

Measuring cotunneling in its wakeOded Zilberberg,^{1,2} Assaf Carmi,² and Alessandro Romito³¹*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland*²*Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100 Israel*³*Dahlem Center for Complex Quantum Systems and Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany*

(Received 11 August 2014; revised manuscript received 24 October 2014; published 11 November 2014)

We introduce a rate formalism to treat classically forbidden electron transport through a quantum dot (cotunneling) in the presence of a coupled measurement device. We demonstrate this formalism for a toy model case of cotunneling through a single-level dot while being coupled to a strongly pinched-off quantum point contact (QPC). We find that the detector generates three types of back-action: the measurement collapses the coherent transport through the virtual state, but at the same time allows for QPC-assisted incoherent transport, and widens the dot level. Last, we obtain the measured cotunneling time from the cross correlation between dot and QPC currents.

DOI: [10.1103/PhysRevB.90.205413](https://doi.org/10.1103/PhysRevB.90.205413)

PACS number(s): 03.65.Ta, 73.23.-b

I. INTRODUCTION

Quantum measurement is a probabilistic process where the detector's outcome is correlated with the system being in a certain state. In turn, the detector's back-action onto the system affects it according to the specific outcome of the measurement [1]. In a projective measurement this is described by the wave-function collapse, in which the outcome of the measurement is an eigenvalue of the measured observable, and the system after the measurement is projected onto the eigenstate corresponding to the obtained eigenvalue [2].

The implications of quantum measurement become particularly interesting when applied to yet another key feature of quantum mechanics—to classically forbidden processes. A striking example of such a forbidden process is that of tunneling under a potential barrier, where a particle appears on the other side of a barrier that it classically could not surmount. Whereas this appearance indicates that a tunneling event occurred, a direct observation of the particle during its virtual passage under the barrier is required as additional verification of this mechanism.

Indeed, such a proposition stands in direct conflict with quantum measurement: if the particle is measured to be under the barrier, it would collapse midway and not tunnel to the other side. Additionally, in order for the particle to collapse in the position of the barrier, it must obtain energy and appear above the barrier. As a result, one may conclude that such a measurement would correspond to effectively increasing the potential barrier, and blocking tunneling altogether. Still there is a constant interest in accessing the properties of a particle tunneling under a barrier (e.g., its traversal time), with a variety of approaches [3–7].

Beyond the scheme of projective measurement, a better suited and more realistic approach in detecting the virtual state under the barrier is that of a modification of the system's state in a continuous process accompanied by a gradual acquisition of information by the detector [8]. In particular, in a weak measurement regime, as opposed to a strong projective measurement, the detector's outcome corresponds to the measured state of the system, but the back-action does not disturb it much. This allows for nontrivial effects in conditional (postselected) measurements, e.g., the appearance of weak values [9], and utilization of the measurement outcome

in quantum feedback circuits [10,11]. Such effects have been successfully employed in practical problems including precision measurements [12–17], quantum state discrimination [18], or quantum state stabilization [11]. Moreover, weak measurements have also been successfully employed in the study of coherent quantum transport under a potential barrier [7,19], as well as transport through many-body virtual states [20].

The detection of a tunneling process via weak measurements can be directly explored in electronic solid-state devices. The typical system under consideration is that of transport between two leads across a quantum dot. By tuning the capacitance of the dot, one can position its eigenenergies relative to the chemical potential of the leads, such that an addition of charge onto the dot is disfavored by Coulomb interactions. As a result, transport through the dot is classically blocked [21]. In this regime, the transport through the dot happens via cotunneling processes with a virtual (classically forbidden) occupation of the dot [22].

Charge detection in quantum dots using a quantum point contact (QPC) that is capacitively coupled to the dot is experimentally well established [23–31]. The QPCs can be tuned to be close to the quantum limit, making them sensitive to the dot's charge fluctuations. Theoretically QPC detectors are well understood, and a proper formalism has been developed to describe the many-electron macroscopic classical signal therein in response to local charges [11,32]. Yet these descriptions are developed for classical transport through the dot [32], or for coherent charge oscillations in isolated systems, e.g., double quantum dots [11,32,33]. In a recent work, measurement of cotunneling was addressed in the regime of weak coupling between a dot and a QPC [20]. Using a weak value approach extended to deal with the interacting dot and a large bandwidth of the QPC, the short cotunneling time was resolved. This led to a vanishingly small cotunneling time in the regime of a diffusive 2D dot with sufficiently closed contacts.

In this work, we extend the standard model of QPC transport [11,32] to treat the partial measurement of virtual occupation of the dot in the cotunneling regime. In our scheme, we treat the interaction between quantum dot (QD) and QPC exactly by utilizing a rate equation formalism,

which is perturbative in the tunneling and exact in the regime of an almost pinched-off QPC. We illuminate and study three mechanisms of back-action that a detector induces onto coherent transport, namely (i) increased phase space for QPC-assisted transport, e.g., inelastic processes in which an electron enters the dot with energy ϵ and leaves the dot with energy $\epsilon' \neq \epsilon$ while another electron crosses the QPC such that the total energy is conserved, (ii) reduced elastic transport (decoherence), and (iii) widening of the dot energy levels. Thus, we can address some of the questions introduced above: we find that due to (i) and contrary to the prediction above, the measurement amplifies the current through the dot and does not block it. Nonetheless, mechanism (ii) implies that transport would be blocked if the dot is incorporated inside an interferometer.

Alongside these main results, we find that the detector signal shows nonmonotonous behavior as a function of the dot variables, which we attribute to sensitivity to the directionality of transport through the dot. Additionally, we determine the correlation between the QPC transport and the successful cotunneling passage through the quantum dot. Interestingly, despite the fact that the regime considered here is different from that of previous approaches which established such a relationship [7,20], we can use the correlated-currents signal to extract the time of the cotunneling process. The obtained time is compared with the simple Heisenberg's uncertainty expectation of $\tau_{\text{cot}} \sim \hbar/\Delta\mathcal{E}$, which is set by the inverse of the "energy-debt" during the virtual transport.

Our results are relevant for contemporary transport experiments [26,29], where different mechanisms of QPC back-action are discussed and an estimation of the cotunneling time is obtained from the linewidth of the differential conductance through the dot. Note that, differently from other approaches dealing with the many-body physics in the cotunneling system and detector [20], our formalism addresses the regime where single cotunneling events are correlated with single electron signals through the QPC.

The paper is structured as follows: in Sec. II, we write down the model of a dot measured by a QPC. In Sec. III, we derive the essential tools and methods for describing the measurement of cotunneling in its wake. Section IV details the resultant interplay between dot and QPC, and its implications on measurable currents on correlations, as well as the ability of measuring the cotunneling time. In Sec. V, we conclude and discuss possible future directions. We provide a comprehensive Appendix A, which details the microscopic calculation required to establish the formalism used in Sec. III.

II. MODEL

Our setup is divided into two components: the system, and the detector that measures it,

$$H = H_{\text{sys}} + H_{\text{det}} + H_{\text{int}}, \quad (1)$$

where H_{sys} describes the system, H_{det} describes the detector, and H_{int} describes the interaction between them. For simplicity, we consider a spinless problem. The results can be directly extended in some regimes to the spinful case. The system consists of a single level quantum dot, that is tunnel-coupled

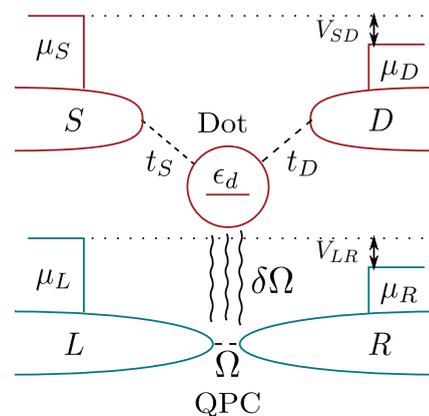


FIG. 1. (Color online) A sketch of the setup: a single-level dot is tunnel-coupled to two leads, S and D with amplitudes t_S and t_D , respectively. The respective dot-leads chemical potentials are μ_S and μ_D , corresponding to a voltage bias $eV_{SD} = \mu_S - \mu_D$. A quantum point contact (QPC) is capacitively coupled to the dot, i.e., transport between the leads L and R is governed by a tunneling amplitude Ω when the dot is empty, and $\Omega - \delta\Omega$ when the dot is occupied. The respective QPC-leads chemical potentials are μ_L and μ_R , corresponding to a voltage bias $eV_{LR} = \mu_L - \mu_R$.

to two electronic leads, the source S , and the drain D (see Fig. 1). By applying a voltage bias, $eV_{SD} = \mu_S - \mu_D$, between the source and the drain chemical potentials (μ_S , μ_D), we can measure the transport properties of the dot. We assume, henceforth, that $eV_{SD} \geq 0$. The quantum dot is capacitively coupled to another lead, the gate lead. Varying the gate voltage V_g on the gate lead controls the number of electrons in the dot. The Hamiltonian that describes the system is

$$H_{\text{sys}} = H_{SD} + H_{\text{dot}} + H_T, \quad (2)$$

where

$$H_{SD} = \sum_{k,\alpha=S,D} \epsilon_{k,\alpha} c_{k,\alpha}^\dagger c_{k,\alpha}, \quad (3)$$

$$H_{\text{dot}} = \epsilon_d d^\dagger d, \quad (4)$$

$$H_T = \sum_{k,\alpha=S,D} t_\alpha c_{k,\alpha}^\dagger d + \text{H.c.} \quad (5)$$

The operator $c_{k,\alpha}$ annihilates an electron with momentum k and energy $\epsilon_{k,\alpha}$ in the lead $\alpha \in \{S, D\}$, d annihilates an electron on the dot with energy ϵ_d , which is modulated by V_g . We have assumed that the tunneling coefficients t_α between the lead α and the dot are independent of the energy.

As long as $\mu_S > \epsilon_d > \mu_D$, transport occurs via sequential tunneling processes through the dot. When the dot level is outside the energy window provided by the leads, $\mu_D < \mu_S < \epsilon_d$ or $\mu_S > \mu_D > \epsilon_d$, and at low temperature, $T \ll \min\{|\mu_D - \epsilon_d|, |\epsilon_d - \mu_S|\}$, sequential tunneling is exponentially suppressed. Nonetheless, a small current is still detected in the drain. This small current is carried by the so-called cotunneling processes, in which electrons from the source virtually tunnel through the dot into the drain [22,34]. By virtually we mean that the tunneling into the dot is classically forbidden by energy conservation, but the overall cotunneling

process is energy conserving. In the present work we focus on this cotunneling regime.

We now turn to describe the detector. The detector is chosen to be a QPC because of its noninvasive nature [10]. The QPC consists of two leads, left and right, that are tunnel-coupled to each other; see Fig. 1. It is described by the following Hamiltonian:

$$H_{\text{det}} = \sum_l \mathcal{E}_l a_l^\dagger a_l + \sum_r \mathcal{E}_r a_r^\dagger a_r + \sum_{l,r} \Omega (a_l^\dagger a_r + \text{H.c.}), \quad (6)$$

where a_l (a_r) annihilates an electron on the left (right) lead with momentum l (r) and energy \mathcal{E}_l (\mathcal{E}_r). The tunneling amplitude between the left and right leads, Ω , is assumed to be energy independent. The QPC is out-of-equilibrium, namely, the chemical potentials of the left and right leads are different, $\mu_L = eV_{LR}/2$, $\mu_R = -eV_{LR}/2$.

The QPC is capacitively coupled to the quantum dot, i.e., the tunneling amplitude Ω is modulated by the charge on the dot,

$$H_{\text{int}} = - \sum_{l,r} \delta\Omega d^\dagger d (a_l^\dagger a_r + \text{H.c.}). \quad (7)$$

The typical time scales for a tunneling event in the QPC to occur are \mathcal{D}^{-1} and $\tilde{\mathcal{D}}^{-1}$ for an empty and an occupied dot, respectively [32]. They are given by $\mathcal{D} = 2\pi\Omega^2\rho_L\rho_R eV_{LR}/\hbar$ and $\tilde{\mathcal{D}} = 2\pi\tilde{\Omega}^2\rho_L\rho_R eV_{LR}/\hbar$ where $\tilde{\Omega} \equiv \Omega - \delta\Omega$. For simplicity, henceforth, we assume the same density of states in the left and right leads, $\rho_L = \rho_R \equiv \rho$.

Note that a single-level quantum dot model corresponds to a quantum dot with appreciable level spacing, larger than the other energy scales in the problem. Such a system can be experimentally realized in semiconductors. Similarly, the QPC detector can be easily tuned to detect single electron tunneling in the sequential limit [35], and to weak coupling in the cotunneling regime [29,30]. Furthermore, this model provides a valid effective description of multilevel quantum dots in the limit where the Thouless energy is the largest energy scale [21].

III. METHODS

The currents through the dot and the QPC are carried by tunneling processes between the source and the drain via the dot, and from the left lead to the right lead in the QPC. In order to obtain these currents and their cross correlations, we use a rate equation formalism [36–38] and calculate the rates for these tunneling processes (see Appendix A). We assume that all the tunnel couplings are weak, namely, $\rho_\alpha |t_\alpha|^2 \ll eV_{SD}$ and $\rho|\Omega|^2 \ll eV_{LR}$, and treat them perturbatively. Within this approach the interaction term, that makes the QPC transport dependent on the dot occupancy, is treated exactly.

We focus on cotunneling rates, in which the occupation of the dot changes only virtually. Cotunneling can occur either between two different leads (e.g., source to drain) or back and forth between a lead and the dot (e.g., source to source). During these processes electrons can tunnel through the QPC. Therefore, the QPC current is sensitive to the virtual changes in the dot occupancy. We denote by $W_{\alpha\alpha'}^n$ the cotunneling rate from lead α to lead α' (where $\alpha, \alpha' \in \{S, D\}$), during which n electrons pass through the QPC.

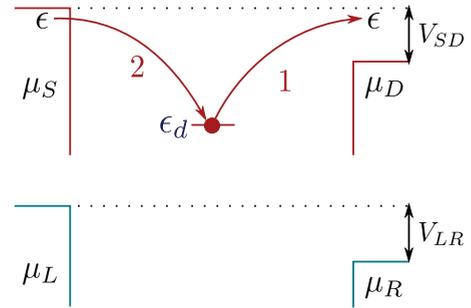


FIG. 2. (Color online) Sketch of the rate W_{SD}^0 of cotunneling through the dot accompanied by no tunneling through the QPC [cf. Eq. (8)]. Similar to Fig. 1, the upper part (brown) depicts the dot system and the lower (blue) the QPC.

Toy model

To study the interplay between the QPC and the dot we introduce a simplified version of our model. In this simplified toy model, we wish to have, at most, one electron that tunnels through the QPC during a cotunneling process through the dot, namely, $W_{\alpha\alpha'}^n$ is negligible for $n > 1$. Thus, only five rates remain relevant in the analysis of the currents and cross correlations: W_{SD}^0 , W_{SD}^1 , W_{DS}^1 , W_{SS}^1 , and W_{DD}^1 [see Figs. 2 and 3(a)–3(d), respectively].

The rate W_{SD}^0 describes processes in which an electron cotunnels through the dot while no tunneling events through the QPC take place. Therefore, this rate contributes only to the current through the dot. The rates W_{SD}^1 and W_{DS}^1 describe processes in which, in addition to cotunneling through the dot, an electron tunnels across the QPC. They contribute both to the current through the dot, and to the current through the QPC. Hence, they generate cross correlations between these currents. The rates W_{SS}^1 and W_{DD}^1 are of processes that contribute to the current through the QPC, but not to the current through the dot. Here, the electron tunnels back and forth between the dot and the same lead.

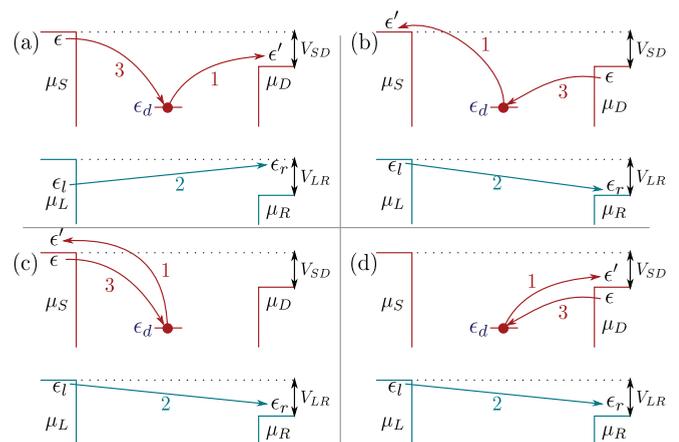


FIG. 3. (Color online) Sketches of the relevant cotunneling rates in our toy model that involve an electron tunneling through the QPC [cf. Eq. (9)]. Similar to Fig. 1, the upper part (brown) depicts the dot system and the lower (blue) the QPC. (a) The rate W_{SD}^1 . (b) The rate W_{DS}^1 . (c) The rate W_{SS}^1 . (d) The rate W_{DD}^1 .

In order to work in the regime of our toy model, we take the following assumptions: First, we work in the zero-temperature limit and assume that no tunneling events occur in the QPC when the dot is occupied, i.e., we set $\tilde{\Omega} = 0$. As a result, transport through the QPC occurs only alongside cotunneling events, i.e., only during the time interval in which the dot is virtually empty, τ_{cot} . We further take the limit where τ_{cot} is much shorter than the typical tunneling time in the QPC, i.e., $D\tau_{\text{cot}} \ll 1$.

The cotunneling time, τ_{cot} , is related by Heisenberg's uncertainty principle to the difference between the energy of the initial (or final) state, and the energy of the virtual interim state, e.g., $\epsilon - \epsilon_d$ and $\epsilon' - \epsilon_d$ in Fig. 3. Of all cotunneling processes, the process with the longest cotunneling time is the one with the lowest energy difference, namely, for $\epsilon' = \mu_D$ (assuming $\mu_S > \mu_D$). Hence, the assumption above implies that $\hbar D \ll \mu_S - \epsilon_d, \mu_D - \epsilon_d$.

Note that this chosen working point of the QPC puts the detector in a regime different from the one typically considered in the context of weak measurements. In a typical weak measurement one assumes to work at finite D and $\tilde{D} = D - \delta D$. The measurement process is then characterized by a measurement time, $\tau_M \sim D/\delta D^2$, needed to resolve the occupancy of the dot. One can then work in a regime with $\tau_{\text{cot}} D \gg 1$ and $\tau_{\text{cot}} \delta D \ll 1$, such that the detector's bandwidth is larger than the cotunneling time we are interested in ($D \gg 1/\tau_{\text{cot}}$), but the measurement during such a time remains weak ($\tau_{\text{cot}} \ll \tau_M$).

In our case $\delta D = D$, and as $\tau_{\text{cot}} D \ll 1$ we do not have a large bandwidth in the detector. Nonetheless, due to the very same choice, $\tilde{D} = 0$, corresponding to a blocked transport through the QPC, the current in the QPC is strongly correlated with a cotunneling through the dot, and one can expect that virtual transport through the dot could be resolved. In other words, the point $\tilde{D} = 0$ is peculiar since, as soon as an electron tunnels through the QPC, one can infer with certainty the happening of a cotunneling event through the quantum dot. This is not a weak measurement in the standard sense, but resembles more a partial-collapse measurement [18,39–42]. Nonetheless, the rarity of such joint tunneling events generates limited back-action, and the detector can be used as a non-destructive detector of the virtual occupation of the dot. In any case, this parameters' regime has the advantage of simplifying the formalism, and rendering the processes involved in the detection clearer [43].

Experimentally, the small bandwidth of the QPC should not serve as a hurdle, as there are sufficiently many cotunneling events in the dot to obtain a measurable signal. We envision that tuning in to this measurement regime corresponds to starting from a conductance plateau in the QPC in the presence of a full dot that is decoupled from leads, i.e., no virtual charge fluctuation on the dot occurs. Increasing the coupling of the dot to its leads allows for cotunneling. If at the same time a QPC current is generated, it should correspond to the mechanism reported here.

1. Derivation of the rates

In Appendix A 2, we rigorously derive the rates $W_{SD}^0, W_{SD}^1, W_{DS}^1, W_{SS}^1$, and W_{DD}^1 . Here, we present a simpler, and intuitive way, to calculate these rates.

We start by incorporating one of the effects of the coupling with the QPC by bestowing a finite width to the energy level of the dot, i.e., the single level gains an effective width $\hbar(D - \tilde{D})/2$, which is proportional to the measurement strength. Note that in our toy model the measurement strength is, therefore, equivalent to $\hbar D/2$. Hence, the cotunneling rate is simply the standard cotunneling rate through a single-level dot with energy ϵ_d and width $\hbar D/2$. This is indeed the effect obtained by the microscopic calculation, which includes the dynamics of the QPC in Appendix A 2.

The total cotunneling rate W_{SD}^0 , is the sum over the rates of all possible cotunneling processes (namely, integration over all incoming energies):

$$\begin{aligned} W_{SD}^0 &= \frac{2\pi}{\hbar} \int_{\mu_D}^{\mu_S} d\epsilon \rho_S \rho_D \left| \frac{t_S t_D}{\epsilon - \epsilon_d - i\hbar D/2} \right|^2 \\ &= \frac{\Gamma_S \Gamma_D}{\pi D} \left[\tan^{-1} \left(\frac{2(\mu_S - \epsilon_d)}{\hbar D} \right) - \tan^{-1} \left(\frac{2(\mu_D - \epsilon_d)}{\hbar D} \right) \right], \end{aligned} \quad (8)$$

where $\Gamma_\alpha \equiv 2\pi \rho_\alpha |t_\alpha|^2 / \hbar$. For later use, we also denote the total coupling $\Gamma = \Gamma_S + \Gamma_D$.

The cotunneling events through the dot, which are accompanied by a tunneling event through the QPC, consist of three stages (see Fig. 3): First, the electron in the dot tunnels out into the source or the drain. Second, a tunneling event occurs in the QPC. Last, a new electron tunnels into the empty dot. Here too, we include an effective width, $\hbar D/2$ to the energy of the dot. The total tunneling rate is given by

$$\begin{aligned} W_{\alpha\alpha'}^1 &= \frac{2\pi}{\hbar} \int_{-\infty}^{\mu_\alpha} d\epsilon \int_{\mu_{\alpha'}}^{\infty} d\epsilon' \int_{-\infty}^{\mu_L} d\epsilon_l \int_{\mu_R}^{\infty} d\epsilon_r \rho_S \rho_D \rho_L \rho_R \left| \frac{t_\alpha t_{\alpha'} \Omega}{(\epsilon' - \epsilon_d - i\hbar D/2)(\epsilon_r + \epsilon' - \epsilon_l - \epsilon_d - i\hbar D/2)} \right|^2 \delta(\epsilon + \epsilon_l - \epsilon' - \epsilon_r) \\ &= \frac{\hbar^2 \Gamma_\alpha \Gamma_{\alpha'} D}{4\pi^2 e V_{LR}} \int_{\mu_{\alpha'} - eV_{LR}}^{\mu_\alpha} d\epsilon \int_{\mu_{\alpha'}}^{\epsilon + eV_{LR}} d\epsilon' \frac{(eV_{LR} + \epsilon - \epsilon')}{[(\epsilon - \epsilon_d)^2 + \frac{\hbar^2 D^2}{4}][(\epsilon' - \epsilon_d)^2 + \frac{\hbar^2 D^2}{4}]}. \end{aligned} \quad (9)$$

The integrals in Eq. (9) are solvable. Out of space consideration, we choose to present these rates in integral form. Here we notice the full extent of the interaction of the QPC with the dot; the QPC affects the cotunneling through the dot in two fashions: First, as previously noted, it gives a finite

width to the energy level of the dot. Second, it allows new QPC-assisted cotunneling processes through the dot, where electrons are emitted from lead α *inelastically* into lead α' , provided that the energy difference is compensated by the QPC.

When the dot and the QPC are decoupled ($\Omega = 0$) all the cotunneling processes are elastic, namely, an electron at energy ϵ in the lead α , can be emitted into lead α' only at the same energy, ϵ . In our case, the dot and the QPC are coupled ($\Omega \neq 0$), and the two systems can exchange energy. Therefore, an electron with energy ϵ in lead α can end up in lead α' having a different energy $\epsilon' \neq \epsilon$. This is enabled by an electron with energy ϵ_l which is transmitted from the left lead of the QPC into the right lead with energy $\epsilon_r \neq \epsilon_l$ keeping the total energy conserved $\epsilon + \epsilon_l = \epsilon' + \epsilon_r$. Examples of inelastic cotunneling processes are schematically depicted in Fig. 3. The total cotunneling rate $W_{\alpha\alpha'}^1$, is the sum over the rates of all possible cotunneling processes (namely, integration over all energies in the relevant leads of the dot and QPC systems).

At zero temperature, the maximal energy that a QPC electron can lose by tunneling is $\mu_L - \mu_R = eV_{LR}$ (assuming that $\mu_L > \mu_R$). Hence, the maximal energy that a dot electron can gain is eV_{LR} , namely, $\epsilon - \epsilon' \leq eV_{LR}$ [see Figs. 3(b)–3(d)]. Similarly, the maximal energy that the dot electron can lose is eV_{SD} [see Fig. 3(a)]. Increasing the applied bias voltages (both on the dot and on the QPC) therefore increases the phase space for QPC-assisted cotunneling processes. Hence, the total rate $W_{\alpha\alpha'}^1$ increases with the applied bias voltages.

In Fig. 4, we plot the rates W_{SD}^0 and W_{SD}^1 as a function of $\hbar\mathcal{D}/(\mu_D - \epsilon_d)$ for different values of $\mu_D - \epsilon_d$ and at fixed V_{SD} , V_{LR} . In this configuration, $\mu_D - \epsilon_d$ affects all possible cotunneling rates. Lowering ϵ_d suppresses the rate of cotunneling processes, leading also to shorter cotunneling times τ_{cot} . This is seen in Fig. 4 by the overall decrease in magnitude of both W_{SD}^0 and W_{SD}^1 as ϵ_d is lowered. As \mathcal{D} increases the probability for an electron to tunnel through the QPC during the time window τ_{cot} grows. Hence, W_{SD}^0 constantly decreases with \mathcal{D} while W_{SD}^1 increases.

At a certain point, W_{SD}^1 peaks and decreases as well. At this point the probability for two electrons to tunnel through the QPC during a cotunneling event becomes relevant. We highlight the regime in which our toy model is valid, i.e., the regime in which W_{SD}^0 slightly decreases while W_{SD}^1 increases but is still far from peaking. Note, however, that even beyond the highlighted regime, in the weak QPC-tunnel-coupling regime the rates W_{SD}^n with $n > 1$ appear as contributions of $O(\Omega^{2n})$.

We have assumed throughout this section an additional finite width $\hbar\mathcal{D}/2$ of the dot energy level. We stress that this is a result of the formal derivation of Eqs. (8) and (9) that appears in Appendix A 2.

2. Currents and correlations

Using the rates, Eqs. (8) and (9), we are able to express the currents and cross correlations of our setup [37,38]. The average currents of the dot and the QPC are given by

$$\langle I_{dot} \rangle = e(W_{SD}^0 + W_{SD}^1 - W_{DS}^0 - W_{DS}^1), \quad (10)$$

$$\langle I_{QPC} \rangle = e(W_{SD}^1 + W_{DS}^1 + W_{SS}^1 + W_{DD}^1). \quad (11)$$

The zero-frequency cross correlation between these currents,

$$S = 2 \int_{-\infty}^{\infty} d\tau [\langle I_{dot}(t + \tau) I_{QPC}(t) \rangle - \langle I_{dot} \rangle \langle I_{QPC} \rangle], \quad (12)$$

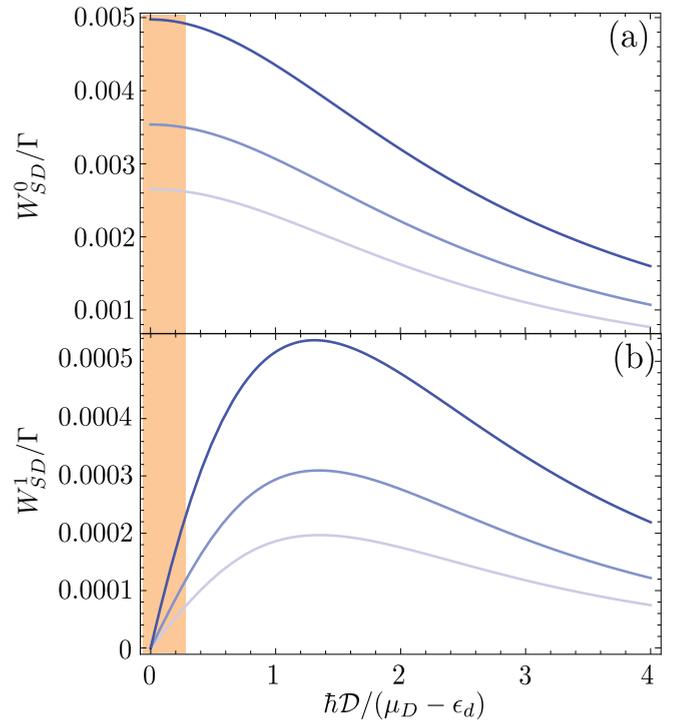


FIG. 4. (Color online) Plots of (a) W_{SD}^0 and (b) W_{SD}^1 as a function of $\hbar\mathcal{D}/(\mu_D - \epsilon_d)$ with bias voltages $eV_{SD} = 4\hbar\Gamma$, $eV_{LR} = 3.9\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. The different curves are for different dot gating ϵ_d such that $\mu_D - \epsilon_d = 4\hbar\Gamma$ (dark blue); $\mu_D - \epsilon_d = 5\hbar\Gamma$ (medium-dark blue); $\mu_D - \epsilon_d = 6\hbar\Gamma$ (bright blue). We highlight the area in which our toy model is relevant, namely, where W_{SD}^0 slightly decreases, W_{SD}^1 is far from peaking, and W_{SD}^n for $n > 1$ are negligible. Two effects are seen here: (i) The higher the dot energy is, the higher the cotunneling rates are; (ii) as $\hbar\mathcal{D}/(\mu_D - \epsilon_d)$ increases it is more probable for electrons to pass through the QPC during the time window in which the dot is empty. This is seen in the decrease in W_{SD}^0 vs the increase in W_{SD}^1 . At the point in which W_{SD}^1 peaks, the probability for two electrons to tunnel through the QPC during a cotunneling event becomes relevant.

is given by

$$S = 2e^2(W_{SD}^1 - W_{DS}^1). \quad (13)$$

IV. RESULTS

In this section, we analyze the resulting currents through the dot and the QPC, and their cross correlations. Also, we relate these quantities to a conditional partial measurement of the occupation of the dot, akin to null weak values [18]. This quantity enables the determination of τ_{cot} .

A. Current through the dot; measurement back-action

The current through the dot is carried by cotunneling processes, in which the occupation of the dot is virtually changed for a short time. The intermediate evolution state, where the occupation of the dot is changed, does not preserve energy. Thus, ideally a strong measurement of the charge on the dot would destroy cotunneling processes. Given that the current through the QPC is affected by the charge on the dot, it can be used to measure this charge. Naively, one would expect

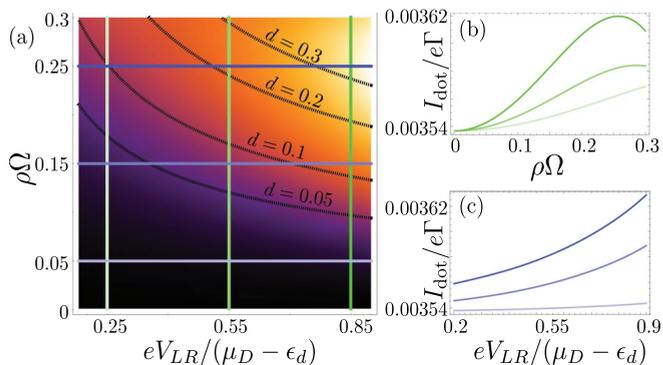


FIG. 5. (Color online) Plots of I_{dot} [cf. Eq. (10)] as a function of $eV_{LR}/(\mu_D - \epsilon_d)$ and $\rho\Omega$ with $eV_{SD} = 4\hbar\Gamma$, $\mu_D - \epsilon_d = 5\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. (a) A density plot of I_{dot} . The I_{dot} along the vertical (green) and horizontal (blue) mesh lines is plotted in (b) and (c), respectively. The dashed lines represent equal- $d \equiv \hbar\mathcal{D}/(\mu_D - \epsilon_d)$ lines. We can see that the current increases with the coupling \mathcal{D} of the QPC. However, the effect does not depend exclusively on \mathcal{D} , but it also directly depends on the QPC parameters: the tunnel-coupling Ω and the voltage-bias V_{LR} .

that turning on a weak coupling between the QPC and the dot will slightly reduce the current through the dot. It turns out that the opposite is true. As seen in Fig. 5, by increasing the coupling between the dot and the QPC (increasing Ω in our toy model), the current through the dot is even slightly *enhanced*.

Figure 5 depicts the current through the dot as a function of the coupling to the QPC, Ω , and as a function of the QPC voltage bias V_{LR} . The current is enhanced both by increasing Ω and by increasing V_{LR} , or alternatively, by increasing \mathcal{D} , and V_{LR} . Note that the regime in which our model is valid is $d \equiv \hbar\mathcal{D}/(\mu_D - \epsilon_d) \lesssim 0.3$, where up to one electron is transferred in the QPC during the virtual cotunneling time of the dot. Therefore, the area in which I_{dot} peaks and decreases is outside the scope of our model's validity.

The current through the dot has two parts, a ‘‘coherent’’ part and ‘‘incoherent’’ part. The coherent part is carried by coherent cotunneling processes, where electrons tunnel through the dot without changing the state of the QPC, the total rate of these processes is W_{SD}^0 . The incoherent part is carried by cotunneling processes through the dot that are accompanied by changes in the state of the QPC, namely, tunneling through the QPC. The rates for these processes are W_{SD}^1 and W_{DS}^1 . As discussed in Sec. III 1, increasing Ω or V_{LR} , and correspondingly \mathcal{D} , increases the probability for tunneling through the QPC, and hence decreases W_{SD}^0 and increases W_{SD}^1 . Additionally, V_{LR} increases also the phase space for inelastic QPC-assisted cotunneling processes through the dot. Hence V_{LR} increases W_{SD}^1 (and W_{DS}^1) even for a fixed \mathcal{D} . Hence, the growth of I_{dot} in Fig. 5 is a result of the availability of phase space for inelastic QPC-assisted cotunneling processes in W_{SD}^1 being higher than the decrease in W_{SD}^0 as a function of these parameters.

The strong measurement nature of the incoherent channel is highlighted by the fact that, in our simplified toy model, having current through the QPC destroys the coherence of a cotunneling process. If the dot would have been embedded into the arm of a Mach-Zehnder interferometer a current in the QPC could serve as a *strong* which-path measurement and would

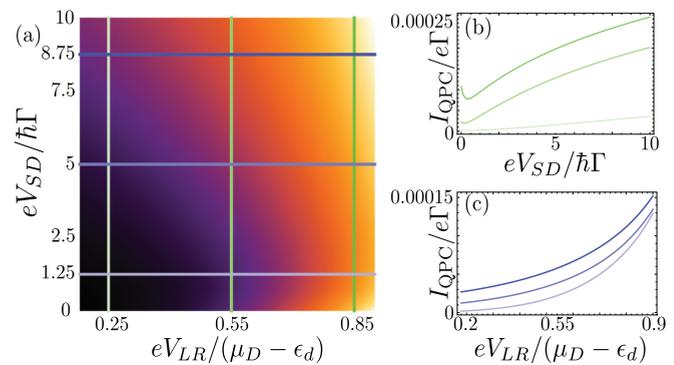


FIG. 6. (Color online) Plots of I_{QPC} [cf. Eq. (11)] as a function of $eV_{LR}/(\mu_D - \epsilon_d)$ and $eV_{SD}/(\hbar\Gamma)$ with $\rho\Omega = 0.15$, $\mu_D - \epsilon_d = 5\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. (a) A density plot of I_{QPC} . The I_{QPC} along the vertical (green) and horizontal (blue) mesh lines is plotted in (b) and (c), respectively. Naturally, the current increases with the the QPC voltage bias V_{LR} . As the voltage on the dot V_{SD} is increased the following occurs: (i) The cotunneling to the drain and back W_{DD}^1 remains constant; (ii) the cotunneling to the source and back W_{SS}^1 becomes less probable; (iii) the cotunneling from drain to source W_{DS}^1 decreases, and disappears once $V_{SD} > V_{LR}$; (iv) additional cotunneling processes from source to drain in W_{SD}^1 are allowed. The latter features dominate the behavior of I_{QPC} . We see initially a slight decrease as W_{DS}^1 vanishes with an overall increase due to W_{SD}^1 .

reduce the interference signal. In this case, the interference signal would be proportional to the decreasing W_{SD}^0 .

B. Current through the QPC; measurement signal

The current through the QPC is sensitive to the virtual changes in the charge on the dot during cotunneling processes. In the discussed toy model, QPC-current pulses (electron tunneling) occur only alongside cotunneling processes in the dot. The cotunneling processes can be either between two leads (from the source to the drain and vice versa), or back and forth between the dot and one of the leads (source-to-source or drain-to-drain). We find, here, a nonmonotonous signal that is dependent on the directionality of the dot transport.

In Fig. 6, we plot I_{QPC} as a function of the two bias voltages: V_{SD} on the dot, and V_{LR} on the QPC. Increasing V_{LR} has two effects: First, it increases the phase space for elastic QPC-tunneling processes, where $\epsilon_l = \epsilon_r$ (see Fig. 3 for clarification of the notations). Second, it increases the phase space for inelastic processes with $\epsilon_l > \epsilon_r$. While the first effect is relevant only for source-to-drain tunneling through the dot, the second effect is relevant for all four possibilities of cotunneling through the dot (from any lead α to any lead α'). The amount of energy that can be lost in the QPC as a compensation for inelastic cotunneling through the dot is bound from above by eV_{LR} . Hence, in cotunneling processes that involve a single lead, raising V_{LR} allows for more energetic particles-hole excitations to appear in the lead [see Figs. 3(c) and 3(d)]. Similarly, drain-to-source cotunneling consumes energy and therefore increasing V_{LR} enlarges the availability of drain-to-source cotunneling [see Fig. 3(b)]. To conclude, the rates $W_{\alpha\alpha'}^1$ (all four combinations) increase with V_{LR} and hence I_{QPC} increases with V_{LR} .

Raising the bias voltage on the dot, V_{SD} , increases the phase space for both elastic and inelastic source-to-drain tunneling, and hence, it increases the rate W_{SD}^1 . The rate W_{DS}^1 however, is suppressed by a raised V_{SD} as the drain-to-source cotunneling processes require higher amount of energy for larger V_{SD} . For a fixed V_{LR} , the available energies for inelastic drain-to-source cotunneling through the dot are bounded from above. Hence, increasing V_{SD} reduces the phase space for such processes, where for $V_{SD} > V_{LR}$ such cotunneling becomes impossible.

The increase of V_{SD} has an additional effect: for a fixed μ_D it reduces the rate W_{SS}^1 . The probability for a cotunneling process to occur depends on the energy difference between the dot and the available energies in the lead. Increasing $\mu_S - \epsilon_d$ increases the difference between the energy of the electron in the dot and the available energies in the source [see Fig. 3(c)], making the source-to-source cotunneling processes less probable. Since μ_D is fixed, the drain-to-drain tunneling processes are not affected by V_{SD} and hence, for a fixed μ_D , W_{DD}^1 is independent of V_{SD} .

To conclude, W_{SD}^1 grows with V_{SD} , while W_{DS}^1 and W_{SS}^1 are reduced by its increase, and W_{DD}^1 is not affected by it. As a result, I_{QPC} generally grows with V_{SD} since the dominant processes are QPC-tunneling processes accompanied by source-to-drain cotunneling through the dot. Yet, for a small bias voltage on the dot, $V_{SD} \ll \mu_D - \epsilon_d$, the four rates W_{SD}^1 , W_{DS}^1 , W_{SS}^1 , and W_{DD}^1 are roughly of the same magnitude and are determined mostly by V_{LR} . Raising V_{SD} in this configuration causes an increase in W_{SD}^1 alongside a decrease in W_{DS}^1 and W_{SS}^1 , causing the current through the QPC to slightly decrease. I_{QPC} starts to increase when V_{SD} is further increased. This can be shown in Fig. 6(b), where for a relatively large V_{LR} , a small decrease in I_{QPC} appears as V_{SD} grows, before the overall current starts to increase.

C. Cross-current correlation; focusing on source-drain processes

We finally arrive to describe the sensing of virtual change in the charge on the dot during cotunneling processes from source to drain. Here, we are interested in the cross-current correlation between the dot and the QPC. This correlation eliminates processes that contribute to the QPC-current from cotunneling processes to a specific lead and back, as they do not generate a current through the dot.

The correlation function S in Eq. (12) is plotted in Fig. 7, as a function of the system's parameters, V_{SD} and $\mu_D - \epsilon_d$. Its behavior stems from the same effects considered above for the currents, I_{dot} and I_{QPC} . The correlation is suppressed at $V_{SD} \ll V_{LR}$ because of counterpropagating QPC-assisted processes. Once the drain-to-source processes (W_{DS}^1) are suppressed for $V_{SD} \geq V_{LR}$, the correlation increases monotonically with V_{SD} because of the increased phase space for cotunneling events. Also, quite intuitively, the deeper the dot's energy level is, the smaller is the probability for a cotunneling event to occur, and thus the smaller S becomes. As to the effects of the detector on the correlation, that we do not plot, a larger coupling to the QPC or a larger V_{LR} leads to a stronger current across the detector, which directly reflects in a larger signal in the correlation.

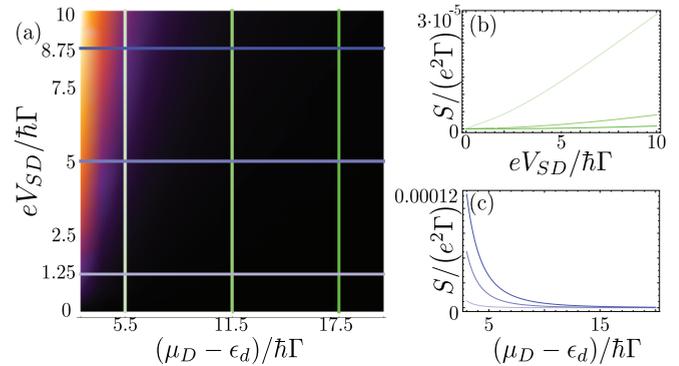


FIG. 7. (Color online) Plots of S [cf. Eqs. (12) and (13)] as a function of $eV_{SD}/(\hbar\Gamma)$ and $(\mu_D - \epsilon_d)/(\hbar\Gamma)$ with $\rho\Omega = 0.15$, $eV_{LR} = 1.5\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. (a) A density plot of S . The S along the vertical (green) and horizontal (blue) mesh lines is plotted in (b) and (c), respectively. As the voltage on the dot V_{SD} is increased the following occurs: (i) The cotunneling from drain to source W_{DS}^1 decreases, and disappears once $V_{SD} > V_{LR}$; and (ii) additional cotunneling processes from source to drain in W_{SD}^1 are allowed. Hence S initially increases with V_{SD} slowly, and only afterward increases as W_{SD}^1 . The deeper the dot level is, the shorter the virtual time in which the dot is empty. As a result, (i) cotunneling processes through the dot become less probable, and (ii) it becomes less probable for a QPC electron to manage to tunnel during this time window, leading to the descent of S .

D. Weak values; cotunneling time

The current through the QPC is generated while the dot is virtually empty, i.e., $I_{QPC} = \sum_i I_{QPC}^{(i)}$ is generated by processes i for which the dot is virtually empty for time $\tau_{\text{cot}}^{(i)}$ with $I_{QPC}^{(i)} \sim e\mathcal{D}\tau_{\text{cot}}^{(i)}$.

Since the correlation function S isolates the contributions to the QPC current arising from cotunneling between the two distinct leads, it encodes information on the physical properties of these cotunneling processes, i.e., one can extract the time the dot is virtually empty, when restricting to current-generating cotunneling events. This can be defined via a weak value procedure [9], where the detector's signal is postselected by retaining only processes of successful cotunneling current from source to drain. In this framework the cotunneling current through the dot plays the role of a postselection operator [20]. This leads to the average over cotunneling times of current-generating events [20] [cf. Eq. (12)],

$$\begin{aligned} \bar{\tau}_{SD} &= \frac{\int_{-\infty}^{\infty} ds [\langle I_{QD}(t)I_{QPC}(t-s) \rangle - \langle I_{QD} \rangle \langle I_{QPC} \rangle]}{e\mathcal{D}\langle I_{QD} \rangle} \\ &= \frac{S}{2e\mathcal{D}\langle I_{QD} \rangle}. \end{aligned} \quad (14)$$

The obtained $\bar{\tau}_{SD}$ is plotted in Fig. 8 as a function of eV_{SD} and $(\mu - \epsilon_d)/(\hbar\Gamma)$.

We note that such a quantity generally depends on the detector's parameters as well. However, the effects of the detector are minimal in the weak measurement regime, which, as discussed in Sec. III, corresponds to $\hbar\mathcal{D} \ll eV_{LR} \ll \mu_D - \epsilon_d$. This allows us to define an intrinsic (dot-dependent only) cotunneling time when $V_{LR} \rightarrow 0$. The evaluation of $\bar{\tau}_{SD}$ that

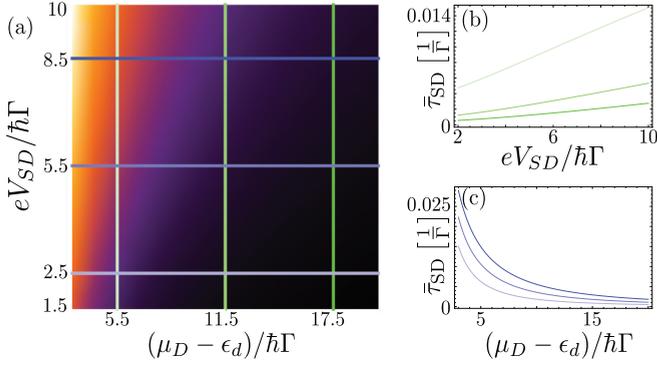


FIG. 8. (Color online) Plots of $\bar{\tau}_{SD}$ [cf. Eq. (14)] as a function of $eV_{SD}/(\hbar\Gamma)$ and $(\mu_D - \epsilon_d)/(\hbar\Gamma)$ with $\rho\Omega = 0.15$, $eV_{LR} = 1.5\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. (a) A density plot of $\bar{\tau}_{SD}$. The $\bar{\tau}_{SD}$ along the vertical (green) and horizontal (blue) mesh lines is plotted in (b) and (c), respectively. Interestingly, as V_{SD} increases, $\bar{\tau}_{SD}$ increases as well, i.e., it appears that the increased phase space adds processes with a slower time into the average. As expected, the deeper the dot level is, the shorter $\bar{\tau}_{SD}$ becomes.

is reported in Fig. 8 is obtained within such a regime, and its values do not depend on the detector parameters much.

In Fig. 8, we see that $\bar{\tau}_{SD}$ increases with V_{SD} , and becomes shorter the lower the dot-level is. The decay of the cotunneling time as a function of $\mu_D - \epsilon_d$ can be intuitively understood in terms of the energy-time uncertainty principle: at equilibrium, the virtual hole that is excited in the dot sets $\hbar/(\mu_D - \epsilon_d)$ as the typical time scale for the occupation of the virtual state.

Discussion

We wish to make a more quantitative evaluation of our obtained cotunneling time [cf. $\bar{\tau}_{SD}$ in Eq. (14) and Fig. 8]. Due to the finite bias voltage on the dot, in Eq. (14) we have a weighted average over the cotunneling times of different cotunneling processes. Within a simple comparative model, this can be taken into account by averaging the cotunneling times predicted by Heisenberg's uncertainty principle, $\tau_h = \hbar/(\epsilon_h - \epsilon_d)$, over all cotunneling processes, each characterized by a different virtual energy, $\epsilon_h - \epsilon_d$, where $\mu_D < \epsilon_h < \mu_S$. Each cotunneling process occurs with probability $P(\epsilon') = 2\pi\rho_S\rho_D|t_{SD}|^2/(\epsilon' - \epsilon_d - i\hbar D/2)^2$ [cf. Eq. (8)]. The averaged cotunneling time is then

$$\bar{\tau}_h = \frac{\int_{\mu_D}^{\mu_S} P(\epsilon')\tau_h(\epsilon')}{\int_{\mu_D}^{\mu_S} P(\epsilon')} = \frac{1}{\hbar D} \frac{\ln[(\hbar D)^2 + 4(\mu_S - \epsilon_d)^2] - \ln[(\hbar D)^2 + 4(\mu_D - \epsilon_d)^2] + 2\ln[\mu_D - \epsilon_d] - 2\ln[\mu_S - \epsilon_d]}{\arctan[2(\mu_S - \epsilon_d)/(\hbar D)] + \arctan[2(\mu_D - \epsilon_d)/(\hbar D)]}. \quad (15)$$

The resulting $\bar{\tau}_h$ is plotted in Fig. 9, where it is compared to $\bar{\tau}_{SD}$. We see the following main differences: (i) there is, approximately, an overall order of magnitude between the two times, which can be attributed to an unknown prefactor in the used $\tau_h(\epsilon')$ from the energy-time uncertainty principle; (ii) whereas $\bar{\tau}_h$ decreases almost as $1/(\mu_D - \epsilon_d)$ (it exhibits almost a linear decrease in a LogLog plot) with a slight dependence on V_{SD} , the $\bar{\tau}_{SD}$ deviates from such a $1/(\mu_D - \epsilon_d)$ descent and shows a stronger dependence on V_{SD} ; (iii) the dependence on V_{SD} is opposite in the two cases. With increasing V_{SD} , in $\bar{\tau}_h$ additional faster cotunneling processes are added into the average and it becomes shorter, whereas in $\bar{\tau}_{SD}$ the average time becomes longer. The latter two differences demonstrate the inherent discrepancy between our simplistic model for $\bar{\tau}_h$, that takes into account only the finite width of the dot level, versus the physical model of $\bar{\tau}_{SD}$ that incorporates also an increased back-action from the detector by adding phase space for slower QPC-assisted processes.

V. CONCLUSIONS AND OUTLOOK

We have presented a model for a direct detection of electron cotunneling through a single-level quantum dot. The detector has been modeled by a charge sensing QPC that is capacitively coupled to the dot. In the regime where the transport through the QPC is fully blocked when the dot is occupied, we obtain a simplified model that allows us to incorporate a microscopic description of the QPC into the rate equation formalism of cotunneling through the quantum dot. We have, thus, determined the current through the QPC (detector's signal),

the cotunneling current in the dot (including the detector's back-action), as well as their correlations (related to weak values and the cotunneling time).

We found that the detector's back-action consists of three different mechanisms: (i) a broadening of the dot energy levels, (ii) a suppression of elastic (coherent) charge-transfer processes, and (iii) an increase of phase space due to QPC-assisted transport. In particular, the latter mechanism is responsible for an increase of the cotunneling current upon increasing the QPC voltage bias. The QPC current generally increases also as a function of the dot voltage bias corresponding to the increased probability of cotunneling processes. However, a counterintuitive *decrease* of I_{QPC} with the voltage bias across the dot is predicted at small voltage biases. In this regime, the dominant effect of the increase of voltage bias is a suppression of the source-to-source, drain-to-drain, and drain-to-source cotunneling processes, which in turn result in a suppressed current through the QPC. Importantly, the current-current correlations allow us, via a weak value based approach, to access the cotunneling time. We find that the cotunneling time obtained from such a direct measurement can be compared to the time estimated from the energy-time uncertainty principle, after taking into account proper averaging of all possible cotunneling events.

Let us, finally, comment on the validity and applicability of our simplified model. Our model essentially relies on the assumptions that (i) transport through the QPC is fully blocked for an occupied dot, and that (ii) only processes of single electron transfer through the QPC during a cotunneling event are considered. The former assumption is used in order to obtain time-independent cotunneling rates in the presence

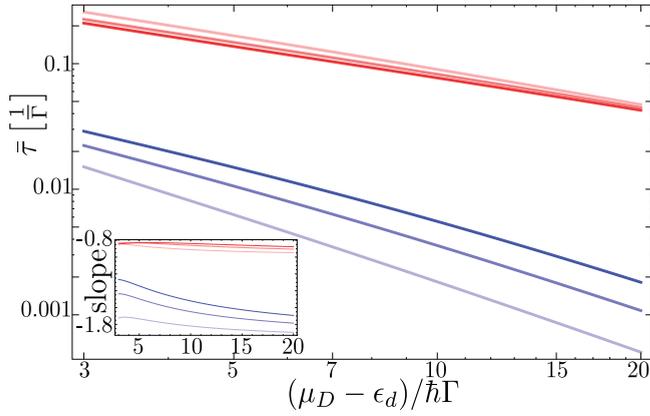


FIG. 9. (Color online) LogLog plots of $\bar{\tau}_{SD}$ and $\bar{\tau}$ [cf. Eqs. (14) and (15)] as a function of $(\mu_D - \epsilon_d)/(\hbar\Gamma)$ with $\rho\Omega = 0.15$, $eV_{LR} = 1.5\hbar\Gamma$, and $\Gamma_S = \Gamma_D$. The plots of $\bar{\tau}_{SD}$ have an overall shorter time (blue) than those of $\bar{\tau}$ (red). For both plotted times the curves are for $V_{SD} = 2.5\hbar\Gamma$ (light), $V_{SD} = 5.5\hbar\Gamma$ (medium-light), and $V_{SD} = 8.5\hbar\Gamma$ (dark). In the inset, we plot the approximate slopes of the LogLog curves, defined as $(\partial \ln \bar{\tau})/[\partial \ln(\mu_D - \epsilon_d)]$. The overall magnitude difference between the two time models can be attributed to an unknown prefactor in the time taken from the energy-time uncertainty principle. The other differences can be attributed to an inherent discrepancy between the two models, as the former incorporates phase-space contributions of slower QPC-assisted processes.

of the QPC. Introducing a Markovian thermalization time in the QPC leads or cutting the infinite time integration in the cotunneling rates by sequential tunneling rates should cure this constraint. The latter condition appears to be quite strong: in order to detect a cotunneling event of duration $\tau_{cot} \sim \hbar/(\mu_D - \epsilon_d)$, the detector is expected to have a large bandwidth $\gtrsim \hbar/\tau_{cot}$, corresponding to a large number of charge-transfer events in the QPC per single cotunneling event [20]. In our model we consider the opposite regime. Nonetheless, the detector is equally sensitive to the short-lived virtual cotunneling states, at the working point defined by (i), since cotunneling processes are the only possible mechanism of activating a current signal in the QPC. This implies that any weak signal in the detector corresponds to obtaining full information on the happening of cotunneling, which makes our measurement of a partial-collapse type (i.e., a strong measurement that happens with a small probability). Hence, unlike in a weak measurement case, the measured cotunneling time is affected by back-action.

Even in its simplified form, it is important to stress that our simplified theoretical model provides a valid description for experiments at the pinched-off working point of the QPC. In this working point, the only approximation is the assumption that the relevant rates are those corresponding to single electron transfer across the QPC. Though this is the key simplification, it is also a realistic physical approximation. What our approximation neglects are the coherences between subsequent electrons tunneling across the QPC, which decay very fast due to the relaxation processes of electrons in the bulk leads. Hence, contributions from a higher number of transported QPC electrons per cotunneling process are small

(cf. Fig. 4). In other words, the virtual cotunneling time is so short that the assumption $\mathcal{D}\tau_{cot} \ll 1$ is physically sensible.

Spinful electrons. In order to highlight the interplay between measurement and coherent transport, we have assumed that the electrons are spinless. Whereas spin physics do not play an important role in QPC transport, in the dot system the Kondo effect [44] could qualitatively change our predictions. Our treatment is immediately applicable for polarized electrons, e.g., in the presence of a strong magnetic field, which could be realized in experiments. Moreover, also in the case of complete spin degeneracy, where Kondo physics takes place, our approach remains valid as long as the Kondo temperature $T_k \sim \sqrt{\rho_\alpha |t_\alpha|^2 U} e^{-U/(2\rho_\alpha |t_\alpha|^2)}$, with U the charging energy on the dot, is smaller than the different energy scales of the system, e.g., $T_k \ll T \ll eV_{SD}$. In this regime, the system is not sensitive to the Kondo physics and our treatment is essentially correct. Note that this specific limit can be achieved by making the tunneling between leads and dot arbitrarily weak. Importantly, extending our analysis to the Kondo regime is extremely interesting, as it features an interplay between transport through a truly many-body correlated virtual state and weak measurement.

Since our model can describe realistic configurations [23–31], it becomes interesting to generalize it to include experimentally relevant effects, e.g., finite temperature, many-level dots, spin physics, and to extend the present results to the more general regime of weak measurement.

ACKNOWLEDGMENTS

We would like to thank S. A. Gurvitz, Y. Oreg, T. Ihn, C. Rössler, K. Ensslin, and D. Zümbuhl for useful discussions. We acknowledge the support from DFG, 4710/1-1, and Swiss National Science Foundation (SNSF).

APPENDIX: DERIVATION OF THE COTUNNELING RATES IN THE PRESENCE OF COUPLED QPC

We present here a rigorous model for the calculation of the cotunneling rates appearing in Eqs. (8) and (9). We start by presenting a general model for the analysis of a QPC current in Appendix A 1. Using this formalism we derive the rates in Appendix A 2.

1. QPC analysis

We present here the microscopic derivation of the tunneling probability amplitudes describing electron transport in the QPC. This derivation is closely related to the one presented in Ref. [32].

The QPC is out-of-equilibrium, namely, the chemical potentials of the left and right leads are different, $\mu_L = eV_{LR}/2$, $\mu_R = -eV_{LR}/2$. Hence, the electron creation and annihilation operators are defined with respect to a vacuum state $|0; e\rangle$ in which all the levels in the left (right) lead are initially filled up to the Fermi energy μ_L (μ_R). We assumed here that the dot is initially empty, denoted by e . Therefore, the vacuum state $|0; e\rangle$ decays exponentially to states with transmitted electrons from the left to the right leads (e.g., $a_r^\dagger a_l |0; e\rangle$, with an electron in the right lead and a hole in the

left lead, $a_r^\dagger a_{r'}^\dagger a_l a_l' |0; e\rangle$ with two electrons in the right lead and two holes in the left lead, etc.).

In the absence of interaction with the dot, the many-body wave function that describes the electron dynamics in the QPC can be written in the following form:

$$|\Psi(t); e\rangle = \left[b_0(t) + \sum_{l_1, r_1} b_{l_1 r_1}(t) a_{r_1}^\dagger a_{l_1} + \sum_{l_1 < l_2, r_1 < r_2} b_{l_1 l_2 r_1 r_2}(t) a_{r_1}^\dagger a_{r_2}^\dagger a_{l_1} a_{l_2} + \dots \right] |0; e\rangle, \quad (\text{A1})$$

where $b_{\mathbf{l}, \mathbf{r}_i}(t)$ with $\mathbf{l}, \mathbf{r}_i = l_1, \dots, l_i, r_1, \dots, r_i$, are the time-dependent probability amplitudes to find the system in a state with i electrons transmitted from left to right. The indices l_i, r_i label the momenta of the transmitted electrons. The initial condition is usually chosen to be $b_{\mathbf{l}, \mathbf{r}_i}(0) = \delta_{i0}$, and the probabilities are summed to 1,

$$\sum_i \sum_{\mathbf{l}, \mathbf{r}_i} |b_{\mathbf{l}, \mathbf{r}_i}(t)|^2 = 1. \quad (\text{A2})$$

The QPC is capacitively coupled to the quantum dot, i.e., the tunneling amplitude Ω is modulated by the charge on the dot. Hence, the wave function of the electrons in the QPC is also modulated by the charge on the dot.

We denote the vacuum state (of the QPC) in the presence of an occupied dot with $|0; f\rangle$. Hence, the many-body wave

function of Eq. (A1) becomes

$$|\tilde{\Psi}(t); f\rangle = \left[\tilde{b}_0(t) + \sum_{l_1, r_1} \tilde{b}_{l_1 r_1}(t) a_{r_1}^\dagger a_{l_1} + \sum_{l_1 < l_2, r_1 < r_2} \tilde{b}_{l_1 l_2 r_1 r_2}(t) a_{r_1}^\dagger a_{r_2}^\dagger a_{l_1} a_{l_2} + \dots \right] |0; f\rangle, \quad (\text{A3})$$

where $\tilde{b}_{\mathbf{l}, \mathbf{r}_i}(t)$ are the time-dependent probability amplitudes in the presence of an occupied dot. These amplitudes satisfy the same normalization condition as the amplitudes $b_{\mathbf{l}, \mathbf{r}_i}(t)$.

The explicit expressions for $b_{\mathbf{l}, \mathbf{r}_i}(t)$ [$\tilde{b}_{\mathbf{l}, \mathbf{r}_i}(t)$] are obtained by substituting Eq. (A1) [Eq. (A3)] into the Schrödinger equation $i\hbar|\dot{\Psi}(t)\rangle = H_{\text{QPC}}|\Psi(t)\rangle$ and choosing appropriate boundary conditions. Note that the time evolution of Eq. (A1) is performed by $H_{\text{QPC}} = H_{\text{det}}$, and the evolution in time of Eq. (A3) is performed by $H_{\text{QPC}} = H_{\text{det}} + H_{\text{int}}$. Therefore, the form of the coefficients $\tilde{b}(t)$ is the same as that of $b(t)$ with Ω replaced by $\tilde{\Omega} \equiv \Omega + \delta\Omega$. Thus, the QPC current can be used to measure the charge of the dot. The strength of the measurement is encoded in the tunneling amplitudes, Ω and $\tilde{\Omega}$. In the following it suffices to present the formalism for the case of an empty dot.

We substitute Eq. (A1) into the Schrödinger equation. Performing the Laplace transform

$$B(\mathcal{E}) = \frac{1}{\hbar} \int_0^\infty e^{i(\mathcal{E}t/\hbar)} b(t) dt, \quad (\text{A4})$$

yields coupled equations for $B(\mathcal{E})$:

$$\mathcal{E} B_0(\mathcal{E}) - \sum_{l_1 r_1} \Omega B_{l_1 r_1}(\mathcal{E}) = i b_0(0), \quad (\text{A5})$$

$$(\mathcal{E} + \mathcal{E}_{l_1} - \mathcal{E}_{r_1}) B_{l_1 r_1}(\mathcal{E}) - \Omega B_0(\mathcal{E}) - \sum_{l_2 r_2} \Omega B_{l_1 l_2 r_1 r_2}(\mathcal{E}) = i b_{l_1 r_1}(0), \quad (\text{A6})$$

$$(\mathcal{E} + \mathcal{E}_{l_1} + \mathcal{E}_{l_2} - \mathcal{E}_{r_1} - \mathcal{E}_{r_2}) B_{l_1 l_2 r_1 r_2}(\mathcal{E}) - \Omega B_{l_1 r_2}(\mathcal{E}) + \Omega B_{l_2 r_2}(\mathcal{E}) - \sum_{l_3 r_3} \Omega B_{l_1 l_2 l_3 r_1 r_2 r_3}(\mathcal{E}) = i b_{l_1 l_2 r_1 r_2}(0), \dots, \quad (\text{A7})$$

where $b_{\mathbf{l}, \mathbf{r}_i}(0)$ are the initial conditions.

Now, a recursive replacement of each of the amplitudes B in the sum, $\sum \Omega B$, by its expression from the subsequent equation, e.g., plugging $B_{l_1 r_1}(\mathcal{E})$ from Eq. (A6) into Eq. (A5), yields

$$\left[\mathcal{E} - \sum_{l_1 r_1} \frac{\Omega^2}{\mathcal{E} + \mathcal{E}_{l_1} - \mathcal{E}_{r_1}} \right] B_0(\mathcal{E}) - \sum_{l_1 l_2 r_1 r_2} \frac{\Omega^2}{\mathcal{E} + \mathcal{E}_{l_1} - \mathcal{E}_{r_1}} B_{l_1 l_2 r_1 r_2}(\mathcal{E}) = i b_0(0) + i\Omega \sum_{l_1 r_1} \frac{b_{l_1 r_1}(0)}{\mathcal{E} + \mathcal{E}_{l_1} - \mathcal{E}_{r_1}}. \quad (\text{A8})$$

We replace sums by integrals, $\sum_{l, r} \rightarrow \int \rho_L(\mathcal{E}_l) \rho_R(\mathcal{E}_r) d\mathcal{E}_l d\mathcal{E}_r$, where ρ_L and ρ_R are the density of states in the left and right leads, respectively. Consequently, the first sum in Eq. (A8) has two contributions: (i) a sum over the singular part that yields $i\pi\Omega^2 \rho_L \rho_R V_{LR}$, and (ii) as sum over the principal value that usually cancels out (unless the pole is close to the Fermi energy which causes a rescaling of the energy levels). Usually, the

second sum in Eq. (A8) can be neglected. Indeed, the probability amplitudes $B(\mathcal{E})$ take the form of Green's functions. Hence, replacing $B_{l_1 l_2 r_1 r_2}(\mathcal{E}) \equiv B(\mathcal{E}, \mathcal{E}_{l_1}, \mathcal{E}_{l_2}, \mathcal{E}_{r_1}, \mathcal{E}_{r_2}) \sim 1/(\mathcal{E} + \mathcal{E}_{l_1} + \mathcal{E}_{l_2} - \mathcal{E}_{r_1} - \mathcal{E}_{r_2})$, the corresponding integral vanishes, provided that $V_{LR} \gg \Omega^2 \rho$. This is true up to corrections due to the initial conditions. In fact, $B(\mathcal{E}, \mathcal{E}_{l_1}, \mathcal{E}_{l_2}, \mathcal{E}_{r_1}, \mathcal{E}_{r_2}) \sim 1/(\mathcal{E} + \mathcal{E}_{l_1} + \mathcal{E}_{l_2} - \mathcal{E}_{r_1} - \mathcal{E}_{r_2}) + b_{l_1 l_2 r_1 r_2}(0)$. If there is a nonvanishing initial condition on $b_{l_1 l_2 r_1 r_2}(0) \neq 0$, we replace the amplitude $B_{l_1 l_2 r_1 r_2}(\mathcal{E})$ by its expression obtained from the subsequent equation. The integral remains vanishing but we take into account the possible initial conditions that may appear in the other coupled equations.

Repeating this procedure for the other equations gives

$$(\mathcal{E} + i\hbar D/2) B_0(\mathcal{E}) = i \sum_{j=0} \left[\Omega^j \sum_{\mathbf{l}, \mathbf{r}_j} \frac{b_{\mathbf{l}, \mathbf{r}_j}(0)}{\prod_{k=1}^j [\mathcal{E} + \sum_{m=1}^k (\mathcal{E}_{l_m} - \mathcal{E}_{r_m})]} \right], \quad (\text{A9})$$

$$(\mathcal{E} + \mathcal{E}_{l_1} - \mathcal{E}_{r_1} + i\hbar\mathcal{D}/2)B_{l_1r_1}(\mathcal{E}) - \Omega B_0(\mathcal{E}) = i \sum_{j=1} \left[\Omega^{j-1} \sum_{\mathbf{l}_j\mathbf{r}_j} \frac{b_{\mathbf{l}_j\mathbf{r}_j}(0)\delta_{l_1\mathbf{l}_j}\delta_{r_1\mathbf{r}_j}}{\prod_{k=2}^j [\mathcal{E} + \sum_{m=1}^k (\mathcal{E}_{l_m} - \mathcal{E}_{r_m})]} \right], \quad (\text{A10})$$

$$(\mathcal{E} + \mathcal{E}_{l_1} + \mathcal{E}_{l_2} - \mathcal{E}_{r_1} - \mathcal{E}_{r_2} + i\hbar\mathcal{D}/2)B_{l_1l_2r_1r_2}(\mathcal{E}) - \Omega B_{l_1r_1}(\mathcal{E}) + \Omega B_{l_2r_2}(\mathcal{E}) = i \sum_{j=2} \left[\Omega^{j-2} \sum_{\mathbf{l}_j\mathbf{r}_j} \frac{b_{\mathbf{l}_j\mathbf{r}_j}(0)\delta_{l_1\mathbf{l}_j}\delta_{r_1\mathbf{r}_j}\delta_{l_2\mathbf{l}_j^2}\delta_{r_2\mathbf{r}_j^2}}{\prod_{k=3}^j [\mathcal{E} + \sum_{m=1}^k (\mathcal{E}_{l_m} - \mathcal{E}_{r_m})]} \right], \quad (\text{A11})$$

...

where $\mathcal{D} = 2\pi\Omega^2\rho_L\rho_R eV_{LR}/\hbar$.

Performing the inverse Laplace transform

$$b(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\frac{\mathcal{E}t}{\hbar}} B(\mathcal{E})d\mathcal{E}, \quad (\text{A12})$$

we can incrementally obtain the time-dependent probability amplitudes. The solutions depend on the initial conditions. Let us first show the solution taking the standard initial condition $b_{\mathbf{l}_i\mathbf{r}_i}(0) = \delta_{i0}$

$$b_0(t) = e^{-(\mathcal{D}/2)t}, \quad (\text{A13})$$

$$b_{l_1r_1}(t) = \frac{\Omega}{\mathcal{E}_{l_1} - \mathcal{E}_{r_1}} e^{-(\mathcal{D}/2)t} [1 - e^{i[(\mathcal{E}_{l_1} - \mathcal{E}_{r_1})t/\hbar]}], \dots \quad (\text{A14})$$

Taking, for example, the initial conditions $b_{\mathbf{l}_i\mathbf{r}_i}(0) = \delta_{i1}\delta_{\mathbf{l}_i\mathbf{l}_i^0}\delta_{\mathbf{r}_i\mathbf{r}_i^0}$, we obtain

$$b_{0|l_1^0r_1^0}(t) = \frac{\Omega}{\mathcal{E}_{l_1^0} - \mathcal{E}_{r_1^0} - i\frac{\mathcal{D}}{2}} e^{-(\mathcal{D}/2)t} [1 - e^{i[(\mathcal{E}_{l_1^0} - \mathcal{E}_{r_1^0} - i(\mathcal{D}/\epsilon))t/\hbar]}], \quad (\text{A15})$$

$$b_{l_1r_1|l_1^0r_1^0}(t) = e^{-(\mathcal{D}/2)t} e^{i[(\mathcal{E}_{l_1} - \mathcal{E}_{r_1})t/\hbar]} \delta_{l_1l_1^0}\delta_{r_1r_1^0} + O(\Omega^2), \dots \quad (\text{A16})$$

We shall see below that these solutions are relevant for the calculation of the rates in the toy model regime.

2. Derivation of the tunneling rates

We use a rate equation formalism in order to describe the transport through the setup of a quantum dot that is coupled to a QPC. The input for the rate equation is tunneling rates. In this Appendix, we provide a full derivation of these rates up to fourth order in the tunneling elements of the dot. We work in the interaction picture with respect to H_T [see Eq. (5)]. As we have seen in the main text, since the QPC dynamics interplays with the charge on the dot [cf. Eqs. (A1) and (A3)], its time evolution is also affected by H_T .

To zeroth order in H_T , there are no tunneling events in the dot and the only dynamics are of the QPC. For example, starting from the initial state with 0 electrons that have tunneled through the QPC and empty dot, $|0; e; t=0\rangle$, the probability amplitude to end up at time t with i electrons that have tunneled through the QPC with momenta $\mathbf{l}_i, \mathbf{r}_i$ is

$$|\mathbf{l}_i, \mathbf{r}_i; \alpha; t|0; e; t=0\rangle^{(0)} = b_{\mathbf{l}_i, \mathbf{r}_i}(t)\delta_{e\alpha}, \quad (\text{A17})$$

where α is the charge on dot. Similarly,

$$|\mathbf{l}_i, \mathbf{r}_i; \alpha; t|0; f; t=0\rangle = \tilde{b}_{\mathbf{l}_i, \mathbf{r}_i}(t)\delta_{f\alpha}. \quad (\text{A18})$$

First order. To the lowest order in H_T , the transition rates of the dot can be calculated using Fermi's "golden rule." The probability amplitude for a sequential tunneling of an electron with energy ϵ from the source to the dot accompanied by tunneling through the QPC is

$$\begin{aligned} |\mathbf{l}_i, \mathbf{r}_i; f; t|0; e; t=0\rangle^{(1)} &= -\frac{i}{\hbar} \int_0^t dt' \langle \mathbf{l}_i, \mathbf{r}_i; f | e^{-i(H_0/\hbar)t'} e^{i(H_0/\hbar)t'} H_T e^{-i(H_0/\hbar)t'} |0; e\rangle \\ &= -\frac{i}{\hbar} t_S^* \int_0^t dt' \sum_j \int_{-\infty}^{\mu_L} d\mathbf{l}_j \rho_L^j \int_{\mu_R}^{\infty} d\mathbf{r}_j \rho_R^j e^{i[(\epsilon_d - \epsilon)t'/\hbar]} \tilde{b}_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_j, \mathbf{r}_j}(t-t') b_{\mathbf{l}_j, \mathbf{r}_j}(t'), \end{aligned} \quad (\text{A19})$$

where t_S, t_D are the tunneling coefficients, $b_{\mathbf{l}_j, \mathbf{r}_j}(t')$ is the probability amplitude that, up to time t' , j electrons with momenta \mathbf{l}_j were transmitted to momenta \mathbf{r}_j , and $\tilde{b}_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_j, \mathbf{r}_j}(t-t')$ is the time-dependent conditional probability amplitude to find the system at time t in a state with i electrons transmitted from left momenta \mathbf{l}_i to right momenta \mathbf{r}_i , given that the aforementioned j electrons were transmitted up to time t' . Similarly, tunneling from the dot to the lead reads

$$\begin{aligned} |\mathbf{l}_i, \mathbf{r}_i; e; t|0; f; t=0\rangle^{(1)} &= -\frac{i}{\hbar} \int_0^t dt' \langle \mathbf{l}_i, \mathbf{r}_i; e | e^{-i(H_0/\hbar)t'} e^{i(H_0/\hbar)t'} H_T e^{-i(H_0/\hbar)t'} |0; f\rangle \\ &= -\frac{i}{\hbar} t_D \int_0^t dt' \sum_j \int_{-\infty}^{\mu_L} d\mathbf{l}_j \rho_L^j \int_{\mu_R}^{\infty} d\mathbf{r}_j \rho_R^j e^{-i[(\epsilon_d - \epsilon)t'/\hbar]} b_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_j, \mathbf{r}_j}(t-t') \tilde{b}_{\mathbf{l}_j, \mathbf{r}_j}(t'). \end{aligned} \quad (\text{A20})$$

Second order. To next order in H_T , the electron can virtually tunnel through the dot. We write the probability amplitude for a cotunneling event where an electron with energy ϵ tunnels from the source to the dot and an electron tunnels from the dot to the

drain with energy ϵ' . During this process i electrons tunnel through the QPC,

$$\begin{aligned}
 \langle \mathbf{l}_i, \mathbf{r}_i ; e ; t | 0 ; e ; t = 0 \rangle^{(2)} &= \left(-\frac{i}{\hbar} \right)^2 \int_0^t dt' \int_0^{t'} dt'' \langle \mathbf{l}_i, \mathbf{r}_i ; e | e^{-i(H_0/\hbar)(t-t')} H_T e^{-i(H_0/\hbar)(t'-t'')} H_T e^{-i(H_0/\hbar)t''} | 0 ; e \rangle \\
 &= -\frac{t_S^* t_D}{\hbar^2} \int_0^t dt' \int_0^{t'} dt'' \sum_{j,k} \int_{-\infty}^{\mu_L} d\mathbf{l}_k \rho_L^k \int_{\mu_R}^{\infty} d\mathbf{r}_k \rho_R^k \int_{-\infty}^{\mu_L} d\mathbf{l}_j \rho_L^j \int_{\mu_R}^{\infty} d\mathbf{r}_j \rho_R^j \\
 &\quad \times e^{-i[(\epsilon_d - \epsilon')t'/\hbar]} e^{-i[(\epsilon_d - \epsilon'')t''/\hbar]} b_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_k, \mathbf{r}_k}(t - t') \tilde{b}_{\mathbf{l}_k, \mathbf{r}_k | \mathbf{l}_j, \mathbf{r}_j}(t' - t'') b_{\mathbf{l}_j, \mathbf{r}_j}(t''), \tag{A21}
 \end{aligned}$$

$b_{\mathbf{l}_j, \mathbf{r}_j}(t'')$ is the probability amplitude that, up to time t'' , j electrons with momenta \mathbf{l}_j were transmitted to momenta \mathbf{r}_j while the dot was empty, $\tilde{b}_{\mathbf{l}_k, \mathbf{r}_k | \mathbf{l}_j, \mathbf{r}_j}(t' - t'')$ is the time-dependent conditional probability amplitude to find the system at time t' in a state with k electrons transmitted from left momenta \mathbf{l}_k to right momenta \mathbf{r}_k after a time evolution from time t'' at a presence of a full dot, given that the aforementioned j electrons were transmitted up to time t'' , and $b_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_k, \mathbf{r}_k}(t - t')$ is the time-dependent conditional probability amplitude to find the system at time t in a state with i electrons transmitted from left momenta \mathbf{l}_i to right momenta \mathbf{r}_i after a time evolution from time t' at a presence of an empty dot, given that the aforementioned k electrons were transmitted up to time t' .

Similarly, the cotunneling event of an electron with energy ϵ to tunnel in and ϵ' tunnels out with an initially full dot is

$$\begin{aligned}
 \langle \mathbf{l}_i, \mathbf{r}_i ; f ; t | 0 ; f ; t = 0 \rangle^{(2)} &= \left(-\frac{i}{\hbar} \right)^2 \int_0^t dt' \int_0^{t'} dt'' \langle \mathbf{l}_i, \mathbf{r}_i ; f | e^{-i(H_0/\hbar)(t-t')} H_T e^{-i(H_0/\hbar)(t'-t'')} H_T e^{-i(H_0/\hbar)t''} | 0 ; f \rangle \\
 &= -\frac{t_S^* t_D}{\hbar^2} \int_0^t dt' \int_0^{t'} dt'' \sum_{j,k} \int_{-\infty}^{\mu_L} d\mathbf{l}_k \rho_L^k \int_{\mu_R}^{\infty} d\mathbf{r}_k \rho_R^k \int_{-\infty}^{\mu_L} d\mathbf{l}_j \rho_L^j \int_{\mu_R}^{\infty} d\mathbf{r}_j \rho_R^j \\
 &\quad \times e^{-i[(\epsilon - \epsilon_d)t'/\hbar]} e^{-i[(\epsilon_d - \epsilon'')t''/\hbar]} \tilde{b}_{\mathbf{l}_i, \mathbf{r}_i | \mathbf{l}_k, \mathbf{r}_k}(t - t') b_{\mathbf{l}_k, \mathbf{r}_k | \mathbf{l}_j, \mathbf{r}_j}(t' - t'') \tilde{b}_{\mathbf{l}_j, \mathbf{r}_j}(t''). \tag{A22}
 \end{aligned}$$

Let us evaluate Eq. (A22) for the case of the toy model described in the main text, i.e., for $\tilde{\Omega} = 0$. Notice that, in this toy model regime, no transport occurs in the QPC when the dot is full. As a result, in Eq. (A22), we can set $j = 0$ and $i = k$, leading to a cancellation of the internal energy integrals in the probability amplitude, which sum over higher-order coherent transport through the QPC.

We begin with considering the probability amplitude for a cotunneling through the dot in the presence of zero electrons passing through the QPC. Using Eq. (A13), Eq. (A22) in this case becomes

$$\begin{aligned}
 \langle 0 ; f ; t | 0 ; f ; t = 0 \rangle^{(2)} \\
 = -\frac{t_S^* t_D}{\hbar^2} \int_0^t dt' \int_0^{t'} dt'' e^{-i[(\epsilon - \epsilon_d)t'/\hbar]} e^{-i[(\epsilon_d - \epsilon'')t''/\hbar]} e^{-(\mathcal{D}/2)t} . \tag{A23}
 \end{aligned}$$

Performing the time integrals and taking the modulo square of the result yields a probability for this process to occur for equal source and drain energies ϵ (different ϵ and ϵ' have

zero probability). Taking the time derivative of this probability results in the integrand of the rate appearing in Eq. (8).

We now consider the probability amplitude for a cotunneling through the dot in the presence of one electron passing through the QPC. Using Eqs. (A13), (A14), and (A16), the expression becomes

$$\begin{aligned}
 \langle 0 ; f ; t | 0 ; f ; t = 0 \rangle^{(2)} \\
 = -\frac{t_S^* t_D}{\hbar^2} \int_0^t dt' \int_0^{t'} dt'' e^{-i[(\epsilon - \epsilon_d)t'/\hbar]} e^{-i[(\epsilon_d - \epsilon'')t''/\hbar]} \\
 \times e^{i[(\mathcal{E}_{l_1} - \mathcal{E}_{r_1})(t-t'')/\hbar]} \frac{\Omega}{\mathcal{E}_{l_1} - \mathcal{E}_{r_1}} \\
 \times e^{-(\mathcal{D}/2)(t-t'')} [1 - e^{i[(\mathcal{E}_{l_1} - \mathcal{E}_{r_1})(t-t'')/\hbar]}] . \tag{A24}
 \end{aligned}$$

Here, too, the evaluation of Eq. (A24) followed by modulo square of the result yields a probability for the process to occur. The obtain rate is, then, integrated upon in Eq. (9).

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
 [2] J. von Neumann, *Mathematical Foundations of Quantum Theory* (Princeton University Press, Princeton, NJ, 1983).
 [3] E. Condon and P. Morse, *Rev. Mod. Phys.* **3**, 43 (1931).
 [4] E. Wigner, *Phys. Rev.* **98**, 145 (1955).
 [5] M. Büttiker and R. Landauer, *Phys. Rev. Lett.* **49**, 1739 (1982).
 [6] D. Sokolovski and L. M. Baskin, *Phys. Rev. A* **36**, 4604 (1987).

[7] A. M. Steinberg, *Phys. Rev. Lett.* **74**, 2405 (1995).
 [8] A. A. Clerk, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, *Rev. Mod. Phys.* **82**, 1155 (2010).
 [9] Y. Aharonov, D. Z. Albert, and L. Vaidman, *Phys. Rev. Lett.* **60**, 1351 (1988).
 [10] A. N. Korotkov, *Phys. Rev. B* **63**, 115403 (2001).
 [11] A. N. Korotkov and D. V. Averin, *Phys. Rev. B* **64**, 165310 (2001).
 [12] O. Hosten and P. Kwiat, *Science* **319**, 787 (2008).

- [13] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, *Phys. Rev. Lett.* **102**, 173601 (2009).
- [14] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, *Phys. Rev. A* **80**, 041803 (2009).
- [15] N. Brunner and C. Simon, *Phys. Rev. Lett.* **105**, 010405 (2010).
- [16] D. J. Starling, P. B. Dixon, N. S. Williams, A. N. Jordan, and J. C. Howell, *Phys. Rev. A* **82**, 011802 (2010).
- [17] O. Zilberberg, A. Romito, and Y. Gefen, *Phys. Rev. Lett.* **106**, 080405 (2011).
- [18] O. Zilberberg, A. Romito, D. J. Starling, G. A. Howland, C. J. Broadbent, J. C. Howell, and Y. Gefen, *Phys. Rev. Lett.* **110**, 170405 (2013).
- [19] M. Büttiker, *Phys. Rev. B* **27**, 6178 (1983).
- [20] A. Romito and Y. Gefen, *Phys. Rev. B* **90**, 085417 (2014).
- [21] I. L. Aleiner, P. W. Brouwer, and L. I. Glazman, *Phys. Rep.* **358**, 309 (2002).
- [22] L. Glazman and M. Pustilnik, in *Nanophysics: Coherence and transport*, Lecture Notes of the Les Houches Summer School 2004 (Elsevier, Amsterdam, 2005).
- [23] M. Field, C. G. Smith, M. Pepper, D. A. Ritchie, J. E. F. Frost, G. A. C. Jones, and D. G. Hasko, *Phys. Rev. Lett.* **70**, 1311 (1993).
- [24] J. M. Elzerman, R. Hanson, J. S. Greidanus, L. H. Willems van Beveren, S. De Franceschi, L. M. K. Vandersypen, S. Tarucha, and L. P. Kouwenhoven, *Phys. Rev. B* **67**, 161308(R) (2003).
- [25] L. DiCarlo, H. J. Lynch, A. C. Johnson, L. I. Childress, K. Crockett, C. M. Marcus, M. P. Hanson, and A. C. Gossard, *Phys. Rev. Lett.* **92**, 226801 (2004).
- [26] D. Harbusch, D. Taubert, H. P. Tranitz, W. Wegscheider, and S. Ludwig, *Phys. Rev. Lett.* **104**, 196801 (2010).
- [27] S. Gasparinetti, M. J. Martinez-Prez, S. de Franceschi, J. P. Pekola, and F. Giazotto, *Appl. Phys. Lett.* **100**, 253502 (2012).
- [28] G. Granger, D. Taubert, C. E. Young, L. Gaudreau, A. Kam, S. A. Studenikin, P. Zawadzki, D. Harbusch, D. Schuh, W. Wegscheider, Z. R. Wasilewski, A. A. Clerk, S. Ludwig, and A. S. Sachrajda, *Nat. Phys.* **8**, 522 (2012).
- [29] B. Küng, C. Rössler, M. Beck, J. Faist, T. Ihn, and K. Ensslin, *New J. Phys.* **14**, 083003 (2012).
- [30] C. Rössler, T. Krähenmann, S. Baer, T. Ihn, K. Ensslin, C. Reichl, and W. Wegscheider, *New J. Phys.* **15**, 033011 (2013).
- [31] D. Maradan, L. Casparis, T.-M. Liu, D. E. F. Biesinger, C. P. Scheller, D. M. Zumbühl, J. D. Zimmerman, and A. C. Gossard, *J. Low Temp. Phys.* **175**, 784 (2014).
- [32] S. A. Gurvitz, *Phys. Rev. B* **56**, 15215 (1997).
- [33] A. Romito, Y. Gefen, and Y. M. Blanter, *Phys. Rev. Lett.* **100**, 056801 (2008).
- [34] D. V. Averin and Y. V. Nazarov, *Phys. Rev. Lett.* **65**, 2446 (1990).
- [35] E. V. Sukhorukov, A. N. Jordan, S. Gustavsson, R. Leturcq, T. Ihn, and K. Ensslin, *Nat. Phys.* **3**, 243 (2007).
- [36] J. J. Sakurai and S. F. Tuan, *Modern Quantum Mechanics* (Addison-Wesley, Reading, MA, 1985), Vol. 1.
- [37] A. N. Korotkov, *Phys. Rev. B* **49**, 10381 (1994).
- [38] J. Koch, F. von Oppen, and A. V. Andreev, *Phys. Rev. B* **74**, 205438 (2006).
- [39] N. Katz, M. Ansmann, R. C. Bialczak, E. Lucero, R. McDermott, M. Neeley, M. Steffen, E. M. Weig, A. N. Cleland, J. M. Martinis, and A. N. Korotkov, *Science (New York)* **312**, 1498 (2006).
- [40] L. P. Pryadko and A. N. Korotkov, *Phys. Rev. B* **76**, 100503 (2007).
- [41] N. Katz, M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, E. Lucero, A. O'Connell, H. Wang, A. N. Cleland, J. M. Martinis, and A. N. Korotkov, *Phys. Rev. Lett.* **101**, 200401 (2008).
- [42] O. Zilberberg, A. Romito, and Y. Gefen, *Phys. Scr.* **T151**, 014014 (2012).
- [43] In Sec. III 1, we see that for our case where $\delta\mathcal{D} = \mathcal{D}$, one of the back-action mechanisms of the QPC on the dot system is to give the dot level a finite width $\hbar\mathcal{D}$. By increasing \mathcal{D} one can give the detector a sufficiently large bandwidth to detect single cotunneling events, $\tau_{\text{cot}}\mathcal{D} \gg 1$, where $\tau_{\text{cot}} \sim \hbar/(\mu_D - \epsilon_d)$ from the Heisenberg's uncertainty principle. This implies that, as soon as \mathcal{D} is large enough to give a large bandwidth to the detector, the induced level broadening becomes comparable with the dot's energetic distance from the leads. Hence the QPC smears the dot level into an incoherent sequential transport regime through QPC-assisted transport.
- [44] A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge, 1997).