All-thermal transistor based on stochastic switching

Sánchez, Rafael; Thierschmann, Holger; Molenkamp, Laurens W.

DOI
10.1103/PhysRevB.95.241401

Publication date
2017

Document Version
Final published version

Published in
Physical Review B (Condensed Matter and Materials Physics)

Citation (APA)
https://doi.org/10.1103/PhysRevB.95.241401

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.
All-thermal transistor based on stochastic switching

Rafael Sánchez, 1 Holger Thierschmann, 2 and Laurens W. Molenkamp 3

1 Instituto Gregorio Millán, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain
2 Kavli Institute of Nanoscience, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
3 Experimentelle Physik 3, Physikalisches Institut, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

(Rceived 23 December 2016; revised manuscript received 8 May 2017; published 1 June 2017)

Fluctuations are strong in mesoscopic systems and have to be taken into account for the description of transport. We show that they can even be used as a resource for the operation of a system as a device. We use the physics of single-electron tunneling to propose a bipartite device working as a thermal transistor. Charge and heat currents in a two-terminal conductor can be gated by thermal fluctuations from a third terminal to which it is capacitively coupled. The gate system can act as a switch that injects neither charge nor energy into the conductor, hence achieving huge amplification factors. Nonthermal properties of the tunneling electrons can be exploited to operate the device with no energy consumption.

DOI: 10.1103/PhysRevB.95.241401

Introduction. Controlling the heat flow at small length scales is a great challenge in present-day electronics and a serious technological issue. On a scientific level, a significant effort is devoted to designing new concepts for nanoscale heat engines and thermoelectric devices [1–4] as well as finding new means to manipulate the flow of heat with devices such as thermal diodes and thermal transistors. This would not only allow for a more efficient removal of waste heat [5], but might also lead to the design of logic circuits operating with heat [6] or noise [7]. In recent years a variety of nanostructures has been identified and demonstrated in experiments as possible thermal diodes [8–12]. For a thermal transistor, however, the situation is more complicated because one has to efficiently switch heat currents by means of temperature.

Recent proposals suggest the use of nonlinearities of a mesoscopic system coupled to environmental modes [13–18]. Usually, a system is considered that is connected to two terminals, emitter E and collector C, and an environment, acting as a base B. A temperature distribution \( \{ T_i \} = (T_E, T_C, T_B) \) generates transport in the system. The aim is to modulate the collector heat current \( J_C \) with a small modulation of the heat injected from the base \( J_B \). This is usually done via inelastic transitions in the system induced by fluctuations in the environment. These can be controlled by tuning the base temperature \( T_B \rightarrow T_B + \Delta T_B \). A thermal transistor effect appears whose amplification factor is defined as [13,15,17]

\[
\alpha_l = \frac{\Delta J_l(\Delta T_B)}{\Delta J_B(\Delta T_B)},
\]

with \( l = E, C \) and \( \Delta J_l(\Delta T_B) = J_l(\Delta T_B) - J_l(0) \). The challenge is now to have a sizable modulation \( \Delta J_l \) out of tiny injected currents \( \Delta J_B \). However, the energy exchange with the environment is inherent to inelastic transitions, thus limiting the performance of the transistor.

Here, we propose an alternative approach. We introduce a nanostructure that mediates the coupling between the system and the environment, which we call a thermal gate [cf. Fig. 1(a)]. Transport in the system then becomes dependent on the state of the gate, which fluctuates at a rate given by the temperature \( T_B \). In particular, the gate can act as a switch if a given (thermally activated) transition blocks the system currents, with no net energy exchange involved. This way, huge amplification coefficients can be achieved.

An additional advantage of our mechanism is that the state of the gate (and hence how the system is coupled to the environment) can be externally controlled at a microscopic level. This scheme furthermore helps to isolate the system from undesired (e.g., phononic) degrees of freedom. The system-gate interaction can be of a different nature, depending on the particular configuration. As an example, the Coulomb interaction has been used in the last years to investigate effects such as mesoscopic Coulomb drag [19–21], energy harvesters [1,22], Maxwell demon refrigerators [3], or heat
engines with no heat absorption [23]. Most remarkably for us, stochastic switching of transport due to coupling to a mesoscopic gate has been recently demonstrated [24].

**Three-state case.** Let us first illustrate the idea with a simple toy model. We assume that the device can exhibit three states which we label |ON⟩, |OFF⟩, and |0⟩ [cf. Fig. 1(b)]. Transport occurs via transitions |0⟩ ↔ |ON⟩ through leads \( l = E, C \) at a rate \( \Gamma_{C}^{-} \). The superindex + (−) accounts for transitions populating (depopulating) the system. The |OFF⟩ state is uncoupled from transport and detuned by \( \Delta E \). Stochastic (Markovian) transitions of the type |ON⟩ ↔ |OFF⟩ switch the system currents. These are due to fluctuations in the gate and are hence assisted by the environment with rates \( \gamma_{ON} \) and \( \gamma_{OFF} = e^{-\Delta E/k_{B}T_{B}}\gamma_{ON} \) obeying detailed balance.

We can write a rate equation for the probability of the system states \( P_{m} \),

\[
\dot{P}_{0} = \Gamma_{C}^{-} P_{ON} - \Gamma_{C}^{+} P_{0} \quad (2)
\]

\[
\dot{P}_{OFF} = \gamma_{OFF} P_{ON} - \gamma_{OFF} P_{OFF}, \quad (3)
\]

with \( l = \sum_{m} P_{m} \) and \( \Gamma_{C}^{\pm} = \sum_{l} \Gamma_{l}^{\pm} \). Solving for the stationary state \( P_{m} = 0 \), we can write the particle currents \( I_{C} = -\dot{E}_{C} = \Gamma_{C}^{-} P_{0} - \Gamma_{C}^{+} P_{ON} \); giving

\[
I_{C} = \frac{\Gamma_{C}^{+} P_{0}}{\Gamma_{C}^{-} + \Gamma_{C}^{+}(1 + e^{-\Delta E/k_{B}T_{B}})} \cdot (4)
\]

It conditions the “uncoupled” current \( I_{ON}^{C} \) on the depopulation of the OFF state, with

\[
P_{OFF}(T) = \frac{e^{-\Delta E/k_{B}T_{B}}\Gamma_{C}^{+}}{\Gamma_{C}^{-} + \Gamma_{C}^{+}(1 + e^{-\Delta E/k_{B}T_{B}})} \cdot (5)
\]

Remarkably, it does not depend on the details of the coupling to the environment, only on its temperature.

In this configuration, the particle current (regardless of the electric or thermal gradient that originates it) is directly proportional to the charge and heat currents \( I_{l} = eI_{l} \) and \( J_{l} = (E_{ON} - \mu_{l})I_{l} \), with \( e \) being the elementary charge, \( E_{ON} \) the energy of the state, and \( \mu_{l} \) the electrochemical potential of \( l \). This way, any current in the conductor can be modulated by tuning \( T_{B} \), in particular,

\[
\Delta I_{C} = -I_{ON}^{C} \Delta P_{OFF}, \quad (6)
\]

with \( \Delta P_{OFF} = P_{OFF}(T_{B} + \Delta T_{B}) - P_{OFF}(T_{B}) \). Remarkably, this modulation takes place without any energy exchange with the environment. Therefore, the amplification factor \( \alpha_{l} \) in Eq. (1) diverges. This makes the device an ideal all-thermal transistor. We note that it could also be used as a perfect and noninvasive thermometer: Changes in the temperature of the environment are measured in the modulation of a (electrically or thermally generated) charge current, \( \Delta I_{l} = -I_{ON}^{C} \Delta P_{OFF} \).

**Implementations.** Let us now specify a possible physical realization of the system/thermal gate partition. After presenting a description of the heat currents, we show how the model can be mapped on the ideal three-state case described above in the appropriate configuration.

The minimal model for transport consists of a single site that connects the two terminals \( l = E, C \). This site can be empty or occupied. The thermal gate consists of a similar site which is coupled to reservoir B. Both subsystems are allowed to fluctuate between two states due to the coupling to three different terminals [Fig. 1(c)].

In ultrasmall devices with strong Coulomb interactions, each site can be assumed to be occupied by up to one electron. Four states \( |N,n⟩ \) are thus relevant, with \( N, n = 0 \) being the charge of the system and of the thermal gate. For simplicity, we consider single-level systems as can be found, for example, in semiconductor quantum dots [1,20,25]. We assume the weak-coupling limit \( \bar{\epsilon}_{n} = 10T_{B,N} = 10 \mu eV/\hbar \). Solid lines mark the degeneracies of the different occupation probabilities \( P_{N,n} \) which cross at triple points. A negative contribution in the rightmost panel is cut off for clarity.

Even if the gating is larger in that region, the amplification is smaller as a function of the interaction \( \alpha_{i} \). Energy-dependent tunneling in the gate reduces the heat current. (c) Amplification factor as a function of the coupling \( V_{l} = T_{B}/2 \). \( \Delta T_{B} = T_{l}/2 \), \( \alpha_{l} = 0.1, \beta_{l} = 0.002 \).

In this model, even for small charge of the system and of the thermal gate, the device charging can be controlled experimentally by means of gate voltages \( V_{S} \) and \( V_{G} \) [cf. Fig. 2(a)]. Then, the energy of the respective site is \( \epsilon_{l} = \epsilon_{l}^{(0)} + \alpha_{l}eV_{l} + \beta_{l}eV_{l}(1 - \delta_{l}) \), where \( \epsilon_{l}^{(0)} \) is the on-site bare energy, and \( \alpha_{l}, \beta_{l} \) are constants given by cross capacitances [22].

We assume the weak-coupling limit \( \bar{\epsilon}_{l} T_{l} \ll k_{B}T_{l} \), where transport is well described by sequential tunneling rates which are quite generally energy dependent. \( \Gamma_{l}^{C} = \Gamma_{l}^{C} f_{l}^{C}(U_{l}^{C}) \). Here, \( U_{l}^{C} = \epsilon_{l}^{(0)} + E_{C} - \mu_{l} \), \( q \) is the occupation of the respective other subsystem, \( f_{l}^{C}(E) = \frac{1}{[1 + e^{(E-\mu_{l})/k_{B}T_{l}}]} \) is the Fermi function, and \( f_{l}^{-}(E) = 1 - f_{l}^{C}(E) \) [19]. For simplicity,
we write $\Gamma_l(U_{lq}) = \Gamma_{lq}$. We are mainly interested in configurations close to at least one triple point (with $|U_{lq}| < k_B T_k$ for every contact) where the current is enabled by charge fluctuations in both dots (cf. Fig. 2). There, charging and uncharging rates are of the same order in all barriers, $\Gamma_{lq}^+ \sim \Gamma_{lq}^-$. Higher-order tunneling effects [27] can be neglected in this limit.

We write four rate equations for the occupation of the different states, $P_{n,m}$ [22,28] whose stationary solution gives the state-resolved particle currents [29],

$$\mathcal{I}_{lq} = \Gamma_{lq}^- P_{l(0,q)} - \Gamma_{lq}^+ P_{l(1,q)}, \quad \text{for } l = \text{E,C}.$$  \hspace{1cm} (8)

For $\mathcal{I}_{Gq}$, replace $P_{l(q)} \rightarrow P_{l(q,i)}$. From them we obtain the charge $I_l = e \sum_q \mathcal{I}_{lq}$ and heat currents $J_l = \sum_q (U_{lq} - \mu_l) \mathcal{I}_{lq}$. We remark that both the energy and rates of each subsystem depend on the charge state of the other one.

In such a device, cyclic transitions exist which transfer an energy $E_C$ between the system and the gate [1,19,20,26]. They are detrimental for our purpose of gating the system with minimal heat exchange. These transitions are of the form $|0,0\rangle \leftrightarrow |1,0\rangle \leftrightarrow |1,1\rangle \leftrightarrow |0,1\rangle \leftrightarrow |0,0\rangle$ including all four charge states. Hence, energy transfer can be suppressed by selecting configurations where transitions in one of the systems are conditioned on the state of the other one, i.e., if $\Gamma_{lq}^0 = 0$ for all terminals $l$ in one partition with the other one being in state $n$. This effectively reduces the system to the three-state case discussed above. We have found two ways how this can be achieved: (i) by filtering some of the transitions in one of the partitions by highly energy-selective tunneling or (ii) by increasing the interaction energy $E_C$.

Let us first consider case (i) with, e.g., $\Gamma_{\text{B0}} = 0$, such that transitions $|0,0\rangle \leftrightarrow |0,1\rangle$ are avoided. This extremely energy-dependent tunneling can be achieved, e.g., by introducing a resonance (a second quantum dot) or a gap in the contact with the base. Then state-resolved currents are conserved: It is clear that $\mathcal{I}_{\text{B0}} = 0$ and (from charge conservation) also $\mathcal{I}_{\text{B1}} = \mathcal{I}_{\text{Cn}}$ [29]. Therefore, from Eq. (8) we have $J_{lq} = 0$, up to higher-order tunneling corrections [27], neglected here.

Note that filtered transitions furthermore suppress the eventual contribution of nonlocal cotunneling of the form $|1,0\rangle \leftrightarrow |0,1\rangle$. Despite the absence of an energy exchange, the state of the system is still sensitive to the occupation in the thermal gate. Thus the particle current reads

$$\mathcal{I}_C = \mathcal{I}_{C\text{iso}}^\text{(E)}(U_{S1})(\hat{n}) + \mathcal{I}_{C\text{iso}}^\text{(E)}(U_{S0})(1 - \langle \hat{n} \rangle), \quad \text{in terms of } \langle \hat{n} \rangle = P_{l(0,1)} + P_{l(1,1)},$$

and the energy-resolved current of an isolated conductor [30],

$$\mathcal{I}_C^\text{(E)} = \left[ \frac{1}{\Gamma_{\text{F}}(E)} + \frac{1}{\Gamma_{\text{C}}(E)} \right]^{-1} \left[ f_C^<(E) - f_{\text{F}}^<(E) \right]. \quad \hspace{1cm} (10)$$

The system switches between two currents depending on the state of the gate: $\mathcal{I}_{C\text{iso}}^\text{(E)}(U_{S1})$, when the gate is occupied, and $\mathcal{I}_{C\text{iso}}^\text{(E)}(U_{S0})$, when the gate is empty. Note that the thermodynamic state of the base reservoir only enters through the average occupation of the thermal gate $\langle \hat{n} \rangle$. Hence we obtain a switching effect in the system which is solely driven by the fluctuations of the gate,

$$\Delta I_l = \left[ \mathcal{I}_{l\text{iso}}^\text{(E)}(U_{S1}) - \mathcal{I}_{l\text{iso}}^\text{(E)}(U_{S0}) \right] \langle \hat{n}(\Delta T_B) - \hat{n}(0) \rangle.$$  \hspace{1cm} (11)

We emphasize that switching takes place with only negligible leakage from the base terminal, thus leading to a large amplification factor.

The effect of thermal gating even improves for the second case (ii), where we control the interaction energy [see Fig. 2(a)]. Increasing $E_C$ can be done experimentally by electrostatic bridging of the two systems [31,32] or in stacked two-dimensional materials, e.g., graphene [20]. We start by considering the strong interaction limit, $E_C \gg k_B T$. We can then choose a configuration close to a triple point for which fluctuations occur in sequences of the form $|N \pm 1,n\rangle \leftrightarrow |N,n\rangle \leftrightarrow |N,n \pm 1\rangle$ with either $N = n = 0 (+)$ or $N = n = 1 (-)$ [24]. Hence we have the seemingly paradoxical consequence that the transferred heat vanishes when the interaction of the two systems is large, $J_B \rightarrow 0$ for $E_C \rightarrow \infty$ [cf. Fig. 2(b)].

Indeed, this limit can be mapped to the ideal three-state system discussed above, if $|\text{ON}| = |N,n\rangle$, $|\text{OFF}| = |N,n \pm 1\rangle$, and $|0\rangle = |N \pm 1,n\rangle$, with $\Delta E = \pm (U_{\text{on}} - \mu_B)$, and $n$ given by the charge occupation of the (ON) state. The switching transitions are now given by $\Gamma_{\pm \text{B0}}$, which obviously satisfy local detailed balance.

We therefore recover the solution obtained in Eq. (4). Then the linear thermal gating can be expressed as

$$\frac{1}{J_l} = \frac{U_{\text{on}} - \mu_B}{k_B T_B^2 (n - \langle \hat{n} \rangle)},$$

for a small gradient $\Delta T_B$ but arbitrary temperature configuration $\{T_i\}$.

Figure 2(b) shows the two ways for minimizing the exchanged energy $J_B$ discussed above: increasing $u = E_C / k_B T$ and introducing level-selected rates. An ideal behavior is found at $u \gg 1$ for strongly coupled systems or at low temperatures. For smaller $u$, undesired cyclic fluctuations involving all the four states start to contribute. These can, however, be filtered by preventing some tunneling transitions, e.g., by making $\Gamma_{\text{B0}} \ll \Gamma_{\text{B1}}$, as discussed above. In Fig. 2(c) we plot the interplay of these two mechanisms. As expected, the amplification factor increases for $\Gamma_{\text{B0}} / \Gamma_{\text{B1}} \rightarrow 0$. For small but finite ratios we observe a nonmonotonic behavior as the interaction $u$ increases, with a larger amplification for small $u$. Note that it remains almost constant for large $u$, i.e., when the base current is suppressed. This is an indication that the thermal gating is purely induced by fluctuations that only depend on the temperature $T_B$. Remarkably, the largest amplification is found in configurations (●) close to one triple point, where sequential tunneling dominates even if one relaxes the weak-coupling assumption.

Energy consumption. Even if no heat flows through the contact to the base terminal, in order to operate the transistor, one has to repeatedly inject (remove) a finite amount of heat in order to increase (reduce) its temperature. In this case the operation speed of the transistor is limited by the time scale of thermalization of the base terminal. However, as thermal gating is driven by charge fluctuations, these can be experimentally frozen by simply closing the tunneling barrier with a plunger gate [1] which does not only allow for much higher switching rates but also can be done at an arbitrarily low energy cost.

Alternatively, the gate system can be coupled to two reservoirs, one at temperature $T$ and the other one at $T + \Delta T_B$, as depicted in Fig. 3. By opening or closing either of the two barriers, the fluctuations will be governed by either of
The control of mesoscopic fluctuations allows for huge amplification factors. Either all-thermal transistors or noninvasive thermometers can be operated by this mechanism, depending on which current (heat or charge) is measured in the collector. Nonthermal fluctuations in the gate system allow for fast switching of the gate temperature at arbitrarily low energy cost. Considering the gate as a separate system is also beneficial in opening the way to gating at a distance and reducing heat leaking. The simplicity of our idea makes it easy exportable to different kinds of systems and interactions (e.g., spin fluctuations [33]). We have particularized the electrostatic interaction of single-electron devices in the sequential tunneling regime, which has recently been demonstrated experimentally [1,3,24,25]. Large amplification is found for configurations where undesired higher-order processes [20,21,27] are marginal, suggesting a way for the modulation of larger currents in stronger-coupling regimes.

Acknowledgments. We acknowledge financial support from the Spanish Ministerio de Economía, Industria y Competitividad via Grants No. MAT2014-58241-P and No. FIS2015-74472-JIN (AEI/FEDER/UE), and the European Research Council Advanced Grant No. 339306 (METIQUM). We also thank the COST Action MP1209 “Thermodynamics in the quantum regime”.


