Maxwell's Refrigerator: An Exactly Solvable Model

Dibyendu Mandal,[1] H. T. Quan,[2,3] and Christopher Jarzynski[2,4]

[1]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[2]*Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA*
[3]*School of Physics, Peking University, Beijing 100871, China*
[4]*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*
(Received 8 April 2013; revised manuscript received 29 May 2013; published 19 July 2013)

We describe a simple and solvable model of a device that—like the ''neat-fingered being'' in Maxwell's famous thought experiment—transfers energy from a cold system to a hot system by rectifying thermal fluctuations. In order to accomplish this task, our device requires a memory register to which it can write information: the increase in the Shannon entropy of the memory compensates the decrease in the thermodynamic entropy arising from the flow of heat against a thermal gradient. We construct the nonequilibrium phase diagram for this device, and find that it can alternatively act as an eraser of information. We discuss our model in the context of the second law of thermodynamics.

DOI: 10.1103/PhysRevLett.111.030602 PACS numbers: 05.70.Ln, 05.40.−a, 05.90.+m, 89.70.Cf

In a thought experiment highlighting the statistical nature of the second law of thermodynamics, Maxwell imagined a tiny creature acting as a gatekeeper between two chambers filled with gases at different temperatures. By preferentially allowing fast-moving molecules to pass from the cold to the hot chamber, and slow ones to pass in the other direction, this creature achieves refrigeration without expending energy. As Maxwell put it, ''the hot system has got hotter and the cold colder and yet no work has been done, only the intelligence of a very observant and neat-fingered being has been employed'' [1].

In this Letter we propose a simple, solvable model of a physical device that accomplishes the same result as Maxwell's intelligent and observant creature: it creates a flow of energy against a thermal gradient, without the input of external work. Our device is a classical two-state system that interacts with a pair of thermal reservoirs and a memory register, which we model as a stream of bits [Fig. 1(a)]. The dynamics consist of stochastic transitions, by means of which the device exchanges energy with the reservoirs and modifies the states of the bits. For appropriate values of the model parameters, these dynamics produce a steady state in which there is a continual flow of energy from the cold reservoir to the hot reservoir, and a record of the system's microscopic evolution is continually written to the stream of bits. Our device is fully autonomous, requiring no intervention by an external agent. Its ability to control the flow of energy between the reservoirs emerges entirely from the microscopic equations of motion.

The term ''Maxwell's demon'' has come to refer not only to the original setting described by Maxwell, but more generally to any situation in which a rectification of microscopic fluctuations produces a decrease of thermodynamic entropy [2,3]. A consensus has emerged that a physical device could achieve such a result, without violating the second law, if it were simultaneously to write information to a memory register [4–8]. In this view, the act of writing increases the information entropy of the memory register, thereby compensating the decrease of thermodynamic entropy produced by the device. If the information is later erased from the memory register, then by Landauer's principle [4,9] there must be an increase in thermodynamic entropy elsewhere. This tidy accounting places the Shannon entropy of a sequence of bits on the same thermodynamic footing as the Clausius entropy, defined in terms of heat and temperature. As long as the sum of these entropies never decreases, the second law remains satisfied. See, however, Refs. [10–13] for dissenting perspectives, which suggest that this consensus is at best an appealing narrative based on the presupposition of the second law, rather than an independent explanation.

Maxwell's demon has recently enjoyed increased attention in a broad range of settings, including artificial molecular machines [14], single photon cooling of atoms [15], biomolecular signal transduction [16], quantum information theory [17] and the feedback control of microscopic fluctuations [18–33]. Maxwell's 19th-century thought experiment has become a touchstone for discussing the thermodynamic implications of information processing by physical systems [34–37]. While the consensus described above has identified and clarified these implications, far less effort has been devoted to uncovering precisely *how* a physical device, acting on its own, might accomplish the same result as Maxwell's hypothetical being [38–43]. To the best of our knowledge, the autonomous model we introduce below is the first to generate a flow of energy against a thermal gradient, effectively acting as a refrigerator without a power supply—just as in the setup considered by Maxwell, but with the intelligent creature replaced by a dumb device. This contrasts with an earlier model of a device that acts as an *engine*, supplying work by extracting heat from a single thermal reservoir [40].

 © 2013 American Physical Society

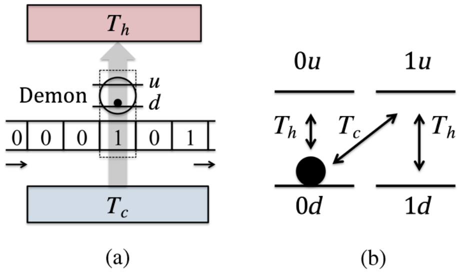

FIG. 1 (color online). (a) The device, or "demon," interacts with a sequence of bits, one at a time, while exchanging energy with two thermal reservoirs. (b) The demon makes intrinsic transitions mediated by the hot reservoir (vertical arrows), and the demon and nearest bit make cooperative transitions $0d \leftrightarrow 1u$ mediated by the cold reservoir (diagonal arrows).

Our autonomous framework also differs from that of Refs. [18–33] (including the experimental realization reported in Ref. [26]), in which external intervention in the form of measurement and feedback is a key element.

In what follows we describe our model and analyze its dynamics. We obtain a nonequilibrium phase diagram for the steady state behavior (Fig. 2), which reveals that our device can act either as a refrigerator, transferring energy from a cold to a hot reservoir, or as an eraser, decreasing the information content of the memory register. Finally, we briefly discuss our model in the context of the second law of thermodynamics.

Our model consists of four components, sketched in Fig. 1(a): a memory register, two thermal reservoirs at temperatures T_c and $T_h > T_c$, and a device that plays the role of Maxwell's demon. The memory register is a sequence of bits (two-state systems) spaced at equal intervals along a tape that slides frictionlessly past the demon. The demon interacts with the nearest bit and with the reservoirs, as we describe in detail in the following paragraphs.

The demon itself is a two-state system, with states u and d characterized by an energy difference $\Delta E = E_u - E_d > 0$. It can make random transitions between these two states by exchanging energy with the hot reservoir, as illustrated by the vertical arrows in Fig. 1(b). We will refer to these as intrinsic transitions, to emphasize that they involve the demon but not the bits. The corresponding transition rates satisfy the requirement of detailed balance [44],

$$\frac{R_{d\to u}}{R_{d\leftarrow u}} = e^{-\beta_h \Delta E}, \tag{1}$$

where $\beta_h = 1/kT_h$ and k is Boltzmann's constant. We parametrize these rates as

$$R_{d\to u} = \gamma(1-\sigma), \quad R_{d\leftarrow u} = \gamma(1+\sigma), \quad \sigma = \tanh\frac{\beta_h \Delta E}{2} \tag{2}$$

where $\gamma > 0$ sets a characteristic rate for these transitions, and $0 < \sigma < 1$.

Each bit has two states, 0 and 1, with equal energies. We assume there are no intrinsic transitions between these two states. That is, the state of the bit can change only via interaction with the demon, as we now discuss.

At any instant in time, the demon interacts only with the nearest bit. As a result, it interacts sequentially with the bits as they pass by. The duration of interaction with each bit is $\tau = l/v$, where l is the spacing between bits and v is the constant speed of the tape. During one such interaction interval, the demon and the nearest bit can make cooperative transitions: if the bit is in state 0 and the demon is in state d, then they can simultaneously flip to states 1 and u, and vice versa [Fig. 1(b), diagonal arrows]. We will use the notation $0d \leftrightarrow 1u$ to denote these transitions, which are accompanied by an exchange of energy with the cold reservoir. The corresponding transition rates again satisfy detailed balance, $R_{0d\to 1u}/R_{0d\leftarrow 1u} = e^{-\beta_c \Delta E}$, where $\beta_c = 1/kT_c$, and we will parametrize them as follows [45]:

$$R_{0d\to 1u} = 1 - \omega, \quad R_{0d\leftarrow 1u} = 1 + \omega, \quad \omega = \tanh\frac{\beta_c \Delta E}{2}, \tag{3}$$

with $0 < \omega < 1$. For later convenience, we also define

$$\epsilon = \frac{\omega - \sigma}{1 - \omega\sigma} = \tanh\frac{(\beta_c - \beta_h)\Delta E}{2}, \tag{4}$$

whose value, $0 < \epsilon < 1$, quantifies the temperature difference between the two reservoirs.

Finally, we assume that the incoming bit stream contains a mixture of 0's and 1's, with probabilities p_0 and p_1, respectively, with no correlations between bits. Let

$$\delta \equiv p_0 - p_1 \tag{5}$$

denote the proportional excess of 0's among incoming bits.

We thus have the following dynamics. When a fresh bit arrives to interact with the demon, its state is 0 or 1. The demon and bit subsequently interact for a time τ, making the transitions shown in Fig. 1(b), thereby exchanging energy with the reservoirs. The state of the bit at the end of the interaction interval is then preserved as the bit joins the outgoing stream, and the next bit in the sequence moves in to have its turn with the demon. The parameters γ, σ, and ω define the intrinsic and cooperative transition rates [Eqs. (2) and (3)], τ gives the duration of interaction with each bit, and δ specifies the statistics of the incoming bits. Under these dynamics, the demon evolves to a periodic steady state, in which its behavior is statistically the same from one interaction interval to the next.

Before proceeding to the solution of these dynamics, we discuss heuristically how our model can achieve the systematic transfer of heat from the cold to the hot reservoir. For this purpose let us assume that each incoming bit is in state 0, hence $\delta = 1$. At the start of a particular interaction interval, the joint state of the demon and newly arrived bit is either $0u$ or $0d$. The demon and bit then evolve together for a time τ, according to the transitions shown in Fig. 1(b). If the joint state at the end of the interaction interval is $0u$ or $0d$, then it must be the case that every transition $0d \rightarrow 1u$ was balanced by a transition $0d \leftarrow 1u$; hence, no net energy was absorbed from the cold reservoir. If the final state is $1u$ or $1d$, then we can infer that there was one net transition from $0d$ to $1u$, and a quantity of energy ΔE was absorbed from the cold reservoir. This amounts to thermal rectification: over the course of one interaction interval, energy can be withdrawn from the cold reservoir but not delivered to it. Moreover, a record of this process is imprinted in the bit stream, as every outgoing bit in state 1 indicates the absorption of energy ΔE from the cold reservoir. Since the demon also exchanges energy with the hot reservoir, and since energy cannot accumulate indefinitely within the demon, in the long run we get a net flux of energy from the cold to the hot reservoir, proportional to the rate at which 1's appear in the outgoing bit stream.

More generally, if the incoming bit stream contains a mixture of 0's and 1's, then an excess of 0's (that is, $\delta > 0$) produces a statistical bias that favors the flow of heat from the cold to the hot reservoir, while an excess of 1's ($\delta < 0$) produces the opposite bias. This bias either competes with or enhances the normal thermodynamic bias due to the temperature difference between the two reservoirs. The demon thus affects the flow of energy between the reservoirs, and modifies the states of the bits in the memory register. We now investigate quantitatively the interplay between these two effects.

Once the demon has reached its periodic steady state, let p'_0 and p'_1 denote the fractions of 0's and 1's in the outgoing bit stream, and let $\delta' = p'_0 - p'_1$ denote the excess of outgoing 0's. Then

$$\Phi \equiv p'_1 - p_1 = \frac{\delta - \delta'}{2} \qquad (6)$$

represents the average production of 1's per interaction interval in the outgoing bit stream, relative to the incoming bit stream. Since each transition $0 \rightarrow 1$ is accompanied by the absorption of energy ΔE from the cold reservoir [Fig. 1(b)], the average transfer of energy from the cold to the hot reservoir, per interaction interval, is given by

$$Q_{c\rightarrow h} = \Phi \Delta E. \qquad (7)$$

A positive value of $Q_{c\rightarrow h}$ indicates that our device pumps energy against a thermal gradient, like the creature imagined by Maxwell.

To quantify the information-processing capability of the demon, let

$$S(\delta) = -\sum_{i=0}^{1} p_i \ln p_i = -\frac{1-\delta}{2}\ln\frac{1-\delta}{2} - \frac{1+\delta}{2}\ln\frac{1+\delta}{2} \qquad (8)$$

denote the information content, per bit, of the incoming bit stream, and define $S(\delta')$ by the same equation, for the outgoing bit stream. Then

$$\Delta S_B \equiv S(\delta') - S(\delta) = S(\delta - 2\Phi) - S(\delta) \qquad (9)$$

provides a measure of the extent to which the demon increases the information content of the memory register. We will interpret a positive value of ΔS_B to indicate that the demon writes information to the bit stream, while a negative value indicates erasure. (More precisely, since $S(\delta')$ neglects the small correlations that arise between the outgoing bits, ΔS_B reflects the change in the Shannon information of the marginal probability distribution of each outgoing bit).

From Eqs. (7) and (9) we see that Φ determines both $Q_{c\rightarrow h}$ and ΔS_B. In the Supplemental Material [46], we show that under the dynamics we have described, the demon reaches a periodic steady state, determined by the model parameters $\Lambda \equiv (\delta, \sigma, \gamma, \omega, \tau)$, in which

$$\Phi(\Lambda) = \frac{\delta - \epsilon}{2}\eta(\Lambda), \qquad \eta > 0 \qquad (10)$$

and

$$Q_{c\rightarrow h}(\beta_h - \beta_c) + \Delta S_B \geq 0. \qquad (11)$$

Equation (11) is a strict inequality when $\delta \neq \epsilon$. An explicit expression for $\eta(\Lambda)$ is given in the Supplemental Material [46], but for our present purposes the crucial point is that the sign of Φ is the same as that of $\delta - \epsilon$. We can think of two effective forces: the bias induced by the incoming bit stream, which favors $\Phi > 0$ when $\delta > 0$ (as discussed above), and the temperature gradient, quantified by ϵ, which favors $\Phi < 0$ [Eq. (7)]. When these compete, the winner is determined by the difference $\delta - \epsilon$.

Equation (10) is obtained by solving for the periodic steady state of the demon, using a linear-algebraic approach. Equation (11) is obtained by constructing a Lyapunov function for the demon and interacting bit. The details of these derivations are provided in the Supplemental Material [46]. Here, we instead use these results to investigate the behavior of our model in the periodic steady state. To that end, we fix γ and ω and construct a phase diagram that illustrates the dependence on δ and ϵ, for various values of τ, shown in Fig. 2. Let us consider the different regions of this diagram, working our way from right to left.

From Eqs. (7) and (10) it follows that $Q_{c\rightarrow h} > 0$ when $\delta > \epsilon$, shown as the most darkly shaded region in Fig. 2. Here, a surplus of incoming 0's prevails over the temperature difference and our demon generates a flow of energy

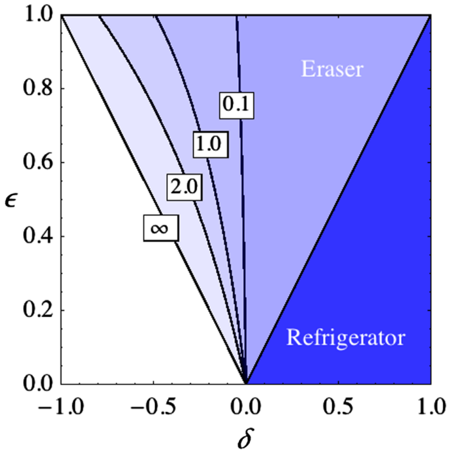

FIG. 2 (color online). Phase diagram of our model at fixed $\gamma = 1$ and $\omega = 1/2$. The parameter δ specifies the incoming bit statistics, and ϵ is a rescaled temperature difference [Eq. (4)]. In the most darkly shaded region the demon acts as a refrigerator ($Q_{c\to h} > 0$), while in the lightly shaded regions it acts as an eraser ($\Delta S_B < 0$). The left boundary of the eraser region is shown for $\tau = 0.1$, 1.0, 2.0, and ∞. In the blank region at the lower left, our model exhibits neither behavior (see text).

from the cold to the hot reservoir. Moreover, Eq. (11) reveals that $\Delta S_B > 0$ in this region (since $\beta_h < \beta_c$). This agrees with the consensus described earlier: in order for a physical device to act in the manner of Maxwell's demon, it must write information to a physical memory register. In this sense, a bit stream with a low information content can be viewed as a thermodynamic resource, which can be expended (by writing to the available memory) in order to achieve refrigeration.

Now consider the region $\epsilon > \delta > 0$, in which the surplus of 0's in the incoming bit stream is not sufficient to overcome the temperature gradient, and energy flows from the hot to the cold reservoir. Since $\Phi < 0$ we get $\delta' > \delta > 0$ [Eq. (6)]. This in turn implies $\Delta S_B < 0$, as $S(\delta)$ is a concave function with a maximum at $\delta = 0$. In this region the demon acts as an eraser, lowering the information content of the bit stream, but the price paid for this erasure is the passage of heat from the hot to the cold reservoir.

In the region $\delta < 0$, energy flows from the hot to the cold reservoir [Eqs. (7) and (10)], but the value of ΔS_B depends on all the model parameters. In Fig. 2, for four different values of τ, we show the line corresponding to $\Delta S_B = 0$. To the right of this line we have $\Delta S_B < 0$ and to the left we have $\Delta S_B > 0$. In the limit $\tau \to \infty$, the boundary between these two behaviors approaches the line $\epsilon = -\delta$.

Examining the phase diagram as a whole, we see that in the shaded regions our model reaches a steady state in which one thermodynamic resource is replenished at the expense of another. Either energy is pumped against a thermal gradient at the cost of writing information to memory (the refrigerator regime), or else memory is made available, by erasure, at the expense of allowing energy to flow from the hot to the cold reservoir (the eraser regime). The boundary between these two behaviors is the line $\delta = \epsilon$. In the unshaded region at the far left, both resources are consumed, as energy flows down the thermal gradient and information is written to the bit stream.

Finally, to place our model within the context of the second law of thermodynamics, note that the first term on the left side of Eq. (11) is the steady-state change in thermodynamic entropy due to the flow of heat, and the second term is the change in information entropy, per interaction interval. Equation (11) can be viewed as a modified Clausius inequality, in which the information entropy of a random sequence of data is explicitly assigned the same thermodynamic status as the physical entropy associated with the transfer of heat. (More precisely, Eq. (11) is a weak version of this inequality, as we neglect correlations among the outgoing bits; see Supplemental Material [46]). Thus our model provides support for the consensus mentioned earlier [4–6], and Eq. (11) is consistent with Landauer's principle [4], which states that a thermodynamic cost must be paid for the erasure of memory. However, in Ref. [4] this cost appears as the dissipation of energy into a single thermal reservoir, whereas in our model it is the transfer of energy from a hot to a cold reservoir.

In summary, we have constructed a simple, solvable model of an autonomous physical system that mimics the behavior of the "neat-fingered being" in Maxwell's thought experiment, generating a systematic flow of energy against a thermal gradient without the input of external work. While Maxwell's creature accomplishes this with intelligence, our inanimate device requires only a memory register to which information can be written. Alternatively, it can harness the flow of energy from hot to cold in order to erase information from the register.

We thank Andy Ballard, Shaon Chakrabarti, Sebastian Deffner, and Zhiyue Lu for useful discussions, and gratefully acknowledge financial support from the National Science Foundation (USA) under Grants No. DMR-0906601, No. ECCS-0925365, and No. DMR-1206971, the University of Maryland, College Park, and Peking University.

[1] *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. F. Rex (Institute of Physics Publishing, Bristol, 2003), p. 4.
[2] J. C. Maxwell, *Theory of Heat* (Longmans, London, 1871).
[3] L. Szilard, Z. Phys. **53**, 840 (1929).
[4] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[5] O. Penrose, *Foundations of Statistical Mechanics: A Deductive Treatment* (Pergamon Press, Oxford, 1970).
[6] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[7] C. H. Bennett and R. Landauer, Sci. Am. **253**, 48 (1985).
[8] O. Maroney, "Information Processing and Thermodynamic Entropy," in The Stanford Encyclopedia of

Philosophy, edited by E. N. Zelta, http://plato.stanford.edu/archives/fall2009/entries/information-entropy/, Fall 2009 ed.
[9] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Nature (London) **483**, 187 (2012).
[10] J. Earman and J. D. Norton, Stud. Hist. Phil. Mod. Phys. **29**, 435 (1998).
[11] J. Earman and J. D. Norton, Stud. Hist. Phil. Mod. Phys. **30**, 1 (1999).
[12] M. Hemmo and O. Shenker, J. Philos. **107**, 389 (2010).
[13] J. D. Norton, Stud. Hist. Phil. Mod. Phys. **42**, 184 (2011).
[14] E. R. Kay, D. A. Leigh, and F. Zerbetto, Angew. Chem., Int. Ed. Engl. **46**, 72 (2007), and references therein.
[15] M. G. Raizen, Sci. Am. **304**, 54 (2011).
[16] Y. Tu, Proc. Natl. Acad. Sci. U.S.A. **105**, 11 737 (2008).
[17] L. del Rio, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, Nature (London) **474**, 61 (2011).
[18] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).
[19] T. Sagawa and M. Ueda, Phys. Rev. Lett. **100**, 080403 (2008).
[20] T. Sagawa and M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009).
[21] K. Jacobs, Phys. Rev. A **80**, 012322 (2009).
[22] F. J. Cao, M. Feito, and H. Touchette, Physica (Amsterdam) **388A**, 113 (2009).
[23] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).
[24] M. Ponmurugan, Phys. Rev. E **82**, 031129 (2010).
[25] J. M. Horowitz and S. Vaikuntanathan, Phys. Rev. E **82**, 061120 (2010).
[26] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
[27] D. Abreu and U. Seifert, Europhys. Lett. **94**, 10 001 (2011).
[28] S. Vaikuntanathan and C. Jarzynski, Phys. Rev. E **83**, 061120 (2011).
[29] H. Dong, D. Z. Xu, C. Y. Cai, and C. P. Sun, Phys. Rev. E **83**, 061108 (2011).
[30] T. Sagawa, http://ptp.ipap.jp/link?PTP/127/1/.
[31] K. Jacobs, Phys. Rev. E **86**, 040106(R) (2012).
[32] D. Abreu and U. Seifert, Phys. Rev. Lett. **108**, 030601 (2012).
[33] T. Sagawa and M. Ueda, Phys. Rev. Lett. **109**, 180602 (2012).
[34] W. H. Zurek, Nature (London) **341**, 119 (1989).
[35] J. Bub, Stud. Hist. Phil. Mod. Phys. **32**, 569 (2001).
[36] K. Maruyama, F. Nori, and V. Vedral, Rev. Mod. Phys. **81**, 1 (2009).
[37] A. Hosoya, K. Maruyama, and Y. Shikano, Phys. Rev. E **84**, 061117 (2011).
[38] H. T. Quan, Y. D. Wang, Y.-X. Liu, C. P. Sun, and F. Nori, Phys. Rev. Lett. **97**, 180402 (2006).
[39] M. Bier and F. J. Cao, Acta Phys. Pol. B **43**, 889 (2012).
[40] D. Mandal and C. Jarzynski, Proc. Natl. Acad. Sci. U.S.A. **109**, 11 641 (2012).
[41] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Phys. Rev. Lett. **110**, 040601 (2013).
[42] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, arXiv:1210.6448v1.
[43] A. C. Barato and U. Seifert, Europhys. Lett. **101**, 60 001 (2013).
[44] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007), Chap. V, 3rd ed.
[45] Note the lack of a rate parameter analogous to γ in Eq. (2). For the cooperative transition rates, we set this parameter to unity by appropriately choosing the unit of time.
[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.030602 for detailed derivations of Eqs. (10) and (11).