Nonequilibrium quantum fluctuations of work

A. E. Allahverdyan
Yerevan Physics Institute, Alikhanian Brothers street 2, Yerevan 375036, Armenia
(Received 14 January 2014; revised manuscript received 21 May 2014; published 26 September 2014)

The concept of work is basic for statistical thermodynamics. To gain a fuller understanding of work and its (quantum) features, it needs to be represented as an average of a fluctuating quantity. Here I focus on the work done between two moments of time for a thermally isolated quantum system driven by a time-dependent Hamiltonian. I formulate two natural conditions needed for the fluctuating work to be physically meaningful for a system that starts its evolution from a nonequilibrium state. The existing definitions do not satisfy these conditions due to issues that are traced back to noncommutativity. I propose a definition of fluctuating work that is free of previous drawbacks and that applies for a wide class of nonequilibrium initial states. It allows the deduction of a generalized work-fluctuation theorem that applies for an arbitrary (out-of-equilibrium) initial state.

DOI: 10.1103/PhysRevE.90.032137 PACS number(s): 05.30.−d, 05.70.Ln

I. INTRODUCTION

The first and second laws of statistical thermodynamics are formulated using the concept of work, i.e., the (average) energy exchanged by a system driven via a time-dependent Hamiltonian [1,2]. In this sense the work is a basic quantity for thermodynamics. It is well-defined both in and out of equilibrium for any (quantum or classical) system interacting with external macroscopic work sources [2].

However, the work as it appears in the first and second laws is an averaged quantity. There are at least two reasons why it is useful to "deaverage" it, i.e., to present it as a random quantity. First, its features are understood better in this way. Recall in this context that the conservation of average energy for an isolated quantum system is just a consequence of conserving energy eigenvalues and their probabilities. Second, the current understanding of the second law is that it has a statistical character and emerges from averaging over fluctuations [3,4]. Hence it is necessary to define fluctuations of work for understanding, e.g., the Thomson's formulation of the second law [4–7]. Both these points are illustrated by fluctuation theorems; see Refs. [7–11] for reviews.

The existing definitions of quantum fluctuations of work can be divided into two groups. Time-global definitions look for the work done between two moments of time, as usual for any transfer quantity [10–21]. Time-local approaches adapt the global definitions infinitesimally along an effective quantum trajectory [22–32].

Here I focus on the time-global approaches (admittedly, they are more fundamental in the quantum case) for a thermally isolated dynamics and note that they do not apply whenever the initial density matrix does not commute with the (initial) Hamiltonian. This limitation is essential, since work-extraction from nonequilibrium (e.g., nondiagonal) states is important both conceptually [5] and practically [33].

The aim of this paper is to present a definition of quantum fluctuations of work that is free of the previous drawbacks. It is based on the Terletsky-Margenau-Hill distribution [34–39]. The definition applies for a class of initial density matrices that do not commute with the (time-dependent) Hamiltonian. It leads to a generalized fluctuation theorem.

However, this definition is neither unique (otherwise there would not be the issue with noncommutativity) nor does it apply for an *arbitrary* initial state, because there it leads to negative probabilities whose physical meaning is not clear. In this context, I formulate two conditions for fluctuating work that are closely linked to its physical meaning as the amount of energy exchanged with the source of work. They need to be satisfied for any definition of the fluctuating work and they hold for the presented one. It remains to be seen whether this is indeed the most convenient definition or there are even better ones to be uncovered in the future.[1]

This paper is organized as follows. Section II defines the system to be studied. The next section reviews previous approaches and explains why specifically they are not applicable out of equilibrium. Section IV proposes two general conditions to be satisfied for any definition of quantum fluctuating work. Section V discusses a new definition of fluctuating work that is free of previous drawbacks. A generalized fluctuation theorem is derived and interpreted in Sec. VI. Section VII discusses certain limitations of the proposed approach. I summarize in the last section. There are two appendices.

II. SETUP

Consider a quantum system with an initial state described by a density matrix ρ. The system is thermally isolated: Its dynamics is described by a time-dependent, Schroedinger

[1]To explain why I decided to focus on the concept of work, I shall compare its features to those of entropy production (EP). For a system coupled to thermal baths, EP amounts to entropy increase of baths [1,2]. This definition does not apply more generally—for nonequilibrium baths or thermally isolated case—since the very definition of entropy is ambiguous there. For those cases, EP is defined as an effective measure of irreversibility that has to be positive and share the heuristics of entropy increase [40–42]. There is some consensus on how to define EP for classical [40,43] and semiclassical systems [11]. But the quantum situation is ambiguous in this respect [44,45]; e.g., Ref. [45] shows that there is a family of EPs associated with different notions of effective phase-space. They lead to different expressions of the (average) EP even for the initially equilibrium (Gibbsian) initial state [46]. These features differ from those of the (average) work, which is well defined for arbitrary (initial) states.

representation Hamiltonian $H(t)$ that generates a unitary evolution operator as follows:

$$U_\tau = \overleftarrow{\exp}\left[-\frac{i}{\hbar}\int_0^\tau dt\, H(t)\right], \tag{1}$$

between the initial time 0 and the final time τ. Here $\overleftarrow{\exp}$ denotes a time-ordered exponent. The (average) work W done on the system reads [1,2]

$$W = \mathrm{tr}[\rho(H_F - H_I)], \tag{2}$$

$$H_I \equiv H(0), \quad H_F = U_\tau^\dagger H(\tau) U_\tau, \tag{3}$$

where H_F (H_I) is the final (initial) Hamiltonian in the Heisenberg representation. The definition of work applies to any initial state: It is the average energy given up by the source of work [1,2]. Due to conservation of energy during the system–work source interaction, W is the average energy transferred to the source of work [1,2]. This is seen explicitly in approaches that deal with system–work source interaction from the first principles; see, e.g., Refs. [47,48]. The intuitive meaning of W is that it is a "high-graded," mechanical energy that can be wholly transferred from one work source to another and dissipated into heat.

The variable W can be observed in several ways, e.g., via the energy of the work source or by measuring the Heisenberg operator $H_F - H_I$ at the final time. Another (more usual) way of observing W is to consider an ensemble of identically prepared systems (described by ρ) and divide it into two (equal) parts. Measuring H_I (H_F) on the first (second) part one recovers $\mathrm{tr}(\rho H_I)$ [$\mathrm{tr}(\rho H_F)$]; see (2). Thus, W is directly observable and manifests the energy conservation (first law) for the present problem. Thus, I take the above definition of the average work W as the basic entity from which the fluctuating work is to be deduced under certain additional assumptions.

Note that, formally, the above thermally isolated setup applies also for an open quantum system interacting with an environment (e.g., thermal baths). Since the work is the energy transferred to the source, one just needs to include the whole environment into a single system interacting with the source. This is, however, a formal procedure, because the environment is normally large and out of control. Thus further research is needed to extend this setup to open systems. In this paper I focus on the thermally isolated setup, also because this is the first step towards understanding the more general (open-system) situation.

III. TWO APPROACHES FOR DEFINING FLUCTUATIONS OF WORK

I now concentrate on two major (and different [18,19]) approaches for defining quantum fluctuations of work. My aim is to compare these definitions to each other and to (2) and understand where specifically they flaw in describing the fluctuating work.

A. Operator of work

The spirit of the Heisenberg representation is that time-dependent operators are analogs of classical, time-dependent random variables. Then the Heisenberg operator $H_F - H_I$ is postulated to be the "observable of work" in the standard sense [2,7,12,17,18,20,25,27,49]: Its eigenvalues are realizations of work and its eigenvectors define the respective probabilities. I stress that (at least formally) only one measurement (that of $H_F - H_I$) is needed to obtain the statistics of work according to this definition.

Now assume that the Schroedinger representation Hamiltonian changes cyclically as follows:

$$H_I = H(0) = H(\tau). \tag{4}$$

One interpretation of (4) is that the system interacts with the source of work only for $0 \leqslant t \leqslant \tau$, i.e., it is strictly isolated for $t < 0$ and $t > \tau$: $H(t < 0) = H(t > \tau) = H_I$.

Now since H_I and H_F have the same eigenvalues, $H_F - H_I$ has eigenvalues of both signs. Since the approach should apply for nonequilibrium initial states, we choose H_I and H_F such that $H_F - H_I$ has an eigenvalue equal to zero. The corresponding eigenvector $|0\rangle$,

$$(H_F - H_I)|0\rangle = 0, \tag{5}$$

is taken as the initial state $|0\rangle\langle 0|$. Due to

$$[H_F, H_I] \equiv H_F H_I - H_I H_F \neq 0, \tag{6}$$

$|0\rangle$ is neither an eigenstate of H_F nor an eigenstate of H_I. Equation (5) implies that $H_F - H_I$ has on the state $|0\rangle\langle 0|$ a definite value equal to zero: For *all single systems* from the ensemble described by $|0\rangle\langle 0|$ no work is done and hence no energy is supposed to be exchanged. But there are examples [18] showing that (5) and (6) are compatible with

$$\langle 0|H_F^m|0\rangle = \langle 0|U_\tau^\dagger H_I^m U_\tau|0\rangle \neq \langle 0|H_I^m|0\rangle \quad \text{for} \quad m > 2. \tag{7}$$

For a system that is strictly isolated for $t < 0$ and $t > \tau$ [recall (4)], the inequality (7) implies that the probabilities of some energies (i.e., the eigenvalues of H_I) do change due to the interaction with the source of work.

Thus, according to this definition it is possible to have energy exchange with strictly zero fluctuations of work. In other words, the link between energy exchange and the work done on a thermally isolated system is generally absent.

Though I tuned $H_F - H_I$ to have a zero eigenvalue, it is clear that the problem is more general, e.g., it persists for $H_F - H_I$ having an eigenvalue close to zero [50].

I opine that due to this problem $H_F - H_I$ cannot be interpreted as the work operator for all initial states. Such an interpretation can be perhaps kept for initial states ρ that commute with H_F or with H_I [18], but it is not clear how to generalize this class of initial states.

B. Two-time measurements of energy

We turn to the second approach [10,11,13–16]. Let the eigenresolution of the Schroedinger representation Hamiltonian $H(t)$ be

$$H(t) = \sum_k \epsilon_k(t)\, \mathcal{E}_k(t), \tag{8}$$

$$\mathcal{E}_k(t)\mathcal{E}_l(t) = \delta_{kl}\mathcal{E}_k(t), \;\; \mathrm{tr}\,\mathcal{E}_k(t) = \mathrm{const}, \tag{9}$$

where $\epsilon_k(t)$ are the eigenvalues of $H(t)$, δ_{kl} is the Kronecker symbol, and $\mathcal{E}_k(t)$ are the projector to the corresponding eigenspace, whose dimension $\mathrm{tr}\,\mathcal{E}_k(t)$ is taken time-independent for simplicity.

Measuring $H(0)$ at $t=0$ produces $\epsilon_k(0)$ with probability $\mathrm{tr}(\rho\,\mathcal{E}_k(0))$ [13–15]. The postmeasurement state has the von Neumann-Luders form $\rho_k \equiv \mathcal{E}_k(0)\rho\,\mathcal{E}_k(0)/\mathrm{tr}(\rho\mathcal{E}_k(0))$; it is then evolved via (1). At the final moment τ one measures $H(\tau)$ and gets $\epsilon_l(\tau)$ with probability $\mathrm{tr}(U_\tau\rho_k U_\tau^\dagger\mathcal{E}_l(\tau))$, which is conditional over the result k of the first measurement.

The fluctuating work is presented as a *classical random variable* with, respectively, realizations and probabilities[2]

$$\epsilon_l(\tau)-\epsilon_k(0), \tag{10}$$

$$\widetilde{p}_{kl} = \mathrm{tr}(\rho\,\mathcal{E}_k(0))\,\mathrm{tr}(U_\tau\rho_k U_\tau^\dagger\mathcal{E}_l(\tau)) \tag{11}$$

$$= \mathrm{tr}(\mathcal{E}_k(0)\,\rho\,\mathcal{E}_k(0)\,U_\tau^\dagger\mathcal{E}_l(\tau)U_\tau). \tag{12}$$

The problem of this definition is that it does not apply to initial states that do not commute with $H(0)$: The average "work" $\widetilde{W}$ reads as follows from (10) and (12):

$$\widetilde{W} = \sum_{kl}\widetilde{p}_{kl}(\,\epsilon_l(\tau)-\epsilon_k(0)\,) = \mathrm{tr}(\,\widetilde{\rho}\,(H_F-H_I)), \tag{13}$$

where

$$\widetilde{\rho} \equiv \sum_k \mathcal{E}_k(0)\,\rho\,\mathcal{E}_k(0). \tag{14}$$

We obtain from (2), (13), and (14) the following:

$$W-\widetilde{W} = \mathrm{tr}((\rho-\widetilde{\rho})H_F)$$

$$= \sum_{k\neq l}\mathrm{tr}(\mathcal{E}_k(0)\,\rho\,\mathcal{E}_l(0)\,H_F). \tag{15}$$

Hence for

$$[\rho,H_I]\neq 0 \quad \text{and} \quad [H_F,H_I]\neq 0, \tag{16}$$

(10) and (12) cannot be related to the work done on the system with initial state ρ by the external source, because $\widetilde{W}\neq W$. The physical reason for this conclusion is that under $[\rho,H_I]=0$ the first energy measurement (at $t=0$) can be said to reveal the pre-existing (but unknown) value of energy. In particular, the postmeasurement density matrix does not change: $\rho=\widetilde{\rho}$ [see (14)]. In contrast, for $[\rho,H_I]\neq 0$ already the first measurement is *invasive*; it leads to an irrversible change $\rho\to\widetilde{\rho}$ of the density matrix that alters its subsequent interaction with the source of work provided that $[H_F,H_I]\neq 0$; see Appendix A for a physical example. Put differently, the reason for inapplicaility of the two-time measurement approach is that it essentially alters the (nonequilibrium) initial state.[3]

[2]While this approach is standardly presented via two sharp measurements of energy, one can naturally wonder whether the same statistics of work can be approached via more feasible measurements; see Ref. [52] for a recent review of this issue.

[3]This point of altering the premeasured state also appears in Ref. [53], where the authors study the energy changes for a system that couples to an external measuring apparatus and is thereby subject to projective measurements of a quantity that does not commute with energy (no work source is supposed to be present). It is expected that in this situation the energy changes of the system will consist of both work and heat; no analysis of this problem is carried out in Ref. [53].

C. Comparing two definitions with each other

We saw that the definition based on the operator of work does always reproduce the average work (2), but it does not account properly the notion of "work = exchanged energy" at least for some initial states. I stress that this definition implies a one-time approach, since one needs to measure the Heisenber operator H_F-H_I at the final time τ. The definition based on the two-time measurements of energy does not reproduce the average work (2) if (16) holds.

It is to be stressed that the drawbacks of both approaches do not show up for $[\rho,H_I]=0$. Hence if one is restricted by such initial states, both approaches perform well, and it is a matter of taste which one is preferred.[4]

Even then the operator definition has an advantage of being time symmetric: In contrast to the two-time energy measurement approach, it applies not only for $[\rho,H_I]=0$ but also for $[\rho,H_F]=0$ [and (16)]; see (3), (5), and (7).

IV. GENERAL CONDITIONS

The above analysis of the two approaches leads to the following general conditions demanded for the proper definition of fluctuations of work.

(i) For cyclic changes of the Hamiltonian [cf. the discussion above (5) and (7)], the zero fluctuations of work should mean no energy exchange:

$$\mathrm{tr}\big(\rho H_I^m\big)=\mathrm{tr}\big(\rho H_F^m\big)\equiv\mathrm{tr}\big(U_\tau\rho\,U_\tau^\dagger\,H_I^m\big) \quad \text{for} \quad m\geqslant 1. \tag{17}$$

(ii) The definition should apply for a possibly wide class of initial states (including initial states that do not commute with the initial Hamiltonian H_I) and it should reproduce the average work (2) for all initial states, where it applies.

As seen above, the first (second) condition does not hold for the first (second) definition of fluctuations.

V. ANOTHER DEFINITION FOR FLUCTUATING WORK

A. Estimation of energies via one measurement

Below I work out a definition that satisfies the above two conditions, and, similarly to (10), it presents the work as a classical random quantity. When discussing the approach based on two measurements, we noted that its drawback stems from the invasive character of the first measurement. It is then natural to *illustrate* a more general approach by avoiding the explicit introduction of the first measurement. Hence at the final time τ we measure [cf. (3) and (8)]

$$H_F=\sum_l \epsilon_l(\tau)\Pi_l,\ \Pi_l\Pi_{l'}=\Pi_l\delta_{ll'}. \tag{18}$$

[4]When discussing this issue with people several times I encountered the viewpoint that Ref. [16] has shown that the quantum fluctuating work is not an operator, i.e., this reference ruled out the first definition. This is not correct: Ref. [16] shows that the work obtained via two-time measurements of energy cannot be (in general) represented as an outcome of an operator. But it does not point out any drawback of the Heisenberg-operator-based definition, far from ruling out all possible definitions of the fluctuating work as an operator.

Given the outcomes of this measurement, and provided that we know ρ and H_I, we follow the ideas of Refs. [54,55] and introduce an approximation $f(H_F)$ of H_I. The unknown function $f(.)$ is sought from minimizing the mean-squared difference (the simplest measure of magnitude) as follows:

$$\mathrm{tr}(\rho\,(f(H_F)-H_I)^2). \tag{19}$$

The minimization is straightforward [cf. (18)] [54,55],

$$f(H_F)=\sum_l f_l\Pi_l,\ f_l\equiv\frac{\mathrm{Re\,tr}(\Pi_l\rho H_I)}{\mathrm{tr}(\Pi_l\rho)}. \tag{20}$$

Now f_l is represented via eigenvalues $\epsilon_k(0)$ of H_I,

$$f_l=\sum_k\epsilon_k(0)p_{k|l},\ p_{k|l}\equiv\frac{\mathrm{Re\,tr}(\Pi_l\mathcal{E}_k\rho)}{\mathrm{tr}(\Pi_l\rho)}, \tag{21}$$

where $\mathcal{E}_k\equiv\mathcal{E}_k(0)$. Provided that $p_{k|l}\geqslant 0$, it can interpreted as a conditional probability for the initial energy to be $\epsilon_k(0)$, provided that the final one is $\epsilon_l(\tau)$. This condition does not hold automatically but rather defines the class of states and Hamiltonians for which it is legitimate to interpret $p_{k|l}$ as probabilities; see Sec. VII.

B. Definition of fluctuating work

Hence fluctuating work is defined as a classical random quantity with, respectively, realizations and probabilities as follows:

$$\epsilon_l(\tau)-\epsilon_k(0),\quad p_{kl}\equiv\mathrm{Re\,tr}(\rho\mathcal{E}_k\Pi_l)\geqslant 0. \tag{22}$$

Now $p_{kl}\geqslant 0$ is interpreted as the joint probability for the eigenvalues of H_I and H_F. We could avoid the reasoning of (19)–(21) and just introduce (22) as a postulate.

Note that p_{kl} has correct marginal probabilities,

$$\sum_k p_{kl}=\mathrm{tr}(\rho\Pi_l),\qquad \sum_l p_{kl}=\mathrm{tr}(\rho\mathcal{E}_k). \tag{23}$$

For $[\rho,H_I]=0$ we revert from (22) to (10) and (12) using $\mathcal{E}_k\rho=\mathcal{E}_k\rho\mathcal{E}_k$.

The Cauchy-Schwartz inequality implies $p_{kl}^2\leqslant 1$ as follows:

$$[\mathrm{Re\,tr}(\rho\mathcal{E}_k\Pi_l)]^2\leqslant|\mathrm{tr}(\sqrt{\rho}\mathcal{E}_k\Pi_l\sqrt{\rho})|^2\leqslant\mathrm{tr}(\rho\Pi_l)\mathrm{tr}(\rho\mathcal{E}_k)\leqslant 1.$$

But for specific choices of ρ, p_{kl} can turn negative for given $\mathcal{E}_k$ and Π_l, and then its interpretation as a joint probability is lost; see Sec. VII. From now on and to (38) we assume that $p_{kl}\geqslant 0$.

Condition (ii) holds, since the first and second moments calculated from (22) are equal, respectively, to the first and second moments of the operator H_F-H_I,

$$\mathrm{tr}(\rho(H_F-H_I)^m)=\sum\nolimits_{k,l}p_{kl}(\epsilon_l(\tau)-\epsilon_k(0))^m, m=1,2. \tag{24}$$

However, already the third moments generally differ, the difference involving a double commutator [cf. (18) and (22)],

$$\mathrm{tr}(\rho(H_F-H_I)^3)-\sum\nolimits_{k,l}p_{kl}(\epsilon_l(t)-\epsilon_k(0))^3 = \mathrm{tr}\left(\rho\left[\frac{H_F+H_I}{2},[H_F,H_I]\right]\right). \tag{25}$$

Let us check that condition (i) holds. For cyclic, $H(t)=H(0)$, change of the Hamiltonian, the zero fluctuations of work mean

$$p_{kl}=\mathrm{Re\,tr}(\rho\mathcal{E}_k\Pi_l)=0\quad\text{for all}\quad k\neq l. \tag{26}$$

Employing $0=\mathrm{Re\,tr}(\rho\mathcal{E}_k\Pi_l)=\mathrm{Re\,tr}(\rho(1-\sum_{k'\neq k}\mathcal{E}_{k'})\Pi_l)$ and rearranging the terms we get

$$\mathrm{tr}(\rho\mathcal{E}_l)=\mathrm{tr}(\rho\Pi_l)=\mathrm{Re\,tr}(\rho\mathcal{E}_l\Pi_l) \tag{27}$$

for all l. The first equality here suffices to establish (17). If ρ does not have zero eigenvalues one can find from (26) stronger conditions, but we shall not dwell on that.

In Sec. III C we noted that the two-time energy measurement approach does not apply when $[\rho,H_F]=0$ but (16) holds. It is now seen that the present definition does not have this drawback: for $[\rho,H_F]=0$ we obtain

$$p_{kl}=\mathrm{tr}(\mathcal{E}_k\Pi_l\rho\Pi_l)\geqslant 0. \tag{28}$$

This expression is intuitive, but (for $[\rho,H_I]\neq 0$) it cannot be obtained from the two-time approach, where one first measures energy at $t=0$ and then at $t=\tau>0$.

C. Discussion

The joint probability p_{kl} for noncommuting variables was introduced in Refs. [34–36] (Terletsky-Margenau-Hill distribution). Though it is one of many possible definitions of joint probabilities for noncommuting variables, it is very convenient in the context of quantum statistical mechanics. This point was made in Refs. [35,36] and we shall confirm it below when deriving the generalized fluctuation theorem. As many other joint probabilities (e.g., the Wigner function), p_{kl} can be measured experimentally [56]. Note that f_l in (20) and (21) corresponds to the generalized weak value [57], which is alternatively known as the locally averaged value of energy [58]; this interpretation was employed in (19) and (21). The relation between f_l and p_{kl} was noted in Refs. [59–61]. Also, the form of p_{kl} leads to the most general consistency condition in the history approach to quantum mechanics [37–39] [49]. p_{kl} behaves expectedly under coarse-graining: when two orthogonal subspaces (e.g., described, respectively, by projectors Π_1 and Π_2) are joined into one space (described by $\Pi_1+\Pi_2$), the probabilities are added as follows:

$$p_{k1}+p_{k2}=\mathrm{Re\,tr}((\Pi_1+\Pi_2)\mathcal{E}_k\rho). \tag{29}$$

Reference [62] derives p_{kl} axiomatically and underlines another deep feature of p_{kl}: It is time symmetric, i.e., invariant with respect to interchanging Π_l with $\mathcal{E}_k$. We already noted this feature around (28).

D. Summary

Let us briefly recall why p_{kl} defined in (22) can be regarded as a joint distribution for initial and final energies:

(1) It emerges from estimating two noncommuting observables via one measurement; see (18)–(21).

(2) Whenever any two among three operators Π_l, $\mathcal{E}_k$, ρ commute, p_{kl} reduces to the expected form $\mathrm{tr}[\,\Pi_l\mathcal{E}_k\rho\,]$.

(3) p_{kl} has correct marginals; see (23). This is in contrast to the two-time measurement approach that generally does not reproduce the correct marginals.

(4) It is time symmetric and linear with respect to projectors Π_l and $\mathcal{E}_k$. Hence p_{kl} is additive, much in the same way as the ordinary probability $\mathrm{tr}[\rho\mathcal{E}_k]$.

VI. GENERALIZED FLUCTUATION THEOREM

A. Derivation and interpretation

Following the logic of the equilibrium fluctuation theorem, we take a parameter β and work out using (22) as follows:

$$\sum_{kl} p_{kl} e^{-\beta(\epsilon_l(\tau)-\epsilon_k(0))} = \mathrm{Re}\sum_{kl} \mathrm{tr}(\rho\Pi_l e^{-\beta\epsilon_l(\tau)}\mathcal{E}_k e^{\beta\epsilon_k(0)}).$$

We get from this the following fluctuation theorem:

$$\langle e^{-\beta(w-\Delta\mathcal{F})}\rangle = \mathrm{Re}\,\mathrm{tr}\big(\sigma\,\rho_{\mathrm{eq}}^{-1}(0)\,\rho\big) \equiv \Upsilon, \tag{30}$$

where $w = \epsilon_l(\tau) - \epsilon_k(0)$ are realizations of the random work, $\langle ...\rangle$ means averaging over p_{kl}, and where

$$\rho_{\mathrm{eq}}(t) \equiv e^{-\beta H(t)} \big/ \mathrm{tr}(e^{-\beta H(t)}), \qquad 0 \leqslant t \leqslant \tau, \tag{31}$$

$$\beta\Delta\mathcal{F} \equiv -\ln\,\mathrm{tr}[e^{-\beta H(t)}] + \ln\,\mathrm{tr}[e^{-\beta H(0)}], \tag{32}$$

$$\sigma \equiv U_\tau^\dagger \rho_{\mathrm{eq}}(\tau) U_\tau. \tag{33}$$

Equation (30) relates to each of the other three processes; see Fig. 1. The first of them is the thermally isolated process we focused on: the system starts from the density matrix ρ and Hamiltonian $H(0)$ and (in the Schroedinger representation) ends at the density matrix $\rho(\tau) = U_\tau \rho U_\tau^\dagger$ and Hamiltonian $H(\tau)$. The work w and averaging $\langle ...\rangle$ in (30) refer to this process.

For the second process we imagine that the system (at some preinitial time) is attached to a thermal bath at temperature $1/\beta$ and relaxes to the Gibbsian equilibrium density matrix $\rho_{\mathrm{eq}}(0)$; cf. (31). Then it follows an isothermal quasiequilibrium process, where the Hamiltonian slowly changes from $H(0)$ to $H(\tau)$ under a weak but fixed coupling with the bath. Since the change is slow, the density matrix during the process equals $\rho_{\mathrm{eq}}(t)$, $0 \leqslant t \leqslant \tau$. The work done in this process is given the equilibrium free-energy difference $\Delta\mathcal{F}$ in (30) and (32) [1,2].

During the third process the system at the end of the previous isothermal process is decoupled from the bath and undergoes the reversal of the first thermally isolated process.

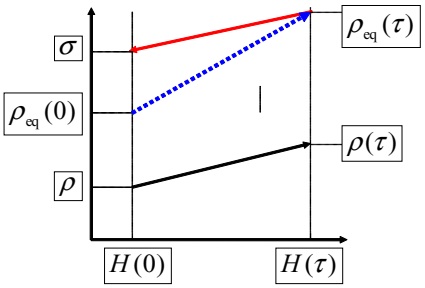

FIG. 1. (Color online) Three processes that appear in (30). They are depicted in a schematic coordinate plane with the x axes (y axes) being Hamiltonian (density matrices) in the Schroedinger representation. Black (lower, full) arrow: The target thermally isolated process. Blue (dashed) arrow: Isothermal process. Red (upper, full) arrow: Another thermally isolated process. Notations refer to (8), (30), and (31).

The final density matrix σ of this process appears in (30) and (33).

In the (initially) equilibrium situation $\rho_{\mathrm{eq}}(0) = \rho$, we revert to the usual (equilibrium) fluctuation theorem,

$$\langle e^{-\beta(w-\Delta\mathcal{F})}\rangle = 1. \tag{34}$$

This theorem relates together characteristics of the first (thermally isolated) and second (isothermal) process. Note that the approach based on two-time measurements of certain observables (not necessarily energy) can also generate fluctuation theorems whose right-hand side is not equal to 1 [an analog of Υ, cf. (30)] [63–65]. There, however, the initial (postmeasurement) state always commutes with the first observable, in contrast to (30), which holds for an arbitrary initial state.

For the equilibrium fluctuation theorem (34) we note that the existence of the bath is necessary for defining the second process (at least when the system is finite, as we assume here). Without the bath, i.e., when the second process is also thermally isolated, the work during the slowest, reversible process is generally *not* given by the free-energy difference [66]. There is a simple way to see this fact explicitly: Any unitary time evolution conserves eigenvalues of the density matrix,

$$\mathrm{Spectrum}[\,U_\tau \rho_{\mathrm{eq}}(0) U_\tau^\dagger] = \mathrm{Spectrum}[\,\rho_{\mathrm{eq}}(0)\,]. \tag{35}$$

Hence $\rho_{\mathrm{eq}}(\tau)$ in (31) cannot be obtained from $\rho_{\mathrm{eq}}(0)$ via a unitary process. Put differently, the equilibrium fluctuation theorem (34) does not generally characterizes the amount of irreversibility (slow versus fast realization) of the thermally isolated process. Instead, it compares two different processes.

Finally, let us again look at (30) and compare it with (34): The equilibrium fluctuation theorem (34) has precisely the same form as the corresponding classical fluctuation relation. This is related to the fact that the equilibrium initial state $\rho_{\mathrm{eq}}(0)$ has classical features with respect to the (initial) energy distribution.[5] In contrast, (30) retains quantum features, since its right-hand side contains noncommutative quantities.

B. Work–free energy relation

Using convexity, $\langle e^x\rangle \geqslant e^{\langle x\rangle}$, we deduce from (22) and (30) a generalization of the usual work–free energy relation,

$$\beta(W - \mathcal{F}) \geqslant \ln\Upsilon \equiv \ln\left(\sum_{kl} \frac{\mu_l}{\nu_k} p_{kl}\right), \tag{36}$$

where μ_l and ν_k are the eigenvalues of σ and $\rho_{\mathrm{eq}}(0)$, respectively. They directly relate to eigenvalues of $H(\tau)$ and $H(0)$. Now $\ln\Upsilon$ can be arbitrarily large, e.g., when one of ν_k is close to zero. Then the equilibrium relation

$$\beta(W - \mathcal{F}) \geqslant 0, \tag{37}$$

carries out to nonequilibrium. Note that $\ln\Upsilon \geqslant 0$ [which guarantees (37)] is not always true; see Appendix B.

[5]Quantum effects are carefully hidden under (34); see, e.g., Ref. [67].

VII. NEGATIVITY OF p_{kl}

The above theory for fluctuations of work was developed under the assumption $p_{kl} \geqslant 0$ [cf. (22)] (though we shall see that, formally, not all results demand this assumption). However, for given projectors $\mathcal{E}_k$ and Π_l with $[\mathcal{E}_k, \Pi_l] \neq 0$, there are ρ's such that

$$p_{kl} = \mathrm{tr}(\rho X_{kl}) < 0, \; X_{kl} \equiv \tfrac{1}{2}(\mathcal{E}_k \Pi_l + \Pi_l \mathcal{E}_k). \tag{38}$$

This is because for $[\mathcal{E}_k, \Pi_l] \neq 0$, X_{kl} has at least one negative eigenvalue [39], e.g., for one-dimensional projectors $\mathcal{E}_k$ and Π_l the nonzero eigenvalues of X_{kl} are

$$\tfrac{1}{2}(\mathrm{tr}(\mathcal{E}_k \Pi_l) \pm \sqrt{\mathrm{tr}(\mathcal{E}_k \Pi_l)}). \tag{39}$$

More generally, for $[\mathcal{E}_k, \Pi_l] \neq 0$ there is a vector $|\psi\rangle$ so $\mathcal{E}_k|\psi\rangle = 0$ but $\mathcal{E}_k \Pi_l |\psi\rangle \neq 0$. Let $X_{kl} = \sum_a x_a |x_a\rangle\langle x_a|$ now be the eigenresolution of X_{kl} and x_1 be the smallest eigenvalue of X_{kl}. We then have

$$x_1 \leqslant \sum_a x_a |\langle x_a|\psi\rangle|^2 = \mathrm{Re}\langle\psi|\mathcal{E}_k \Pi_l|\psi\rangle = 0. \tag{40}$$

This proves that at least the smallest eigenvalue of X_{kl} is negative, since for $[\mathcal{E}_k, \Pi_l] \neq 0$ the inequality in (40) is strict. (It turns into equality for $[\mathcal{E}_k, \Pi_l] = 0$, in which case $x_1 = 0$.) The magnitude of this negativity can be estimated from[6]

$$-\tfrac{1}{8} \leqslant X_{kl} \leqslant 1, \tag{41}$$

where 1 is the unit operator, and, e.g., $X_{kl} \leqslant 1$ means that the eigenvalues of $1 - X_{kl}$ are non-negative. Thus the smallest eigenvalue of X_{kl} is not smaller than $-\frac{1}{8}$ [this is consistent with (39)].

Whenever $p_{kl} < 0$, the usual probability interpretation for p_{kl}, and hence the presented definition of fluctuations of work, do not apply. Nevertheless, the expression (24) for the first and second moments of work, as well as the fluctuation theorem (30), still apply *formally*, i.e., their derivations do not require the validity of $p_{kl} \geqslant 0$. This condition is demanded, e.g., for (36).

However, the positive eigenvalues of X_{kl} are larger than the negative one(s), e.g., due to $\mathrm{tr}[\, X_{kl}] = \mathrm{tr}[\, \Pi_l \mathcal{E}_k \Pi_l \,] \geqslant 0$. Also, in certain cases of $p_{kl} < 0$ we can follow the reasoning of (29) and still define positive probabilities by coarse-graining p_{kl}.

VIII. SUMMARY

This paper is started by studying the applicability of the existing definitions of fluctuating work to nonequilibrium initial states of a quantum system subject to a thermally isolated process. The approach based on two-time energy measurements do not apply for initial states that do not commute with the initial Hamiltonian, because it does not properly reproduce the average work; see Sec. III B. The applicability domain of the operator definition of work is wider, but it is still limited, because this definition does not support (for nonequilibrium initial states) the relation between the work and energy change; see Sec. III A.

The route to defining quantum fluctuating work goes via formulating necessary physical conditions which possible definition should hold. I proposed in Sec. IV that there are (at least) two such restrictions: the fluctuating work should relate to energy change and it should respect the definition (2) of the average work.

I worked out in Sec. V (what seems to me) the simplest definition of the fluctuating work that holds the above two features. This definition does apply to the class of nonequilibrium initial states. Its applicability domain is clearly defined by the non-negativity $p_{kl} \geqslant 0$ of joint probabilities; see (22) and Sec. (VII).

This definition employes only one measurement [by analogy to the definition of work based on the Heisenberg operator (3); see Sec. III A]. For initial states that commute with the initial Hamiltonian this definition reduces to what is obtained with the two-time energy measurements.

I believe that this definition of fluctuating work does advance our understanding of nonequilibrium statistical mechanics, e.g., it allows us to derive a generalized fluctuation theorem which connects together three related processes; see Sec. VI.

ACKNOWLEDGMENTS

I acknowledge discussions with K. Hovhannisyan, E. Sherman, D. Sokolovskii, M. Campisi, and M. Perernau-Llobet. I was supported by COST network MP1209.

APPENDIX A: FLUCTUATIONS OF WORK FOR RABI'S MODEL

1. Rabi's model

The purpose of this section is to illustrate the definition of work fluctuations (proposed in the main text) for Rabi's model: a two-level system driven by an oscillating external field [69,70]. I also compare different definitions of fluctuating work.

There are several reasons why I choose to illustrate the theory of fluctuating work with this specific model. They are as follows:

(1) The model is basic for several fields [quantum optics, NMR or electron spin resonsnce (ESR) physics, etc.].

(2) It is exactly solvable.

(3) The nonequilibrium initial states for this model are theoretically natural and experimentally realizable.

The time-dependent Hamiltonian of the model reads [69,70]

$$H(t) = \frac{\omega}{2}\sigma_z + \frac{g}{2}[\sigma_x \cos(\omega t) + \sigma_y \sin(\omega t)], \tag{A1}$$

where $\omega > 0$ is the (free) frequency of the two-level system, while g quantifies the coupling with the external field.

Here $\sigma_{x,y,z}$ are Pauli matrices. We shall write them in the representation of up $|\uparrow\rangle$ and down $|\downarrow\rangle$ spin states,

$$\sigma_z = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|, \; \sigma_x = |\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|, \tag{A2}$$

$$\sigma_y = -i(|\uparrow\rangle\langle\downarrow| - |\downarrow\rangle\langle\uparrow|). \tag{A3}$$

[6]Inequalities in (41) were derived in Ref. [68] for a slightly more general case of two non-negative operators (not necessarily projections). The first [second] inequality follows from $(\mathcal{E}_k + \Pi_l - \frac{1}{2})^2 \leqslant 1$ $[(\mathcal{E}_k - \Pi_l)^2 \geqslant 0]$ using $\mathcal{E}_k = \mathcal{E}_k^2 \leqslant 1$ and $\Pi_l = \Pi_l^2 \leqslant 1$.

Equation (A1) assumes the resonant case, where the frequencies of the two-level system and external are both equal ω (we take $\hbar = 1$) [69,70].

The eigenvalues

$$\epsilon_1 = \tfrac{1}{2}\sqrt{g^2+\omega^2},\ \ \epsilon_2 = -\tfrac{1}{2}\sqrt{g^2+\omega^2}, \tag{A4}$$

of $H(t)$ do not depend on time. The orthogonal and normalized eigenvectors of $H(t)$ read ($|\epsilon_{1,2}(t)\rangle$ are row vectors)

$$\langle\epsilon_1(t)| = \frac{((\omega+\sqrt{\omega^2+g^2})\,e^{i\omega t},\, g)}{\sqrt{g^2+(\omega+\sqrt{\omega^2+g^2})^2}}, \tag{A5}$$

$$\langle\epsilon_2(t)| = \frac{((\omega-\sqrt{\omega^2+g^2})\,e^{i\omega t},\, g)}{\sqrt{g^2+(\omega-\sqrt{\omega^2+g^2})^2}}. \tag{A6}$$

The unitary operator generated by (A1) is [69,70]

$$U_t = \exp\left[\frac{-i\omega t}{2}\,\sigma_z\right]\exp\left[\frac{-igt}{2}\,\sigma_x\right]. \tag{A7}$$

It satisfies the evolution equation $i\,\partial_t U_t = H(t)U_t$, as verified by direct substitution.

2. Fluctuations of work

The advantage of this situation is that the up and down initial states and their mixtures are natural initial states [69,70]. We thus take the initial state as

$$\rho = \frac{1-\delta}{2}\,|\uparrow\rangle\langle\uparrow| + \frac{1+\delta}{2}\,|\downarrow\rangle\langle\downarrow|, \tag{A8}$$

where $|\delta| \leqslant 1$ is a parameter. For applications in NMR or ESR physics, $|\delta|$ is a small dimensionless number, e.g., $|\delta| \sim 10^{-2}$ [69]; it can be significantly larger in quantum optics [70].

Clearly, the initial state ρ does not commute with the initial Hamiltonian $H(0)$ (excluding the case $\delta = 0$) [cf. (A1), (A5), and (A6)], so we are in the situation described in the main text.

According to (16) of the main text, we get from (A4) three values of the the fluctuating work (A9), (A11), and (A13) with their respective probabilities (A10), (A12), and (A14) (with an obvious adaptation of notations),

$$\sqrt{g^2+\omega^2}, \tag{A9}$$

$$p_+ = \mathrm{Re}\{\langle\epsilon_2(0)|U_t^\dagger|\epsilon_1(t)\rangle\,\langle\epsilon_1(t)|U_t\rho|\epsilon_2(0)\rangle\}; \tag{A10}$$

$$-\sqrt{g^2+\omega^2}, \tag{A11}$$

$$p_- = \mathrm{Re}\{\langle\epsilon_1(0)|U_t^\dagger|\epsilon_2(t)\rangle\,\langle\epsilon_2(t)|U_t\rho|\epsilon_1(0)\rangle\}; \tag{A12}$$

$$0, \tag{A13}$$

$$p_0 = 1 - p_+ - p_-. \tag{A14}$$

Equations (A5), (A6), (A7), (A10), (A12), and (A14) imply

$$p_+ = \frac{\omega^2\sin^2[gt/2]}{2(g^2+\omega^2)}\left(1+\delta\sqrt{\frac{g^2}{\omega^2}+1}\right), \tag{A15}$$

$$p_- = \frac{\omega^2\sin^2[gt/2]}{2(g^2+\omega^2)}\left(1-\delta\sqrt{\frac{g^2}{\omega^2}+1}\right), \tag{A16}$$

$$p_0 = \frac{g^2+\omega^2\cos^2[gt/2]}{g^2+\omega^2}. \tag{A17}$$

Note that $p_+ - p_-$ and δ have the same sign, and this agrees with the logics of the second law (even though the initial state is not in equilibrium): $\delta > 0$ means the lower (down) initial state is more populated [cf. (A8)] and hence the probability of energy increase is larger: $p_+ > p_-$.

The average work agrees with (A15) and (A16) as follows:

$$W = \mathrm{tr}(U_t\rho U_t^\dagger H(t) - \rho H(0)) = \sqrt{g^2+\omega^2}(p_+ - p_-)$$
$$= \delta\omega\sin^2[gt/2]. \tag{A18}$$

Eqs. (A15) and (A16) are non-negative, and hence qualify as probabilities, for

$$1 \geqslant |\delta|\sqrt{\frac{g^2}{\omega^2}+1}, \tag{A19}$$

i.e., for a sufficiently mixed initial state ($|\delta|$ is not close to 1), and/or for a sufficiently small $\frac{g^2}{\omega^2}$ (relatively weak influence on the two-level system). Condition (A19) does not hold, and hence either p_{12} or p_{21} is negative for $|\delta| = 1$ (initially pure state).

3. Two-time measurements of energy

Now the two-time measurement approach produces the same three realizations ($\pm\sqrt{g^2+\omega^2}$, 0), but their probabilities differ as follows:

$$\sqrt{g^2+\omega^2}, \tag{A20}$$

$$\widetilde{p}_+ = \langle\epsilon_2(0)|\rho|\epsilon_2(0)\rangle\,\langle\epsilon_2(0)|U_t^\dagger|\epsilon_1(t)\rangle\,\langle\epsilon_1(t)|U_t|\epsilon_2(0)\rangle, \tag{A21}$$

$$-\sqrt{g^2+\omega^2}, \tag{A22}$$

$$\widetilde{p}_- = \langle\epsilon_1(0)|\rho|\epsilon_1(0)\rangle\langle\epsilon_1(0)|U_t^\dagger|\epsilon_2(t)\rangle\langle\epsilon_2(t)|U_t|\epsilon_1(0)\rangle, \tag{A23}$$

$$0, \tag{A24}$$

$$1 - \widetilde{p}_- - \widetilde{p}_+. \tag{A25}$$

The difference between (A10) and (A12) and then (A21) and (A23) is best visible without working out (A21) and (A23) but looking directly to the average produced by (A21) and (A23),

$$\sqrt{g^2+\omega^2}\,(\widetilde{p}_+ - \widetilde{p}_-) = \frac{\delta\omega^3\sin^2[gt/2]}{g^2+\omega^2}. \tag{A26}$$

It is seen that (A26) does differ from the average work (A18), and hence the approach based on the two-time measurements of energy does not apply.

4. Operator of work

Let us now turn to the operator of work approach. This operator is given as

$$\Delta H(t) = U_t^\dagger H(t)U_t - H(0). \tag{A27}$$

As follows from (A1) and (A7), $\Delta H(t)$ has eigenvalues and (respective) eigenvectors as follows:

$$\Delta_1 = w\sin[gt/2], \tag{A28}$$

$$\langle\Delta_1| = \frac{(-i(\sin[gt/2]-1),\,\cos[gt/2])}{\sqrt{2(1-\sin[gt/2])}}, \tag{A29}$$

$$\Delta_2 = -w \sin[gt/2], \tag{A30}$$

$$\langle \Delta_2| = \frac{(-i(\sin[gt/2]+1), \cos[gt/2])}{\sqrt{2(1+\sin[gt/2])}}. \tag{A31}$$

On the initial state ρ each of these eigenvalues is realized with probabilities

$$\langle \Delta_1|\rho|\Delta_1\rangle = \frac{1+\delta \sin[gt/2]}{2}, \tag{A32}$$

$$\langle \Delta_2|\rho|\Delta_2\rangle = \frac{1-\delta \sin[gt/2]}{2}. \tag{A33}$$

Now the average work (A18) is expectedly reproduced from (A27)–(A33):

$$W = \sum_{k=1,2} \Delta_k \langle \Delta_k|\rho|\Delta_k\rangle. \tag{A34}$$

We compare predictions of the operator of work approach with (A9)–(A14). According to (A9), (A11), and (A13) there are three time-independent realizations of work, while in (A28) and (A30) there are two time-dependent realizations $\pm\omega \sin[gt/2]$. Note that the eigenvalues of the operator of work $\pm\omega \sin[gt/2]$ nullify simultaneously with probabilities (A10) and (A12) for nonzero values. Also, p_- in (A12) can be zero due to $1 = \delta\sqrt{1+g^2/\omega^2}$, indicating that the fluctuations of work are strictly non-negative, while $\pm\omega \sin[gt/2]$ can still assume negative values with nonzero probability.

APPENDIX B: A LOWER BOUND FOR Υ

The factor Υ is defined by (26) of the main text.

To derive a lower bound for Υ we minimize it over ν_k under the constraint $\sum_k \nu_k = 1$ using Lagrange multipliers (recall that ν_k and μ_l are probabilities). This produces the following:

$$\Upsilon \geqslant \left[\sum\nolimits_k p_k^{1/2} \left(\sum\nolimits_l p_{k|l}\mu_l\right)^{1/2}\right]^2. \tag{B1}$$

This lower bound is achievable and its right-hand side is smaller than 1, because it is a squared overlap of two probability vectors: p_k and $\sum_l p_{k|l}\mu_l$. Hence $\ln \Upsilon$ in (26) (of the main text) can be negative.

[1] R. Balian, *From Microphysics to Macrophysics*, Vol. I (Springer, Berlin, 1992).
[2] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (D. Reidel, Dordrecht, 1983).
[3] H. S. Leff and A. F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Princeton University Press, Princeton, NJ, 1990).
[4] I. M. Bassett, Phys. Rev. A **18**, 2356 (1978).
[5] A. Lenard, J. Stat. Phys. **19**, 575 (1978).
[6] W. Thirring, *A Course in Mathematical Physics 4: Quantum Mechanics of Large Systems* (Springer, Vienna, 1983).
[7] G. N. Bochkov and Yu. E. Kuzovlev, Physica A **106**, 443 (1981); **106**, 480 (1981).
[8] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[9] C. Maes, Séminaire Poincaré **2**, 29 (2003).
[10] M. Campisi, P. Hanggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
[11] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).
[12] S. Yukawa, J. Phys. Soc. Jpn. **69**, 2370 (2000).
[13] J. Kurchan, arXiv:cond-mat/0007360 (2000).
[14] H. Tasaki, arXiv:cond-mat/0009244 (2000).
[15] S. Mukamel, Phys. Rev. Lett. **90**, 170604 (2003).
[16] P. Talkner, E. Lutz, and P. Hanggi, Phys. Rev. E **75**, 050102(R) (2007).
[17] T. Monnai and S. Tasaki, arXiv:cond-mat/0308337.
[18] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. E **71**, 066102 (2005).
[19] A. Engel and R. Nolte, Europhys. Lett. **79**, 10003 (2007).
[20] M. F. Gelin and D. S. Kosov, Phys. Rev. E **78**, 011116 (2008).
[21] M. Campisi, New J. Phys. **15**, 115008 (2013).
[22] F. W. J. Hekking and J. P. Pekola, Phys. Rev. Lett. **111**, 093602 (2013).
[23] J. P. Pekola, P. Solinas, A. Shnirman, and D. V. Averin, New J. Phys. **15**, 115006 (2013).
[24] W. De Roeck and C. Maes, Phys. Rev. E **69**, 026115 (2004).
[25] V. Chernyak and S. Mukamel, Phys. Rev. Lett. **93**, 048302 (2004); V. Chernyak, F. Sanda, and S. Mukamel, Phys. Rev. E **73**, 036119 (2006).
[26] Y. Subasi and B. L. Hu, Phys. Rev. E **85**, 011112 (2012).
[27] H. Li and J.-S. Wang, arXiv:1304.6286 (2013).
[28] B. Leggio, A. Napoli, A. Messina, and H.-P. Breuer, Phys. Rev. A **88**, 042111 (2013).
[29] B. Leggio, A. Napoli, H.-P. Breuer, and A. Messina, Phys. Rev. E **87**, 032113 (2013).
[30] J. M. Horowitz, Phys. Rev. E **85**, 031110 (2012).
[31] M. Campisi, P. Talkner, and P. Hanggi, Phys. Rev. Lett. **105**, 140601 (2010).
[32] Fei Liu, Phys. Rev. E **86**, 010103(R) (2012).
[33] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, Europhys. Lett. **66**, 419 (2004).
[34] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933); Y. P. Terletsky, Zh. Eksp. Teor. Fiz **7**, 1290 (1937); P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
[35] A. O. Barut, Phys. Rev. **108**, 565 (1957).
[36] H. Margenau and R. N. Hill, Progr. Theor. Phys. **26**, 722 (1961).
[37] S. Goldstein and D. N. Page, Phys. Rev. Lett. **74**, 3715 (1995).
[38] L. Diosi, Phys. Rev. Lett. **92**, 170401 (2004).
[39] J. B. Hartle, Phys. Rev. A **70**, 022104 (2004).
[40] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[41] T. Hatano and S.-i. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[42] H. Ge and H. Qian, Phys. Rev. E **81**, 051133 (2010).
[43] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[44] W. De Roeck, Compt. Rend. Phys. **8**, 674 (2007).
[45] S. Deffner, Europhys. Lett. **103**, 30001 (2013).
[46] S. Deffner (private communication).
[47] C. R. Willis and R. H. Picard, Phys. Rev. A **9**, 1343 (1974).
[48] J. Gemmer and G. Mahler, Eur. Phys. J. D **17**, 385 (2001).
[49] Two recent papers [27, 26] studied fluctuations of work—by employing what effectively amounts to the operator of work definition—time-locally along effective classical trajectories.

For defining those trajectories they employed the notion of decoherent histories. Note that the drawback pointed out at (5) is just formulated on a consistent history for the Heisenberg operator of work.
[50] Similar drawbacks of the difference between two Heisenberg operators were noted in the context of simultaneous quantum measurements [51].
[51] M. Ozawa, arXiv:quant-ph/0310072 (2003).
[52] G. Watanabe, B. P. Venkatesh, and P. Talkner, Phys. Rev. E **89**, 052116 (2014).
[53] J. Yi and Y. W. Kim, Phys. Rev. E **88**, 032105 (2013).
[54] S. Luo, Int. J. Theor. Phys. **41**, 1713 (2002).
[55] M. J. W. Hall, Phys. Rev. A **69**, 052113 (2004).
[56] C. Bamber and J. S. Lundeen, Phys. Rev. Lett. **112**, 070405 (2014).
[57] J. Dressel, S. Agarwal, and A. N. Jordan, Phys. Rev. Lett. **104**, 240401 (2010).
[58] L. Cohen, Phys. Lett. A **212**, 315 (1996).
[59] D. Sokolovski, Phys. Rev. A **76**, 042125 (2007).
[60] M. Ozawa, AIP Conf. Proc. **1363**, 53 (2011).
[61] T. Sagawa, arXiv:0901.4212 (2009).
[62] L. M. Johansen, arXiv:0804.4379 (2008).
[63] D. Kafri and S. Deffner, Phys. Rev. A **86**, 044302 (2012).
[64] A. E. Rastegin, J. Stat. Mech.: Theo. Exp (2013) P06016.
[65] T. Albash, D. A. Lidar, M. Marvian, and P. Zanardi, Phys. Rev. E **88**, 032146 (2013).
[66] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. E **71**, 046107 (2005).
[67] V. Vedral, arXiv:1204.5559 (2012).
[68] F. Zhang, *Matrix Theory* (Springer, New York, 2011).
[69] W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill, San Francisco, 1964).
[70] V. Vedral, *Modern Foundations of Quantum Optics* (Imperial College Press, London, 2005).