

## Powerful Coulomb-drag thermoelectric engine

A.-M. Daré\* and P. Lombardo

Aix Marseille Univ, Univ Toulon, CNRS, IM2NP UMR 7334, 13397 Marseille, France

(Received 13 April 2017; revised manuscript received 4 July 2017; published 8 September 2017)

We investigate a thermoelectric nanoengine whose properties are steered by Coulomb interaction. The device whose design decouples charge and energy currents is made up of two interacting quantum dots connected to three different reservoirs. We show that, by tailoring the tunnel couplings, this setup can be made very attractive for energy-harvesting prospects, due to a delivered power that can be of the order of the quantum bound [R. S. Whitney, *Phys. Rev. Lett.* **112**, 130601 (2014); *Entropy* **18**, 208 (2016)], with a concomitant fair efficiency. To unveil its properties beyond the sequential quantum master equation, we apply a nonequilibrium noncrossing approximation in the Keldysh Green's function formalism, and a quantum master equation that includes cotunneling processes. Both approaches are rather qualitatively similar in a large operating regime where sequential tunneling alone fails.

DOI: [10.1103/PhysRevB.96.115414](https://doi.org/10.1103/PhysRevB.96.115414)

### I. INTRODUCTION

A growing interest in nanodevices for energy harvesting from a temperature difference has expanded recently (see, e.g., [1–3]). In addition to the prevalent two-reservoir–one-dot device, a three-terminal thermoelectric (TTTE) setup, as shown in Fig. 1, emerged a short time ago [4]. One of its assets is its ability to decouple charge and heat currents. In this system, the working principle relies on Coulomb interaction, which triggers the output power. The interest in such systems is not only theoretical but also experimental [5–8]. Energy harvesting from voltage fluctuations in this type of device has also been established experimentally [9].

Other agents for energy harvesting in multiterminal systems have also been proposed [10] based on electron-phonon or electron-photon interactions; see, e.g., Ref. [11] and references therein. The present TTTE setup has been realized experimentally as an engine [6] and also as a system for thermal gating, that is, steering of a charge current by the temperature of a remote reservoir [7]. It has also been investigated theoretically in the refrigerator mode [12]. Theory for this system is not yet very advanced, the only approach being the  $T$ -matrix-based sequential quantum master equation (QME), which has its own range of validity—weak dot-lead coupling—and can account for the experimental results of Refs. [6,7]. It can also be supplemented with full-counting statistics to assess current fluctuations [13–15]. Incidentally, Coulomb drag [16] has received much attention recently in related circuits, for which the QME up to cotunneling processes has been investigated [17–19], however this was outside the thermoelectric context and in the absence of any temperature bias.

Our goal is to enlarge the parameter range in which the present setup was previously studied [2]. We apply, therefore, two radically different approaches to go beyond the sequential QME. The first one is an out-of-equilibrium generalization [20,21] of the noncrossing approximation (NCA) that enables the computation of nonlinear charge and energy currents. This technique has been previously used in a thermoelectric context in the case of a double-orbital correlated single dot connected to two terminals [22], for which it was shown that the Kondo

physics can be opportune to boost the device performance, ensuring a fair power output without eroding efficiency. However, we use the NCA here for rather high temperatures, for which the Kondo physics is not expected to occur. The second approach is a QME theory that includes nonlocal cotunneling processes. We show that the comparisons between the cotunneling QME and the NCA give satisfactory agreement, especially for charge-current stopping voltage, in some parameter range where sequential tunneling alone is failing.

The NCA allows us to show that, due to its topology, this TTTE device outperforms the properties of a two-terminal setup, delivering an output power that can be of the order of the quantum bound [23–25] concomitantly with a significant fraction of the Carnot efficiency.

The paper is organized as follows: after a presentation of the formalism and techniques (whose details are relegated to the Appendixes), the results are displayed, after which we conclude.

### II. THE MODEL

#### A. Hamiltonian and parameters

The dots—index  $t$  or  $b$  for top or bottom—are described by a nondegenerate orbital each, and they are coupled by a nonlocal Coulomb repulsion  $U$ ; this is depicted graphically in Fig. 1 by a capacitive coupling. The three reservoirs [top ( $t$ ), left bottom ( $L_b$ ), and right bottom ( $R_b$ ), respectively] are supposed to be noninteracting Fermi seas, with possibly their own chemical potentials and temperatures. The Hamiltonian describing the present device reads  $H = H_0 + H_T$ , where the disconnected part for dots and leads is, in obvious notations,

$$H_0 = \epsilon_t \hat{n}_t + \epsilon_b \hat{n}_b + U \hat{n}_t \hat{n}_b + \sum_{\alpha=t, L_b, R_b} H_{0\alpha}, \quad (1)$$

with  $H_{0\alpha} = \sum_k \epsilon_{k\alpha} \hat{n}_{k\alpha}$ . Hybridization between dots and leads reads

$$H_T = \sum_k (V_{kt} c_{kt}^\dagger d_t + \text{H.c.}) + \sum_{\beta=L_b, R_b} \sum_k (V_{k\beta} c_{k\beta}^\dagger d_b + \text{H.c.}) \quad (2)$$

The hybridization parameters will depend on the wave vector only through energy:  $V_{k\alpha} = V_\alpha(\epsilon_{k\alpha})$  [26]. They will enter the

\*Anne-Marie.Dare@univ-amu.fr

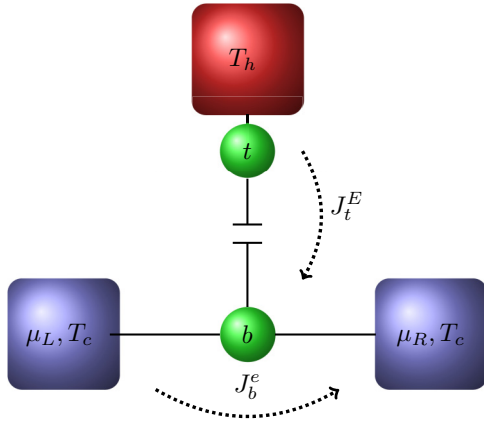


FIG. 1. Schematic representation of the device and of currents of interest: The top dot is connected to a hot reservoir and coupled by Coulomb interaction to a second dot (bottom). The latter is connected to two cold reservoirs through which a bias voltage can be applied. The energy current  $J_t^E$  flows from the top dot, while the electric current  $J_b^e$  flows through the bottom one.

present calculations through  $\Gamma_\alpha(\epsilon) = 2\pi\rho_\alpha(\epsilon)|V_\alpha(\epsilon)|^2$  with  $\rho_\alpha(\epsilon)$  the  $\alpha$ -lead density of states.

The three-terminal device performance is driven by charge fluctuations of the two dots [4]. The nondegenerate-orbital occupancy fluctuates, and for fermions the fluctuations are maximized at half-filling due to  $\langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle(1 - \langle n \rangle)$ . To promote this favorable situation, we choose henceforth for the two dots the following energies:  $\epsilon_t = \epsilon_b \equiv \epsilon_d = -U/2$ , which guarantees half-filled dots. Without loss of generality, the chemical potential of the top lead will be  $\mu_t = 0$ .  $V$ , the voltage bias across the bottom part of the device will be applied symmetrically, with  $\mu_L = -\frac{eV}{2}$  and  $\mu_R = \frac{eV}{2}$ . The temperature  $T_t = T_h$  of the hot-top reservoir will be supposed higher than the temperature of the cold lower ones:  $T_{L_b} = T_{R_b} = T_c$ . In the case of half-filling and symmetrically applied bias, we have the following equality between Fermi functions of the two cold reservoirs:  $f_{R_b}(\epsilon_d + U) = 1 - f_{L_b}(\epsilon_d)$ .

To be run as a thermoelectric engine [4], one needs nonuniform and dissymmetrical hybridizations between the  $b$ -dot and its reservoirs. We choose

$$\begin{aligned} \Gamma_t(\epsilon) &= \Gamma_t, \\ \Gamma_{R_b}(\epsilon) &= \Gamma_b \theta(\epsilon) \kappa(\epsilon), \\ \Gamma_{L_b}(\epsilon) &= \Gamma_{R_b}(-\epsilon), \end{aligned} \quad (3)$$

where the function  $\kappa(\epsilon)$  will be specified later.

### B. Currents in the Keldysh formalism and the NCA

In the Keldysh formalism, stationary charge [26] and energy currents flowing outside an  $\alpha$ -terminal into a  $d$ -dot can be expressed as

$$\begin{aligned} \begin{pmatrix} J_\alpha^e \\ J_\alpha^E \end{pmatrix} &= \frac{i}{\hbar} \int \frac{d\epsilon}{2\pi} \begin{pmatrix} e \\ 1 \end{pmatrix} \Gamma_\alpha(\epsilon) \\ &\times [f_\alpha(\epsilon)G_d^>(\epsilon) + [1 - f_\alpha(\epsilon)]G_d^<(\epsilon)], \end{aligned} \quad (4)$$

where  $G_d^<(\epsilon)$  are the lesser and greater  $d$ -dot Green's functions,  $f_\alpha(\epsilon)$  is the Fermi function of the  $\alpha$ -reservoir,  $e > 0$  is the

elementary charge, and hybridization parameters  $\Gamma_\alpha(\epsilon)$  were previously defined. In the preceding integrands, the two terms can be interpreted as a balance between in and out transfers; indeed, for fermions  $iG_d^>(\epsilon) \geq 0$ , whereas  $iG_d^<(\epsilon) \leq 0$  [27].

We define the electric current flowing through the bottom dot from its symmetric expression  $J_b^e \equiv J_{L_b}^e = -J_{R_b}^e = (J_{L_b}^e - J_{R_b}^e)/2$ , which leads to

$$\begin{aligned} J_b^e &= \frac{ie}{2\hbar} \int \frac{d\epsilon}{2\pi} [G_b^>(\epsilon)(\Gamma_{L_b}(\epsilon)f_{L_b}(\epsilon) - \Gamma_{R_b}(\epsilon)f_{R_b}(\epsilon)) \\ &+ G_b^<(\epsilon)(\Gamma_{L_b}(\epsilon)[1 - f_{L_b}(\epsilon)] - \Gamma_{R_b}(\epsilon)[1 - f_{R_b}(\epsilon)])]. \end{aligned} \quad (5)$$

In the thermoelectric engine regime, for properly chosen nonproportional hybridization functions  $\Gamma_{\beta b}(\epsilon)$ , the above current can be finite even if the Fermi functions  $f_{L_b}(\epsilon)$  and  $f_{R_b}(\epsilon)$  are equal, that is, in the absence of any bias applied to the lower part of the device. In the stationary regime, the electric current flowing into the top dot must vanish,

$$J_t^e = \frac{ie}{\hbar} \int \frac{d\epsilon}{2\pi} \Gamma_t(\epsilon) [f_t(\epsilon)G_t^>(\epsilon) + [1 - f_t(\epsilon)]G_t^<(\epsilon)] = 0, \quad (6)$$

while the energy current flowing from the top dot to the bottom one reads

$$J_t^E = \frac{i}{\hbar} \int \frac{d\epsilon}{2\pi} \epsilon \Gamma_t(\epsilon) [f_t(\epsilon)G_t^>(\epsilon) + [1 - f_t(\epsilon)]G_t^<(\epsilon)]. \quad (7)$$

The thermal engine efficiency is defined by the ratio

$$\eta = \frac{\mathcal{P}}{J_t^E} = \frac{J_b^e V}{J_t^E}, \quad (8)$$

with  $\mathcal{P}$  the output power.

As previously mentioned in Eq. (3),  $\Gamma_{L_b}(\epsilon)$  and  $\Gamma_{R_b}(\epsilon)$  are not proportional, hence one needs to evaluate two Green's functions for each dot. Indeed, the usual simplification [26] that allows to calculate only the difference  $i[G^>(\epsilon) - G^<(\epsilon)]$  no longer applies.

Equations (6) and (7) look very similar, but in the present study the first cancels while the second is expected to be the thermal energy supply. This constitutes a key constraint for relevant approaches to study the problem at hand. For example, the many-body Hubbard-I approximation [28,29] is too simple: It can be shown that within this method, the term in square brackets in Eq. (6) vanishes, and as a consequence the energy current [Eq. (7)] too, due to an inadequate treatment of fluctuations. The Ng method [30], which can handle Kondo physics [31], would also be inefficient to account for the present device properties.

This led us to develop a variant of the NCA for the present system. In this approach, the Green's functions of the two fermions residing, respectively, on the top and bottom dots are expressed in terms of fictitious particle Green's functions [32]. The approximation of the NCA essentially lies in the self-energy choice for these four fictitious particles, two fermions and two bosons. The details of these self-energies, as well as the two real-fermion Green's function expression, are given in Appendix A.

The NCA is a perturbative approach that presumes  $U > \Gamma$ ; however, contrary to the QME, there is no restriction on temperatures, except at very low temperature. We use the finite  $U$  version of the NCA, which has been shown, in its equilibrium version, to have some drawbacks [33], especially in the particle-hole symmetric case, which is the situation at hand. If so, the two fictitious fermions are degenerate, as well as the two fictitious bosons. As a result, the two real fermions, residing on the top and bottom dots, are described by the same Green's functions, notwithstanding the fact that the surroundings of the two dots may be quite different. We argue that this caveat is not prohibitive: first of all, this overstated symmetry exists also in the QME when including cotunneling processes, as proved even for  $\Gamma_t \neq \Gamma_b$ . Second, in the following we consider rather high temperatures, for which the problem is less acute.

As discussed in Refs. [33,34], the easier way to cure this symmetry drawback would be to include vertex corrections in the fictitious-particle self-energies, an approximation called the one crossing approximation (OCA). If it was undertaken at equilibrium [33], it would be a formidable task to adapt this idea in the present out-of-equilibrium case for many reasons: First, in the present calculations, the lesser and greater self-energies of the four fictitious particles are needed, and not only the retarded ones. Second, at the OCA level, the self-energies are products of five Green's functions. Hence it would be very cumbersome to apply the Langreth rules, which are needed for out-of-equilibrium Green's functions. Last but not least, in the NCA or the OCA, the interdependent Green's functions are obtained from self-consistent numerical evaluations. Accessing reliable results for eight coupled functions would probably be an elusive goal. Finally, although we are in a different configuration, with dissymmetrical lead couplings, the temperatures investigated in the present paper are rather high [35], higher than those below which a qualitative difference was found in the conductance evaluation in the NCA and the OCA for the two-terminal device [34].

### C. $T$ -matrix QME

Up to now, to the best of our knowledge, only the sequential QME has been investigated for the present device [2,6,7,11]. The  $T$ -matrix cotunneling has already been worked out, but in Coulomb drag systems without temperature bias [17–19]. Cotunneling has also been developed in the presence of a temperature bias [36], but another method, namely a real-time diagrammatic method, was used to obtain the transition rates.

In the QME, the reservoir degrees of freedom are traced out, hence one focuses on the four two-dot states, noted in a self-explanatory manner, from empty to full:  $|0\rangle$ ,  $|t\rangle$ ,  $|b\rangle$ , and  $|2\rangle$ . The corresponding probabilities  $P_n$  for  $n = 0, t, b, 2$  are governed by transfer rates  $\gamma_{mn}$  from state  $m$  to state  $n$ . With the weak-coupling assumption, the rates for tunneling-induced transition can be obtained by the generalized Fermi golden rule [37],

$$\gamma_{mn} = \frac{2\pi}{\hbar} \sum_{i'f'} W_{i'} |\langle f'n | T | i'm \rangle|^2 \delta(E_f - E_i), \quad (9)$$

where  $|i\rangle = |m\rangle \otimes |i'\rangle$  is the initial product state for dots and reservoirs,  $|f\rangle = |n\rangle \otimes |f'\rangle$  is the final one, and  $E_i$  and  $E_f$  are

their respective energy. More precisely,  $|n\rangle$  ( $|m\rangle$ ) is the double-dot state, and the corresponding energy is  $\epsilon_n$  ( $\epsilon_m$ ). The states  $|i'\rangle$  and  $|f'\rangle$  are shorthand notations for tensorial products of three reservoir states.  $W_{i'}$  is the probability of the state  $|i'\rangle$ , assuming equilibrium for each reservoir, i.e., it is given by the grand-canonical Gibbs distribution with appropriate temperatures and chemical potentials. Up to the second order in  $H_T$ , the  $T$ -matrix can be written [37]

$$T = H_T + H_T \frac{1}{E_i - H_0 + i\eta} H_T, \quad (10)$$

where  $H_0$  and  $H_T$  were previously defined in Eqs. (1) and (2). The first term in the  $T$ -matrix expression represents the sequential tunneling processes, while the second term includes the cotunneling ones. The nature of the  $T$ -matrix QME requires that the energies of the double dot states are discrete, implying  $\Gamma \ll T, U$ , to be relevant.

In the stationary regime, the probabilities  $P_n$  are obtained from the following system:

$$\begin{aligned} \frac{dP_0}{dt} &= -(\gamma_{0t} + \gamma_{0b} + \tilde{\gamma}_{02})P_0 + \gamma_{t0}P_t + \gamma_{b0}P_b + \tilde{\gamma}_{20}P_2 = 0, \\ \frac{dP_t}{dt} &= \gamma_{0t}P_0 - (\gamma_{t0} + \tilde{\gamma}_{tb} + \gamma_{t2})P_t + \tilde{\gamma}_{bt}P_b + \gamma_{2t}P_2 = 0, \\ \frac{dP_b}{dt} &= \gamma_{0b}P_0 + \tilde{\gamma}_{tb}P_t - (\gamma_{b0} + \tilde{\gamma}_{bt} + \gamma_{b2})P_b + \gamma_{2b}P_2 = 0, \\ \frac{dP_2}{dt} &= \tilde{\gamma}_{20}P_0 + \gamma_{t2}P_t + \gamma_{b2}P_b - (\tilde{\gamma}_{20} + \gamma_{2t} + \gamma_{2b})P_2 = 0, \end{aligned} \quad (11)$$

where the notation highlights the distinction between cotunneling terms ( $\tilde{\gamma}_{mn}$ ) and sequential ones ( $\gamma_{mn}$ ). In the particle-hole symmetric case,  $\epsilon_t = \epsilon_b \equiv \epsilon_d = -U/2$ , simplifications occur as detailed in Appendix B. One then gets  $P_t = P_b$  and  $P_0 = P_2$ , leading by normalization to  $P_0 + P_t = \frac{1}{2}$ . Furthermore, using  $\Gamma_t(\epsilon) = \Gamma_t$  and  $\Gamma_{L_b}(\epsilon) = \theta(-\epsilon)\Gamma_b = \Gamma_{R_b}(-\epsilon)$ , we obtain

$$P_t = \frac{\hbar}{2} \frac{\gamma_{0t} + \gamma_{0b}}{\Gamma_t + \Gamma_b}, \quad (12)$$

indicating that the probabilities  $P_i$  are not affected by cotunneling processes for the present parameters. It stems from the particle-hole symmetry, indeed, as detailed in the Appendixes,  $\tilde{\gamma}_{02} = \tilde{\gamma}_{20} = \tilde{\gamma}_{tb} = \tilde{\gamma}_{bt}$ , thus the cotunneling does not affect the dot occupancy. However, it will modify the currents. One can note that the dots are half-filled as expected, since  $\langle n_t \rangle = \langle n_b \rangle = P_t + P_2 = 1/2$ .

To express the electric current of interest, a more detailed version of transfer rates is needed with the specification of relevant reservoirs. Sequential terms imply only one reservoir, and the superscript in  $\gamma_{nm}^\alpha$  indicates the concerned one. For the drag current, cotunneling processes involve two reservoirs, one of which is always the  $t$  one; thus only the second reservoir needs to be specified: For example,  $\tilde{\gamma}_{bt}^{R_b}$  is the transfer rate from the  $|b\rangle$  state to the  $|t\rangle$  one, with the fermion on the bottom dot hopping to the  $R_b$  reservoir. The detailed expressions for all transfers in the present particle-hole symmetric model, as well as the regularization scheme, are postponed to Appendix B.

In a general case, one should have also to consider transfer rates of the form  $\tilde{\gamma}_{nn}^{\alpha\beta}$  reflecting processes in which the two-dot

states do not change, however two-electron hops from and to different reservoirs are implied in the process. These terms do not enter the  $P_i$  expressions but would enter the currents for some of them. They are intradot cotunneling processes, and they cancel or do not contribute to the currents in our case, because of the present structureless dots, and because of our choice for hybridization functions  $\Gamma_\alpha(\epsilon)$ . The only cotunneling terms implied in our calculations are delocalized processes implying the two dots, as in Ref. [17].

Finally, the charge currents involving the bottom dot read

$$\begin{aligned} J_{Lb}^e &= e[(\gamma_{0b}^{Lb} + \tilde{\gamma}_{02}^{Lb})P_0 + (\gamma_{t2}^{Lb} + \tilde{\gamma}_{tb}^{Lb})P_t \\ &\quad - (\gamma_{b0}^{Lb} + \tilde{\gamma}_{bt}^{Lb})P_b - (\gamma_{2t}^{Lb} + \tilde{\gamma}_{20}^{Lb})P_2], \\ J_{Rb}^e &= e[(\gamma_{0b}^{Rb} + \tilde{\gamma}_{02}^{Rb})P_0 + (\gamma_{t2}^{Rb} + \tilde{\gamma}_{tb}^{Rb})P_t \\ &\quad - (\gamma_{b0}^{Rb} + \tilde{\gamma}_{bt}^{Rb})P_b - (\gamma_{2t}^{Rb} + \tilde{\gamma}_{20}^{Rb})P_2]. \end{aligned} \quad (13)$$

In the stationary regime,  $J_b^e \equiv J_{Lb}^e = -J_{Rb}^e$ . Gathering the results of Appendix B, we get

$$J_b^e = \frac{e\Gamma_b}{\hbar} \frac{1}{2(1 + \Gamma_b/\Gamma_t)} (f_{Lb}(\epsilon_d) - f_t(\epsilon_d)) + \frac{e}{2}\Delta, \quad (14)$$

where the first right-side term is the sequential current, whereas the second one is a purely cotunneling contribution whose detailed expression is given in the Appendix B.

For the present parameters [special choice of  $\Gamma_{Lb}(\epsilon)$ ,  $\Gamma_{Rb}(\epsilon)$ ,  $\epsilon_d$ , and  $U$ ], we can evaluate the energy current supplied by the  $t$ -reservoir without additional calculations. This is not true in the general case in which evaluating energy or heat current is not a simple issue [36]. First in the  $T$ -matrix QME, the dot spectral functions are  $\delta$ -functions, which are located at  $\epsilon_d$  (negative in our calculations) and  $\epsilon_d + U$  (positive here), and there is neither broadening nor shift [38], unlike in some other QME approaches; see, e.g., Ref. [36]. Second,  $\Gamma_{Lb}(\epsilon)$  cancels for positive energy, such that carriers entering the bottom dot from the left reservoir can only have an energy  $\epsilon_d$ , such that  $J_{Lb}^E = \frac{\epsilon_d}{e} J_{Lb}^e$ . On the other hand,  $\Gamma_{Rb}(\epsilon)$  is nonzero only for positive energy, leading to  $J_{Rb}^E = \frac{\epsilon_d + U}{e} J_{Rb}^e$ . By energy conservation in the stationary regime, we have  $J_t^E + J_{Lb}^E + J_{Rb}^E = 0$ . Using  $J_{Lb}^e = -J_{Rb}^e$ , the so-called tight-coupling relation emerges:

$$J_t^E = \frac{U}{e} J_b^e, \quad (15)$$

which leads to an engine efficiency

$$\eta_{\text{QME}} = \frac{\mathcal{P}}{J_t^E} = \frac{eV}{U}, \quad (16)$$

which was found in the sequential regime [2], and it persists at the cotunneling level for the present parameters.

### III. RESULTS

We begin by quantitative comparisons between the NCA and the QME. We take first  $\kappa(\epsilon) = 1$  in the hybridization functions [Eqs. (3)]. Furthermore, we choose the following equality:  $\Gamma_t = \Gamma_b \equiv \Gamma$ . In Fig. 2 the output power is displayed for parameters for which the QME is supposed to be relevant [2]. Noting the Carnot efficiency  $\eta_C = 1 - T_c/T_h$ , the sequential current vanishes for  $eV = \eta_C U$ , as can be readily

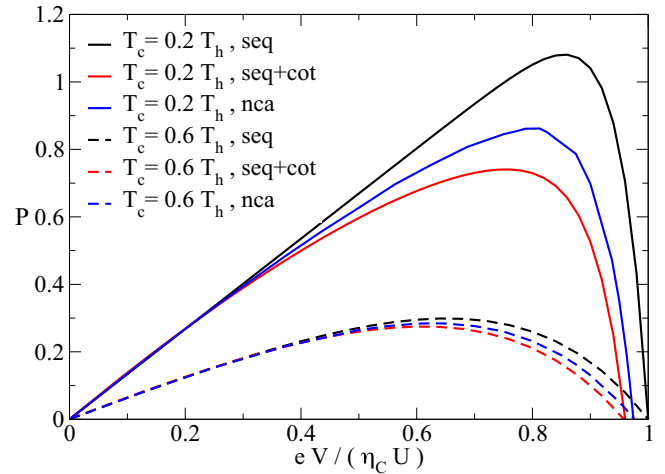


FIG. 2. Comparison between the NCA and the QME with (seq + cot) and without (seq) cotunneling processes: The output power  $\mathcal{P}$ , in  $\Gamma^2/\hbar$  units, is displayed as a function of bias voltage, for  $U = 10T_h$ ,  $\epsilon_d = -U/2$ ,  $\Gamma = T_h/100$ ,  $T_c = 0.2T_h$  (solid lines), or  $T_c = 0.6T_h$  (dashed lines).

seen by looking at the first right-side term in Eq. (14). For high bias  $eV \gtrsim \frac{1}{2}\eta_C U$ , taking into account cotunneling processes matters, especially for low  $T_c$  as  $T_c = 20\Gamma$ . The NCA results are closer to these predictions, particularly concerning the stopping voltage, which corresponds to the bias voltage needed to cancel the current. As in the QME, for the present parameters in the NCA, the tight-coupling relation between energy and charge currents is nearly fulfilled, leading to an efficiency  $\eta_{\text{NCA}} \simeq eV/U$ .

We then consider a parameter range where sequential processes alone are not supposed to be suitable, indeed  $T > \Gamma$  is not fulfilled. In Fig. 3, the electric and renormalized energy currents are plotted as a function of renormalized bias voltage. The cotunneling processes, which boost the electric current at low or null voltage, rapidly erode it for a moderate bias and reduce the stopping voltage by a factor 2.5. There is about a factor 2 between the NCA and the cotunneling QME current amplitudes. However, they nearly coincide concerning the stopping voltage ( $\sim 1.83\Gamma$  in the NCA,  $\sim 1.95\Gamma$  in the QME). The qualitative agreement between the QME and the NCA is worse for the energy current (up to a  $U/e$  factor, red and green curves). Indeed, interestingly, for the present parameters there is a significant difference in the NCA between the electric drag current and the renormalized thermal current  $eJ_t^E/U$ . Not only are  $J_b^e$  and  $J_t^E$  not proportional to each other, but also they do not even vanish for the same bias. At low voltage, as long as the electric current exceeds  $eJ_t^E/U$ , the efficiency will be higher than  $eV/U$ . For a higher  $V$ , the situation is reversed. When the drag charge current vanishes, the energy current corresponding to heat extracted from the hot source is distributed—equitably, as can be demonstrated for the symmetric case—to the cold reservoirs, without charge current flowing through the  $b$ -dot. This transfer is driven by charge fluctuations in between the bottom dot and each cold reservoir separately. For this to happen, a more elaborate structure for the Green's functions is needed, beyond the  $\delta$ -function shape, as occurs in the  $T$ -matrix QME. In the lower panel of Fig. 3, the



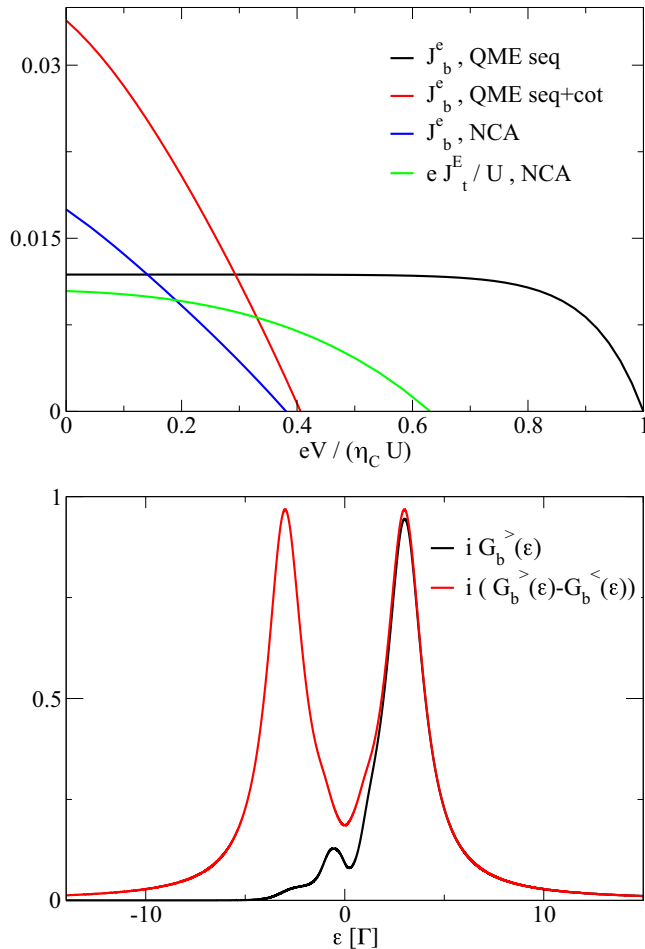


FIG. 3. Top panel: Charge current in  $eV/h$  units, evaluated in the QME and in the NCA, for  $U = 6\Gamma$ ,  $T_h = \Gamma$ , and  $T_c = 0.2T_h$ . The renormalized energy current evaluated in the NCA,  $\frac{eJ_t^E}{U}$ , is also presented. Bottom: Green's functions for the same parameters, and for  $V$  equal to the upper panel stopping voltage.

Green's functions  $iG_b^>(\epsilon)$  and  $i[G_b^>(\epsilon) - G_b^<(\epsilon)]$  are plotted, and they reveal two broad peaks located around  $\epsilon_d$  and  $\epsilon_d + U$ , as well as a low-energy finer structure in between. One has  $iG_b^>(-\epsilon) = -iG_b^<(\epsilon)$  for the particle-hole symmetric case.

In the sequential QME, due to the  $\delta$ -function peaks in the spectral functions, the charge current is unaffected by a moderate bias, as seen in Fig. 3. Only a bias roughly of the order of  $U$  (it depends also on temperatures) can decrease sensibly the drag current, and as a consequence the efficiency can attain the Carnot efficiency. On the contrary, in the NCA—as well as in the QME, including cotunneling processes but for different reasons—the voltage rapidly reduces the current, thus bounding the output power and efficiency. To contain this power erosion, we consider a family of hybridization functions in Eqs. (3). Namely, we choose shifted Heaviside functions

$$\kappa(\epsilon) = H(\epsilon - \delta). \quad (17)$$

For the same parameters as in Fig. 3, as shown in the upper panel in Fig. 4, varying  $\delta$  from 0 to  $\frac{U}{2}$  erodes the current at

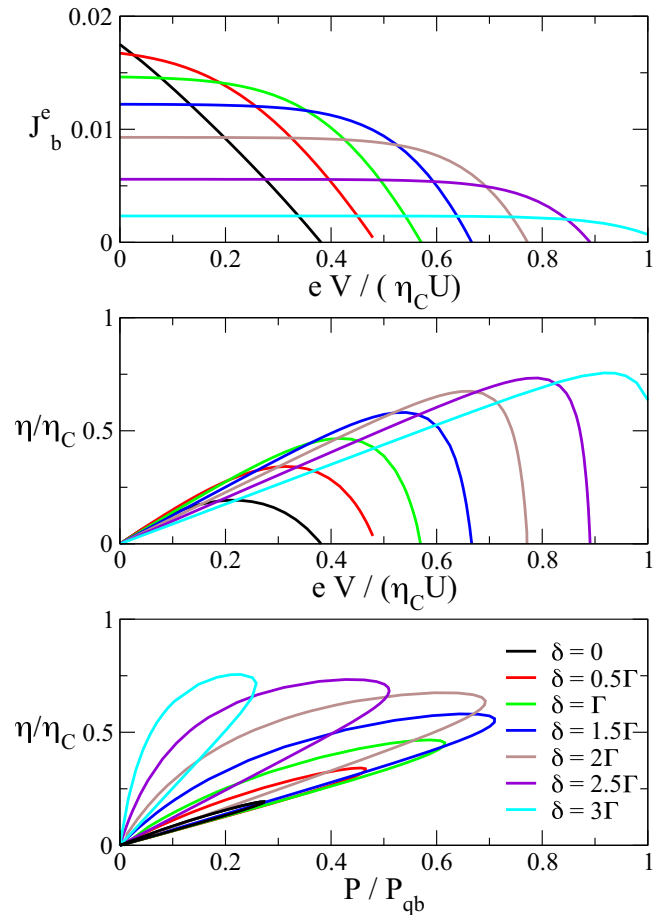


FIG. 4. NCA calculations for the same parameters as in Fig. 3:  $U = 6\Gamma$ ,  $\epsilon_d = -U/2$ ,  $T_h = \Gamma$ , and  $T_c = 0.2T_h$ . The hybridization functions are shifted by  $\delta$ . Top panel: electric current in  $eV/h$  units, middle panel: ratio of device to Carnot efficiencies as a function of bias, bottom: ratio of device to Carnot efficiencies as a function of renormalized power. See text for details.

null bias, but it ensures a quasi-independence from voltage in a larger range as  $\delta$  grows. This can be very advantageous for efficiency and output power as revealed in the middle and lower panels of Fig. 4.

Despite a sensible reduction of the zero-bias current, its quasi-independence against voltage proves beneficial to maximum power and corresponding efficiency: Going from  $\delta = 0$  to  $\delta = 1.5\Gamma$ , the first one grows by a factor 2.6, while the second one is enhanced by a factor 2.9. These tailored hybridization functions make this device very attractive. It is tempting to make a comparison with the maximum power that can be delivered by one channel connecting two heat sources. Whitney uncovered in Ref. [23] the following upper quantum bound on output power for a two-terminal engine:

$$\mathcal{P}_{qb} = A_0 \frac{\pi^2}{h} k_B^2 \Delta T^2 \quad (18)$$

with  $A_0 \simeq 0.0321$ . Later this bound was extended to the three-terminal setup [25]. With the temperatures used in Fig. 4 ( $T_h = \Gamma$ ,  $T_c = 0.2\Gamma$ ), this bound, Eq. (18), attains 0.0322

$\Gamma^2/\hbar$ , and the lower part of Fig. 4 reveals that the present device achieves a maximum power of about  $0.7 \mathcal{P}_{\text{qb}}$  for  $\delta = 1.5\Gamma$ .

Furthermore, in the two-terminal setup, the quantum bound can be reached for a transmission function of boxcar type, for which the efficiency can be evaluated. One finds

$$\eta_{\mathcal{P}_{\text{qb}}} = \frac{1}{(\bar{T}/\Delta T) D + 1/2}, \quad (19)$$

where  $\bar{T} = \frac{T_h + T_c}{2}$ ,  $\Delta T = T_h - T_c$ , and  $D \simeq 2.872$ . For the present parameters, it leads to  $\eta_{\mathcal{P}_{\text{qb}}}/\eta_C \simeq 0.47$ , which is surpassed in the present device. It appears that the large power is not achieved at the expense of efficiency. We can also compare the present efficiency at maximum power  $\eta_{\mathcal{P}_{\text{max}}} = 0.448$  with the Curzon-Ahlborn [39] one for the present temperatures:  $\eta_{\text{CA}} = 0.55$ .

The quantum bound settled by Whitney was first obtained for a two-terminal device. In this case, the Meir-Wingreen-Landauer exact formula [Eq. (3) of Ref. [20]] applies with the only hypothesis of proportionate left and right lead-dot couplings, and noninteracting leads, but potentially in the presence of in-dot Coulomb interaction. Thus the output power from the interacting system is forced by this quantum bound. To fix an order of magnitude in the two-terminal case, it can be mentioned, for example, that for the problem of two orbitals in the Kondo regime (see Ref. [22], Fig. 2), the maximum power per channel was about one-tenth of the quantum bound. For a noninteracting Lorentzian transmission function, it would be even much smaller [40]. The present three-terminal setup in the Coulomb-blockade regime might not be constrained by this quantum bound, which was obtained in a noninteracting theory [25]. Our numerical results do not exceed the bound but almost reach it, making the TTTE setup very attractive from an energy-harvesting perspective.

The regime  $\Gamma > T_h$  would be an interesting issue to address. On the one hand, one could expect a less favorable situation than the one presented in Fig. 4, due to a likely broadening of Green's functions as  $\Gamma$  increases, usually detrimental to transport. On the other hand, as in the two-terminal case, one cannot exclude for lower temperatures emerging sharp structures reminiscent of the Kondo resonance, which grows in the two-terminal device [22] and is beneficial to power. Unfortunately in the case of dissymmetrical hybridizations, we encounter numerical difficulties [35] that prevent low-temperature study and would need more numerical investment to be settled; this is left for future work.

#### IV. CONCLUSIONS

We have shown that the TTTE setup presented in Fig. 1 not only offers the interesting feature of decoupling charge and heat currents, but it also overcomes the performance of the one-dot–two-terminal setup. It had been claimed to be “optimal” [4] because of the possibility to attain Carnot efficiency when the sequential QME is reliable. Expanding the parameter range for this device by developing two different strategies to explore its properties, we have shown that it can also yield an outstanding power due to its topology.

The two techniques employed are very different and are successfully compared in a regime in which the sequential QME alone fails. The NCA is presumed to be valid on a

wider parameter range than the QME. However, the observed qualitative agreement—better for the electric current—may be reminiscent of the concordance found between the cotunneling QME and experiments in Coulomb coupled quantum dots in an unexpected parameter range [18].

*Note added.* During the review process, we became aware of some related work [41] that addressed the issue of calculating the heat current in more general situations than the one studied here, up to the cotunneling order.

#### ACKNOWLEDGMENT

We gratefully acknowledge C. Mattei for detecting a few misprints.

#### APPENDIX A: NCA EQUATIONS

In the NCA, the true fermions are replaced by pseudoparticles [32]. Assuming the following notations for fermion creation operators:  $d_\sigma^\dagger$ , with  $\sigma = -\bar{\sigma} = \uparrow$  or  $\downarrow$ , respectively, for the  $b$ - or  $t$ -dot orbital, we define the pseudofermion  $f_\sigma^\dagger$  and the pseudoboson  $a^\dagger$  and  $b^\dagger$  creation operators by  $d_\sigma^\dagger = f_\sigma^\dagger b + \sigma a^\dagger f_{\bar{\sigma}}$ .  $b^\dagger$  is the bosonic operator creating the empty two-dot state, while  $a^\dagger$  corresponds to the creation of the doubly occupied two-dot state.

The present top and bottom fermion Green's functions can be expressed in terms of the pseudoparticle lesser and greater Green's functions. The relations between the Fourier-transformed Green's functions are

$$\begin{aligned} G_{d\sigma}^<(\omega) &= i \int \frac{d\omega'}{2\pi} [G_{f\sigma}^<(\omega') D_b^>(\omega' - \omega) \\ &\quad - G_{f\bar{\sigma}}^>(\omega' - \omega) D_a^<(\omega')], \\ G_{d\sigma}^>(\omega) &= i \int \frac{d\omega'}{2\pi} [G_{f\sigma}^>(\omega') D_b^<(\omega' - \omega) \\ &\quad - G_{f\bar{\sigma}}^<(\omega' - \omega) D_a^>(\omega')]. \end{aligned} \quad (A1)$$

The pseudoparticle Green's functions obey Keldysh equations:

$$\begin{aligned} D_{a(b)}^{\gtrless}(\omega) &= D_{a(b)}^R(\omega) \Pi_{a(b)}^{\gtrless}(\omega) D_{a(b)}^A(\omega), \\ G_{f\sigma}^{\gtrless}(\omega) &= G_{f\sigma}^R(\omega) \Sigma_{f\sigma}^{\gtrless}(\omega) G_{f\sigma}^A(\omega), \end{aligned} \quad (A2)$$

where  $D_{a(b)}^{R(A)}$  are retarded (advanced) boson Green's functions,  $\Pi_{a(b)}^{\gtrless}$  are the boson self-energies, and  $\Sigma_{f\sigma}^{\gtrless}(\omega)$  are the pseudofermion self-energies.

The NCA mainly lies in the choice for the pseudoboson and -fermion self-energies. Using a diagrammatic technique, the following expressions are obtained:

$$\begin{aligned} \Sigma_{f\uparrow}^<(\omega) &= - \int \frac{d\omega'}{2\pi} A_b(\omega - \omega') D_b^<(\omega') \\ &\quad - \int \frac{d\omega'}{2\pi} B_t(\omega' - \omega) D_a^<(\omega'), \\ \Sigma_{f\downarrow}^<(\omega) &= - \int \frac{d\omega'}{2\pi} A_t(\omega - \omega') D_b^<(\omega') \\ &\quad - \int \frac{d\omega'}{2\pi} B_b(\omega' - \omega) D_a^<(\omega'), \end{aligned}$$

$$\begin{aligned}
\Pi_b^<(\omega) &= - \int \frac{d\omega'}{2\pi} B_b(\omega' - \omega) G_{f\uparrow}^<(\omega') \\
&\quad - \int \frac{d\omega'}{2\pi} B_t(\omega' - \omega) G_{f\downarrow}^<(\omega'), \\
\Pi_a^<(\omega) &= - \int \frac{d\omega'}{2\pi} A_t(\omega - \omega') G_{f\uparrow}^<(\omega') \\
&\quad - \int \frac{d\omega'}{2\pi} A_b(\omega - \omega') G_{f\downarrow}^<(\omega'), \quad (\text{A3})
\end{aligned}$$

whereas

$$\begin{aligned}
\Sigma_{f\uparrow}^>(\omega) &= \int \frac{d\omega'}{2\pi} B_b(\omega - \omega') D_b^>(\omega') \\
&\quad + \int \frac{d\omega'}{2\pi} A_t(\omega' - \omega) D_a^>(\omega'), \\
\Sigma_{f\downarrow}^>(\omega) &= \int \frac{d\omega'}{2\pi} B_t(\omega - \omega') D_b^>(\omega') \\
&\quad + \int \frac{d\omega'}{2\pi} A_b(\omega' - \omega) D_a^>(\omega'), \\
\Pi_b^>(\omega) &= \int \frac{d\omega'}{2\pi} A_b(\omega' - \omega) G_{f\uparrow}^>(\omega') \\
&\quad + \int \frac{d\omega'}{2\pi} A_t(\omega' - \omega) G_{f\downarrow}^>(\omega'), \\
\Pi_a^>(\omega) &= \int \frac{d\omega'}{2\pi} B_t(\omega - \omega') G_{f\uparrow}^>(\omega') \\
&\quad + \int \frac{d\omega'}{2\pi} B_b(\omega - \omega') G_{f\downarrow}^>(\omega'). \quad (\text{A4})
\end{aligned}$$

In the preceding equations,

$$\begin{aligned}
B_b(\omega) &= \Gamma_{L_b}(\omega)[1 - f_{L_b}(\omega)] + \Gamma_{R_b}(\omega)[1 - f_{R_b}(\omega)], \\
B_t(\omega) &= \Gamma_t(\omega)[1 - f_t(\omega)], \\
A_b(\omega) &= \Gamma_{L_b}(\omega)f_{L_b}(\omega) + \Gamma_{R_b}(\omega)f_{R_b}(\omega), \\
A_t(\omega) &= \Gamma_t(\omega)f_t(\omega). \quad (\text{A5})
\end{aligned}$$

We explicitly checked, as was done also in Ref. [42], that these self-energies ensure charge and energy current conservation,  $J_{L_b}^e = -J_{R_b}^e$  and  $J_{L_b}^E + J_{R_b}^E + J_t^E = 0$ .

## APPENDIX B: QME TRANSFER RATES

In the particle-hole symmetric model  $\epsilon_t = \epsilon_b \equiv \epsilon_d = -U/2$ , choosing  $\Gamma_t(\epsilon) = \Gamma_t$ ,  $\Gamma_{L_b}(\epsilon) = \theta(-\epsilon)\Gamma_b = \Gamma_{R_b}(-\epsilon)$ , the sequential terms are readily evaluated. One finds

$$\begin{aligned}
\gamma_{0t} &= \frac{1}{\hbar} f_t(\epsilon_d)\Gamma_t, \quad \gamma_{t0} = \frac{1}{\hbar} [1 - f_t(\epsilon_d)]\Gamma_t, \\
\gamma_{0b} &= \gamma_{0b}^{L_b} + \gamma_{0b}^{R_b} \quad \text{with} \quad \gamma_{0b}^{L_b} = \frac{1}{\hbar} f_{L_b}(\epsilon_d)\Gamma_b, \quad \gamma_{0b}^{R_b} = 0, \\
\gamma_{b0} &= \gamma_{b0}^{L_b} + \gamma_{b0}^{R_b} \quad \text{with} \quad \gamma_{b0}^{L_b} = \frac{1}{\hbar} [1 - f_{L_b}(\epsilon_d)]\Gamma_b, \\
\gamma_{b0}^{R_b} &= 0, \quad \gamma_{t2} = \gamma_{t2}^{L_b} + \gamma_{t2}^{R_b} \quad \text{with} \quad \gamma_{t2}^{L_b} = 0, \\
\gamma_{t2}^{R_b} &= \frac{1}{\hbar} f_{R_b}(\epsilon_d + U)\Gamma_b, \quad \gamma_{2t} = \gamma_{2t}^{L_b} + \gamma_{2t}^{R_b}, \quad \gamma_{2t}^{L_b} = 0,
\end{aligned}$$

$$\begin{aligned}
\gamma_{2t}^{R_b} &= \frac{1}{\hbar} [1 - f_{R_b}(\epsilon_d + U)]\Gamma_b, \quad \gamma_{b2} = \frac{1}{\hbar} f_t(\epsilon_d + U)\Gamma_t, \\
\gamma_{2b} &= \frac{1}{\hbar} [1 - f_t(\epsilon_d + U)]\Gamma_t. \quad (\text{B1})
\end{aligned}$$

The cotunneling terms request more calculations and a regularization procedure. They are of two types: some of them are local cotunneling (intradot), while the others are nonlocal cotunneling (interdot) [17]. In the following, we only consider transition rates that may affect  $P_i$  and/or currents of interest. For example, for three terminals but for an arbitrary hybridization function between dots and reservoirs, one obtains

$$\begin{aligned}
\tilde{\gamma}_{02} &= \frac{1}{\hbar} \sum_{\beta_b} \int \frac{d\epsilon}{2\pi} \Gamma_t(\epsilon) \Gamma_{\beta_b}(\epsilon_t + \epsilon_b + U - \epsilon) \left| \frac{1}{\epsilon - \epsilon_t + i\eta} \right. \\
&\quad \left. + \frac{1}{\epsilon_t + U - \epsilon + i\eta} \right|^2 f_i(\epsilon) f_{\beta_b}(\epsilon_t + \epsilon_b + U - \epsilon). \quad (\text{B2})
\end{aligned}$$

The other cotunneling terms are of the same kind. For our particular choice of  $\Gamma_\alpha(\epsilon)$  and in the particle-hole symmetric case, one obtains

$$\begin{aligned}
\tilde{\gamma}_{02} &= \tilde{\gamma}_{02}^{L_b} + \tilde{\gamma}_{02}^{R_b}, \quad \tilde{\gamma}_{20} = \tilde{\gamma}_{20}^{L_b} + \tilde{\gamma}_{20}^{R_b}, \\
\tilde{\gamma}_{tb} &= \tilde{\gamma}_{tb}^{L_b} + \tilde{\gamma}_{tb}^{R_b}, \quad \tilde{\gamma}_{bt} = \tilde{\gamma}_{bt}^{L_b} + \tilde{\gamma}_{bt}^{R_b}, \quad (\text{B3})
\end{aligned}$$

with

$$\begin{aligned}
\tilde{\gamma}_{02}^{L_b} &= \tilde{\gamma}_{20}^{R_b} = \tilde{\gamma}_{tb}^{L_b} = \tilde{\gamma}_{bt}^{R_b} = \tilde{R}_1 + I_1, \\
\tilde{\gamma}_{02}^{R_b} &= \tilde{\gamma}_{20}^{L_b} = \tilde{\gamma}_{tb}^{R_b} = \tilde{\gamma}_{bt}^{L_b} = \tilde{R}_2 + I_2. \quad (\text{B4})
\end{aligned}$$

In the preceding expressions,

$$\begin{aligned}
\tilde{R}_1 &= \frac{\Gamma_b \Gamma_t}{\hbar} \int_0^{+\infty} \frac{d\epsilon}{2\pi} \frac{1}{(\epsilon - \epsilon_d)^2} f_t(\epsilon) [1 - f_{R_b}(\epsilon)], \\
\tilde{R}_2 &= \frac{\Gamma_b \Gamma_t}{\hbar} \int_0^{+\infty} \frac{d\epsilon}{2\pi} \frac{1}{(\epsilon - \epsilon_d)^2} f_{R_b}(\epsilon) [1 - f_t(\epsilon)], \\
I_1 &= \frac{\Gamma_b \Gamma_t}{\hbar} \int_0^{+\infty} \frac{d\epsilon}{2\pi} \frac{1}{(\epsilon + \epsilon_d)^2 + \eta^2} \left[ 1 - 2 \frac{\epsilon + \epsilon_d}{\epsilon - \epsilon_d} \right] \\
&\quad \times f_t(\epsilon) [1 - f_{R_b}(\epsilon)], \\
I_2 &= \frac{\Gamma_b \Gamma_t}{\hbar} \int_0^{+\infty} \frac{d\epsilon}{2\pi} \frac{1}{(\epsilon + \epsilon_d)^2 + \eta^2} \left[ 1 - 2 \frac{\epsilon + \epsilon_d}{\epsilon - \epsilon_d} \right] \\
&\quad \times f_{R_b}(\epsilon) [1 - f_t(\epsilon)]. \quad (\text{B5})
\end{aligned}$$

$\tilde{R}_1$  and  $\tilde{R}_2$  are well-behaved expressions whereas  $I_1$  and  $I_2$  diverge, and one needs to extract manually these diverging terms. The divergence that needs to be cured is inherent to the present method  $T$ -matrix-based QME [43]. In the current calculation, one needs the difference  $I_1 - I_2$ , which can be

written

$$I_1 - I_2 = \int_{\epsilon_d}^{+\infty} d\epsilon \frac{F(\epsilon)}{\epsilon^2 + \eta^2} \quad (\text{B6})$$

with

$$F(\epsilon) = \frac{\Gamma_t \Gamma_b}{2\pi \hbar} \left( 1 - 2 \frac{\epsilon}{\epsilon - 2\epsilon_d} \right) (f_t(\epsilon - \epsilon_d) - f_{R_b}(\epsilon - \epsilon_d)). \quad (\text{B7})$$

We cannot proceed analytically further as was done, for example, in Refs. [17,19], because in the preceding equation the two Fermi functions are characterized by different temperatures. We thus proceed along the lines of Ref. [44], throwing away the divergent part, which is in fact a sequential contribution,

by making the substitution

$$I_1 - I_2 \rightarrow \mathcal{P} \int_{\epsilon_d}^{+\infty} d\epsilon \frac{F(\epsilon) - F(0)}{\epsilon^2}, \quad (\text{B8})$$

where  $\mathcal{P}$  indicates the principal part. Finally, the current expression requires the term  $\tilde{R}_1 - \tilde{R}_2 + I_1 - I_2$ , which reads after regularization

$$\Delta = \frac{\Gamma_b \Gamma_t}{\hbar} \int_0^{+\infty} \frac{d\epsilon}{2\pi (\epsilon - \epsilon_d)^2} [f_t(\epsilon) - f_{R_b}(\epsilon)] + \mathcal{P} \int_{\epsilon_d}^{+\infty} d\epsilon \frac{F(\epsilon) - F(0)}{\epsilon^2}. \quad (\text{B9})$$

- 
- [1] M. Zebarjadi, K. Esfarjani, M. S. Dresselhaus, Z. F. Ren, and G. Chen, *Energy Environ. Sci.* **5**, 5147 (2012).
- [2] B. Sothmann, R. Sánchez, and A. N. Jordan, *Nanotechnology* **26**, 032001 (2015).
- [3] G. Benenti, G. Casati, K. Saito, and R. S. Whitney, *Phys. Rep.* **694**, 1 (2017).
- [4] R. Sánchez and M. Büttiker, *Phys. Rev. B* **83**, 085428 (2011).
- [5] B. Roche, P. Roulleau, T. Jullien, Y. Jompol, I. Farrer, D. A. Ritchie, and D. C. Glatti, *Nat. Commun.* **6**, 6738 (2015).
- [6] H. Thierschmann, R. Sánchez, B. Sothmann, F. Arnold, C. Heyn, W. Hansen, H. Buhmann, and L. W. Molenkamp, *Nat. Nanotechnol.* **10**, 854 (2015).
- [7] H. Thierschmann, F. Arnold, M. Mittermüller, I. Maier, C. Heyn, W. Hansen, H. Buhmann, and L. W. Molenkamp, *New J. Phys.* **17**, 113003 (2015).
- [8] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, *Phys. Rev. Lett.* **115**, 260602 (2015).
- [9] F. Hartmann, P. Pfeffer, S. Höfling, M. Kamp, and L. Worschech, *Phys. Rev. Lett.* **114**, 146805 (2015).
- [10] O. Entin-Wohlman, Y. Imry, and A. Aharony, *Phys. Rev. B* **82**, 115314 (2010).
- [11] H. Thierschmann, R. Sánchez, B. Sothmann, H. Buhmann, and L. W. Molenkamp, *C. R. Phys.* **17**, 1109 (2016).
- [12] Y. Zhang, G. Lin, and J. Chen, *Phys. Rev. E* **91**, 052118 (2015).
- [13] R. Sánchez and M. Büttiker, *Europhys. Lett.* **100**, 47008 (2012).
- [14] Y. V. Nazarov and Y. M. Blanter, *Quantum Transport, Introduction to Nanoscience* (Cambridge University Press, New York, 2009).
- [15] R. Sánchez, B. Sothmann, A. N. Jordan, and M. Büttiker, *New J. Phys.* **15**, 125001 (2013).
- [16] B. N. Narozhny and A. Levchenko, *Rev. Mod. Phys.* **88**, 025003 (2016).
- [17] K. Kaasbjerg and A.-P. Jauho, *Phys. Rev. Lett.* **116**, 196801 (2016).
- [18] A. J. Keller, J. S. Lim, D. Sánchez, R. López, S. Amasha, J. A. Katine, H. Shtrikman, and D. Goldhaber-Gordon, *Phys. Rev. Lett.* **117**, 066602 (2016).
- [19] J. S. Lim, R. Lopez, and D. Sánchez, *arXiv:1612.06627*.
- [20] N. S. Wingreen and Y. Meir, *Phys. Rev. B* **49**, 11040 (1994).
- [21] M. H. Hettler, J. Kroha, and S. Hershfield, *Phys. Rev. B* **58**, 5649 (1998).
- [22] J. Azema, A.-M. Daré, S. Schäfer, and P. Lombardo, *Phys. Rev. B* **86**, 075303 (2012).
- [23] R. S. Whitney, *Phys. Rev. Lett.* **112**, 130601 (2014).
- [24] R. S. Whitney, *Phys. Rev. B* **91**, 115425 (2015).
- [25] R. S. Whitney, *Entropy* **18**, 208 (2016).
- [26] A.-P. Jauho, N. S. Wingreen, and Y. Meir, *Phys. Rev. B* **50**, 5528 (1994).
- [27] H. J. W. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*, 2nd ed. (Springer-Verlag, Berlin, 2008), Sec. 3.4.
- [28] J. Hubbard, *Proc. R. Soc. London, Ser. A* **276**, 238 (1963).
- [29] A. C. Hewson, *Phys. Rev.* **144**, 420 (1966).
- [30] T. K. Ng, *Phys. Rev. Lett.* **76**, 487 (1996).
- [31] B. Dong and X. L. Lei, *J. Phys.: Condens. Matter* **14**, 11747 (2002).
- [32] K. Haule, S. Kirchner, J. Kroha, and P. Wölfle, *Phys. Rev. B* **64**, 155111 (2001).
- [33] C. N. Sposetti, L. O. Manuel, and P. Roura-Bas, *Phys. Rev. B* **94**, 085139 (2016).
- [34] L. Tosi, P. Roura-Bas, A. M. Llois, and L. O. Manuel, *Phys. Rev. B* **83**, 073301 (2011).
- [35] Low temperatures ( $T_h \ll \Gamma$ ) cannot be reached in the case of nonproportionate hybridization functions within our algorithm: the self-consistent NCA equations do not converge.
- [36] N. M. Gergs, C. B. M. Hörig, M. R. Wegewijs, and D. Schuricht, *Phys. Rev. B* **91**, 201107(R) (2015).
- [37] H. Bruus and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction* (Oxford University Press, New York, 2016).
- [38] K. Kaasbjerg and W. Belzig, *Phys. Rev. B* **91**, 235413 (2015).
- [39] F. L. Curzon and B. Ahlborn, *Am. J. Phys.* **43**, 22 (1975).
- [40] X. Li (private communication).
- [41] N. Walldorf, A.-P. Jauho, and K. Kaasbjerg, *Phys. Rev. B* **96**, 115415 (2017).
- [42] L. Tosi, P. Roura-Bas, and A. A. Aligia, *Phys. Rev. B* **88**, 235427 (2013).
- [43] S. Koller, M. Grifoni, M. Leijnse, and M. R. Wegewijs, *Phys. Rev. B* **82**, 235307 (2010).
- [44] M. Turek and K. A. Matveev, *Phys. Rev. B* **65**, 115332 (2002).