

# Quantum Bochkov–Kuzovlev work fluctuation theorems

BY MICHELE CAMPISI, PETER TALKNER\* AND PETER HÄNGGI

*Institut für Physik, Universität Augsburg, Universitätsstrasse 1,  
D-86135 Augsburg, Germany*

The quantum version of the Bochkov–Kuzovlev identity is derived on the basis of the appropriate definition of work as the difference of the measured internal energies of a quantum system at the beginning and the end of an external action on the system given by a prescribed protocol. According to the spirit of the original Bochkov–Kuzovlev approach, we adopt the ‘exclusive’ viewpoint, meaning that the coupling to the external work source is *not* counted as part of the internal energy. The corresponding canonical and microcanonical quantum fluctuation theorems are derived as well, and are compared with the respective theorems obtained within the ‘inclusive’ approach. The relations between the quantum inclusive work  $w$ , the exclusive work  $w_0$  and the dissipated work  $w_{\text{dis}}$ , are discussed and clarified. We show by an explicit example that  $w_0$  and  $w_{\text{dis}}$  are distinct stochastic quantities obeying different statistics.

## 1. Introduction

One of the main objectives of *non-equilibrium thermodynamics* is the study of the response of physical systems to applied external perturbations. Around the middle of the last century, major advancements were obtained in this field with the development of linear-response theory by several authors, among which we mention Callen & Welton [1], Green [2] and Kubo [3]. This theory inspired by the works of Einstein [4] on the Brownian movement and of Johnson [5] and Nyquist [6] on noise in electrical circuits, established that, under certain circumstances, the linear response to small perturbations is determined by the equilibrium fluctuations of the system. In particular, the linear-response coefficients are proportional to two-point correlation functions for Hamiltonian systems [3], as well as for stochastic, generally non-equilibrium systems [7]. In principle, an infinite hierarchy of higher order fluctuation–dissipation relations connects the  $n$ th order response coefficients to  $(n + 1)$ -point correlation functions.

In contrast, fluctuation theorems are compact relations that provide information about the complete *nonlinear response*. Accordingly, fluctuation–dissipation relations of all orders can be derived therefrom. Bochkov & Kuzovlev [8,9] were the first to put forward one such complete nonlinear fluctuation

\*Author for correspondence (peter.talkner@physik.uni-augsburg.de).

theorem. These authors noticed that, for a *classical* system, their general fluctuation theorem implies the following extremely simple non-equilibrium identity:

$$\langle e^{-\beta W_0} \rangle = 1, \quad (1.1)$$

where  $W_0$  is the work done on the system by the external perturbation during one specific realization thereof,  $\langle \cdot \rangle$  denotes the average over many realizations of the same perturbation and  $\beta = (k_B T)^{-1}$ , with  $T$  being the initial temperature of the system and  $k_B$  being Boltzmann's constant. Owing to the properties of convexity of the exponential function, an almost immediate consequence of equation (1.1), is the second law of thermodynamics in the form  $\langle W_0 \rangle \geq 0$ ; i.e. when a system is perturbed from an initial thermal equilibrium, on average, it can only absorb energy.

The works of Bochkov & Kuzovlev [8,9] have recently re-gained a great deal of attention, after Jarzynski [10] derived, within the framework of classical mechanics, a salient result similar to equation (1.1),

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}, \quad (1.2)$$

which, in contrast to equation (1.1), allows the extraction of an equilibrium property of the system, i.e. its free energy (difference)  $F$ , from non-equilibrium fluctuations of work  $W$ . Evidently, the definitions of work adopted by Jarzynski [10] and Bochkov & Kuzovlev [8,9] (denoted here, respectively, as  $W$  and  $W_0$ ) do not coincide. The relationships between these two work definitions and the corresponding non-equilibrium identities, equations (1.1) and (1.2), were discussed in a very clear and elucidating manner in Jarzynski & Horowitz [11] and Jarzynski [12], which, for the sake of clarity, we shall summarize below.

Let us express the time-dependent Hamiltonian of the driven classical system as the sum of the unperturbed system Hamiltonian  $H_0$  and the interaction energy stemming from the coupling of the external time-dependent perturbation  $X(t)$  to a certain system observable  $Q$ ,

$$H(\mathbf{q}, \mathbf{p}; t) = H_0(\mathbf{q}, \mathbf{p}) - X(t)Q(\mathbf{q}, \mathbf{p}). \quad (1.3)$$

We restrict ourselves to the simplest case of a protocol governed by a single 'field'  $X(t)$  coupling to the conjugate generalized coordinate  $Q$ . Generalization to several fields  $X_i(t)$  coupling to different generalized coordinates  $Q_i$  is straightforward.

The definition of work  $W$ , according to Jarzynski [10], stems from an *inclusive* viewpoint, where one counts the term  $X(t)Q$  as being a part of the system internal energy. By contrast, the definition of work  $W_0$ , according to Bochkov & Kuzovlev [8,9], belongs to an *exclusive* viewpoint where instead, this interaction term is not counted as part of the energy of the system. More explicitly, if  $\mathbf{q}_0, \mathbf{p}_0$  is a certain initial condition that evolves to  $\mathbf{q}_f, \mathbf{p}_f$  in a time  $t_f - t_0$ , according to the Hamiltonian evolution generated by  $H$ , then the two different definitions of work become

$$W \doteq H(\mathbf{q}_f, \mathbf{p}_f; t_f) - H(\mathbf{q}_0, \mathbf{p}_0; t_0) \quad (1.4)$$

and

$$W_0 \doteq H_0(\mathbf{q}_f, \mathbf{p}_f) - H_0(\mathbf{q}_0, \mathbf{p}_0). \quad (1.5)$$

It is important to stress that Bochkov & Kuzovlev [8,9] only obtained equation (1.1) in the *classical* case, notwithstanding the fact that they developed a quantum version of their theory, as well. This difficulty is related to the fact that work was identified by Bochkov & Kuzovlev [8,9] with the quantum expectation of a pretended *work operator*, given by the difference of final and initial Hamiltonian in the Heisenberg representation. To be more clear, if the quantum Hamiltonian reads

$$H(t) = H_0 - X(t)Q, \quad (1.6)$$

where now  $H, H_0$  and  $Q$  are Hermitian operators, the work operator was defined by Bochkov & Kuzovlev [8,9] as<sup>1</sup>

$$W_0 \doteq H_0^{\text{H}}(t_f) - H_0, \quad (1.7)$$

where the superscript ‘H’ denotes the Heisenberg picture. A similar approach was also employed within the inclusive viewpoint, with work defined as [13]

$$W = H^{\text{H}}(t_f) - H_0. \quad (1.8)$$

As pointed out clearly by some of us before with the work by Talkner *et al.* [14], the Jarzynski equality (1.2) *cannot* be obtained on the basis of the work operator (1.8). Likewise, the Bochkov–Kuzovlev identity (1.1) cannot be obtained on the basis of equation (1.7). It is by now clear that the impossibility of extending the classical results (1.1) and (1.2) on the basis of quantum work operators (1.7) and (1.8), respectively, is related to the fact that work characterizes a process, rather than a state of the system, and consequently cannot be represented by an observable that would yield work as the result of a single projective measurement. In contrast, energy must be measured twice in order to determine work, once before the protocol starts and a second time immediately after it has ended. The difference of the outcomes of these two measurements then yields the work performed on the system in a particular realization [14].

In this paper, we will adopt the exclusive viewpoint of Bochkov & Kuzovlev [8,9], but use the proper definition of work as the difference between the outcomes of two projective measurements of  $H_0$ , to obtain the quantum version of equation (1.1). Indeed, we will develop the theory of quantum work fluctuations within the exclusive two-point measurements viewpoint in great generality. In §2, we study the characteristic function of work. In §§3 and 4, we derive the exclusive versions of the Tasaki–Crooks quantum fluctuation theorem [15–17], and of the microcanonical quantum fluctuation theorem [18], respectively. Section 5 closes the paper with some remarks concerning the relationships between the inclusive work, the exclusive work and the dissipated work.

## 2. Characteristic function of work

As mentioned in §1, work is properly defined in quantum mechanics as the difference of the energies measured at the beginning and the end of the protocol, i.e. at times  $t_0$  and  $t_f > t_0$ , respectively. According to the exclusive viewpoint

<sup>1</sup>Bochkov & Kuzovlev [8] defined the ‘operator of energy absorbed by the system’  $E = \int_{t_0}^{t_f} X(\tau)Q^{\text{H}}(\tau) d\tau$ , where  $Q^{\text{H}}(\tau)$  is the operator  $Q$  in the Heisenberg representation. It is not difficult to prove that  $E$  coincides with  $W_0$  in equation (1.7).

which we adopt here, this energy is given by the unperturbed Hamiltonian  $H_0$ . Let  $e_n$  and  $|n, \lambda\rangle$  denote the eigenvalues and eigenvectors of  $H_0$ ,

$$H_0|n, \lambda\rangle = e_n|n, \lambda\rangle. \quad (2.1)$$

Here,  $n$  is the quantum number labelling the eigenvalues of  $H_0$  and  $\lambda$  denotes further quantum numbers needed to specify uniquely the state of the system, in case of degenerate energies. A measurement of  $H_0$  at time  $t_0$  gives a certain eigenvalue  $e_n$ , while a subsequent measurement of  $H_0$  at time  $t_f$  gives another eigenvalue  $e_m$ , so that

$$w_0 = e_m - e_n. \quad (2.2)$$

Evidently,  $w_0$  is a stochastic variable owing to the intrinsic randomness entailed in the quantum measurement processes and possibly in the statistical mixture nature of the initial preparation. In the following, we derive the statistical properties of  $w_0$ , in terms of its probability density function (pdf), and the associated characteristic function of work.

Let the system be prepared at time  $t < t_0$  in a certain state described by the density matrix  $\rho(t_0)$ . We further assume that the perturbation  $X(t)$  is switched on at a time  $t_0$ . At the same time, the first measurement of  $H_0$  is performed, with outcome  $e_n$ . This occurs with probability

$$p_n = \sum_{\lambda} \langle n, \lambda | \rho(t_0) | n, \lambda \rangle = \text{Tr } P_n \rho(t_0), \quad (2.3)$$

where  $P_n$  is the projector onto the eigenspace spanned by the eigenvectors belonging to the eigenvalue  $e_n$ ,

$$P_n = \sum_{\lambda} |n, \lambda\rangle \langle n, \lambda|, \quad (2.4)$$

and  $\text{Tr}$  denotes the trace over the Hilbert space. According to the postulates of quantum mechanics, immediately after the measurement, the system is found in the state

$$\rho_n = \frac{P_n \rho(t_0) P_n}{p_n}. \quad (2.5)$$

For times  $t > t_0$ , the system evolves according to

$$\rho_n(t) = U_{t,t_0} \rho_n U_{t,t_0}^\dagger, \quad (2.6)$$

with  $U_{t,t_0}$  denoting the unitary time-evolution operator obeying the Schrödinger equation governed by the full time-dependent Hamiltonian (equation (1.6))

$$i\hbar \partial_t U_{t,t_0} = H(t) U_{t,t_0}, \quad U_{t_0,t_0} = \mathbb{1}. \quad (2.7)$$

At time  $t_f$ , the second measurement of  $H_0$  is performed, and the eigenvalue  $e_m$  is obtained with probability

$$p(m|n) = \text{Tr } P_m \rho_n(t_f). \quad (2.8)$$

Therefore, the probability density to observe a certain value of work  $w_0$  is given by

$$p_{t_f,t_0}^0(w_0) = \sum_{m,n} \delta(w_0 - [e_m - e_n]) p(m|n) p_n. \quad (2.9)$$

We use the superscript ‘0’ throughout this paper to indicate the exclusive viewpoint. The same symbols without the superscript ‘0’ denote the respective quantities within the inclusive viewpoint.

The Fourier transform of the probability density of work gives the characteristic function of work

$$G_{t_f, t_0}^0(u) = \int dw_0 p_{t_f, t_0}^0(w_0) e^{i u w_0}, \quad (2.10)$$

which allows quick derivations of fluctuation theorems and non-equilibrium equalities. Performing calculations analogous to those reported by Talkner *et al.* [18], we find the characteristic function of work in the form of a two-point *quantum correlation function*

$$G_{t_f, t_0}^0(u) = \text{Tr} e^{i u H_0^{\text{H}}(t_f)} e^{-i u H_0} \bar{\rho}(t_0) \equiv \langle e^{i u H_0^{\text{H}}(t_f)} e^{-i u H_0} \rangle, \quad (2.11)$$

where  $\bar{\rho}(t_0)$  is defined as

$$\bar{\rho}(t_0) = \sum_n p_n \rho_n = \sum_n P_n \rho(t_0) P_n, \quad (2.12)$$

and the superscript ‘H’ stands for the Heisenberg representation, i.e.

$$H_0^{\text{H}}(t_f) = U_{t_f, t_0}^\dagger H_0 U_{t_f, t_0}. \quad (2.13)$$

This exclusive-work characteristic function  $G_{t_f, t_0}^0$  should be compared with the inclusive-work characteristic function  $G_{t_f, t_0}$  that is obtained when looking at the difference  $w$  of the outcomes  $E_n(t_0)$  and  $E_m(t_f)$  of measurements of the *total* time-dependent Hamiltonian  $H(t)$ . In this case, one obtains [14,18]

$$G_{t_f, t_0}(u) = \text{Tr} e^{i u H^{\text{H}}(t_f)} e^{-i u H_0} \bar{\rho}(t_0) \equiv \langle e^{i u H^{\text{H}}(t_f)} e^{-i u H_0} \rangle. \quad (2.14)$$

The difference lies in the distinct fact that  $H_0^{\text{H}}(t_f)$  appears in the exclusive approach in place of the full  $H^{\text{H}}(t_f)$ .

### (a) Reversed protocol

Next, consider the reversed protocol

$$\tilde{X}(t) = X(t_f + t_0 - t), \quad (2.15)$$

which consecutively assumes values as if time was reversed. Let  $\tilde{H}(t)$  be the resulting Hamiltonian,

$$\tilde{H}(t) = H_0 - \tilde{X}(t) Q. \quad (2.16)$$

The characteristic function of work now reads

$$\tilde{G}_{t_f, t_0}^0(u) = \text{Tr} e^{i u \tilde{H}_0^{\text{H}}(t_f)} e^{-i u H_0} \bar{\rho}(t_0) \equiv \langle e^{i u \tilde{H}_0^{\text{H}}(t_f)} e^{-i u H_0} \rangle, \quad (2.17)$$

where

$$\tilde{H}_0^{\text{H}}(t_f) = \tilde{U}_{t_f, t_0}^\dagger H_0 \tilde{U}_{t_f, t_0} \quad (2.18)$$

and  $\tilde{U}_{t_f, t_0}$  is the time-evolution operator generated by  $\tilde{H}(t)$ ,

$$i \hbar \partial_t \tilde{U}_{t, t_0} = \tilde{H}(t) \tilde{U}_{t, t_0}, \quad \tilde{U}_{t_0, t_0} = \mathbb{1}. \quad (2.19)$$

Assuming that the Hamiltonian  $H(t)$  is invariant under time reversal i.e.<sup>2</sup>

$$\Theta H(t) \Theta^{-1} = H(t), \quad (2.20)$$

where  $\Theta$  is the antiunitary time-reversal operator [20], the time-evolution operators associated to the forward and backward protocols are related by the following important relation (see appendix A):

$$U_{t_0, t_f} = U_{t_f, t_0}^\dagger = \Theta \tilde{U}_{t_f, t_0} \Theta^{-1}. \quad (2.21)$$

In §3, we will derive the quantum version of equation (1.1) and its associated work fluctuation theorem. This will be accomplished by choosing the initial density matrix to be a Gibbs canonical state. In §4, we will, instead, assume an initial microcanonical state.

### 3. Canonical initial state

For a system staying at time  $t_0$  in a canonical Gibbs state,

$$\rho(t_0) = \bar{\rho}(t_0) = \frac{e^{-\beta H_0}}{Z_0}, \quad (3.1)$$

where  $Z_0 = \text{Tr} e^{-\beta H_0}$  and  $\bar{\rho}(t_0)$  coincides with  $\rho(t_0)$  because the latter is diagonal with respect to the eigenbasis of  $H_0$  (see equation (2.12)). Plugging equation (3.1) into equation (2.11), we obtain

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr} e^{iuH_0^H(t_f)} e^{-iuH_0} e^{-\beta H_0}}{Z_0}, \quad (3.2)$$

where for completeness we have listed the dependence upon  $\beta$  among the arguments of  $G_{t_f, t_0}^0$ . The quantum version of equation (1.1) immediately follows by setting  $u = i\beta$ ,

$$\langle e^{-\beta w_0} \rangle = G_{t_f, t_0}^0(\beta; i\beta) = \frac{\text{Tr} e^{-\beta H_0^H(t_f)}}{Z_0} = \frac{\text{Tr} e^{-\beta H_0}}{Z_0} = 1, \quad (3.3)$$

where in the third equation we have used equation (2.13), the cyclic property of the trace and the unitarity of the time-evolution operator:  $U_{t_f, t_0}^\dagger U_{t_f, t_0} = \mathbb{1}$ .

Moreover, we find the following important relation between  $G_{t_f, t_0}^0$  and  $\tilde{G}_{t_f, t_0}^0$  (see appendix B):

$$G_{t_f, t_0}^0(\beta; u) = \tilde{G}_{t_f, t_0}^0(\beta; -u + i\beta). \quad (3.4)$$

<sup>2</sup>Here, we assume that the Hamiltonian does not depend on any odd parameter, e.g. a magnetic field. Treating that case is straightforward and amounts to reversing the sign of the odd parameter on the right-hand side of equation (2.20), see [19].

By means of inverse Fourier transform, the following quantum Bochkov–Kuzovlev fluctuation relation between the forward and backward work pdfs is obtained:

$$\frac{p_{t_f, t_0}^0(\beta; w_0)}{\tilde{p}_{t_f, t_0}^0(\beta; -w_0)} = e^{\beta w_0}. \quad (3.5)$$

This must be compared with the quantum Tasaki–Crooks relation that is obtained within the inclusive viewpoint [17],

$$\frac{p_{t_f, t_0}(\beta; w)}{\tilde{p}_{t_f, t_0}(\beta; -w)} = e^{\beta(w - \Delta F)}, \quad (3.6)$$

where, in contrast to equation (3.5), the term  $\Delta F = -\beta^{-1}[\ln \text{Tr} e^{-\beta H(t)} - \ln \text{Tr} e^{-\beta H_0}]$ , appears.

(a) *Remarks*

Equations (3.3) and (3.5) constitute original quantum results that do not appear in the works of Bochkov & Kuzovlev [8,9]. In the *classical* case, they found a fluctuation theorem similar to equation (3.5) reading

$$\frac{P[Q(\tau); X(\tau)]}{P[\varepsilon \tilde{Q}(\tau); \varepsilon \tilde{X}(\tau)]} = \exp \left[ \beta \int_{t_0}^{t_f} d\tau X(\tau) \dot{Q}(\tau) \right], \quad (3.7)$$

where  $P[Q(\tau); X(\tau)]$  is the probability density *functional* to observe a certain trajectory  $Q(\tau)$  given a certain protocol  $X(\tau)$ . Here,  $Q(\tau)$  is a shorthand notation for  $Q(\mathbf{q}(\mathbf{q}_0, \mathbf{p}_0, \tau), \mathbf{p}(\mathbf{q}_0, \mathbf{p}_0, \tau))$ , see equation (1.3), where  $(\mathbf{q}(\mathbf{q}_0, \mathbf{p}_0, \tau), \mathbf{p}(\mathbf{q}_0, \mathbf{p}_0, \tau))$  is the evolved initial condition  $\mathbf{q}_0, \mathbf{p}_0$  at some time  $\tau \in [t_0, t_f]$ , for a certain protocol  $X(\tau)$ . The symbol  $\varepsilon$  denotes the parity of the observable  $Q$  under time reversal (assumed to be equal to 1 in this paper). The symbol ‘ $\sim$ ’ denotes quantities referring to the reversed protocol. The *classical* probability of work  $p_{t_f, t_0}^{\text{cl}, 0}(W_0)$  is obtained from the  $Q$ -trajectory probability density functional  $P[Q(\tau); X(\tau)]$  via the formula

$$p_{t_f, t_0}^{\text{cl}, 0}(W_0) = \int \mathcal{D}Q(\tau) P[Q(\tau); X(\tau)] \delta \left[ W_0 - \int_{t_0}^{t_f} d\tau X(\tau) \dot{Q}(\tau) \right], \quad (3.8)$$

where the integration is a functional integration over all possible trajectories, such that  $\int_{t_0}^{t_f} d\tau X(\tau) \dot{Q}(\tau) = W_0$ . With this formula, one finds from equation (3.7) the exclusive version of the classical Crooks fluctuation theorem for the work probability densities [11]

$$p_{t_f, t_0}^{\text{cl}, 0}(\beta; W_0) = \tilde{p}_{t_f, t_0}^{\text{cl}, 0}(\beta; -W_0) e^{\beta W_0}. \quad (3.9)$$

Notably, a quantum version of equation (3.7) does not exist because: ‘ $\dots$  in the quantum case it is impossible to introduce unambiguously a [ $\dots$ ] probability functional’ [9]. It is only by giving up the idea of true quantum trajectories and embracing instead the two-point measurement approach that the quantum exclusive fluctuation theorem equation (3.5) can be obtained, and has been obtained here, for the first time. For protocols that run in the presence of multiple measurements quantum fluctuation theorems of the form of equation (3.7) yet exist [21].

#### 4. Microcanonical initial state

Next, we consider an initial microcanonical state of energy  $E$ , that can formally be expressed as

$$\rho(t_0) = \bar{\rho}(t_0) = \frac{\delta(H_0 - E)}{\Omega_0(E)}, \quad (4.1)$$

wherein  $\Omega_0(E) = \text{Tr} \delta(H_0 - E)$ . Actually, one has to replace the singular Dirac function  $\delta(x)$  by a smooth function sharply peaked around  $x=0$ , but with infinite support. A normalized Gaussian with arbitrarily small width serves this purpose well.

With this choice of initial condition, the characteristic function of work reads

$$\begin{aligned} G_{t_f, t_0}^0(E; u) &= \frac{\text{Tr} e^{iuH_0^H(t_f)} e^{-iuH_0} \delta(H_0 - E)}{\Omega_0(E)} \\ &= \frac{\text{Tr} e^{iu[H_0^H(t_f) - E]} \delta(H_0 - E)}{\Omega_0(E)}, \end{aligned} \quad (4.2)$$

where, for completeness, we listed the dependence upon  $E$  among the arguments of  $G_{t_f, t_0}^0$ . By applying the inverse Fourier transform, we obtain

$$p_{t_f, t_0}^0(E; w_0) = \frac{\text{Tr} \delta(H_0^H(t_f) - E - w_0) \delta(H_0 - E)}{\Omega_0(E)}. \quad (4.3)$$

Likewise, for the reversed protocol,

$$\tilde{p}_{t_f, t_0}^0(E; w_0) = \frac{\text{Tr} \delta(\tilde{H}_0^H(t_f) - E - w_0) \delta(H_0 - E)}{\Omega_0(E)} \quad (4.4)$$

is found.

We then find the following relation between the forward and backward work probability densities (see appendix C):

$$\Omega_0(E) p_{t_f, t_0}^0(E; w_0) = \Omega_0(E + w_0) \tilde{p}_{t_f, t_0}^0(E + w_0; -w_0). \quad (4.5)$$

Then, the quantum microcanonical fluctuation theorem reads, within the exclusive viewpoint,

$$\frac{p_{t_f, t_0}^0(E; w_0)}{\tilde{p}_{t_f, t_0}^0(E + w_0; -w_0)} = \frac{\Omega_0(E + w_0)}{\Omega_0(E)}. \quad (4.6)$$

This must be compared with the quantum microcanonical fluctuation theorem, obtained within the inclusive viewpoint [18],

$$\frac{p_{t_f, t_0}(E; w)}{\tilde{p}_{t_f, t_0}(E + w; -w)} = \frac{\Omega_f(E + w)}{\Omega_0(E)}. \quad (4.7)$$

The difference lies in the fact that within the exclusive viewpoint, the density of states at the final energy  $E + w_0$  is determined by the unperturbed Hamiltonian, i.e.  $\Omega_0(E + w_0) = \text{Tr} \delta(H_0 - (E + w_0))$ , whereas it results from the total Hamiltonian in the inclusive approach:  $\Omega_f(E + w) = \text{Tr} \delta(H(t_f) - E - w)$ .



Equation (4.7) was first obtained within the classical framework by Cleuren *et al.* [22]. It is not difficult to see that equation (4.6) holds classically as well.

(a) *Remarks*

Just as for equation (3.5), this equation (4.6) is a new result that was not reported before by Bochkov & Kuzovlev [8,9]. It is very interesting to notice, however, that those authors already put forward a *classical* fluctuation theorem for the microcanonical ensemble, which can be recast in the form [9]

$$\frac{P[I(\tau); X(\tau); E]}{P[-\varepsilon\tilde{I}(\tau); \varepsilon\tilde{X}(\tau); E + W_0]} = \frac{\Omega_0(E + W_0)}{\Omega_0(E)}, \quad (4.8)$$

where  $P[I(\tau); X(\tau); E]$  is the probability density functional to observe a certain trajectory  $I(\tau)$  given a certain protocol and an initial microcanonical ensemble of energy  $E$ . Here,

$$I(\tau) = \dot{Q}(\mathbf{q}(\mathbf{q}_0, \mathbf{p}_0, \tau), \mathbf{p}(\mathbf{q}_0, \mathbf{p}_0, \tau)) \quad (4.9)$$

denotes the *current*. By functional integration, the classical microcanonical theorem for the pdf of work,

$$\frac{p_{t_f, t_0}^{\text{cl}, 0}(E, W_0)}{\tilde{p}_{t_f, t_0}^{\text{cl}, 0}(E + W_0, W_0)} = \frac{\Omega_0(E + W_0)}{\Omega_0(E)}, \quad (4.10)$$

is obtained from equation (4.8) in the same way as equation (3.5) follows from equation (3.7). However, the quantum version of equation (4.8) does not exist, and the derivation of the quantum microcanonical fluctuation theorem (4.6) is indeed only possible if the two-point measurement approach is adopted.

The fluctuation relations of equations (4.6) and (4.7) can be further expressed in terms of entropy, according to the rules of statistical mechanics. Following Gibbs [23], two different prescriptions are found in textbooks to obtain the entropy associated with the microcanonical ensemble,

$$s(E) = k_B \ln \Omega(E) = \text{Tr } \delta(H - E) \quad (4.11)$$

and

$$S(E) = k_B \ln \Phi(E) = \text{Tr } \theta(H - E). \quad (4.12)$$

The two definitions coincide for large systems with short-range interactions among their constituents, but may substantially differ if the size of the system under study is small. By now it is clear that, of the two, only the second—customarily called ‘Hertz entropy’—is the fundamentally correct one [24–31].<sup>3</sup> Using the microcanonical expression for the temperature

<sup>3</sup>It is interesting to notice that Einstein was well aware of the works of Hertz [28,29], which he praised as excellent (‘vortrefflich’) [32].

$k_B T(E) = (\partial S(E)/\partial E)^{-1} = \Phi(E)/\Omega(E)$ , we can re-express the quantum microcanonical Bochkov–Kuzovlev fluctuation relation in terms of entropy and temperature as

$$\frac{p_{t_f, t_0}^0(E; w_0)}{\tilde{p}_{t_f, t_0}^0(E + w_0; -w_0)} = \frac{T_0(E)}{T_0(E + w_0)} e^{[S_0(E+w_0) - S_0(E)]/k_B}, \quad (4.13)$$

where the subscript ‘0’ in  $T$  and  $S$  denotes that these quantities are calculated for the unperturbed Hamiltonian  $H_0$ . Likewise, adopting the inclusive viewpoint, one obtains

$$\frac{p_{t_f, t_0}(E; w)}{\tilde{p}_{t_f, t_0}(E + w; -w)} = \frac{T_0(E)}{T_f(E + w)} e^{[S_f(E+w) - S_0(E)]/k_B}, \quad (4.14)$$

where the subscript ‘ $f$ ’ in  $T$  and  $S$  denotes that these quantities are calculated for the total final Hamiltonian  $H(t_f)$ .<sup>4</sup>

## 5. Discussion

We derived the quantum Bochkov–Kuzovlev identity as well as the quantum canonical and microcanonical work fluctuation theorems within the exclusive approach, and have elucidated their relations to the original works of Bochkov & Kuzovlev [8,9]. The extension of the corresponding classical theorems to the quantum regime is only possible due to the proper definition of work as a two-time quantum observable. We close with the following comments:<sup>5</sup>

- For a cyclic process,  $X(t_f) = X(t_0)$ , inclusive- and exclusive-work fluctuation theorems coincide. However, in no way is it true that the exclusive approach of Bochkov & Kuzovlev, adopted here, is restricted to cyclic processes, as some authors have suggested [13,34,35]. As stressed in §1, the difference of the two approaches originates from the different definitions of work, and is not related to whether the process under study is cyclic or not.
- Within the inclusive approach, it is natural to define the dissipated work as  $w_{\text{dis}} = w - \Delta F$  [36,37]. Then, the Jarzynski equality (1.2) can be rewritten as  $\langle e^{-\beta w_{\text{dis}}} \rangle = 1$ . This might make one believe that the exclusive work  $w_0$  coincides with the dissipated work  $w_{\text{dis}}$ . This, though, would be generally wrong. The dissipated work  $w_{\text{dis}}$  is a stochastic quantity whose statistics, given by  $p_{t_f, t_0}^{\text{dis}}(w_{\text{dis}}) = p_{t_f, t_0}(w_{\text{dis}} + \Delta F)$ , in general does not coincide with the statistics of exclusive work  $w_0$ , given by  $p_{t_f, t_0}^0(w_0)$ . See appendix D for a counterexample. Only for a cyclic process, for which  $\Delta F = 0$ , does the dissipated work  $w_{\text{dis}}$  coincide with the inclusive work  $w$ , which in turn coincides with the exclusive work  $w_0$ .

<sup>4</sup>If instead of the microcanonical ensemble (4.1), the modified microcanonical ensemble  $\varrho(t_0) = \theta(E - H_0)/[\text{Tr } \theta(E - H_0)]$  [33] is used as the initial equilibrium state, then the fluctuation theorem assumes the same form as in equation (4.14), but without the ratio of temperatures [18].

<sup>5</sup>Similar remarks were made also within the classical framework by Jarzynski [12].

The authors gratefully acknowledge financial support from the German Excellence Initiative via the *Nanosystems Initiative Munich* (NIM), the Volkswagen Foundation (project I/80424) and the DFG via the collaborative research centre SFB-486, Project A-10, via the DFG project no. HA1517/26-2.

### Appendix A. Derivation of equation (2.21)

The time-evolution operator  $\tilde{U}_{t_f, t_0}$  can be expressed as a time-ordered product [38],

$$\tilde{U}_{t_f, t_0} = \lim_{N \rightarrow \infty} e^{-i\tilde{H}(t_N)\tau} e^{-i\tilde{H}(t_{N-2})\tau} \dots e^{-i\tilde{H}(t_1)\tau}, \quad (\text{A } 1)$$

where  $\tau = (t_f - t_0)/N$  and  $t_\nu = t_0 + \nu\tau$ , for  $\nu = 0, \dots, N$  (note that  $t_N = t_f$ ). Owing to equation (2.15) and (2.16),  $\tilde{H}(t) = H(t_f + t_0 - t)$ , then

$$\tilde{U}_{t_f, t_0} = \lim_{N \rightarrow \infty} e^{-iH(t_1)\tau} e^{-iH(t_2)\tau} \dots e^{-iH(t_N)\tau}. \quad (\text{A } 2)$$

Therefore,

$$\Theta \tilde{U}_{t_f, t_0} \Theta^{-1} = \lim_{N \rightarrow \infty} \Theta e^{-iH(t_1)\tau} \Theta^{-1} \Theta e^{-iH(t_2)\tau} \Theta^{-1} \dots \Theta e^{-iH(t_N)\tau} \Theta^{-1}, \quad (\text{A } 3)$$

where we inserted  $\Theta^{-1}\Theta = \mathbb{1}$ ,  $N - 1$  times. Owing to the property (2.20) and the antilinearity of  $\Theta$ ,

$$\Theta e^{-iH(t)\tau} \Theta^{-1} = e^{iH(t)\tau}. \quad (\text{A } 4)$$

Using this equation, we find

$$\Theta \tilde{U}_{t_f, t_0} \Theta^{-1} = \lim_{N \rightarrow \infty} e^{iH(t_1)\tau} e^{iH(t_2)\tau} \dots e^{iH(t_N)\tau} \quad (\text{A } 5)$$

$$= \lim_{N \rightarrow \infty} [e^{-iH(t_N)\tau} \dots e^{-iH(t_2)\tau} e^{-iH(t_1)\tau}]^\dagger \quad (\text{A } 6)$$

$$= U_{t_f, t_0}^\dagger = U_{t_0, t_f}. \quad (\text{A } 7)$$

In a similar way, we also obtain

$$U_{t_f, t_0} = \Theta \tilde{U}_{t_f, t_0}^\dagger \Theta^{-1}. \quad (\text{A } 8)$$

### Appendix B. Derivation of equation (3.4)

According to equation (2.11), the characteristic function reads

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr } U_{t_f, t_0}^\dagger e^{iuH_0} U_{t_f, t_0} e^{-iuH_0} e^{-\beta H_0}}{Z_0}. \quad (\text{B } 1)$$

Using equations (A 7) and (A 8), we obtain

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr } \Theta \tilde{U}_{t_f, t_0} \Theta^{-1} e^{iuH_0} \Theta \tilde{U}_{t_f, t_0}^\dagger \Theta^{-1} e^{-iuH_0} e^{-\beta H_0} \Theta \Theta^{-1}}{Z_0},$$

where we have inserted  $\Theta\Theta^{-1} = \mathbb{1}$  at the right-hand side. By multiplying equation (A 4) by  $\Theta^{-1}$  from the left and by  $\Theta$  from the right, we have

(replacing  $\tau$  with  $u$ )

$$\Theta^{-1} e^{iH(t)u} \Theta = e^{-iH(t)u^*}; \quad (\text{B 2})$$

therefore,

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr } \Theta \tilde{U}_{t_f, t_0} e^{-iu^* H_0} \tilde{U}_{t_f, t_0}^\dagger e^{iu^* H_0} e^{-\beta H_0} \Theta^{-1}}{Z_0}. \quad (\text{B 3})$$

The antilinearity of  $\Theta$  implies, for any trace class operator  $A$ ,

$$\text{Tr } \Theta A \Theta^{-1} = \text{Tr } A^\dagger. \quad (\text{B 4})$$

Therefore,

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr } e^{-\beta H_0} e^{-iu H_0} \tilde{U}_{t_f, t_0} e^{iu H_0} \tilde{U}_{t_f, t_0}^\dagger}{Z_0}. \quad (\text{B 5})$$

Using the cyclic property of the trace finally leads to

$$G_{t_f, t_0}^0(\beta; u) = \frac{\text{Tr } \tilde{U}_{t_f, t_0}^\dagger e^{i(-u+i\beta)H_0} \tilde{U}_{t_f, t_0} e^{-i(-u+i\beta)H_0} e^{-\beta H_0}}{Z_0} \quad (\text{B 6})$$

$$= \tilde{G}_{t_f, t_0}^0(\beta; -u + i\beta). \quad (\text{B 7})$$

### Appendix C. Derivation of equation (4.5)

According to equation (4.3), the microcanonical exclusive-work pdf reads

$$p_{t_f, t_0}^0(E; w_0) = \frac{\text{Tr } \delta(H_0^H(t_f) - E - w_0) \delta(H_0 - E)}{\Omega_0(E)} \quad (\text{C 1})$$

$$= \frac{\text{Tr } U_{t_f, t_0}^\dagger \delta(H_0 - E - w_0) U_{t_f, t_0} \delta(H_0 - E)}{\Omega_0(E)}. \quad (\text{C 2})$$

Employing equation (A 7) and (A 8), we obtain

$$\Omega_0(E) p_{t_f, t_0}^0(E; w_0) = \text{Tr } \Theta \tilde{U}_{t_f, t_0} \Theta^{-1} \delta(H_0 - E - w_0) \Theta \tilde{U}_{t_f, t_0}^\dagger \Theta^{-1} \delta(H_0 - E) \Theta \Theta^{-1}, \quad (\text{C 3})$$

where we have inserted  $\Theta \Theta^{-1} = \mathbb{1}$  at the end. As the Dirac delta function is a *real* function, we have

$$\Theta^{-1} \delta(H_0 - E) \Theta = \delta(H_0 - E) \quad (\text{C 4})$$

because  $H_0$  is assumed to be invariant under time reversal. Then,

$$\Omega_0(E) p_{t_f, t_0}^0(E; w_0) = \text{Tr } \Theta \tilde{U}_{t_f, t_0} \delta(H_0 - E - w_0) \tilde{U}_{t_f, t_0}^\dagger \delta(H_0 - E) \Theta^{-1}. \quad (\text{C 5})$$

Using equation (B 4), we obtain

$$\Omega_0(E) p_{t_f, t_0}^0(E; w_0) = \text{Tr } \delta(H_0 - E) \tilde{U}_{t_f, t_0} \delta(H_0 - E - w_0) \tilde{U}_{t_f, t_0}^\dagger. \quad (\text{C 6})$$

Thanks to the cyclic property of the trace, one finally arrives at

$$\Omega_0(E) p_{t_f, t_0}^0(E; w_0) = \text{Tr } \tilde{U}_{t_f, t_0}^\dagger \delta(H_0 - E) \tilde{U}_{t_f, t_0} \delta(H_0 - E - w_0) \quad (\text{C 7})$$

$$= \Omega_0(E + w_0) \tilde{p}_{t_f, t_0}^0(E + w_0; -w_0). \quad (\text{C 8})$$

### Appendix D. Comparison between dissipated-work and exclusive-work probability density functions

In this appendix, we provide an example that shows that the dissipated work  $w_{\text{dis}}$  and the inclusive work  $w_0$  are distinct stochastic quantities with different statistical properties. To this end, we show that their pdfs may have different first and second moments. We consider a driven quantum harmonic oscillator of unit mass and unit angular frequency,

$$H(t) = \frac{p^2}{2} + \frac{q^2}{2} - X(t)q. \quad (\text{D } 1)$$

For simplicity, we assume  $t_0 = 0$ ,  $X(t_0) = 0$ , and we chose units in such a way that  $\hbar = 1$ . Let  $|n, t\rangle$  denote the instantaneous eigenvectors of  $H(t)$  corresponding to the instantaneous eigenvalues  $E_n(t) = (n + 1/2) - X^2(t)/2$ .

#### (a) The probability density of dissipated work

The pdf of inclusive work, corresponding to an initial canonical state, is

$$p_{t,0}(w) = \sum_{mn} \frac{\delta(w - m + n + X^2(t)/2) |a_{mn}|^2 e^{-\beta(n+1/2)}}{Z(0)}, \quad (\text{D } 2)$$

where  $Z(0) = \sum_n e^{-\beta(n+1/2)}$  is the initial partition function, and  $|a_{nm}|^2$  are the probabilities to make a transition between two eigenstates of the total Hamiltonian

$$|a_{mn}|^2 = |\langle m, t | U_{t,0} | n, 0 \rangle|^2, \quad (\text{D } 3)$$

where we have set  $t_0 = 0$  and  $t_f = t$ . According to Talkner *et al.* [39,40], the mean value and the variance of the inclusive-work pdf (D 2) are given by

$$\langle w \rangle = \int dx x p_{t,0}(x) = L(t) - \frac{X^2(t)}{2} \quad (\text{D } 4)$$

and

$$\langle \Delta w^2 \rangle = \int dx [x - \langle w \rangle]^2 p_{t,0}(x) = 2UL(t), \quad (\text{D } 5)$$

where  $U = \sum_n (n + 1/2) e^{-\beta(n+1/2)} / Z_0$  is the initial average energy, and<sup>6</sup>

$$L(t) = \frac{C(t)^2}{2} + \frac{[S(t) - X(t)]^2}{2}, \quad (\text{D } 6)$$

where

$$S(t) = \int_0^t ds X(s) \sin(t - s) \quad \text{and} \quad C(t) = \int_0^t ds X(s) \cos(t - s). \quad (\text{D } 7)$$

<sup>6</sup>In Talkner *et al.* [39,40],  $L$  is given as  $L(t) = |\int_0^t ds \dot{f}(s) e^{is}|^2$ , where  $f = -X/\sqrt{2}$ . It is a matter of elementary calculus to check that this expression coincides with equation (D 6).

The partition function of work at the final time  $t$  is  $Z(t) = Z(0) e^{\beta X^2(t)/2}$ ; therefore, the free-energy difference  $\Delta F = -\beta^{-1} \ln Z(t)/Z_0$  is given by  $\Delta F = -X^2(t)/2$  [39]. Hence, the dissipated work is

$$w_{\text{dis}} = w + \frac{X^2(t)}{2}. \quad (\text{D } 8)$$

Accordingly, the dissipated work pdf is

$$\begin{aligned} p_{t,0}^{\text{dis}}(w_{\text{dis}}) &= p_{t,0}(w) = p_{t,0}\left(w_{\text{dis}} - \frac{X^2(t)}{2}\right) \\ &= \sum_{mn} \frac{\delta(w_{\text{dis}} - m + n) |a_{mn}|^2 e^{-\beta(n+1/2)}}{Z(0)}. \end{aligned} \quad (\text{D } 9)$$

It immediately follows that

$$\langle w_{\text{dis}} \rangle = \int dx x p_{t,0}^{\text{dis}}(x) = L(t) \quad (\text{D } 10)$$

and

$$\langle \Delta w_{\text{dis}}^2 \rangle = \int dx [x - L(t)]^2 p_{t,0}^{\text{dis}}(x) = 2UL(t). \quad (\text{D } 11)$$

Note that, as it should be,  $\langle w_{\text{dis}} \rangle \geq 0$ .

(b) *The probability density of exclusive work*

The exclusive-work pdf is given by

$$p_{t,0}^0(w_0) = \sum_{mn} \frac{\delta(w_0 - m + n) |a_{mn}^0|^2 e^{-\beta(n+1/2)}}{Z(0)}, \quad (\text{D } 12)$$

where  $|a_{mn}^0|^2$  denotes the probability to make a transition between two states of the unperturbed Hamiltonian

$$|a_{mn}^0|^2 = |\langle m, 0 | U_{t,0} | n, 0 \rangle|^2. \quad (\text{D } 13)$$

It is known [41,42] that the transition probabilities  $|a_{mn}|^2$  depend on the time  $t$  at which the second measurement is performed, via the function  $L(t)$ , that is the  $|a_{mn}|^2$  are of the form  $|a_{mn}|^2 = f_{nm}[L(t)]$ , for certain functions  $f_{nm}$  that need not be specified here. Using Wigner functions to calculate the transition probabilities as in Campisi [41], we notice that the transition probabilities  $|a_{mn}^0|^2$  are obtained from the same expression as of  $|a_{mn}|^2$ , with the only difference that  $L(t)$  is replaced by

$$L_0(t) = \frac{C^2(t)}{2} + \frac{S^2(t)}{2}, \quad (\text{D } 14)$$

that is  $|a_{mn}^0|^2 = f_{nm}[L_0(t)]$ . Therefore, the exclusive-work pdf (D 12) is obtained from the dissipated-work pdf (D 9) simply by replacing  $L(t)$  with  $L_0(t)$ . It follows

immediately that

$$\langle w_0 \rangle = \int dx x p_{t,0}^0(x) = L_0(t) \quad (\text{D } 15)$$

and

$$\langle \Delta w_0^2 \rangle = \int dx [x - L_0(t)]^2 p_{t,0}^0(x) = 2UL_0(t). \quad (\text{D } 16)$$

Note that, as expected,  $\langle w_0 \rangle \geq 0$ .

For the specific protocol

$$X(t) = 2 \sin(t), \quad (\text{D } 17)$$

we find

$$L(t) - L_0(t) = t \sin(2t), \quad (\text{D } 18)$$

which is apparently different from zero, except for integer multiples of  $\pi/2$ . Thus, for any duration  $t$  of the protocol (D 17) that is not an integer multiple of  $\pi/2$ ,  $L_0 \neq L$ . Accordingly, the first and second moments of  $p_{t,0}^{\text{dis}}$  and  $p_{t,0}^0$  differ, meaning that  $w_{\text{dis}}$  and  $w_0$  are distinct stochastic variables with different statistical properties.

It should be stressed that analogous results are found also for a classical driven harmonic oscillator. The statistics of dissipated work and of exclusive work generally *differ*, this fact holds true both quantum mechanically and classically.

## References

- 1 Callen, H. B. & Welton, T. A. 1951 Irreversibility and generalized noise. *Phys. Rev.* **83**, 34–40. (doi:10.1103/PhysRev.83.34)
- 2 Green, M. S. 1952 Markoff random processes and the statistical mechanics of time-dependent phenomena. *J. Chem. Phys.* **20**, 1281. (doi:10.1063/1.1700722)
- 3 Kubo, R. 1957 Statistical mechanical theory of irreversible processes I. *J. Phys. Soc. Jpn* **12**, 570–586. (doi:10.1143/JPSJ.12.570)
- 4 Einstein, A. 1926 *Investigations on the theory of Brownian movement*. London, UK: Methuen. (Reprinted by Dover 1956.)
- 5 Johnson, J. B. 1928 Thermal agitation of electricity in conductors. *Phys. Rev.* **32**, 97–109. (doi:10.1103/PhysRev.32.97)
- 6 Nyquist, H. 1928 Thermal agitation of electric charge in conductors. *Phys. Rev.* **32**, 110–113. (doi:10.1103/PhysRev.32.110)
- 7 Hänggi, P. & Thomas, H. 1982 Stochastic processes: time-evolution, symmetries and linear response. *Phys. Rep.* **88**, 207–319. (doi:10.1016/0370-1573(82)90045-X)
- 8 Bochkov, G. N. & Kuzovlev, Y. E. 1977 General theory of thermal fluctuations in nonlinear systems. *Sov. Phys. JETP* **45**, 125–130. (*Zh. Eksp. Theor. Fiz.* **72**, 238–247.)
- 9 Bochkov, G. N. & Kuzovlev, Y. E. 1981 Nonlinear fluctuation–dissipation relations and stochastic models in nonequilibrium thermodynamics I. Generalized fluctuation-dissipation theorem. *Physica A* **106**, 443–479. (doi:10.1016/0378-4371(81)90122-9)
- 10 Jarzynski, C. 1997 Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693. (doi:10.1103/PhysRevLett.78.2690)
- 11 Jarzynski, C. & Horowitz, J. 2007 Comparison of work fluctuation relations. *J. Stat. Mech.* P11002. (doi:10.1088/1742-5468/2007/11/P11002)
- 12 Jarzynski, C. 2007 Comparison of far-from-equilibrium work relations. *C. R. Phys.* **8**, 495–506. (doi:10.1016/j.crhy.2007.04.010)
- 13 Allahverdyan, A. E. & Nieuwenhuizen, T. M. 2005 Fluctuations of work from quantum subensembles: the case against quantum work-fluctuation theorems. *Phys. Rev. E* **71**, 066102. (doi:10.1103/PhysRevE.71.066102)

- 14 Talkner, P., Lutz, E. & Hänggi, P. 2007 Fluctuation theorems: work is not an observable. *Phys. Rev. E* **75**, 050102(R). (doi:10.1103/PhysRevE.75.050102)
- 15 Campisi, M., Talkner, P. & Hänggi, P. 2009 Fluctuation theorem for arbitrary open quantum systems. *Phys. Rev. Lett.* **102**, 210401. (doi:10.1103/PhysRevLett.102.210401)
- 16 Talkner, P., Campisi, M. & Hänggi, P. 2009 Fluctuation theorems in driven open quantum systems. *J. Stat. Mech.* **2009**, P02025. (doi:10.1088/1742-5468/2009/02/P02025)
- 17 Talkner, P. & Hänggi, P. 2007 The Tasaki–Crooks quantum fluctuation theorem. *J. Phys. A* **40**, F569–F571. (doi:10.1088/1751-8113/40/26/F08)
- 18 Talkner, P., Hänggi, P. & Morillo, M. 2008 Microcanonical quantum fluctuation theorem. *Phys. Rev. E* **77**, 051131. (doi:10.1103/PhysRevE.77.051131)
- 19 Andrieux, D., Gaspard, P., Monnai, T. & Tasaki, S. 2009 The fluctuation theorem for currents in open quantum systems. *New J. Phys.* **11**, 043014. (doi:10.1088/1367-2630/11/4/043014)
- 20 Messiah, A. 1962 *Quantum mechanics*. Amsterdam, The Netherlands: North-Holland.
- 21 Campisi, M., Talkner, P. & Hänggi, P. 2010 Fluctuation theorems for continuously monitored quantum fluxes. *Phys. Rev. Lett.* **105**, 140601. (doi:10.1103/PhysRevLett.105.140601)
- 22 Cleuren, B., Van den Broeck, C. & Kawai, R. 2006 Fluctuation and dissipation of work in a Joule experiment. *Phys. Rev. Lett.* **96**, 050601. (doi:10.1103/PhysRevLett.96.050601)
- 23 Gibbs, J. 1902 *Elementary principles in statistical mechanics*. Yale, CT: Yale University Press. (Reprinted by Dover 1960.)
- 24 Campisi, M. 2005 On the mechanical foundations of thermodynamics: the generalized Helmholtz theorem. *Stud. Hist. Phil. Mod. Phys.* **36**, 275–290. (doi:10.1016/j.shpsb.2005.01.001)
- 25 Campisi, M. 2008 Statistical mechanical proof of the second law of thermodynamics based on volume entropy. *Stud. Hist. Phil. Mod. Phys.* **39**, 181–194. (doi:10.1016/j.shpsb.2007.09.002)
- 26 Campisi, M. & Kobe, D. H. 2010 Derivation of the Boltzmann principle. *Am. J. Phys.* **78**, 608–615. (doi:10.1119/1.3298372)
- 27 Dunkel, J. & Hilbert, S. 2006 Phase transitions in small systems: microcanonical vs. canonical ensembles. *Physica A* **370**, 390–406. (doi:10.1016/j.physa.2006.05.018)
- 28 Hertz, P. 1910 Über die mechanischen Grundlagen der Thermodynamik. *Ann. Phys.* **338**, 225–274. (doi:10.1002/andp.19103381202)
- 29 Hertz, P. 1910 Über die mechanischen Grundlagen der Thermodynamik. *Ann. Phys.* **338**, 537–552. (doi:10.1002/andp.19103381305)
- 30 Pearson, E. M., Halicioglu, T. & Tiller, W. A. 1985 Laplace-transform technique for deriving thermodynamic equations from the classical microcanonical ensemble. *Phys. Rev. A* **32**, 3030–3039. (doi:10.1103/PhysRevA.32.3030)
- 31 Schlüter, A. 1948 Zur Statistik klassischer Gesamtheiten. *Z. Naturforsch.* **3A**, 350–360. [Transl., see <http://people.chem.byu.edu/rbshirts/research/schluter1948translation.doc>].
- 32 Einstein, A. 1911 Bemerkungen zu den P. Hertzschen Arbeiten: ‘Über die mechanischen Grundlagen der Thermodynamik’. *Ann. Phys.* **339**, 175–176. (doi:10.1002/andp.19113390111)
- 33 Ruelle, D. 1969 *Statistical mechanics*. New York, NY: W. A. Benjamin.
- 34 Andrieux, D. & Gaspard, P. 2008 Quantum work relations and response theory. *Phys. Rev. Lett.* **100**, 230404. (doi:10.1103/PhysRevLett.100.230404)
- 35 Cohen, E. G. D. & Mauzerall, D. 2005 The Jarzynski equality and the Boltzmann factor. *Mol. Phys.* **103**, 2923–2926. (doi:10.1080/00268970500151536)
- 36 Kawai, R., Parrondo, J. M. R. & Van den Broeck, C. 2007 Dissipation: the phase-space perspective. *Phys. Rev. Lett.* **98**, 080602. (doi:10.1103/PhysRevLett.98.080602)
- 37 Vaikuntanathan, S. & Jarzynski, C. 2009 Dissipation and lag in irreversible processes. *EPL* **87**, 60005. (doi:10.1209/0295-5075/87/60005)
- 38 Schleich, W. P. 2001 *Quantum optics in phase space*. Berlin, Germany: Wiley.
- 39 Talkner, P., Burada, P. S. & Hänggi, P. 2008 Statistics of work performed on a forced quantum oscillator. *Phys. Rev. E* **78**, 011115. (doi:10.1103/PhysRevE.78.011115)
- 40 Talkner, P., Burada, P. S. & Hänggi, P. 2009 Erratum: statistics of work performed on a forced quantum oscillator. *Phys. Rev. E* **79**, 039902. (doi:10.1103/PhysRevE.79.039902)
- 41 Campisi, M. 2008 Increase of Boltzmann entropy in a forced harmonic oscillator. *Phys. Rev. E* **78**, 051123. (doi:10.1103/PhysRevE.78.051123)
- 42 Husimi, K. 1953 Miscellanea in elementary quantum mechanics, II. *Prog. Theor. Phys.* **9**, 381–402. (doi:10.1143/PTP.9.381)