^r T_1^g \eta_l \widetilde{f}_l}{\{T_2^r + \eta_l[T_1^g \widetilde{f}_l + T_1^r(1-\widetilde{f}_l)]\}^2}. \tag{61}$$

It is straightforward to see that, because of the term $\eta_l T_1^r(1-f_l)$ added to the denominator of (61), γ cannot be smaller than 1/2 [a closer look at (61) could also prove that γ cannot be smaller than $1-\widetilde{f}_l/2$].

Equation (61) becomes more readable if we consider particular cases. First, let us suppose that all available states in the well are above the conduction band edge of the left electrode so that we have $\eta_l = 1$. This case corresponds to applied voltages V smaller than the first peak voltage of the I-V characteristic. If, in addition, ρ_w' has a narrow peak for $E = E_R$, we also have $T_1^g = T_1^r = T_1(E_R)$, and $T_2^r = T_2(E_R)$, that, after substitution in (61), yields

$$\gamma = 1 - \frac{2T_1(E_R)T_2(E_R)\widetilde{f}_l}{[T_2(E_R)+T_1(E_R)]^2} = 1 - \widetilde{f}_l \frac{T_{\text{cohe}}^{\text{pk}}}{2}, \tag{62}$$

where $T_{\text{cohe}}^{\text{pk}}$ is the peak tunneling probability of the double barrier in the case of completely coherent transport, i.e., $T_{\text{cohe}}^{\text{pk}} = 4T_1T_2/(T_1+T_2)^2$. However, the presence of $T_{\text{cohe}}^{\text{pk}}$ does not mean that complete coherence is required for the applicability of (62): it is simply required that the longitudinal density of states in the well has a resonance narrow enough that all quantities involved in the calculation of transition rates are practically constant in the energy range of the resonant peak.

Equation (62) shows the way the product $\widetilde{f}_l T_{\text{cohe}}^{\text{pk}}$ affects the value of γ: maximum suppression ($\gamma = 1/2$) appears when both $\widetilde{f}_l$ and $T_{\text{cohe}}^{\text{pk}}$ are 1, i.e., when there is large charge accumulation at the emitter (high $\widetilde{f}_l$), and symmetric barrier transmission probabilities ($T_{\text{cohe}}^{\text{pk}} \approx 1$). The devices characterized in Ref. 20 were designed to meet these conditions near the current peak, where they actually exhibit a noise factor $\gamma \approx 1/2$.

If the Ω_l states that can be transmitted are completely filled, we have $\widetilde{f}_l = 1$, and we recover the well-known result[3–6,8,11,19]

$$\gamma = 1 - \frac{T_{\text{cohe}}^{\text{pk}}}{2}. \tag{63}$$

2. Temperature dependence

Recent experimental measurements of shot noise in double barrier diodes as a function of temperature[20] have shown a reduced suppression with increasing temperature. This fact has received little theoretical attention. A simple explanation of this effect is provided by our model: as temperature increases, inelastic collisions in the well become more frequent; in other words, the effective mean-free path gets shorter, resulting in a lower and wider resonant peak in the density of states and a lower peak-to-valley ratio in the I-V characteristics.

We now have to remove the hypothesis of narrow density of states ρ_w', while keeping $\eta_l = 1$. Let us suppose the condition of high accumulation at the emitter, in other words, let $\widetilde{f}_l = 1$. From (61) we obtain

$$\gamma = 1 - \frac{2T_2^r T_1^g}{(T_2^r + T_1^g)^2}. \tag{64}$$

In order to better explain the meaning of (64), let us consider the following situation: a constant voltage is applied between the electrodes, and the temperature is progressively raised. Suppose that the barrier dimensions are such that $T_1(E_R) \approx T_2(E_R)$ (this is the case, for example, of the devices fabricated and studied by Ciambrone *et al.*[20]). For very low temperatures the hypothesis of narrow density of states is valid, because inelastic collisions are rare, therefore Eq. (63) is applicable, and a suppression factor of one half is expected. When the temperature increases, the peak of ρ_w' widens, and T_1^g and T_2^r start to differ from $T_1(E_R) = T_2(E_R)$. In particular, we would have $T_1^g > T_1(E_R) \approx T_2(E_R) > T_2^r$, because collisions with phonons make electrons in the well relax to lower energy states, so that generation occurs more easily at higher energies (because higher energy states in the well are depopulated), while recombination occurs at lower energies (because electrons occupy lower energy states). It is straightforward to see from (64) that γ depends only on the ratio T_2^r/T_1^g, and is minimum when that ratio is unity. As temperature increases, this ratio decreases, and γ approaches unity.

This interpretation is supported by the experimental results shown in Ref. 20, where, at temperatures up to 155 K, shot-noise suppression is smoothly dependent on temperature, while it rapidly vanishes at higher temperatures.

We would like to emphasize that, if $T_1(E_R)<T_2(E_R)$, it is possible to observe reduced shot noise with increasing temperature: as a matter of fact, since $T_1^g>T_1(E_R)$ and $T_2(E_R)>T_2^r$, at some temperature we could have $T_1^g\approx T_2^r$, and have a value of γ close to one half. Let us point out that the possibility of both a positive and negative dependence of shot noise on temperature is one of the relevant prediction of our approach.

We also want to stress the point that shot noise cannot be used as a probe for measuring the coherence of transport in resonant-tunneling diodes. In fact, in the so-called ''sequential regime''[26,33] all electrons entering the well are inelastically scattered, but the density of states in the well is still narrow enough that the shot-noise factor γ is not affected by the collision rates and is simply given by (62), as in the case of completely coherent transport.

3. Second resonance

A very small noise reduction is to be expected after the first current peak: electrons in the well relax towards the resonant peak of ρ_w', and therefore fall below the conduction band edge of Ω_l, leaving higher states in the well mostly empty, so that Pauli exclusion is poorly effective in preventing generation from barrier 1. Many states in the well now have longitudinal energies smaller then $E_{\rm cbl}$, so that we have η_l appreciably smaller than unity. Let us also consider the case of high accumulation, i.e., $\tilde{f}_l=1$: (61) becomes

$$\gamma=1-\frac{2T_2^rT_1^g\eta_l}{(T_2^r+T_1^g\eta_l)^2}. \tag{65}$$

What happens now is that maximum suppression can only be obtained if $T_2^r=T_1^g\eta_l$. If energy relaxation cannot be disregarded, most electrons in the well relax towards the first resonant peak of ρ_w', and therefore fall below the conduction band edge of Ω_l, leaving higher states in the well empty, which means $\eta_l\approx 0$. In this case, unless barrier transmission probabilities differ by many orders of magnitude, no noise suppression should be obtained. In other words, Pauli exclusion is no more effective in preventing generation from barrier 1. This is the case, for example, of the diodes studied by Ciambrone *et al.*[20]

As we have said in the Introduction, noise characterization can provide information about transport in such devices complementary to that given by dc and ac characteristics. For example, if the density of states in the well ρ_w' has a second-resonant level, shot-noise measurements can provide useful insights into the coherence of transport. In fact, in the case of completely coherent transport, electrons do not relax to the lower resonant peak of ρ_w' and η_l is close to one, leading to a one half shot-noise factor, if the barriers have an equal transmission probability at the second-resonant energy. Otherwise, even a small rate of incoherent processes is sufficient to relax electrons to lower levels, depopulating the second-resonant level and leading to an η_l much smaller than 1 and to a shot-noise factor close to 1.

Moreover, looking at the dependence of shot noise on temperature, we could also determine whether Pauli exclusion or Coulomb repulsion is the dominant cause of correlation between different pulses: in fact, while the effectiveness of Pauli exclusion is strongly dependent on the collision rate in the well (i.e., on temperature), that of Coulomb interaction [not accounted for in (61)] depends mainly on the total charge accumulated in the well.

C. Short time of observation and noise at higher frequencies

Suppose that the time of observation T is much smaller than the characteristic times for generation and recombination through either barriers, i.e., $\tau\gg T$. In this situation, consecutive subpulses corresponding to a single electron traversing the device are separated by a time longer than T and do not appear to be correlated. This is easily the case, for example, if the well region has macroscopic dimensions, i.e., several electron diffusion lengths; in such a case, noise corresponding to two single barrier diodes in series is expected.

In the previous sections we have dealt with the case $\omega\tau\ll 1$, which is not applicable now. Rather, we are in the opposite limit of $\omega\tau\gg 1$, where $h_{\alpha\beta}(t)=\langle\alpha\rangle$, ($\alpha,\beta=r_1$, r_2, g_1, g_2), and $H_{\alpha\beta}(\omega)=2\pi\delta(\omega)\langle\alpha\rangle$, as is readily obtained from (A6)–(A10). By substituting such a form of $H_{\alpha\beta}$ in (B1) and then in (31), we have

$$\frac{S(\omega)}{2q^2}=\lambda_1^2\langle g_1+r_1\rangle+\lambda_2^2\langle g_2+r_2\rangle. \tag{66}$$

The full shot noise for the single barrier m ($m=1,2$) in the case of uncorrelated pulses is given by $S_m(\omega)\equiv 2q(q\langle g_m\rangle+q\langle r_m\rangle)$, therefore (66) can be written as

$$S(\omega)=\lambda_1^2S_1(\omega)+\lambda_2^2S_2(\omega). \tag{67}$$

Let us recall that λ_m is the ratio of the voltage drop across barrier m to the voltage drop across the whole device (i.e., if we refer to Fig. 3, $\lambda_m=\epsilon_m/qV$). Near equilibrium, if R_m is the differential resistance of barrier m, and $R=R_1+R_2$ is the total device resistance, we have $\lambda_m=R_m/R$ and $S_m=4k_B\Theta/R_m$ [as can be obtained from (B8)], which, substituted in (67), yields the Johnson-Nyquist result for $S(\omega)$, as expected.

If the bias is high enough that transport occurs in only one direction, i.e., for example, $\langle r_1\rangle=\langle g_2\rangle=0$, we have $S_1=S_2=2q\langle i\rangle$, with $\langle i\rangle=q\langle g_1\rangle=q\langle r_2\rangle$ and

$$\gamma=\frac{S(\omega)}{2q\langle i\rangle}=\lambda_1^2+\lambda_2^2\geqslant\frac{1}{2}, \tag{68}$$

where the equal sign holds if $\lambda_1=\lambda_2=1/2$, i.e., again, if the structure is symmetric.

In order to write the dependence of $S(\omega)$ on frequency, let us indicate the noise spectral density at low frequencies ($\omega\tau\ll 1$) given by (35) with $S_{\rm LF}$ and that at high frequencies ($\omega\tau\gg 1$) given by (66) with $S_{\rm HF}$. It is straightforward to see that if we put the complete expressions (A7)–(A10) into (B1), we obtain the term $(1+\omega^2\tau^2)$ at the denominator. Therefore, the expression of $S(\omega)$ contains a pole in $1/\tau$ and a zero, and can be written as

$$S(\omega)=S_{\rm LF}\frac{1+\omega^2\tau_z^2}{1+\omega^2\tau^2}, \tag{69}$$

where τ_z is obtained as

$$\tau_z^2=\frac{S_{\mathrm{HF}}}{S_{\mathrm{LF}}}\tau^2. \tag{70}$$

Equation (69) is applicable for frequencies much smaller than the inverse of the tunneling time of either barriers.

V. SUMMARY

We have shown that our approach has a definite advantage over those existing in the literature, because it can be applied in the whole range of transport regimes. We have seen that the results of Refs. 4–6, 8, 11, 19, which are valid in the case of completely coherent transport, are recovered, as well as those of Davies *et al.*,[11] which have been obtained on the basis of a semiclassical model, i.e., for a high rate of incoherent processes in the well. Our model has enabled us to explain how similar results could be obtained from quite different models, and even if the correlation between consecutive subpulses was discarded. Moreover, we have included in our model the combined effects of Pauli exclusion and of Coulomb repulsion on the suppression of shot noise, and concluded that in practical devices a suppression in excess of one half is not to be expected.

We have also studied shot noise in different operating conditions. At equilibrium we have recovered the Johnson-Nyquist noise, which is not a new result, but a good test for the validity of our model. At subpeak voltages we have predicted smaller suppression of shot noise, due to the empty states at the cathode. At voltages higher than that for the first peak, except for particular cases implying strongly asymmetric barriers, we recover full shot noise, because time correlations between transitions into and from the well region are reduced due to electron thermalization. In particular, for second-resonance biases, we have shown that the Pauli exclusion plays no role in reducing shot noise, if collisions are effective in establishing a thermal-equilibrium energy distribution; in that case, the study of noise suppression in that case helps us to determine the role played by Coulomb repulsion.

The dependence of shot-noise suppression on temperature has been observed in experiments, and has been addressed theoretically in this paper. We have shown that our model simply explains the experimental results. Finally, we have shown the dependence of shot noise on frequency, up to the inverse of the tunneling time of each barrier.

In the future, numerical simulations of realistic resonant-tunneling structures based on our model will be performed, in order to relax some of the approximations required to treat the problem analytically and to make a comparison with experimental results. In addition, we plan to perform experiments based on some predictions of our model: in particular, double barrier diodes with barriers of equal transparency at the second-resonance voltage peak, and strongly asymmetric diodes which could exhibit enhanced shot-noise suppression with increasing temperature.

ACKNOWLEDGMENTS

This work has been supported by the Ministry for the University of Scientific and Technological Research and by the Italian National Research Council (CNR) of Italy.

APPENDIX A: CORRELATION FUNCTIONS

In this appendix we shall derive the correlation functions $h_{\alpha\beta}(t)$ needed in Sec. III, following Davies *et al.*[11] Let us start by calculating the correlation function $h_{r_l r_m}(t)$, $(l,m=1,2)$. We need to define the conditional probability $p(N,t|M,0)$ that N electrons are in the well at time t, given that there were M electrons at time 0. From it, we can write the conditional probability per unit time $r_l(t|M,0)$ that an r_l transition occurs at time t, given that M electrons were in the well at time 0, in the form

$$r_l(t|M,0)=\sum_{N=0}^{N_0} r_l(N)p(N,t|M,0), \tag{A1}$$

where N_0 is the total number of states in the well.

We already introduced the probability $p_0(M)$ of having M electrons in the well, therefore we can write

$$\langle r_m\rangle=\sum_{M=0}^{N_0} r_m(M)p_0(M), \tag{A2}$$

from which we obtain $p_{r_m}(M)$, i.e., the probability that, when an r_m hop occurs, M electrons are left in the well, as

$$p_{r_m}(M)\equiv\frac{p_0(M+1)r_m(M+1)}{\langle r_m\rangle}. \tag{A3}$$

The probability per unit time that an r_l hop occurs at time t, given that an r_m hop occurred at time 0, i.e., $h_{r_l r_m}(t)$, is then given by

$$h_{r_l r_m}(t)=\sum_{M=0}^{N_0} p_{r_m}(M)r_l(t|M,0). \tag{A4}$$

From (14) we can write (A1) as

$$\begin{aligned} r_l(t|M,0)&=\sum_{N=0}^{N_0}\left(\langle r_l\rangle+\frac{\Delta N}{\tau_{rl}}\right)p(N,t|M,0)\\ &=\langle r_l\rangle+\frac{M-\langle N\rangle}{\tau_{rl}}e^{-|t|/\tau}, \end{aligned} \tag{A5}$$

where the second equality comes from the fact that $p(N,t|M,0)$ is normalized to unity and that the number of electrons in the well relax exponentially with time constant τ to the mean value $\langle N\rangle$,[11] as we know from the rate equation $dN/dt=g(N)-r(N)$.

From (14), (A2)–(A5), we now write

$$\begin{aligned} h_{r_l r_m}(t)&=\frac{1}{\langle r_m\rangle}\sum_{M=0}^{N_0} p_0(M+1)\left(\langle r_m\rangle+\frac{M+1-\langle N\rangle}{\tau_{rm}}\right)\\ &\quad\times\left(\langle r_l\rangle+\frac{M-\langle N\rangle}{\tau_{rl}}e^{-|t|/\tau}\right)\\ &=\langle r_l\rangle-\frac{1}{\tau_{rl}}\left(1-\frac{\mathrm{var}(N)}{\tau_{rm}\langle r_m\rangle}\right)e^{-|t|/\tau}; \end{aligned} \tag{A6}$$

sums are easily evaluated if we notice that $\Sigma_{N=0}^{N_0}p_0(N)\Delta N=0$ and $\Sigma_{N=0}^{N_0}p_0(N)(\Delta N)^2=\mathrm{var}(N)$. The Fourier transform $H_{r_l r_m}(\omega)$ of (A6) is

$$H_{r_l r_m}(\omega)=2\pi\delta(\omega)\langle r_l\rangle-\frac{2\tau}{\tau_{rl}}\left(1-\frac{\mathrm{var}(N)}{\tau_{rm}\langle r_m\rangle}\right)\frac{1}{1+\omega^2\tau^2}. \tag{A7}$$

Analogously, we can obtain the correlation functions for all the other processes, i.e., for $l,m=1,2$:

$$H_{g_l g_m}(\omega)=2\pi\delta(\omega)\langle g_l\rangle-\frac{2\tau}{\tau_{gl}}\left(1-\frac{\mathrm{var}(N)}{\tau_{gm}\langle g_m\rangle}\right)\frac{1}{1+\omega^2\tau^2}, \tag{A8}$$

$$H_{r_l g_m}(\omega)=2\pi\delta(\omega)\langle r_l\rangle+\frac{2\tau}{\tau_{rl}}\left(1-\frac{\mathrm{var}(N)}{\tau_{gm}\langle g_m\rangle}\right)\frac{1}{1+\omega^2\tau^2}, \tag{A9}$$

$$H_{g_l r_m}(\omega)=2\pi\delta(\omega)\langle g_l\rangle+\frac{2\tau}{\tau_{gl}}\left(1-\frac{\mathrm{var}(N)}{\tau_{rm}\langle r_m\rangle}\right)\frac{1}{1+\omega^2\tau^2}. \tag{A10}$$

APPENDIX B: USEFUL FORMULAS

1. Calculation of $S(\omega)$

From (30), by using (34) and considering the fact that $\langle\beta\rangle H_{\alpha\beta}=\langle\alpha\rangle H_{\beta\alpha}$ $(\alpha,\beta=g_1,g_2,r_1,r_2)$, we have

$$\begin{aligned}\frac{\langle|I(\omega)|^2\rangle}{q^2T}=&\lambda_1^2\{\langle g_1\rangle[1+H_{g_1,g_1}(\omega)]\\&-2\langle g_1\rangle H_{r_1,g_1}(\omega)+\langle r_1\rangle[1+H_{r_1,r_1}(\omega)]\}\\&+2\lambda_1\lambda_2[-\langle g_1\rangle H_{g_2,g_1}(\omega)\\&+\langle g_1\rangle H_{r_2,g_1}(\omega)+\langle r_1\rangle H_{g_2,r_1}(\omega)\\&-\langle r_1\rangle H_{r_2,r_1}(\omega)]+\lambda_2^2\{\langle g_2\rangle[1+H_{g_2,g_2}(\omega)]\\&-2\langle g_2\rangle H_{r_2,g_2}(\omega)+\langle r_2\rangle[1+H_{r_2,r_2}(\omega)]\}.\end{aligned} \tag{B1}$$

Substitution of (A7)–(A10) in (B1) and then in (31) yields, for $\omega\tau\ll 0$,

$$\begin{aligned}\frac{S(\omega)}{2q^2}=&\lambda_1^2\left[\langle g_1+r_1\rangle\left(1-\frac{2\tau}{\tau_1}\right)+\frac{2\tau}{\tau_1^2}\mathrm{var}(N)\right]\\&+2\lambda_1\lambda_2\left(\frac{\tau}{\tau_1}\langle g_2+r_2\rangle+\frac{\tau}{\tau_2}\langle g_1+r_1\rangle-\mathrm{var}(N)\frac{2\tau}{\tau_1\tau_2}\right)\\&+\lambda_2^2\left[\langle g_2+r_2\rangle\left(1-\frac{2\tau}{\tau_2}\right)+\frac{2\tau}{\tau^2}\mathrm{var}(N)\right],\end{aligned} \tag{B2}$$

but, if we put (20) in (B2), by writing $\langle g\rangle$ as $\langle g_1+r_1+g_2+r_2\rangle/2$, we obtain

$$\frac{S(\omega)}{2q^2}=(\lambda_1^2+2\lambda_1\lambda_2+\lambda_2^2)\left(\frac{\tau^2\langle g_1+r_1\rangle}{\tau_2^2}+\frac{\tau^2\langle g_2+r_2\rangle}{\tau_1^2}\right), \tag{B3}$$

which reduces to (35) if we simply remember that $\lambda_1+\lambda_2=1$. Note that, as expected, the dependence of noise at low frequencies upon the relative sizes of current pulses due to traversal of the two barriers is canceled out.

Now, from (36) and the fact that $r=r_1+r_2$, we can write

$$\begin{aligned}q\langle g_1+r_1\rangle&=\left(\frac{2\langle r_1\rangle}{\langle r_2\rangle}+1\right)I_l+I_r,\\q\langle g_2+r_2\rangle&=\left(\frac{2\langle r_2\rangle}{\langle r_1\rangle}+1\right)I_r+I_l.\end{aligned} \tag{B4}$$

By substituting (B4) and (25) in (35), we straightforwardly get Eq. (37).

2. Derivation of Eq. (39)

Each state $|\alpha_s\rangle$ $(s=l,w,r)$ is characterized by its total energy E_s^T and a set of other parameters β_s. Electron transitions between regions conserve energy, hence we can write $|M_{1lw}|^2=|M_1(E_w^T,\beta_w,\beta_l)|^2\delta(E_w-E_l)$. From (3) and (5) we can write

$$\begin{aligned}g_1-r_1=&\frac{2\pi}{\hbar}\int\int\int|M_1(E^T,\beta_w,\beta_l)|^2\rho_l(E^T,\beta_l)\rho_w(E^T,\beta_w)\\&\times[f_l(E^T,\beta_l)-f_w(E^T,\beta_w)]d\beta_w d\beta_l dE^T.\end{aligned} \tag{B5}$$

Differentiation of (B5) with respect to $\epsilon_1=E_{fl}-E_{fw}$, given that at equilibrium $f_l=f_w$, yields

$$\begin{aligned}\left.\frac{d(g_1-r_1)}{d\epsilon_1}\right|_{\epsilon_1=0}=&\frac{2\pi}{\hbar}\int\int\int|M_1(E^T,\beta_w,\beta_l)|^2\rho_l(E^T,\beta_l)\\&\times\rho_w(E^T,\beta_w)\left.\frac{df_w}{dE_{fw}}\right|_{\epsilon_1=0}d\beta_w d\beta_l dE^T,\end{aligned} \tag{B6}$$

in which, according to the Fermi-Dirac statistics which holds at thermal equilibrium, we have

$$\left.\frac{df_w}{dE_{fw}}\right|_{\epsilon_1=0}=\frac{f_{w0}(1-f_{w0})}{k_B\Theta}, \tag{B7}$$

where k_B, Θ, and f_{w0} are the Boltzmann constant, the temperature, and Fermi-Dirac occupation factor, respectively. Substitution of (B7) in (B6) finally yields

$$\left.\frac{d(g_1-r_1)}{d\epsilon_1}\right|_{\epsilon_1=0}=\frac{g_1}{k_B\Theta}. \tag{B8}$$

*Fax number: +39-50-568522. Electronic address: ianna@pimac2.iet.unipi.it

[1] G. B. Lesovik, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 513 (1989) [JETP Lett **49**, 592 (1989)].

[2] A. van der Ziel, *Noise in Solid State Devices and Circuits* (Wiley, New York, 1986), pp. 16 and 17.

[3] Y. P. Li, A. Zaslavsky, D. C. Tsui, M. Santos, and M. Shayegan, Phys. Rev. B **41**, 8388 (1990).

[4] B. Yurke and G. P. Kochanski, Phys. Rev. B **41**, 8184 (1990).
[5] M. Büttiker, Phys. Rev. Lett. **65**, 2901 (1990).
[6] L. Y. Chen and C. S. Ting, Phys. Rev. B **43**, 4534 (1991). We must point out that this paper has a serious flaw in the application of the Ramo-Shokley theorem in the beginning; actually, the formula $I=(I_L+I_R)/2$ at the end of the first page holds true only when the voltage drops across each barrier are equal. However, in the zero frequency limit, in which we are interested now, their error has no effect on the final result, which is independent on the detailed contribution of the current through each barrier to the total current, as we have shown in Eq. (B3) of Appendix B.
[7] S. Hershfield, Phys. Rev. B **46**, 7061 (1992).
[8] M. Büttiker, Phys. Rev. B **46**, 12 485 (1992).
[9] Th. Martin and R. Landauer, Phys. Rev. B **45**, 1742 (1992).
[10] L. Y. Chen and C. S. Ting, Phys. Rev. B **46**, 4717 (1992).
[11] J. H. Davies, P. Hyldgaard, S. Hershfield, and J. W. Wilkins, Phys. Rev. B **46**, 9620 (1992).
[12] J. C. Egues, S. Hershfield, and J. W. Wilkins, Phys. Rev. B **49**, 13 517 (1994).
[13] E. R. Brown, IEEE Trans. Electron. Dev. **39**, 2686 (1992).
[14] L. Y. Chen, Phys. Rev. B **48**, 4914 (1993).
[15] K. M. Hung and G. Y. Wu, Phys. Rev. B **48**, 14 687 (1993).
[16] H. Birk, M. J. M. de Jong, and C. Schönenberger, Phys. Rev. Lett. **75**, 1610 (1995).
[17] S. Hershfield, J. H. Davies, P. Hyldgaard, C. J. Stanton, and J. W. Wilkins, Phys. Rev. B **47**, 1967 (1993).
[18] P. Ciambrone, M. Macucci, G. Iannaccone, B. Pellegrini, M. Lazzarino, L. Sorba, and F. Beltram, in *Proceedings of the 7th Vilnius Conference on Fluctuation Phenomena in Physical Systems,* edited by V. Palenskis (Vilnius University Press, Vilnius, 1994).
[19] H. C. Liu, J. Li, G. C. Aers, C. R. Leavens, M. Buchanan, and L. R. Wasilewski, Phys. Rev. B **51** 5116 (1995).
[20] P. Ciambrone, M. Macucci, G. Iannaccone, B. Pellegrini, L. Sorba, M. Lazzarino, and F. Beltram, Electron. Lett. **31**, 503 (1995).
[21] M. Reznikov, M. Heiblum, H. Shtrikman, and D. Mahalu, Phys. Rev. Lett. **75**, 3340 (1995).
[22] A. Kumar, L. Saminadayar, D. C. Glattli, Y. Jin, and B. Etienne, Phys. Rev. Lett. **76**, 2778 (1996).
[23] However the consensus on this point is not unanimous: Brown, in Ref. 13 predicted and measured minimum shot noise ≈ 0.35 of the full value; Egues *et al*., in Ref. 12 predicted minimum shot noise ≈ 0.45 of the full value.
[24] M. Macucci and B. Pellegrini, in *Noise in Physical Systems and* 1/*f Fluctuations*, edited by P. H. Handel and A. L. Chung (American Institute of Physics, New York, 1993), p. 284.
[25] J. B. Johnson, Phys. Rev. B **29**, 367 (1927); H. Nyquist, *ibid*. **32**, 110 (1928).
[26] G. Iannaccone and B. Pellegrini, Phys. Rev. B **52**, 17 406 (1995).
[27] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
[28] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Redwood City, 1985), p. 322.
[29] G. Iannaccone and B. Pellegrini, Phys. Rev. B **53**, 2020 (1996).
[30] N. W. Aschroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976), pp. 6, 244, 316.
[31] A. D. Stone and P. A. Lee, Phys. Rev. Lett. **54**, 1196 (1985).
[32] M. Jonson and A. Grincwajg, Appl. Phys. Lett. **51**, 1729 (1987).
[33] M. Büttiker, IBM J. Res. Dev. **32**, 63 (1988).
[34] R. Landauer, Phys. Rev. B **47**, 16 427 (1993).
[35] A. Van der Ziel, in *Noise in Electron Devices*, edited by L. D. Smullin and H. A. Haus (Chapman and Hall, London, 1959), pp. 320–324. Though the rate equation is widely used for the study of noise processes, we wish to point out that it is rigorously true only if one considers the average over an ensemble of systems, or a time average if the system is ergodic. On the left side we have the time derivative of the actual number of particles in the well, while on the right side we have probabilities per unit time.
[36] S. Ramo, Proc. IRE **27**, 584 (1939); W. Shockley, J. Appl. Phys. **9**, 635 (1938).
[37] B. Pellegrini, Phys. Rev B **34**, 5921 (1986).
[38] B. Pellegrini, Il Nuovo Cimento **15**, 855 (1993); **15**, 881 (1993).
[39] B. Pellegrini and M. Macucci, in *Noise in Physical Systems and* 1/*f Fluctuations*, edited by P. H. Handel and A. L. Chung (American Institute of Physics, New York, 1993), p. 300.
[40] We use a definition of Fourier transform [J. Mathews, R. L. Walker, *Mathematical Methods of Physics* (Addison Wesley, Redwood City, 1970), p. 102], which is the complex conjugate of that used in Ref. 9.
[41] C. J. Stanton and J. W. Wilkins, Physica **134B**, 255 (1985).
[42] R. Landauer, Physica D **38**, 226 (1989).
[43] R. Sarpeshkar, T. Delbrück, and C. A. Mead, Circuits Dev. **9** (6), 23 (1993).
[44] M. Büttiker, Phys. Rev. B **33**, 3020 (1986).