Fluctuations of work from quantum subensembles: The case against quantum work-fluctuation theorems

Allahverdyan, A.E.; Nieuwenhuizen, T.M.

DOI
10.1103/PhysRevE.71.066102

Publication date
2005

Document Version
Final published version

Published in
Physical Review E

Fluctuations of work from quantum subensembles: The case against quantum work-fluctuation theorems

A. E. Allahverdyan[1,2] and Th. M. Nieuwenhuizen[1]
[1]*Institute for Theoretical Physics, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands*
[2]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
(Received 31 August 2004; published 2 June 2005)

We study how Thomson's formulation of the second law of thermodynamics (no work is extracted from an equilibrium ensemble by a cyclic process) emerges in the quantum situation through the averaging over fluctuations of work. The latter concept is carefully defined for an ensemble of quantum systems, the members of which interact with macroscopic sources of work. The approach is based on splitting a mixed quantum ensemble into pure subensembles, which according to quantum mechanics are maximally complete and irreducible. The splitting is done by filtering the outcomes of a measurement process. The approach is corroborated by comparing to relevant experiments in quantum optics. A critical review is given of two other approaches to fluctuations of work proposed in the literature. It is shown that in contrast to those, the present definition (i) is consistent with the physical meaning of the concept of work as mechanical energy lost by the macroscopic sources, or, equivalently, as the average energy acquired by the ensemble; (ii) applies to an arbitrary nonequilibrium state. There is no direct generalization of the classical work-fluctuation theorem to the proper quantum domain. This implies nonclassical scenarios for the emergence of the second law.

DOI: 10.1103/PhysRevE.71.066102 PACS number(s): 05.30.−d, 05.70.Ln

I. INTRODUCTION

A. Classical situation

The second law of thermodynamics was deduced in the 19th century, and formulated for a single closed system, in a way resembling the laws of mechanics [1–6]. It was, however, already the insight of Maxwell [7,8] and Gibbs [9] that this law has in fact a statistical character, and refers to averages over an ensemble of identically prepared systems, rather than to a single system. This viewpoint became widely accepted since the beginning of the 20th century, when the first robust observations of fluctuations were made.[1] Together with the theoretical work of Boltzmann in the kinetic theory of gases and of Smoluchowski, Fokker, Planck, and Einstein in the physics of Brownian motion, they formed a consistent picture of the second law as emerging from microphysics through averaging over fluctuations. A detailed summary of this activity is presented in the book by Epstein [1], while Tolman [2] discusses theoretical aspects of the situation. Since then, the statistical understanding of the second law has entered several modern books of statistical physics and thermodynamics [3,4]. The current perspectives on the classical and quantum Brownian motion in the context of the second law can be found in Refs. [13–15].

At the end of the 1970s several groups independently gave a derivation of Thomson's formulation of the second law [16–22], *no work can be extracted from an initially canonical equilibrium system by means of a cyclic thermally isolated process*, starting directly from quantum or classical Hamiltonian equations of motion. The very possibility of getting this thermodynamical result directly from equations of motion is due to the fact that work is a transparent quantity unambiguously defined both in and out of equilibrium for any (quantum or classical) system interacting with external macroscopic work sources.[2] As the main consequence, Thomson's formulation is the only one that is valid for both finite and infinite systems which do start in equilibrium, but can be driven arbitrarily far from it by external sources (see [23] and Secs. II C and VII for more details).

The standard understanding of the second law and fluctuations is based on Einstein's formula relating entropy to the probability of a fluctuation around equilibrium [1–3,5]. This suffices for the purposes of near-equilibrium thermodynamics of macroscopic bodies, in particular, because all the formulations of the second law are equivalent for them and entropy is defined unambiguously. In the more general case of finite systems and/or systems driven strongly out of their initial equilibrium, relations between the second law and fluctuations ought to be studied anew for each meaningful formulation of the law separately.

The purpose of the present paper is to understand how Thomson's formulation of the second law in the quantum situation emerges through the averaging over fluctuations. More specifically, if the (average) work done on the initially equilibrium ensemble during a cyclic process is always nonnegative, what are fluctuations of this work, and how do they behave? There are definite answers to these questions in the classical situation: the definition of fluctuations of work is straightforward, and model-independent information on them is given by an equality first derived by Bochkov and Kuzovlev in 1977 [16] (BK equality). Later on, this equality, sometimes also called the work-fluctuation theorem, was extended to noncyclic processes [24], and has undergone various

[1]It was thus rather surprising to see recent claims on "violations of the second law" [10] or "transient violations of the second law'' [11] due to fluctuations; see in this context our comment [12].

[2]These features of work are in contrast to those of entropy, whose meaning is too closely tied to equilibrium states of macroscopic bodies.

generalizations.[3] The basic messages of the classical BK equality are recollected and reviewed in Sec. II E.

B. Quantum situation

While these developments concern the classical situation, a number of recent works are devoted to quantum extensions of the BK equality [27–32]. The first definition of fluctuations of work in the quantum situation (and a quantum extension of BK equality) was proposed by Bochkov and Kuzovlev themselves [16,18]. It is based on constructing a certain operator in the Heisenberg representation, associating it with an operator of work, and thus treating work as an ordinary quantum mechanical observable pertaining to the system and not to the work source.

Another extension was initiated by Kurchan [28], based on two-time measurements of energy. This second approach is closely tied to the Schrödinger representation.

There are therefore two different approaches to the definition of fluctuations of work and to quantum extensions of the BK equality; both of them attracted attention recently [27,29–32], and are reviewed below in Sec. VI. However, the fact that in the quantum situation these two approaches for defining fluctuations of work are different [32] is already calling for attention to the situation.[4]

C. General conditions on fluctuations of work

Our objective is to propose a third definition of fluctuations of work, which is motivated by the fundamental physics of quantum (sub)ensembles. The definition is guided by the following observation. Since the usual work is now presented as an average of a random quantity—for the moment we leave unspecified whether this is a random classical quantity or an operator—it is natural to require the following two conditions on its (fluctuating) realizations and on its average.

(1) Once the average work is unambiguously defined for any quantum system starting in an arbitrary initial state and interacting with a macroscopic source of work, the same should hold for fluctuations of work. In particular, the general definition should *not* be restricted to (initially) equilibrium states of the system, since one part of statistical thermodynamics deals with work extraction from nonequilibrium systems [3,33], and one should, of course, be able to define fluctuations of work in this most general situation.

(2) Realizations of the random quantity work should have the same physical meaning of mechanical (high-graded) energy as the usual (average) work. In particular, if one happens to extract some work from a single realization, one should be able—at least in principle—to use precisely this amount for the standard purposes, e.g., for driving motors. (Basic features of work are recalled in Sec. II B.)

Both these conditions are naturally satisfied by the classical definition, and in our opinion without them the very program of studying the emergence of the second law in the quantum situation becomes ill defined.

It appears to the present authors that, as we discuss below in Sec. VI, neither of the existing two quantum approaches—in the way they stand presently—can be viewed as providing a proper definition of fluctuations of work in the quantum situation. Both approaches fail out of equilibrium (no first condition), while even for an initially equilibrium state it is not clear that the second condition is satisfied.[5]

D. A different approach to fluctuations of work in the quantum situation

These are the reasons to introduce in the present paper a *possible* approach to quantum fluctuations of work that will satisfy the above two conditions. It starts with explicitly respecting the first condition, that is, always defining realizations of (the random quantity) work as some average energy given off by the macroscopic source of work. If the corresponding ensemble of physical systems already consists of subensembles, nontrivial realizations can be defined via the average energy exchange of each subensemble with the source.[6] For a classical ensemble each single member completely characterizes a subensemble, and the classical definition of fluctuations of work follows naturally. In contrast, a quantum equilibrium state is described by a homogeneous quantum ensemble, the Gibbsian ensemble, which by itself does not consist of subensembles. This prevents us from proceeding as such. First, the Gibbsian ensemble has to split with the help of a selective quantum measurement into a set of (homogeneous) subensembles. Thus, the initial Gibbsian ensemble is transformed into an inhomogeneous ensemble (with the same density overall matrix) which already *consists* of the subensembles. The obtained structure of these subensembles does depend on the type of measurement, and as a consequence the resulting fluctuations of work in the quantum situation appear to be context dependent. The role of contextuality in quantum physics has been strressed over and over again; see, e.g., [34,35]. Second, systems from each subensemble interact with the work source which realizes on them the same process. Realizations of work are defined via the average energy[7] received by each subensemble. Each realization has its probability naturally determined by the

[3]A rather complete account of various generalizations of the classical work-fluctuation theorem, as well as its relation with other fluctuation theorems, e.g., those describing entropy production, is given in Refs. [25,26]. Local versions of the fluctuation theorems are also discussed there.

[4]The difference in viewpoints is not completely unexpected, since the work as it appears in statistical thermodynamics [2–5] is an essentially classical quantity (mechanical energy transferred from a classical source of work).

[5]Neither of these points was discussed in papers which support those definitions; see, e.g., [16,18,27,28,30–32].

[6]As with any exchange process, this is operationally characterized by measurements at two different times.

[7]Normally this averaging is done either by letting many identically prepared systems interact with the work source, or by operating with a single system but repreparing its state after each interaction period. Both these ways are feasible and are realized experimentally; see Sec. III G for more details.

weight of the corresponding subensemble in the overall (inhomogeneous) ensemble.

Within this approach each realization of the random quantity work is already a partially averaged quantity: it refers to a subensemble rather than to a single system. Recall that already in classical statistical physics there are two strategies for studying fluctuations [3]. In the first case, the studied quantity (e.g., energy) has a direct mechanical meaning. In this situation it can be measured directly, and its fluctuations are represented by outcomes that vary from one measurement to another. In other cases the quantity of interest (e.g., the temperature) has itself only a statistical meaning. Then some averaging has to be done before realizations of this random quantity can be obtained. In practice, it will often happen that nature is doing this for us, for example, when we measure the temperature of a liquid. There are cases, however, where the averaging has to be done by hand, which is close to the typical coarse-graining done in sociological experiments (e.g., the height of persons being in a certain interval).

In quantum mechanics a similar situation shows up: when measuring the energy of a particular member of an ensemble, this member can be said to have that value of the energy after the measurement, since quickly repeated measurements will give the same answer. The approach of this paper will be to put forward the idea that fluctuations of work in the quantum situation have the same nature as fluctuations of temperature in the classical situation: subensembles are needed to define (partially averaged) values of work.

The presented approach, deduced from conceptual consistency, appears to be related to certain quantum optical experiments, which (though never interpreted in this way) have been realized by observing fluctuations of work; see Sec. III G.

E. Implications for the emergence of the second law

As one of the main results of our approach, the second law in Thomson's formulation—whose statement reads in the same way in both quantum and classical regimes—has in those two situations rather different scenarios of emergence. The basic qualitative difference is that, in contrast to classics, the fluctuations of work in the proper quantum situation are not controlled by any direct analog of the BK equality. More specifically, in classics the structure of work as a random quantity is such that there have to be realizations that provide work (i.e., that are active). In the quantum case, however, there need not be any active realization (active subensemble).

We have taken the simplest situation that allows one to study Thomson's formulation of the second law and fluctuations of work, that is, we consider a finite quantum or classical system interacting with external sources of work. The restriction to finite—though possibly large—systems is at any rate natural for studying fluctuations, and allows us to focus on the conceptual issues connected to fluctuations of work. The approach is generalized directly for systems coupled to an environment (e.g., thermal baths), and under certain natural conditions allows us to express fluctuations of work via quantities referring to the open system only (see Sec. III F).

F. Guidelines for reading this paper

The paper is intended to be self-consistent and is organized as follows. In Sec. II we recall the definition of fluctuations of work in the classical situation and review the BK equality and its consequences relevant for the emergence of the second law. In Sec. III we present the definition of fluctuations of work in the quantum situation. This is the central section of the paper which addresses the definition of work and physics of quantum ensembles, and finishes with a detailed discussion of the physical meaning of the approach. This section also outlines generalizations of the approach and connects it to relevant experiments. The dispersion of work is studied in Sec. IV. In Sec. V we show that fluctuations of work in the quantum situation are not controlled by any direct analog of the classical BK equality. An anticlassical scenario for the emergence of the second law in Thomson's formulation is described in Sec. IV.

In Sec. VI we make a comparison with the two known approaches on fluctuations of work in the quantum situation (no preliminary knowledge of these subjects is assumed). These approaches offer different extensions of the classical BK equality. We do not intend to imply that these approaches do not have a physical meaning or that they cannot be useful for their own sake. We only state that—in the way they stand presently—they do not describe fluctuations of work in the proper quantum situation. Sections III and VI can be read independently from the rest of the paper.

We close with a summary of our results. Some details are worked out in Appendixes.

II. CLASSICAL FLUCTUATIONS OF WORK AND BK EQUALITY

A. The setup

Consider an ensemble $\mathcal{E}$ of identical classical systems $\mathcal{S}$ which are thermally isolated [3,4]: they move according to their own dynamics and interact with an external macroscopic work source $\mathcal{W}$. This interaction is described via the time dependence of some parameters $R(t)=\{R_1(t), R_2(t),\ldots\}$ of the system's Hamiltonian $H(t)=H\{R(t)\}$; see Refs. [3,4].

The parameters move along a certain trajectory $R(t)$ which at some initial time $t=0$ starts from $R(0)$, and ends at $R(\tau)$ at the final time $t=\tau$. Cyclic thermally isolated processes are defined by $R(0)=R(\tau)$ and thus

$$H\{R(\tau)\}=H\{R(0)\}\equiv H. \tag{1}$$

At the initial time the ensemble is in equilibrium, that is, the common probability distribution $\mathcal{P}(x,p;t=0)\equiv\mathcal{P}(x,p)$ of all its canonically conjugated coordinates $x=(x_1,\ldots,x_n)$ and momenta $p=(p_1,\ldots,p_n)$ is given by the Gibbs distribution with the initial Hamiltonian $H(x,p)$ and temperature $T=1/\beta\geqslant 0$:

$$\mathcal{P}(x,p) = \frac{e^{-\beta H(x,p)}}{Z}, \quad Z = \int dx\, dp\, e^{-\beta H(x,p)}. \tag{2}$$

This equilibrium distribution can be prepared by means of a thermal bath coupled with the system $\mathcal{S}$ for $t<0$. It is assumed that for times $0 \leqslant t \leqslant \tau$ the system $\mathcal{S}$ is decoupled from the bath—an alternative assumption would be that its coupling to the bath is so weak that it can be neglected in the considered time interval—and the evolution of the ensemble is described by the Liouville equation for $\mathcal{P}(x,p;t)$:

$$\partial_t \mathcal{P}(x,p;t) = \frac{\partial H(x,p,t)}{\partial p}\frac{\partial \mathcal{P}(x,p;t)}{\partial x} - \frac{\partial H(x,p,t)}{\partial x}\frac{\partial \mathcal{P}(x,p;t)}{\partial p}. \tag{3}$$

B. Work

In statistical thermodynamics there are two alternative definitions of work [2–4,6,36]. Both are necessary for the proper understanding of its physical meaning [4,36,37]. The first reads as follows. The work W is the average energy gained by $\mathcal{S}$ during a thermally isolated system-work-source interaction with $\mathcal{W}$ [3,4]:

$$W = \int dx\, dp[\mathcal{P}(x,p;\tau)H(x,p;\tau) - \mathcal{P}(x,p)H(x,p;0)]. \tag{4}$$

Due to conservation of energy, W is equal to the average energy lost by the work source $\mathcal{W}$. This definition was (implicitly) proposed by Caratheodory [36]. A concise history of various definitions of work is given in [38], while various perspectives of work in classical mechanics are reviewed in [39].

For cyclic processes Eq. (4) takes a simpler form,

$$W = \int dx\, dp[\mathcal{P}(x,p;\tau) - \mathcal{P}(x,p)]H(x,p;0). \tag{5}$$

There is a second, alternative definition going back to Gibbs and Planck [36,38]: The negative work $-W$ is the energy transferred to the work source $\mathcal{W}$. Its distinguishing feature with respect to other forms of energy is that it can, in principle, be transferred with 100% efficiency to other work sources via interactions of the system-work-source type. In particular, it can be retransferred to collective degrees of freedom that perform *classical deterministic* motion generated by a suitable Hamiltonian. These degrees of freedom are thus purely mechanical and serve as prototypes of macroscopic mechanical devices (such as a motor, piston, turbine, etc.). For them the differential work can be calculated in the usual way of ordinary mechanics, that is, multiplying the external force by the corresponding displacement [4].

Both these definitions of work are expected to be equivalent at least for sufficiently ideal work sources [4,36,37].

There is yet another, equivalent formula for the work W: the integral of the rate of energy change,

$$W = \int_0^\tau dt \int dx\, dp\, \mathcal{P}(x,p;t)\frac{\partial H(x,p;t)}{\partial t}. \tag{6}$$

To get from here to Eq. (5) one performs integration by parts, uses the standard boundary conditions, that is, $\mathcal{P}(x,p;t)$ decays for $x \to \pm\infty$ or $p \to \pm\infty$, and employs Eq. (3). This formula for W is more general and can be applied to processes that involve an environment.

C. Fluctuations of work

Though the ensemble $\mathcal{E}$ is described by the probability distribution $\mathcal{P}(x,p)$, each single system $\mathcal{S}$ from this ensemble has at a given moment of time explicit values for all its dynamical variables. These values may vary from one single system to another due to the distribution of initial conditions.

Each single member of the ensemble is then coupled to the external source of work that realizes on it a unique thermally isolated process (the same for all members). In other words, the same parameters $R(t)$ of the Hamiltonian are varied in the same way for each member. The motion of the single system is described by Eq. (3) with now $\mathcal{P}(x,p;t)$ being a product of two δ functions $\delta(x-x(t))\delta(p-p(t))$, which are probability densities concentrated at the solutions of the canonical equations of motion:

$$\dot{p} = -\partial_x H(x,p;t), \quad \dot{x} = \partial_p H(x,p;t). \tag{7}$$

The trajectories generated by (7), together with their initial conditions distributed according to Eq. (2), serve as *realizations* of the random process given by Eq. (3).

The work $w(x,p)$ exchanged in each thermally isolated process can then be calculated consistently with Eq. (5):

$$w(x,p) = H(x(\tau),p(\tau);\tau) - H(x,p) \tag{8}$$

$$= H(x(\tau),p(\tau)) - H(x,p), \tag{9}$$

where $H(x(\tau),p(\tau);\tau)$ is the value of the Hamiltonian on the trajectory that started at $t=0$ from (x,p), with $x(\tau)$ and $p(\tau)$ being the corresponding solutions of (7). This work can be observed as the energy decrease of the mechanical degree of freedom of the macroscopic work source, or alternatively via energy increase of the system $\mathcal{S}$. In this latter scenario the energy of $\mathcal{S}$ has to be measured twice, at the moments $t=0$ and τ.

The work $w(x,p)$ for a single system is a random quantity, since it varies from one single system to another. It can be positive or negative. Its probability distribution $P(w)$ is determined by $\mathcal{P}(x,p)$, since this is the probability with which each single system enters in the ensemble:

$$P(w) = \int dx\, dp\, \mathcal{P}(x,p)\delta(w - w(x,p)). \tag{10}$$

There being used no special features of the initial equilibrium distribution function, the same definition for the work in a single realization can be given for any initial ensemble. It is seen that the two desired conditions for fluctuations of work formulated in Sec. I are naturally satisfied: the initial

distribution may be arbitrary and "work for a single realization" has the same physical meaning as average work.

D. Derivation of BK equality

One now derives the BK equality in the classical situation for a closed cycle [16,18,24]:

$$\langle e^{-\beta w}\rangle \equiv \int dw\, P(w)e^{-\beta w} = \int dx\, dp\, \mathcal{P}(x,p;0)e^{-\beta w(x,p)}$$
$$= \frac{1}{Z(0)}\int dx\, dp\, e^{-\beta H(x,p)-\beta w(x,p)}$$
$$= \frac{1}{Z(0)}\int dx\, dp\, e^{-\beta H(x(\tau),p(\tau);\tau)}$$
$$= \frac{1}{Z(0)}\int dx(\tau)dp(\tau)e^{-\beta H(x(\tau),p(\tau);\tau)} \quad (11)$$

$$= \frac{Z(\tau)}{Z(0)} = 1, \quad (12)$$

where we used the Liouville theorem $dx\,dp = dx(\tau)dp(\tau)$ and Eqs. (2), (9), and (10). The last equality in Eq. (12) is due to the assumed cyclic feature of the process.

E. Qualitative messages of the BK equality

The BK equality is by itself an exact mathematical relation. Several important qualitative results can be deduced from it.

1. The second law

As the exponential function is convex, one gets directly $1=\langle e^{-\beta w}\rangle \geqslant e^{-\beta\langle w\rangle}$, and then $W=\langle w\rangle \geqslant 0$, which is the statement of the second law in Thomson's formulation: no work can be extracted from an equilibrium system by means of a cyclic process. This formulation of the second law is well known and has an independent and more general derivation in both the classical and quantum situations [16,19–22,40].

2. Active realizations

To satisfy $1=\langle e^{-\beta w}\rangle$ directly leads to the following observation: for any cyclic thermally isolated process there are realizations which are active, that is, for which work is extracted after the process: $w(x,p)<0$. The relative weight of such active realizations can be estimated via the Cauchy inequality:

$$1 = \left(\int dx\, dp \sqrt{\mathcal{P}(x,p)}\sqrt{\mathcal{P}(x,p)}e^{-\beta w(x,p)}\right)^2$$
$$\leqslant \int dx\, dp\, \mathcal{P}(x,p)\int dx\, dp\, \mathcal{P}(x,p)e^{-2\beta w(x,p)}, \quad (13)$$

which can be written as

$$\langle e^{-2\beta w}\rangle \geqslant 1. \quad (14)$$

A stronger relation is obtained using the generalized Cauchy inequality (see Appendix A)

$$\langle e^{-2\beta w}\rangle \geqslant 1 + \frac{[\langle (f-\langle f\rangle)e^{-\beta w}\rangle]^2}{\langle (f-\langle f\rangle)^2\rangle} > 1, \quad (15)$$

where $f(x,p)$ is an arbitrary integrable function in the phase space, and where

$$\langle f\rangle \equiv \int dx\, dp\, \mathcal{P}(x,p)f(x,p). \quad (16)$$

Equation (15) is stronger than Eq. (14), since now $\langle e^{-2\beta w}\rangle$ is strictly larger than 1. Equations (14) and (15) allow us to understand how relevant the active realizations are with respect to both their probability and the amount of extracted work.

3. Dispersion of work

For sufficiently high temperatures one can make a cumulant expansion

$$1 = \exp\left(-\beta\langle w\rangle + \frac{\beta^2}{2}(\langle w^2\rangle - \langle w\rangle^2) + \cdots\right) \quad (17)$$

which shows that for sufficiently high temperatures the ratio of the dispersion of work $\langle w^2\rangle - \langle w\rangle^2$ and its average increases with temperature:

$$\frac{\langle w^2\rangle - \langle w\rangle^2}{\langle w\rangle} = 2T. \quad (18)$$

A detailed survey of various cumulant expansion-based results derivable from the BK equality is contained in Refs. [16–18].

F. Noncyclic processes

For noncyclic processes there is an analog of the equality (12), which is derived in a similar way with the conclusion [24] $\langle e^{-\beta w}\rangle = e^{-\beta[F(\tau)-F(0)]}$, where $F(\tau)=-T\ln e^{-\beta H(\tau)}$ is the corresponding free energy. This relation allows us to calculate differences of free energy via (nonequilibrium) measurements of work.[8]

This generalized equality is not directly relevant for our present purposes, because here we are interested in the second law in Thomson's formulation which refers to cyclic processes.

III. QUANTUM ENSEMBLES AND THE DEFINITION OF FLUCTUATIONS OF WORK

A. The setup

The quantum setup for studying thermally isolated processes is a straightforward extension of the classical one. (We denote all operators by a care.)

An ensemble $\mathcal{E}$ of identically prepared quantum systems $\mathcal{S}$ is described at $t=0$ by a density matrix $\hat{\rho}(0)=\hat{\rho}$. The eigenresolutions of $\hat{\rho}$ and of the Hamiltonian $\hat{H}$ read

[8] A number of issues related to this point were discussed in a recent exchange of opinions [41,42].

$$\hat{\rho} = \sum_{k=1}^{n} p_k |p_k\rangle\langle p_k|, \tag{19}$$

$$\hat{H} = \sum_{k=1}^{n} \varepsilon_k |\varepsilon_k\rangle\langle \varepsilon_k|, \tag{20}$$

where $\{|\varepsilon_k\rangle\}_{k=1}^n$ and $\{|p_k\rangle\}_{k=1}^n$ with $\langle \varepsilon_k|\varepsilon_l\rangle = \langle p_k|p_l\rangle = \delta_{kl}$ are the eigenvectors of $\hat{H}$ and $\hat{\rho}$, respectively, which form bases in the n-dimensional Hilbert space $\mathcal{H}$, and where ε_k and p_k are the corresponding eigenvalues.

Frequently, but not always, we will consider initially Gibbsian density matrices:

$$\hat{\rho}(0) = \hat{\rho} = \frac{e^{-\beta\hat{H}}}{Z}, \quad Z = \mathrm{tr}\, e^{-\beta\hat{H}}, \tag{21}$$

$$p_k = \frac{e^{-\beta\varepsilon_k}}{\sum_{k=1}^{n} e^{-\beta\varepsilon_k}}, \quad |p_k\rangle = |\varepsilon_k\rangle, \quad k = 1, \ldots, n, \tag{22}$$

where $T = 1/\beta \geqslant 0$ is the temperature of the ensemble. We shall order the eigenvalues of $\hat{H}$ as

$$\varepsilon_1 \leqslant \varepsilon_2 \leqslant \cdots \leqslant \varepsilon_n. \tag{23}$$

Then according to (22), the eigenvalues of $\hat{\rho}$ will be ordered as

$$p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n > 0. \tag{24}$$

For the Gibbsian density matrix all eigenvalues are strictly positive.

Analogously to the classical case, the Gibbsian state (21) is prepared for $t < 0$ by letting $\mathcal{S}$ interact with a macroscopic thermal bath, and then decoupling it from the bath, so that the interaction is absent for $t > 0$. There is, however, a relevant difference between quantum and classical: in the quantum situation the coupling of $\mathcal{S}$ with the bath has to be weak for the stationary state of $\mathcal{S}$ to be Gibbsian.[9] A detailed analysis of this and similar differences between the Gibbs distribution in quantum and classical situations is presented in [14,15].

At $t = 0$ $\mathcal{S}$ starts to interact with an external macroscopic work source $\mathcal{W}$. The resulting evolution of $\mathcal{S}$ is generated by (an effective) Hamiltonian $\hat{H}\{R(t)\}$, which is time dependent via classical (c-number) parameters $R(t)$.[10] The evolution of $\mathcal{S}$ is thus unitary and has the same general features of reversibility as the dynamics of a completely isolated $\mathcal{S}$. It is well known that in general a Hamiltonian evolution of the complete system $\mathcal{S}+\mathcal{W}$ does not reduce to a Hamiltonian evolution for the state of $\mathcal{S}$. However, in the present case this is achieved owing to the *macroscopic* character of $\mathcal{W}$, as discussed in [4].

A cyclic process at the moment $t = \tau$ is defined in the same way as in classics, $R(\tau) = R(0)$, leading to

$$\hat{H}(\tau) = \hat{H}(0) = \hat{H}. \tag{25}$$

The Hamiltonian $\hat{H}(t)$ generates a unitary evolution:

$$i\hbar \frac{d}{dt}\hat{\rho}(t) = [\hat{H}(t), \hat{\rho}(t)], \tag{26}$$

$$\hat{\rho}(t) = \hat{U}_t \hat{\rho}(0) \hat{U}_t^{\dagger}, \tag{27}$$

$$\hat{U}_t = \overleftarrow{\exp}\left(-\frac{i}{\hbar}\int_0^t ds\, \hat{H}(s) \right), \tag{28}$$

where $\overleftarrow{\exp}$ and $\overrightarrow{\exp}$ denote time-ordered and time-antiordered exponents, respectively.

B. Work

The whole discussion in Sec. II B directly applies in the quantum situation, except that $\mathcal{S}$ is now a quantum system, and Eqs. (4) and (5) should be substituted by their quantum analogs (i.e., $\mathcal{P} \to \hat{\rho}$, $H \to \hat{H}$, and $\int dx\, dp \to \mathrm{tr}$). In particular, the work W done by the external source $\mathcal{W}$ is identified with the average energy change [3,4]

$$W = \mathrm{tr}[\hat{\rho}(\tau)\hat{H} - \hat{\rho}\hat{H}] = \mathrm{tr}\, \hat{\rho}\hat{\Omega}, \tag{29}$$

where we denoted

$$\hat{\Omega} \equiv \hat{U}_\tau^{\dagger}\hat{H}(\tau)\hat{U}_\tau - \hat{H} = \hat{U}_\tau^{\dagger}\hat{H}\hat{U}_\tau - \hat{H}. \tag{30}$$

Here $\hat{U}_\tau^{\dagger}\hat{H}(\tau)\hat{U}_\tau$ is the Hamiltonian operator in the Heisenberg representation at the end time τ of the cyclic process. This operator $\hat{\Omega}$ is sometimes called the "operator of work" [6,16,27]. We shall show, however, in Sec. IV A that it is not clear whether it satisfies all criteria to deserve this identification. Moreover, the much weaker interpretation of $\hat{\Omega}$—by analogy to the classical expression (8)—as the "energy difference operator in the Heisenberg representation" is also incorrect in general; see Sec. IV A. In our approach $\hat{\Omega}$ will always appear inside averages over density matrices as in the definition of work (29); we do not need any particular interpretation of $\hat{\Omega}$.

The remarks we made after Eq. (5) for the classical situation are valid in the quantum case as well. W is equal to the average energy decrease of the work source $\mathcal{W}$. This is a classical, mechanical energy which can be transferred with 100% efficiency to an other work source, and, in particular, it can be transferred to another mechanical degree of freedom performing classical deterministic motion. In that respect both the classical and quantum definitions are consistent and can be indistinguishable from the viewpoint of this mechanical degree. This property is the underlying reason why phenomenological thermodynamics, where no (quantum or classical) identification of $\mathcal{S}$ is given, can and does exist.

[9] Due to weak coupling to the bath, the energy costs for switching the interaction on and off become negligible. This holds in both the quantum and the classical situations [14,15].

[10] Note that this time dependence is in the Schrödinger representation. To avoid confusion we do not deal with the implicit Heisenberg representation.

The work is typically observed via suitable (classical) measurements done on the work source, or, alternatively, by measuring the initial and final average energies on the ensemble $\mathcal{E}$. Both these ways are routinely employed in practice, e.g., in NMR and ESR physics, where the system $\mathcal{S}$ corresponds to spin $\frac{1}{2}$ under the influence of external magnetic fields [43].

Finally, the quantum analog of formula (6) reads

$$W = \int_0^\tau dt \, \mathrm{tr}\left(\hat{\rho}(t) \frac{d\hat{H}(t)}{dt} \right). \tag{31}$$

Equation (29) can be recovered from this formula upon integration by parts and using (26).

C. Quantum ensembles

The definition of fluctuations of work in the classical situation was based on the distinction between classical ensembles of systems described by a probability distribution versus a single member of that ensemble. It should not be surprising that fluctuations of work in the quantum situation are closely tied to the meaning of what is a quantum ensemble.

Thus, for our further purposes we need an account of various features of quantum ensembles and their differences with respect to the classical ones. There are several sources in literature [34,44–48] where this type of question is studied with special attention.[11]

1. Statistical interpretation of quantum mechanics

Within this interpretation of quantum mechanics a quantum "state" is described by a density matrix $\hat{\rho}$ [34,44–48]. Any state, including a pure state $|\psi\rangle\langle\psi|$, describes an ensemble of identically prepared systems. For instance, in an ideal Stern-Gerlach experiment all particles of the upper beam together are described by the wave function $|\uparrow\rangle$ or the pure density matix $|\uparrow\rangle\langle\uparrow|$. The description is optimal, in the sense that all particles have $\sigma_z = +1$, but incomplete in the sense that σ_x and σ_y are unknown: upon measuring either of them, one will get ± 1 with equal probabilities.

This interpretation suffices for describing experiments, including those done on a single system [34,48,52,53]. As compared to other interpretations of quantum mechanics, the statistical interpretation is dealing more succesfully with a number of conceptual problems, including the quantum measurement problem [33,34,54].

2. Homogeneous ensembles

In general, a density matrix $\hat{\rho}$ can be applied to describe two types of quantum ensembles, *homogeneous* and *inhomogeneous*.

For a homogeneous ensemble $\mathcal{E}(\hat{\rho})$ only the density matrix $\hat{\rho}$ is given and no further specification is made about a single system $\mathcal{S}$ from that ensemble. A typical example is an ensemble prepared by thermalization, that is, by letting each single system $\mathcal{S}$ interact weakly with an equilibrium thermal bath, and waiting sufficiently long until the equilibrium state of $\mathcal{S}$ is established.

Let us study the features of homogeneous ensembles in more detail. We start by comparing them to classical ensembles. In the classical situation, the description of an ensemble by means of a probability distribution still implies that each single system has definite values for *all* its variables. For a homogeneous quantum ensemble $\mathcal{E}(\hat{\rho})$, only those observables (Hermitian operators existing in the Hilbert space $\mathcal{H}$) $\hat{A}$ that are dispersionless on $\mathcal{E}(\hat{\rho})$,

$$[\mathrm{tr}(\hat{A}\hat{\rho})]^2 = \mathrm{tr}(\hat{A}^2\hat{\rho}), \tag{32}$$

can be said to have definite values for all single systems $\mathcal{S}$ from $\mathcal{E}(\hat{\rho})$. Indeed, it is shown in Appendix C that dispersionless observables satisfy

$$\hat{A}\hat{\rho} = \alpha\hat{\rho}, \tag{33}$$

where α is a c number. This implies $\mathrm{tr}(\hat{A}^m\hat{\rho}) = [\mathrm{tr}\,\hat{A}\hat{\rho}]^m$, with $m = 0, 1, 2, 3\ldots$, and the above statement follows. For a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$, we return from Eq. (33) to the standard notion of $|\psi\rangle$ being an eigenstate of $\hat{A}$.

Any other, nondispersionless observable $\hat{B}$—even if it commutes with the density matrix $\hat{\rho}$—does not have a definite value in a single system $\mathcal{S}$ from $\mathcal{E}(\hat{\rho})$. It is true that for $[\hat{\rho}, \hat{B}] = 0$, $\mathcal{E}(\hat{\rho})$ can be prepared by mixing[12] pure states ensembles $\{\mathcal{E}(|p_k\rangle\langle p_k|)\}_{k=1}^n$ with probabilities $\{p_k\}_{k=1}^n$, where $\{|p_k\rangle\}_{k=1}^n$ and $\{p_k\}_{k=1}^n$ are, respectively, the common eigenvectors of $\hat{\rho}$ and $\hat{B}$ and the eigenvalues of $\hat{\rho}$. If $\mathcal{E}(\rho)$ is *known* to be prepared in such a way, then $\hat{B}$ has indeed definite values for each single member of $\mathcal{E}$. However, in general this need not apply, since there are (infinitely) many other ways to prepare the same ensemble $\mathcal{E}(\hat{\rho})$ via mixing N subensembles with density matrices $\{|\psi_\alpha\rangle\langle\psi_\alpha|\}_{\alpha=1}^N$ and probabilities $\{\lambda_\alpha\}_{\alpha=1}^N$. They correspond to the (infinitely) many ways in which the Hermitian operator $\hat{\rho}$ can be decomposed as [34,46–48]

[11] Though the theory of quantum ensembles is almost as old as quantum mechanics itself, it still attracts lively discussions; see, e.g., [49–51]. It is interesting to note that the basic differences between classical and quantum ensembles were correctly understood by Elsasser as early as in 1937 [44].

[12] Mixing ensembles $\mathcal{E}(\hat{\rho}_1)$ and $\mathcal{E}(\hat{\rho}_2)$ with probabilities p_1 and p_2, respectively, means that one throws a dice with probabilities of outcomes equal to p_1 and p_2, and depending on the outcome one picks up a system from $\mathcal{E}(\hat{\rho}_1)$ or $\mathcal{E}(\hat{\rho}_2)$, keeping no information on where the system came from. Alternatively, one can join together Np_1 systems from $\mathcal{E}(\hat{\rho}_1)$ and Np_2 systems from $\mathcal{E}(\hat{\rho}_2)$ ($N \gg 1$), so that no information is kept on where a single system came from. Then any subensemble of M systems ($N \gg M$) is described by the density matrix $\hat{\rho} = p_1\hat{\rho}_1 + p_2\hat{\rho}_2$. Note that the restriction $N \gg M$ is important, see, e.g., [50], and some confusion arose in the literature by not taking it into account.

$$\hat{\rho} = \sum_{\alpha=1}^{N} \lambda_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|, \quad \lambda_\alpha \geqslant 0, \quad \sum_{\alpha=1}^{N} \lambda_\alpha = 1, \qquad (34)$$

where $|\psi_\alpha\rangle$ are some normalized—but in general not orthogonal—vectors existing in the same n-dimensional Hilbert space $\mathcal{H}$,[13] and where $|\psi_\alpha\rangle\langle\psi_\alpha|$ are distinct.

The eigenresolution (19) is only a particular case of (34), and if now the ensemble $\mathcal{E}(\hat{\rho})$ is prepared by one of the ways corresponding to (34) with nonorthogonal $|\psi_\alpha\rangle$, the constituents of $\mathcal{E}(\hat{\rho})$ come from the subensembles $\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha|)\}$ and the observable $\hat{B}$ has in general no definite value for these subensembles.

We conclude with three related features of a homogeneous ensemble: (1) The ensemble cannot be thought to consist of definite subensembles; (2) a single system from such an ensemble does not by itself define a subensemble; (3) There are thus no homogeneous ensembles in classical statistical physics, since a single system is known to have definite values of all its variables.

3. Pure-state ensembles

The description of a homogeneous ensemble via pure density matrices, $\hat{\rho}^2 = \hat{\rho}$, has several special features.

First of all, it is seen from Eq. (34) that for a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$ in the right-hand side (RHS) of representation (34) only one term shows up: $|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|$.[14] Thus, pure-state ensembles cannot be prepared via mixing of other ensembles of the system $\mathcal{S}$, or, put differently, first, pure-state ensembles are irreducible; and second, this description is the maximally *complete* one possible in quantum mechanics.

The latter known thesis can be substantiated as follows. First one notes from Eqs. (32) and (33) that for a fixed $\hat{\rho}$ dispersionless observables form a linear space: if two operators are dispersionless, so is their sum, and multiplication by a number conserves the dispersionless feature.

From Eq. (33) and Appendix C one sees that if the mixed density matrix $\hat{\rho}$ has k, $1 \leqslant k \leqslant n$, nonzero eigenvalues ($n$ being the dimension of the Hilbert space $\mathcal{H}$), then the dimension of the linear space formed by the corresponding dispersionless observables is equal to

$$N_k = (n-k)^2 + 1. \qquad (35)$$

This number is maximal for $k=1$, that is, for pure density matrices. In other words, pure density matrices provide definite values for a larger set of observables than mixed density matrices.[15] For a mixed state all dispersionless observables have to be degenerate.

Though the features of irreducibility and completeness create a conceptual difference between pure and mixed density matrices, this should not be taken as an invitation to prescribe pure density matrices to a single system, reserving the mixed ones for ensembles; further reasons for this are analyzed in Refs. [34,45–48,54].[16]

4. Inhomogeneous ensembles

A mixed density matrix $\hat{\rho}$ can also describe inhomogeneous ensembles. Such an ensemble $\mathcal{E}_i$ is a collection of homogeneous subensembles $\{\mathcal{E}(\hat{\rho}_\alpha)\}_{\alpha=1}^N$ with probabilities $\{\lambda_\alpha\}_{\alpha=1}^N$, so that each single system from $\mathcal{E}_i$ is known to be taken from the ensemble $\mathcal{E}(\hat{\rho}_\alpha)$ with probability λ_α, $\alpha = 1, \ldots, N$. Obvious cases are when the subensembles $\mathcal{E}(\hat{\rho}_\alpha)$ are separated in space—as happens for the two beams of the Stern-Gerlach experiment—or in time, or by means of some other classical quantity.

Inhomogeneous ensembles are typically prepared by means of selective measurements.[17] In that case the above classical quantity is the corresponding record of the macroscopic apparatus by which this measurement was done. Below in Sec. III E we describe in detail how an initially homogeneous ensemble can be separated into subensembles by means of a measurement.

The inhomogeneous ensemble $\mathcal{E}_i$ is still described by the overall density matrix $\hat{\rho} = \Sigma_{\alpha=1}^N \lambda_\alpha \hat{\rho}_\alpha$, but in contrast to the homogeneous situation this is not the full description. The latter is provided by the list

$$\{\lambda_\alpha, \hat{\rho}_\alpha\}_{\alpha=1}^N. \qquad (36)$$

So more information is known about the inhomogeneous ensemble $\mathcal{E}_i$ than only $\hat{\rho}$. If the inhomogeneous ensemble is just a combination of homogeneous ones, this is obvious. If the inhomogeneous ensemble was prepared by means of a measurement, then the above information results from the measurement carried out and from selection of the outcomes (see more details in Sec. III E below).

[13] Normalization and belonging to $\mathcal{H}$ are necessary for $|\psi_\alpha\rangle\langle\psi_\alpha|$ to describe some ensemble of the systems $\mathcal{S}$.

[14] This can also be deduced from a more general result: any $|\psi_\alpha\rangle$ that can appear in Eq. (34) is orthogonal to the linear space formed by the eigenvectors of $\hat{\rho}$ corresponding to eigenvalue zero. Indeed, let $|0\rangle$ be one such eigenvector, then $\langle 0|\hat{\rho}|0\rangle = \Sigma_\alpha \lambda_\alpha |\langle 0|\psi_\alpha\rangle|^2 = 0$; thus $\langle 0|\psi_\alpha\rangle = 0$ for $\lambda_\alpha > 0$.

[15] For $k=n$ we get $N_k = 1$, since in this case only operators proportional to unity are dispersionless. For $n=2$ and $k=1$, $N_k = 2$: all dispersionless observables for a two-dimensional pure density matrix $|\psi\rangle\langle\psi|$ can be represented as $\alpha|\psi\rangle\langle\psi| + \beta|\psi_\perp\rangle\langle\psi_\perp|$, where $\langle\psi|\psi_\perp\rangle = 0$, and where α and β are two independent real numbers.

[16] Among the reasons we find convincing is the analysis of the quantum measurement process [33,54].

[17] These measurements need not be done on the system $\mathcal{S}$ directly; they can be indirect as well. Imagine an ensemble of two spin-1/2 particles described by pure density matrix $|\psi\rangle\langle\psi|$, where $|\psi\rangle = (1/\sqrt{2})(|+\rangle_1 \otimes |+\rangle_2 + |-\rangle_1 \otimes |-\rangle_2)$, and where $|\pm\rangle_{1,2}$ are the eigenvectors of $\hat{\sigma}_z^{(1,2)}$ with eigenvalues ± 1 for the first and second particle, respectively. One can now measure $\hat{\sigma}_z^{(1)}$, and keep both the results of these measurements and the order of their appearance (thus, one keeps a sequence of random numbers ± 1). For the subensemble of the second spin this amounts to preparation of inhomogeneous ensemble $\{1/2, |+\rangle_2\, {}_2\langle +|; 1/2, |-\rangle_2\, {}_2\langle -|\}$.

5. Prescribed ensemble fallacy

This fallacy rests on forgetting the difference between homogeneous and inhomogeneous ensembles [47,55], that is, it rests on neglecting that the overall (postmeasurement) density matrix $\hat{\rho}=\Sigma_{\alpha=1}^{N}\lambda_\alpha\hat{\rho}_\alpha$ is physically different from the one before the measurement, even though it is mathematically still the same (in a Stern-Gerlach experiment the initial beam has been split into two parts). The fact that many mathematical splittings are possible just agrees with the fact that many experiments are possible in principle. "Switching to another representation," as is often done in theoretical considerations that commit the prescribed ensemble fallacy, is by itself impossible, unless one makes a second measurement setup. In any given situation; however, once the experimental setup is determined, there is no choice; instead, the splitting is unique, physical, and contextual.

In spite of explicit warnings [3], the fallacy frequently (re)appears in applications and interpretations of quantum statistical physics. Consider, for example, the basic tool of statistical physics, the equilibrium ensemble described by the Gibbsian density matrix (21). It is typically obtained by thermalization process, that is, due to interaction with a thermal bath. One sometimes hears with respect to this ensemble that it represents the system being in states of definite energy with the corresponding probabilities p_k. This is a valid description of the ensemble only after the measurement of energy $\hat{H}$ has been done, something which is not typical in applications. Moreover, as we recalled above and below, one can choose to make a different measurement, and then the interpretation in terms of definite energies will be explicitly wrong. The reason why some applications—though starting from the above incorrect premise—do not lead to contradictions is clear: they use this premise merely for "explanation of what actually happens," while in real calculations and comparisons with experiment only the density matrix (21) is employed.

D. Fluctuations of work

Once the properties of quantum ensembles are clarified, we can proceed with the quantum definition of fluctuations of work. The most reasonable way to define this concept in the quantum situation is to proceed along the same lines as in classics, taking into account when needed the differences between quantum and classical ensembles.

It is convenient to separate the definition into the following steps.

(1) The initial ensemble $\mathcal{E}(\hat{\rho})$ is homogeneous, since it was prepared by means of a thermal bath. With help of a suitable measurement (see Sec. III E for details), one separates $\mathcal{E}(\hat{\rho})$ into irreducible, maximally complete subensembles $\{\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha|)\}_{\alpha=1}^{N}$ with probabilities $\{\lambda_\alpha\}_{\alpha=1}^{N}$, so that the resulting inhomogeneous ensemble is still described by the same density matrix $\hat{\rho}$ and thus (34) is valid.

In the quantum situation irreducible, maximally complete subensembles are described by pure density matrices $|\psi\rangle\langle\psi|$. The important point is that *these subensembles play here the same role as the single systems for the classical definition of fluctuations of work.*

Note that once it is understood that the initial ensemble $\mathcal{E}(\hat{\rho})$ is homogeneous and that measurements are anyhow needed to make it inhomogeneous, we have to admit any measurement which will produce pure-state ensembles, even those with nonorthogonal $|\psi_\alpha\rangle$'s.

Recall that the present step of preparing an inhomogeneous ensemble out of the initial homogeneous one is absent in the classical situation, simply because there are no essentially homogeneous classical ensembles (i.e., each single system can be viewed as defining a subensemble).

(2) This step almost literally repeats its classical analog. The single systems from each subensemble $\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha|)$ interacts with the work source which realizes the same thermally isolated process on each single system from each subensemble.

The evolution of the corresponding subensemble during the cyclic process between times 0 and τ is given by the von Neumann equation

$$i\hbar\frac{d}{dt}\hat{\rho}_\alpha(t)=[\hat{H}(t),\hat{\rho}_\alpha(t)],\quad \hat{\rho}_\alpha(0)=|\psi_\alpha\rangle\langle\psi_\alpha|, \tag{37}$$

$$\hat{\rho}_\alpha(\tau)=\hat{U}_\tau\hat{\rho}_\alpha(0)\hat{U}_\tau^{\dagger}. \tag{38}$$

(3) In analogy with the corresponding classical step we define the work w_α done on the subensemble α via Eq. (5), or alternatively via Eq. (31):

$$w_\alpha=\mathrm{tr}(\hat{\Omega}|\psi_\alpha\rangle\langle\psi_\alpha|)=\langle\psi_\alpha(\tau)|\hat{H}|\psi_\alpha(\tau)\rangle-\langle\psi_\alpha(0)|\hat{H}|\psi_\alpha(0)\rangle, \tag{39}$$

$$=\int_0^\tau dt\,\mathrm{tr}\left[\hat{\rho}_\alpha(t)\frac{d\hat{H}(t)}{dt}\right]. \tag{40}$$

This is the average energy decrease of the mechanical degree of freedom of the work source due its interaction with the corresponding subensemble. Thus w_α has the meaning of work by itself, but it is a quantity that had to be averaged over the subensemble. The probability of w_α is equal to λ_α, since, as seen from Eq. (34), this is the probability by which the corresponding pure subensemble enters the overall ensemble described by $\hat{\rho}$.

Thus we defined a random *c*-number quantity of work *w* with realizations w_α and probabilities λ_α:

$$w=\{w_\alpha,\lambda_\alpha\}_{\alpha=1}^{N}. \tag{41}$$

As follows from Eqs. (29) and (34) the work done on the overall ensemble is equal to the weighted average over the pure subensembles:

$$W=\sum_{\alpha=1}^{N}\lambda_\alpha w_\alpha. \tag{42}$$

Equation (42) remains true for any initial ensemble. This definition of fluctuations of work can be applied to any initial ensemble and not only to that described by the Gibbsian density matrix (21).

The fluctuations of work do depend on the pure ensembles $\{|\psi_\alpha\rangle\langle\psi_\alpha|\}_{\alpha=1}^{N}$, which are defined uniquely once the

measurement separating the overall ensemble into pure subensembles is specified. What we defined as fluctuations are thus the ones between subensembles (intersubensemble fluctuations).

If $[\hat{\rho},\hat{H}]=0$, then the ensemble described by $\hat{\rho}$ is stationary: all (one-time) averages are time independent. Now note that the stationary ensemble can be decomposed into nonstationary subensembles, since in general $[|\psi_\alpha\rangle\langle\psi_\alpha|,\hat{H}]\neq 0$. This fact implies nothing pathologic, since work is defined for any initial ensemble, not only for stationary ones. It is checked from Eq. (39) that if there is no interaction with the work sources, then $\Omega\equiv 0$, and all possible realizations of work are zero.

(4) Note that for macroscopic systems it is not realistic to have available measurements producing pure-state subensembles, since the directly available measurements are only those of macroscopic quantities which are typically degenerate. In this case we may need to apply a coarse-grained definition of fluctuations of work, where the initial mixed ensemble is separated into mixed subensembles described by density matrices $\hat{\sigma}_\gamma$ $(\hat{\sigma}_\gamma^2\neq\hat{\sigma}_\gamma)$

$$\hat{\rho}=\sum_\gamma \nu_\gamma\hat{\sigma}_\gamma,\quad \nu_\gamma\geqslant 0,\quad \sum_\gamma \nu_\gamma=1. \tag{43}$$

The definition then proceeds as above, changing $|\psi_\alpha\rangle\langle\psi_\alpha|\rightarrow\hat{\sigma}_\gamma$ in Eq. (39).

This is a coarse-grained definition, since the realizations of work $\mathrm{tr}(\hat{\Omega}\hat{\sigma}_\gamma)$ can be reduced to more fundamental ones, i.e., each of them can be presented as a convex sum of $\mathrm{tr}(|\psi_\alpha\rangle\langle\psi_\alpha|\hat{\Omega})$. As a consequence fluctuations of work—as quantified, e.g., by dispersion of work defined and discussed in Sec. IV—are maximal for pure-state decompositions (more details on this are found in Sec. IV A).

E. Separation of a homogeneous ensemble into pure subensembles by filtering outcomes of a positive operator valued measurement

1. Positive operator valued measurements

It is now our purpose to discuss how precisely one separates (with the help of measurements and subsequent filtering) an initial homogeneous ensemble $\mathcal{E}(\hat{\rho})$ into a mixed ensemble consisting of pure (necessarily homogeneous) subensembles.

The most general type of quantum measurement that allows to produce pure postmeasurement states corresponds to a positive operator valued measure (POVM) [34,47] defined via N operators $\hat{G}_\alpha$—not necessarily orthogonal—existing in the n-dimensional Hilbert space $\mathcal{H}$ and satisfying the completeness relation

$$\sum_{\alpha=1}^{N}\hat{G}_\alpha^\dagger\hat{G}_\alpha=\hat{1}. \tag{44}$$

The standard projective measurements of an observable $\hat{A}$ living in the n-dimensional Hilbert space $\mathcal{H}$ and having nondegenerate spectrum $\{a_\alpha\}_{\alpha=1}^n$ are included in Eq. (44), since now $N=n$ and $\{\hat{G}_\alpha\}_{\alpha=1}^n=\{|a_\alpha\rangle\langle a_\alpha|\}_{\alpha=1}^n$, the latter being the set of orthonormal eigenvectors of $\hat{A}$. If the spectrum of $\hat{A}$ happens to have degeneracies, so that each eigenvalue a_α has multiplicity n_α, then $\hat{G}_\alpha$ is the n_α-dimensional projector on the subspace formed by n_α linearly independent eigenvectors of $\hat{A}$ which correspond to the eigenvalue a_α. Here $N\leqslant n$ is equal to the number of distinct eigenvalues of $\hat{A}$.

If the measurement described by Eq. (44) is done on the ensemble described by a density matrix $\hat{\rho}$, then the result α is found with probability

$$\lambda_\alpha=\mathrm{tr}(\hat{G}_\alpha^\dagger\hat{G}_\alpha\hat{\rho})=\mathrm{tr}(\hat{G}_\alpha\hat{\rho}\hat{G}_\alpha^\dagger), \tag{45}$$

where $\lambda_\alpha\geqslant 0$ and $\Sigma_{\alpha=1}^N\lambda_\alpha=1$, due to Eq. (44). After selecting results of the measurements referring to the outcome α one has the (sub)ensemble of systems described by a density matrix

$$\hat{\rho}'_\alpha=\frac{\hat{G}_\alpha\hat{\rho}\hat{G}_\alpha^\dagger}{\mathrm{tr}(\hat{G}_\alpha^\dagger\hat{G}_\alpha\hat{\rho})}. \tag{46}$$

This subensemble occurs with probability λ_α as given by Eq. (45), simply because this is the probability of the outcome α. The overall postmeasurement inhomogeneous ensemble thus *consists of N* subensembles each of which has a density matrix (46) and probability (45). The density matrix of the overall postmeasurement ensemble is[18]

$$\hat{\rho}'=\sum_{\alpha=1}^{N}\lambda_\alpha\hat{\rho}'_\alpha. \tag{47}$$

POVMs are closely related to more usual projective measurements [34,47,56,57]. The detailed outline of this connection is given in Appendix D. Here we recall the main items.

(1) Assume that the system $\mathcal{S}$ is coupled to another system $\mathcal{G}$ initially in a pure state, and then some (Hermitian) observable pertaining to $\mathcal{G}$ is measured. From the viewpoint of the system $\mathcal{S}$ this then amounts to some POVM.

(2) Any given POVM can be realized in the above way, upon the proper choice of the initial state of $\mathcal{G}$, the interaction Hamiltonian, and the Hermitian observable pertaining to $\mathcal{G}$.

Note from Eq. (46) that provided the initial density matrix $\hat{\rho}$ is pure, $\hat{\rho}=|\psi\rangle\langle\psi|$, the postmeasurement ensembles are described by pure density matrices as well: $\hat{\rho}'_\alpha=\hat{G}_\alpha|\psi\rangle\langle\psi|\hat{G}_\alpha^\dagger$. In this sense a POVM measurement does not introduce noise into the postmeasurement ensembles. More general, noisy measurements are discussed in Appendix D.

2. Separation of mixed quantum ensemble

Applying a POVM measurement, one now wishes to separate the mixed quantum ensemble described by the den-

[18]Let us stress again that a specific splitting has occurred by choosing the measurement: In the ideal Stern-Gerlach situation, by choosing the magnets in the z direction, the original beam is split in the z direction with z components of the spins according to the beams.

sity matrix $\hat{\rho}$ into pure subensembles. The density matrix $\hat{\rho}'$ of the overall post-measurement ensembles should then coincide with $\hat{\rho}$ given in (19) or (21), while $\hat{\rho}'_\alpha$ appearing in (46) should be pure:

$$\hat{\rho}'_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|. \tag{48}$$

Then the density matrix (21) is decomposed as in (34). It will prove useful to write (34) in an equivalent way

$$\hat{\rho} = \sum_{\alpha=1}^{N} \widetilde{|\psi_\alpha\rangle}\widetilde{\langle\psi_\alpha|}, \quad \widetilde{|\psi_\alpha\rangle} \equiv \sqrt{\lambda_\alpha}|\psi_\alpha\rangle, \tag{49}$$

since it will allow us to focus on $\widetilde{|\psi_\alpha\rangle}$, remembering that the probabilities λ_α can be recovered via $\lambda_\alpha = \widetilde{\langle\psi_\alpha|}\widetilde{|\psi_\alpha\rangle}$.

Let us first see which $\{\lambda_\alpha\}_{\alpha=1}^N$ and $\{|\psi_\alpha\rangle\langle\psi_\alpha|\}_{\alpha=1}^N$ can enter in Eq. (34), and then we shall discuss which specific measurements should be done to achieve the separation. According to the classification theorem [58–61], if one has

$$\widetilde{|\psi_\alpha\rangle} = \sum_{k=1}^{n} M_{\alpha k}\sqrt{p_k}|p_k\rangle, \tag{50}$$

$$\lambda_\alpha = \widetilde{\langle\psi_\alpha|}\widetilde{|\psi_\alpha\rangle} = \sum_{k=1}^{n} |M_{\alpha k}|^2 p_k, \tag{51}$$

where $\{p_k\}_{k=1}^n$ and $\{|p_k\rangle\}_{k=1}^n$ are the eigenvalues and eigenfunctions of the density matrix $\hat{\rho}$, and where $M_{\alpha k}$ are complex numbers satisfying

$$\sum_{\alpha=1}^{N} M_{\alpha k}M^*_{\alpha j} = \delta_{ij}, \quad k,j = 1, \ldots, n; \tag{52}$$

then Eq. (49) becomes $\hat{\rho} = \Sigma_{k=1}^n p_k|p_k\rangle\langle p_k|$, as it should.[19] The converse appears to be true as well: any decomposition (49) admits a representation (50) with some complex numbers $M_{\alpha k}$ satisfying Eq. (52).[20]

As seen from Eq. (52), the very possibility of writing Eq. (49) implies

$$N \geqslant n, \tag{53}$$

since $M_{\alpha k}$ can be viewed as n different N-component orthonormal vectors. The rectangular matrix $\{M_{\alpha k}\}_{\alpha=1}^N, {}_{k=1}^n$ can be completed to a unitary $N \times N$ matrix by adding suitable elements.

It is now straightforward to see which POVM achieves the decomposition (49). Take, for example,

$$\hat{G}_\alpha = \frac{\widetilde{|\psi_\alpha\rangle}\widetilde{\langle\psi_\alpha|}\hat{\rho}^{-1/2}}{\sqrt{\widetilde{\langle\psi_\alpha|}\widetilde{|\psi_\alpha\rangle}}} = \sqrt{\lambda_\alpha}|\,\psi_\alpha\rangle\langle\psi_\alpha|\hat{\rho}^{-1/2}, \tag{54}$$

where $|\psi_\alpha\rangle$ is defined in Eq. (49). Note that the converse appears to be true as well. For a given POVM (44) with

$$\hat{G}^\dagger_\alpha\hat{G}_\alpha = |\pi_\alpha\rangle\langle\pi_\alpha|, \tag{55}$$

and $|\pi_\alpha\rangle$ satisfying

$$\hat{1} = \sum_{\alpha=1}^{N} |\pi_\alpha\rangle\langle\pi_\alpha|, \tag{56}$$

which need be neither orthogonal, nor normalized,[21] one can construct a representation (49) and (34) of $\hat{\rho}$ as

$$\hat{\rho} = \sum_{\alpha=1}^{N} \hat{\rho}^{1/2}|\pi_\alpha\rangle\langle\pi_\alpha|\hat{\rho}^{1/2}. \tag{57}$$

Thus, we have seen how all possible decompositions of a mixed ensemble into pure subensenbles can be constructed via suitable measurements. We stress that the decomposition into a specific set of subensembles is related to a physical measurement, rather than to a mathematical choice (prescribed ensemble fallacy).

3. Preparation versus measurements

To avoid possible confusions we recall once again that the above separation procedure corresponds to *preparation* of the inhomogeneous ensemble $\{\lambda_\alpha, \mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha|)\}_{\alpha=1}^N$ with $\hat{\rho} = \Sigma_\alpha\lambda_\alpha|\psi_\alpha\rangle\langle\psi_\alpha|$, starting from the initial homogeneous ensemble $\mathcal{E}(\hat{\rho})$. Though this preparation was based on a suitable measurement process, we were not interested in some aspects usually associated with it. For example, we did not keep track of the pointer variable of the measuring apparatus, which obviously should be the main goal of any measurement process studied for its own purposes [54]. We were more interested by the influence of the measurement process on the final state of the system $\mathcal{S}$, which is the basic characteristic feature of the preparation process in quantum mechanics [34].

As we mentioned above, POVMs are related to more usual projective measurements; see [34,47,56,57] and Appendix D for more details. Therefore, the main difference between POVMs and projective measurements lies in the aspect of the postmeasurement state preparation.

[19] Note that any vector $\widetilde{|\psi_\alpha\rangle}$ having $\widetilde{\langle\psi_\alpha|}\widetilde{|\psi_\alpha\rangle} < 1$ and existing in the Hilbert space formed by the eigenvectors of $\hat{\rho}$ corresponding to its nonzero eigenvalues, can appear in at least one separation (49) of $\hat{\rho}$. This follows from (50).

[20] To prove this part of the statement, recall footnote 14, expand $\widetilde{|\psi_\alpha\rangle}$ over the eigenbase $|p_k\rangle$ of $\hat{\rho}$, $\widetilde{|\psi_\alpha\rangle} = \Sigma_{k=1}^n\langle p_k|\widetilde{\psi_\alpha}\rangle|p_k\rangle$, substitute this into Eq. (49), and then deduce Eq. (52) using the orthonormality and completeness of the above base in the Hilbert space $\mathcal{H}$: $\Sigma_{\alpha=1}^N\langle p_k|\widetilde{\psi_\alpha}\rangle\widetilde{\langle\psi_\alpha}|p_l\rangle = \langle p_k|\hat{\rho}|p_l\rangle = \delta_{kl}p_k$. Thus, any decomposition (49) and (34) can be constructed via Eq. (50) and $M_{\alpha k} = \langle p_k|\widetilde{\psi_\alpha}\rangle/\sqrt{p_k}$ satisfying Eq. (50). If some eigenvalues of $\hat{\rho}$ are equal to zero, then the above construction should be restricted to eigenvectors of $\hat{\rho}$ corresponding to its nonzero eigenvalues.

[21] If one assumes in Eq. (56) that $|\pi_\alpha\rangle$ are normalized, $\langle\pi_\alpha|\pi_\alpha\rangle = 1$, then this leads to orthogonality: $\langle\pi_\beta|\pi_\alpha\rangle = \delta_{\alpha\beta}$. Indeed, denoting $\hat{\Pi}_\alpha = |\pi_\alpha\rangle\langle\pi_\alpha|$, one gets $\Sigma_{\alpha\neq\beta}^N(\hat{\Pi}_\alpha\hat{\Pi}_\beta)^\dagger(\hat{\Pi}_\alpha\hat{\Pi}_\beta) = \Sigma_{\alpha\neq\beta}^N\hat{\Pi}_\beta\hat{\Pi}_\alpha\hat{\Pi}_\alpha\hat{\Pi}_\beta = \hat{\Pi}_\beta(1-\hat{\Pi}_\beta)\hat{\Pi}_\beta = 0$. Since $(\hat{\Pi}_\alpha\hat{\Pi}_\beta)^\dagger(\hat{\Pi}_\alpha\hat{\Pi}_\beta)$ is non-negative by construction, one concludes that $\hat{\Pi}_\alpha\hat{\Pi}_\beta = 0$ for $\alpha \neq \beta$.

4. An example

For a realistic example illustrating POVM measurements versus projective ones, let us consider a harmonic oscillator—within the optical realization this is a single mode of an electromagnetic field—with Hamiltonian

$$\hat{H} = \hbar\omega\hat{a}^\dagger\hat{a}, \quad [\hat{a},\hat{a}^\dagger] = 1, \tag{58}$$

where $\hat{a}$ and $\hat{a}^\dagger$ are, respectively, annihilation and creation operators, and where ω is the frequency. The corresponding energy levels are $\varepsilon_k = \hbar\omega k$, $k = 0,1,2,\ldots$. Consider for this system an initial homogeneous Gibbsian ensemble $\mathcal{E}(\hat{\rho})$, described by the density matrix (21), $\rho \propto e^{-\beta\hat{H}}$, at temperature T.

We shall outline two realistic ways to separate the initial homogeneous ensemble $\mathcal{E}(\hat{\rho})$ into pure subensembles.

The first one amounts to measuring the energy $\hat{H}$ of the oscillator. Within the optical realization this corresponds to measuring the number of photons $\hat{a}^\dagger\hat{a}$. The postmeasurement ensembles are described by pure density matrices $\{|\varepsilon_k\rangle \times\langle\varepsilon_k|\}_{k=0}^{\infty}$, and the corresponding separation of $\hat{\rho}$ is given by the eigenrepresentation (19).

The second one is equally well known, especially in quantum optics [34,48,56,57]. It consists of making a heterodyne measurement described by POVM (44) with

$$\hat{G}_\alpha = \frac{1}{\sqrt{\pi}}|\alpha\rangle\langle\alpha|, \qquad \int d\,\mathrm{Re}(\alpha)d\,\mathrm{Im}(\alpha)\hat{G}_\alpha^\dagger\hat{G}_\alpha = 1, \tag{59}$$

where $|\alpha\rangle$ is the coherent state of the harmonic oscillator, and where $\alpha = \mathrm{Re}(\alpha) + i\,\mathrm{Im}(\alpha)$ is a complex index. The last relation in (59) is the continuous analog of (44). Recall that the coherent states are not orthogonal [56], $|\langle\alpha|\beta\rangle|^2 = \exp(-|\alpha-\beta|^2)$, that is, the heterodyne measurement is a nontrivial POVM. It is standardly realized by coupling the original mode (oscillator) with a probe mode, and then measuring a Hermitian observable of the combined system; see [57] for relevant details. The corresponding decomposition of the equilibrium density matrix $\hat{\rho}$ now reads [56] (the so-called P representation of the thermal density matrix)

$$\hat{\rho} = \frac{e^{-\beta\hat{H}}}{Z} = \int \frac{d\,\mathrm{Re}(\alpha)d\,\mathrm{Im}(\alpha)}{\pi\overline{n}} e^{-|\alpha|^2/\overline{n}}|\alpha\rangle\langle\alpha|, \tag{60}$$

where $\overline{n} = (e^{\hbar\omega\beta}-1)^{-1}$ is the average number of photons.

As for another example of physically realizable POVM, note that for photon polarization—one of the simplest cases of a two-dimensional Hilbert space—*all* POVM measurements can be realized via linear optics elements (e.g., beam splitters) [62].

F. Fluctuations of work for a system coupled to the environment

The above definition of fluctuations of work generalizes straightforwardly to situations when in the relevant time interval $(0,\tau)$ the system $\mathcal{S}$ interacts with an arbitrary environment $\mathcal{B}$ (e.g., thermal baths). While it is obvious that the above definition can be applied directly to the whole system $\mathcal{S}+\mathcal{B}$, our objective is to show that it can be applied in such a way that realizations of work can be deduced by following $\mathcal{S}$ only. This is one of the basic features of the average work, and this feature is relevant, since by the very definition of the environment, its state is (at least partially) out of observation and control.

The total Hamiltonian of $\mathcal{S}+\mathcal{B}$ now reads

$$\hat{H}_{\mathcal{S}+\mathcal{B}}(t) = \hat{H}(t) + \hat{H}_{\mathcal{B}+\mathcal{I}}, \tag{61}$$

where $\hat{H}(t)$ is the Hamiltonian of $\mathcal{S}$ describing its interactions with the external fields, and where $\hat{H}_{\mathcal{B}+\mathcal{I}}$ stands for the Hamiltonian of $\mathcal{B}$ and its interaction with $\mathcal{S}$. Note that the coupling between $\mathcal{S}$ and $\mathcal{B}$ need not be weak. The situation where $\mathcal{B}$ is a thermal bath, and where the action of the work source is very slow, corresponds to the (usual) quasistatic thermodynamical processes.

Assume that $\mathcal{S}+\mathcal{B}$ is initially in some (in general correlated) state $\hat{\rho}_{\mathcal{S}+\mathcal{B}}$. One now applies to the system $\mathcal{S}$ a POVM measurement, for concreteness the one given by Eq. (54). The corresponding operators $\hat{G}_\alpha$ generalize as

$$\hat{G}_\alpha = \sqrt{\lambda_\alpha}|\psi_\alpha\rangle\langle\psi_\alpha|\hat{\rho}_{\mathcal{S}}^{-1/2} \otimes \hat{1}_{\mathcal{B}}, \tag{62}$$

where $\hat{\rho}_{\mathcal{S}} = \mathrm{tr}_{\mathcal{B}}\,\hat{\rho}_{\mathcal{S}+\mathcal{B}}$, is the reduced density matrix of $\mathcal{S}$, $|\psi_\alpha\rangle\langle\psi_\alpha|$ are some pure density matrices existing in the Hilbert space of $\mathcal{S}$, $\{\lambda_\alpha\}_{\alpha=1}^N$ are the probabilities of various outcomes of the measurement, and $\hat{1}_{\mathcal{B}}$ is the unity operator acting in the space of the environment.

Once the measurement is carried out and the outcomes are filtered, the postmeasurement states of $\mathcal{S}+\mathcal{B}$ read

$$|\psi_\alpha\rangle\langle\psi_\alpha| \otimes \hat{\rho}_{\alpha,\mathcal{B}}, \quad \hat{\rho}_{\alpha,\mathcal{B}} = \frac{\hat{\rho}_{\mathcal{S}}^{-1/2}\hat{\rho}_{\mathcal{S}+\mathcal{B}}\hat{\rho}_{\mathcal{S}}^{-1/2}}{\lambda_\alpha}, \tag{63}$$

with $\hat{\rho}_{\alpha,\mathcal{B}}$ being the postmeasurement state of $\mathcal{B}$. In general it differs from the premeasurement state $\hat{\rho}_{\mathcal{B}} = \mathrm{tr}_{\mathcal{S}}\,\hat{\rho}_{\mathcal{S}+\mathcal{B}}$, due to initial correlations between $\mathcal{S}$ and $\mathcal{B}$.

Now the work source starts to act on each single system from the subensemble $\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha| \otimes \hat{\rho}_{\alpha,\mathcal{B}})$. Though the source is acting only via $\mathcal{S}$, as seen from Eq. (61), it influences the environment, since the latter couples to $\mathcal{S}$. For the same reason, the state of $\mathcal{S}+\mathcal{B}$ will (in general) not remain factorized as in Eq. (63). Still the work w_α done on the subensemble $\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha| \otimes \hat{\rho}_{\alpha,\mathcal{B}})$ can be expressed via quantities referring to $\mathcal{S}$ only: as follows from Eq. (31),

$$w_\alpha = \int_0^\tau dt\,\mathrm{tr}\left[\hat{\rho}_{\alpha,\mathcal{S}}(t)\frac{d\hat{H}(t)}{dt}\right], \tag{64}$$

$$\hat{\rho}_{\alpha,\mathcal{S}}(t) = \mathrm{tr}_{\mathcal{B}}\Big[\overleftarrow{e}^{-(i/\hbar)\int_0^t ds\,\hat{H}_{\mathcal{S}+\mathcal{B}}(s)}(|\psi_\alpha\rangle\langle\psi_\alpha| \otimes \hat{\rho}_{\alpha,\mathcal{B}})\overrightarrow{e}^{(i/\hbar)\int_0^t ds\,\hat{H}_{\mathcal{S}+\mathcal{B}}(s)}\Big], \tag{65}$$

where $\hat{\rho}_{\alpha,\mathcal{S}}(t)$ is the state of $\mathcal{S}$ at moment t, provided $\mathcal{S}+\mathcal{B}$ starts from the initial state (63). We recall that the (anti)ordered exponents are defined by Eq. (28).

Thus the work—done on the whole system $\mathcal{S}+\mathcal{B}$ through the work source coupled to $\mathcal{S}$ only—was defined as a random quantity with realizations $\{w_\alpha\}_{\alpha=1}^N$ and probabilities $\{\lambda_\alpha\}_{\alpha=1}^N$.

G. On the experimental demonstration of fluctuations of work

The above approach to fluctuations of work was based on the standard knowledge of quantum statistical mechanics and on a number of reasonable consistency requirements the sought concept was supposed to satisfy. Later on we realized that there were already several experiments in quantum optics that quite closely correspond to the present scheme of studying fluctuations of work (not unexpectedly, the experimentalists did not think in terms of fluctuations of work).

Let us start with one example of such experiments [52]. It is especially interesting, since fluctuations of work are not extracted as a by-product, but, on the contrary, the very experiment is realized by their detection.

The experiment was carried out with a single $^{199}Hg^{+}$ ion confined in a linear rf trap. The same experiment was repeated with many (from 5 to 400) independent ions in the trap. $^{199}Hg^{+}$ has spin $\frac{1}{2}$, and its ground state energy level is split by hyperfine interactions into two levels $|\varepsilon_0\rangle$ and $|\varepsilon_1\rangle$, $\varepsilon_0<\varepsilon_1$, having the total angular momenta 0 and 1, respectively. The difference $\varepsilon_1-\varepsilon_0$ is of the order of the radio frequency photon energy. The levels are stable: both the spontaneous decay and the magnetic dipole decay $|\varepsilon_1\rangle\rightarrow|\varepsilon_0\rangle$ can be neglected. There is also the third excited level $|\varepsilon_2\rangle$ which is *highly* unstable, and decays to $|\varepsilon_1\rangle$ by spontaneous emission of one optical photon.

A laser beam is applied in resonance to the transition $|\varepsilon_1\rangle\Leftrightarrow|\varepsilon_2\rangle$. The transitions $|\varepsilon_2\rangle\Leftrightarrow|\varepsilon_0\rangle$ and $|\varepsilon_1\rangle\Leftrightarrow|\varepsilon_0\rangle$ can be neglected: no dynamics is generated if the ion is initially in the state $|\varepsilon_0\rangle$ (i.e., it belongs to the ensemble described by the density matrix $|\varepsilon_0\rangle\langle\varepsilon_0|$). In contrast, if the ion is initially in the state $|\varepsilon_1\rangle$, the dynamics is that of two energy levels $|\varepsilon_1\rangle$ and $|\varepsilon_2\rangle$ driven by the classical laser field (work source) and interacting with vacuum electromagnetic modes which induce spontaneous transition $|\varepsilon_2\rangle\rightarrow|\varepsilon_1\rangle$ (weakly coupled zero-temperature thermal bath).

If after some time the laser field is turned off (cyclic process), the ion is back to the state $|\varepsilon_1\rangle$, and the work has gone from the source of the laser field to the bath. It is proportional to the total intensity of the spontaneously generated radiation, and can be observed via measuring this quantity with help of a photodetector. This work can also be calculated via Eq. (64), where $\hat{\rho}_{1,\mathcal{S}}(0)=|\varepsilon_1\rangle\langle\varepsilon_1|$, and where τ is the duration of the laser field action. It is important to note that during the driving by the laser the ion is repeatedly prepared in the state $|\varepsilon_1\rangle$, due to the spontaneous transition $|\varepsilon_2\rangle\rightarrow|\varepsilon_1\rangle$. If the laser field acts sufficiently long, the *single* ion simulates the behavior of the ensemble $\mathcal{E}(|\varepsilon_1\rangle\langle\varepsilon_1|)$.

Assume that initially the single ion belongs to the ensemble $\mathcal{E}(\hat{\rho})$,

$$\hat{\rho}=p_0|\varepsilon_0\rangle\langle\varepsilon_0|+p_1|\varepsilon_1\rangle\langle\varepsilon_1|,\quad p_0+p_1=1. \tag{66}$$

Such states are stable and can be prepared by optical pumping. After switching the laser field on and off, one of two values for the intensity of the spontaneously generated radiation—thus one of two values $w_1>0$ and $w_0=0$ of work—is observed: a nonzero (zero) value corresponds to the energy level ε_1 (ε_0), and is realized with probability p_1 (p_0) upon repeating the experiment with other single systems from the same ensemble $\mathcal{E}(\hat{\rho})$. The measurement of the energy

$$\hat{H}=\varepsilon_0|\varepsilon_0\rangle\langle\varepsilon_0|+\varepsilon_1|\varepsilon_1\rangle\langle\varepsilon_1| \tag{67}$$

is thus deduced via observing fluctuations of work [52]. Its fluctuating values are $w_0=0$ with probability $\lambda_0=p_0$ and $w_1>0$ with probability $\lambda_1=p_1$. Thus the average work is $W=p_1w_1$.[22] Later we shall pay attention to the dispersion of work. Here it equals $\delta w^2=p_1(1-p_1)w_1^2$. Notice that there is no 'active' subenemble with $w_\alpha<0$, as would have to occur classically, see section II E. As we explain further on, such indeed needs not occur in the general quantum situation.

The second experiment is worth discussing, since it realized the first master device (the ammonia-beam maser) ever to be operated [78]. The ammonia molecule NH_3 has two closely located low-lying energy levels $|\varepsilon_0\rangle$ and $|\varepsilon_1\rangle$—with the gap $\varepsilon_1-\varepsilon_0$ being 23 870 MHz—which form an effective two-level system.

Initially one starts with a high-temperature beam (ensemble) of ammonia molecules, such that the above two levels are almost equally populated. The members of this ensemble are then let to interact with an analog of the Stern-Gerlach measuring device: An energy measurement is realized with help of a strong dc quadrupole electric field. The resulting subensembles described by the density matrices $|\varepsilon_0\rangle\langle\varepsilon_0|$ and $|\varepsilon_1\rangle\langle\varepsilon_1|$ are separated in space: the first subensemble is lost, since it was not interesting for the purposes of the experiment, while the second ensemble is directed to high-Q microwave cavity resonant at $\varepsilon_1-\varepsilon_0=23\ 870$ MHz. A resonant emf field of the cavity realizes a π-pulse, such that the state of the subensemble at the exit of the cavity is $|\varepsilon_0\rangle\langle\varepsilon_0|$ and the resulting energy $\varepsilon_1-\varepsilon_0=-w_1$ per molecule has been transferred to the emf field thereby amplifying it. This corresponds to one realization w_1 of the random quantity work. Since no work was exchanged with the first subensemble, the corresponding realization is $w_2=0$.

H. Discussion

There are several questions on the physical meaning of the proposed definition of fluctuations of work that we decided to discuss separately. Some of these questions were asked by ourselves, while others came from our colleagues.

Question 1. Among all decompositions (34) of the Gibbsian density matrix $\hat{\rho}$, there is a unique one (up to accidental degeneracies of the spectrum) given by the eigenvectors of $\hat{\rho}$ and realized via measurement of the Hamiltonian $\hat{H}$. Then the energy has a definite value on each subensemble. Should not one therefore restrict the definition of fluctuations of work to this separation only?

Answer 1. There are at least two reasons why the answer is no. First, even if the energy has a definite value initially, it

[22] This work can be calculated theoretically with the help of quantum optical master equations; see, e.g., [48] for solution of similar problems.

will in general not have any definite value at the final moment, since an eigenvector of the initial Hamiltonian may evolve into a superposition of eigenvectors. Second, more general separations are anyhow necessary to define fluctuations of work for an arbitrary ensemble, which cannot be decomposed into subensembles with each of them having a definite value of energy.

Question 2. Is the orthogonal separation not special by the fact that various ensembles are described by orthogonal pure density matrices, and can thus be discriminated unambiguously?

Answer 2. By definition any POVM is connected with an unambiguous descrimination of its different outcomes. This can be additionally clarified by looking at the example of the projective realization of a POVM presented in Appendix D, where various subensembles constructed after the measurement are seen to be described by orthogonal wave functions in the composite Hilbert space $\mathcal{H}\otimes\mathcal{H}'$. The above question mixes the present situation with a different one, where one is given a single system coming from one of two ensembles having nonorthogonal density matrices, and is requested to determine by means of a measurement from which ensemble it is coming. Then, indeed, no measurements can ensure unambiguous discrimination [47].

Question 3. Can an experimentalist come up with a concrete realization for a nontrivial POVM measurement? Measurements are something you have to do in a real life, not just mathematically.

Answer 3. As we discussed around Eq. (59), in quantum optics at least one nontrivial POVM is routinely realized and employed. More generally, many *indirect* projective measurement corresponds to a POVM, as far as the postmeasurement preparation of the target system is concerned (the aspect we are mainly interested in); see [34,47,48,56,57] and Appendix D for more details.

Question 4. The authors prescribe the viewpoint that even pure density matrices (wave functions) describe an ensemble of quantum systems and not a single system, as some people like to think. How will the proposed definition change, if one wished to insist on the latter interpretation of quantum mechanics?

Answer 4. The necessity of prescribing even the pure density matrices as ensembles of quantum systems was stressed in [34,45–47]. In particular, it is needed for the consistent solution of the quantum measurement problem [34,54]. But it is also known that with respect to certain aspects of quantum theory the prescription of pure density matrices to a single system is relatively harmless.[23] We do not have space to discuss in detail what are those aspects and what precisely is meant by "relatively harmless." We may mention that the definition of fluctuations of work remains then basically unchanged, but even becomes conceptually closer to its classical analog, since now in defining fluctuations of work one assumes operations with single systems in both quantum and classical situations.

[23]The price to pay is that the relevant information can anyhow be obtained only by doing measurements on an ensemble of identically prepared systems. Ensembles enter anyhow.

Question 5. Is the above ensemble interpretation consistent with the very idea of thermodynamical fluctuations? When we talk of fluctuations in thermodynamics we think of fluctuations of temperature or pressure of the single object in front of us—changes in time.

Answer 5. We mentioned already (in particular, when discussing experimental realizations) that the ensemble (or statistical) interpretation is capable of describing experiments with single systems, which are viewed as members of an ensemble. A single system is even capable of simulating the behavior of the whole ensemble provided it is reprepared in the same initial state in the course of this evolution.

As for the second question, let us imagine we are interested by fluctuations of energy of a single system in front of us. Upon measuring the energy of this system we find a definite result, but the subsequent measurements of energy will record the very same value. If we are interested in the energy statistics of the system in the original state, we either have to have an ensemble of identically prepared systems, or we have to reprepare the system after each measurement. In this sense the use of ensembles seems to be inevitable.

Question 6. How natural is it that the fluctuations of work are not uniquely defined, since the very separation of a mixed ensemble into pure subensembles is not unique?

Answer 6. To repeat: a homogeneous quantum ensemble described by a mixed density matrix $\hat{\rho}$ *does not consist of pure subensembles*. One needs some measurements to achieve its separation—that is, to gain some knowledge on single systems—and this is a specific physical process performed on the ensemble. Fluctuations of work are contextual, since they depend on the type of measurement employed.

Question 7. The presented definition of fluctuations of work refers to *intersubensemble* fluctuations, i.e., to a random quantity which changes from one subensemble to another. Should not a reasonable definition of such fluctuations refer to a random quantity which changes from one *single* system to another?

Answer 7. Let us start with a general remark. In statistical physics there are two types of fluctuating quantities [3]. Fluctuations of quantities having a direct mechanical meaning, e.g., energy, are defined straightforwardly. These fluctuations are indeed something which (in the quantum case) changes from one measurement done on a single system to another measurement done on another single system from the same ensemble. In contrast, the definition of fluctuations of quantities such as entropy and temperature is far less trivial. Indeed, the very notion "temperature of a single system," so natural in the everyday life, is based on neglecting fluctuations, which allows us to identify an ensemble with its single member. Strictly speaking, both temperature and entropy characterize the ensemble and not a single system. Consistent definitions proposed in the theoretical literature [3]—and for fluctuations of temperature confirmed by experiment [63]—employ a finite ensemble of systems such that the standard thermodynamic relations can still be

applied.[24] In this way, the fluctuating temperature (or entropy) has the same standard meaning.

Continuing our answer, we recall that it is the necessity to keep the physical meaning of work that led us to a definition of its fluctuations which has an intersubensemble character. It remains to stress that within the standard quantum theory we do not know how to define fluctuations of work inside of an irreducible subensemble. There were in the literature some attempts in this direction, which are described in Sec. VI. However, they do not satisfy the natural conditions on fluctuations of work, as outlined in the Introduction (arbitrary initial state; proper physical meaning). In particular, this concerns the approach based on the operator of work.[25]

IV. DISPERSION OF WORK

In Eq. (39) we defined the work $w_\alpha = \mathrm{tr}(\hat{\Omega}|\psi_\alpha\rangle\langle\psi_\alpha|)$ done on the subensemble $\mathcal{E}(|\psi_\alpha\rangle\langle\psi_\alpha|)$. Here we study how the realizations w_α of the random quantity work are spread around their mean $W=\Sigma_{\alpha=1}^{N}\lambda_\alpha w_\alpha$. The most direct quantity that characterizes this spreading is the (intersubensemble) dispersion

$$\delta w^2 = \sum_{\alpha=1}^{N} \lambda_\alpha [\langle\psi_\alpha|\hat{\Omega}|\psi_\alpha\rangle - \mathrm{tr}(\hat{\Omega}\hat{\rho})]^2 \tag{68}$$

$$= \sum_{\alpha=1}^{N} \lambda_\alpha (w_\alpha - W)^2 = \sum_{\alpha=1}^{N} \lambda_\alpha w_\alpha^2 - W^2. \tag{69}$$

In contrast to W, this quantity depends explicitly on the subensembles used to define w_α in (39). So it depends explicitly on the physical process that separated the initial ensemble into subensembles (contextuality).

It is useful to determine the maximal $\delta w^2_{\max}$ and the minimal $\delta w^2_{\min}$ values of δw^2 over all possible decompositions $\{|\psi_\alpha\rangle\langle\psi_\alpha|,\lambda_\alpha\}_{\alpha=1}^{N}$ corresponding to the fixed density matrix $\hat{\rho}=\Sigma_{\alpha=1}^{N}\lambda_\alpha|\psi_\alpha\rangle\langle\psi_\alpha|$. According to Eq. (57) these extremizations can be carried out over all possible decompositions of unity in our n-dimensional Hilbert space,

$$\sum_{\alpha=1}^{N} \hat{\Pi}_\alpha = \hat{1}, \quad \hat{\Pi}_\alpha = |\pi_\alpha\rangle\langle\pi_\alpha|, \tag{70}$$

where $\{|\pi_\alpha\rangle\}_{\alpha=1}^{N}$ have in general to be neither normalized nor orthogonal. The practical realization of the above extrema would require rather specific measurement setups.

Note that dispersions similar to (68), with Ω corresponding to some other relevant observable, where introduced and studied in quantum optics [66], where separation of an ensemble by means of (continuous) measurements are well known and were studied both experimentally and theoretically; see [48] for a review. The results we present below on the minimal and maximal values of the dispersion δw^2 do not depend on the details of Ω and can thus be useful in general.

A. Maximal dispersion of work

The maximization of δw^2 over all possible separations (34) and (70) for given $\hat{\rho}$ and $\hat{\Omega}$ is carried out in Appendix G. The result reads

$$\delta w^2_{\max} = \sum_{i,k=1}^{n} \frac{2p_i p_k}{p_i + p_k} |\langle p_k|\hat{\Omega}|p_i\rangle|^2 - W^2 \tag{71}$$

$$= 2\int_0^\infty ds\, \mathrm{tr}[(\hat{\Omega}\hat{\rho}e^{-s\hat{\rho}})^2] - W^2. \tag{72}$$

This maximum is reached on the set $\{|\pi_\alpha\rangle\}_{\alpha=1}^{n}$, the eigenvectors of the Hermitian operator being

$$\hat{X} = \sum_{i,k=1}^{n} \frac{2p_i p_k \langle p_i|\hat{\Omega}|p_k\rangle}{p_i + p_k} |p_i\rangle\langle p_k|, \tag{73}$$

where p_k and $|p_k\rangle$ are the eigenvalues and eigenvectors of $\hat{\rho}$, as defined by Eq. (19).

Only when $\hat{\rho}$ and $\hat{\Omega}$ commute, $[\hat{\rho},\hat{\Omega}]=0$, does the maximal dispersion (71) reduce to the more usual expression $\mathrm{tr}[\hat{\rho}\hat{\Omega}^2]-[\mathrm{tr}(\hat{\rho}\hat{\Omega})]^2$. This and related questions are discussed in more detail around Eqs. (112) and (113).

1. Coarse-grained situation

The maximal dispersion (71) and (72) provides an upper bound for the dispersion of work defined in a coarse-grained way; see the discussion around Eq. (43). Indeed, according to that discussion the coarse-grained dispersion of work defined with respect to separation of $\mathcal{E}(\hat{\rho})$ to mixed-state subensembles reads

$$\delta w^2_{\mathrm{cg}} = \sum_{\gamma} \nu_\gamma [\mathrm{tr}(\hat{\sigma}_\gamma \hat{\Omega}) - W]^2. \tag{74}$$

Note the decomposition of $\hat{\sigma}_\gamma$ into some set of pure-state subensembles, $\hat{\sigma}_\gamma=\Sigma_\alpha \mu_\alpha^{(\gamma)}|\psi_\alpha^{(\gamma)}\rangle\langle\psi_\alpha^{(\gamma)}|$, where $\mu_\alpha^{(\gamma)}$ are the corresponding probabilities with $\Sigma_\alpha \mu_\alpha^{(\gamma)}=1$. One now finds that the dispersion δw^2 defined as in Eqs. (68) and (69), that is, via the separation of the ensemble $\mathcal{E}(\hat{\rho})$ into pure-state subensembles $\hat{\rho}=\Sigma_{\alpha,\gamma}\nu_\gamma\mu_\alpha^{(\gamma)}|\psi_\alpha^{(\gamma)}\rangle\langle\psi_\alpha^{(\gamma)}|$, is never smaller than δw^2_{cg}:

$$\begin{aligned}\delta w^2 - \delta w^2_{\mathrm{cg}} &\geqslant \sum_{\alpha,\gamma} \nu_\gamma \mu_\alpha^{(\gamma)} (\langle\psi_\alpha^{(\gamma)}|\hat{\Omega}|\psi_\alpha^{(\gamma)}\rangle - W)^2 - \delta w^2_{\mathrm{cg}} \\ &= \sum_{\alpha,\gamma} \nu_\gamma \mu_\alpha^{(\gamma)} \Big(\langle\psi_\alpha^{(\gamma)}|\hat{\Omega}|\psi_\alpha^{(\gamma)}\rangle \\ &\quad - \sum_{\beta} \mu_\beta^{(\gamma)} \langle\psi_\beta^{(\gamma)}|\hat{\Omega}|\psi_\beta^{(\gamma)}\rangle \Big)^2 \geqslant 0. \end{aligned} \tag{75}$$

[24] This is not the only definition of temperature fluctuations. The one proposed in [64] follows a different idea, but again refers to an incomplete ensemble.

[25] Thus if these fluctuations exist, and we assume they do, their description seems to be outside of today's theories. It might be of some interest to see whether more detailed definitions of fluctuations of work can be given in theories of subquantum mechanics, e.g., Bohmian or Nelsonian mechanics. A rather natural situation seems to exist in stochastic electrodynamics, but so far this theory has its own problems [65].

2. The behavior of the maximal dispersion δw^2_{max} for high and low temperatures

With $\hat{\rho}$ given by the Gibbs distribution (21), (19), and (22), one gets from Eq. (71)

$$\delta w^2_{\max} \to 0 \quad \text{for } T \to 0, \tag{76}$$

where T is the temperature of the Gibbsian ensemble. This is a natural result, as for a finite system $\mathcal{S}$ and $T \to 0$ one gets $\hat{\rho} \to |\varepsilon_0\rangle\langle\varepsilon_0|$, where according to Eqs. (20) and (19), $|\varepsilon_0\rangle$ is the common eigenvector of $\hat{\rho}$ and $\hat{H}$ corresponding to the lowest energy (assuming that the latter is not degenerate). As no separation of a pure state into subensembles is possible, the work can take only one value. It is obvious that this is a general feature: by construction, the random quantity work as defined by Eq. (39) does not fluctuate if the initial ensemble is pure. In the same way as in classics, fluctuations of work are present for mixed ensembles only. In this respect the dispersion of work is similar to the von Neumann entropy $S_{\rm vN} = -\mathrm{tr}\,\hat{\rho}\ln\hat{\rho}$, which is also equal to zero for pure density matrices $\hat{\rho}$.

For very high temperatures, where $\hat{\rho} \simeq \hat{1}/n$ so that $p_i \simeq 1/n$, one gets from Eq. (71) (using that $W \to 0$)

$$\delta w^2_{\max} = \frac{1}{n}\,\mathrm{tr}(\hat{\Omega}^2). \tag{77}$$

It is seen that for high temperatures the maximal dispersion may be $O(1)$, provided that the (positive) eigenvalues of $\hat{\Omega}^2$ are finite and do not scale with n.

B. The minimal dispersion of work vanishes

Here we show that there are decompositions into subensembles such that for any $\alpha = 1, \ldots, N$,

$$w_\alpha = \langle\psi_\alpha|\hat{\Omega}|\psi_\alpha\rangle = \sum_{\beta=1}^{N} \lambda_\beta w_\beta = W, \tag{78}$$

that is, the work does not fluctuate at all. In particular, this means that the dispersion δw^2 attains its minimal value equal to zero. This fact is in contrast to the classical situation, where according to points *b* and *c* in Secs. II E 2 and II E 3, $w(x,p)$ should be negative at least for some values of (x,p), and the dispersion of work is large at least for sufficiently high temperatures.

Recall that due to the parametrization (55), (57), and (70), Eq. (78) states that for each α

$$\frac{\langle\pi_\alpha|\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}|\pi_\alpha\rangle}{\langle\pi_\alpha|\hat{\rho}|\pi_\alpha\rangle} = \mathrm{tr}(\hat{\rho}\hat{\Omega}), \tag{79}$$

where $\{|\pi_\alpha\rangle\}_{\alpha=1}^{N}$ with $N \geqslant n$ have to satisfy Eq. (70). This is equivalent to

$$0 = \langle\pi_\alpha|\hat{Y}|\pi_\alpha\rangle, \tag{80}$$

$$\hat{Y} \equiv \hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2} - \mathrm{tr}[\hat{\Omega}\hat{\rho}]\hat{\rho}, \tag{81}$$

where $\hat{Y}$ is Hermitian and (in general) nondiagonal in the eigenrepresentation of $\hat{\rho}$ and traceless:

$$\mathrm{tr}\,\hat{Y} = 0. \tag{82}$$

We now intend to show that in the Hilbert space $\mathcal{H}$ there are orthonormal bases $\{|\pi_i\rangle\}_{i=1}^{n}$ which for the given $\hat{Y}$ do satisfy Eqs. (78) and (80).

1. Some concepts from majorization theory

To this end, let us recall some concepts from the mathematical theory of majorization [67–70]. For two real vectors $x = (x_1 \geqslant \cdots \geqslant x_n)$ and $y = (y_1 \geqslant \cdots \geqslant y_n)$, with their components arranged in nonincreasing way, y is said to majorize x,

$$x \prec y, \tag{83}$$

if the following conditions are satisfied

$$\sum_{i=1}^{k} x_i \leqslant \sum_{i=1}^{k} y_i, \quad k = 1, \ldots, n-1, \tag{84}$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \tag{85}$$

Due to Horn's theorem [67–70], Eq. (83) implies the existence of an $n \times n$ unitary matrix Q_{ij} such that

$$x_i = \sum_{k=1}^{n} y_j |Q_{ij}|^2. \tag{86}$$

The proof of this statement is recalled in Appendix F. This proof is constructive, since it allows us to determine Q_{ij}, starting from given x and y.

2. The minimal dispersion of work is zero

Now denote by $(y_1 \geqslant \cdots \geqslant y_n)$ the eigenvalues of the Hermitian matrix $\hat{Y}$ arranged in a nonincreasing way. Denote by $\{|y_i\rangle\}_{i=1}^{n}$ the corresponding eigenvectors. As follows from Eqs. (82), (84), and (85)

$$(y_1, \ldots, y_n) \succ (0, \ldots, 0). \tag{87}$$

According to Eq. (86) there exists a unitary operator $\hat{Q}$ in the Hilbert space $\mathcal{H}$ such that

$$0 = \sum_{j=1}^{n} y_j |\langle y_j|\hat{Q}|y_i\rangle|^2 = \langle y_i|\hat{Q}^\dagger\hat{Y}\hat{Q}|y_i\rangle. \tag{88}$$

By denoting

$$\hat{Q}|y_i\rangle = |\pi_i\rangle, \quad i = 1, \ldots, n, \tag{89}$$

and identifying labels i and α, we see that Eq. (88) and the desired statement (80) are equivalent.

C. Dispersion of work averaged over all possible separations of the ensemble

We have obtained the maximal and the minimal values of the dispersion of work δw^2. It is useful to have a third characteristic value of δw^2, the dispersion of work for a randomly

chosen separation of the initial ensemble described by $\hat{\rho}$ into pure subensembles. Such a quantity will not depend explicitly on the specific measurement used for separation (i.e., it relates to all measurements that could possibly be done), and thus will help to understand how typical are the maximal and the minimal values of δw^2.

Note from Eqs. (50) and (52) that for a given separation of $\hat{\rho}$, that is, for a given representation (34), the pure density matrices $|\psi_\alpha\rangle\langle\psi_\alpha|$ are expressed via elements $M_{\alpha i}$ of an $N\times N$ unitary matrix M [see the remark after Eq. (53)]. We shall define the average dispersion δw^2_{av} by assuming that M is random, and then integrating $\delta w^2\{M_{\alpha i}\}$ over all possible unitary $N\times N$ matrices. Since there are no reasons for introducing *a priori* biases, we shall assume for the above integration the most uniform, unitary-invariant measure (Haar's measure):

$$\delta w^2_{\mathrm{av}} = \frac{\int \prod_{i,\alpha=1}^{N} d\,\mathrm{Re}\,M_{\alpha i} d\,\mathrm{Im}\,M_{\alpha i}\Theta\{M_{\alpha i}\}\delta w^2\{M_{\alpha i}\}}{\int \prod_{i,\alpha=1}^{N} d\,\mathrm{Re}\,M_{\alpha i} d\,\mathrm{Im}\,M_{\alpha i}\Theta\{M_{\alpha i}\}}, \tag{90}$$

where $\Theta\{M_{\alpha i}\}$ comes due to the unitarity constraint

$$\Theta\{M_{\alpha i}\} = \prod_{\alpha=1}^{N}\delta\left[\sum_{i=1}^{N}|M_{\alpha i}|^2 - 1\right]\prod_{\alpha<\beta}^{N}\delta\left[\sum_{i=1}^{N}M_{\alpha i}M^*_{\beta i}\right].$$

The rows (or, equivalently, the columns) of the matrix M are thus assumed to be a set of N orthonormalized, uniformly random vectors. The quantity δw^2_{av} is calculated in Appendix H:

$$\delta w^2_{\mathrm{av}} = \int_0^\infty ds\left[\prod_{k=1}^{n}\frac{1}{1+sp_m}\right]\left[\sum_{i=1}^{n}\left(\frac{p_i\langle\varepsilon_i|\hat{\Omega}|\varepsilon_i\rangle}{1+sp_i}\right)^2 + \left(\sum_{i=1}^{n}\frac{p_i\langle\varepsilon_i|\hat{\Omega}|\varepsilon_i\rangle}{1+sp_i}\right)^2\right] - W^2. \tag{91}$$

Note that δw^2_{av} (fortunately) depends neither on N nor on the off-diagonal elements $\langle\varepsilon_i|\hat{\Omega}|\varepsilon_j\rangle$ of $\hat{\Omega}$. (Recall that N is the number of pure subensembles constructed from the original homogeneous ensemble, that can be chosen at will, while n is the fixed number of nonzero eigenvalues of $\hat{\rho}$.)

For $\hat{\rho}$ having the Gibbsian form (21), (19), and (22), δw^2_{av} has the following features for low and high temperatures T. It goes to zero for $T\to 0$ for the same reasons as $\delta w^2_{\mathrm{max}}$ does. In contrast, for very high temperatures, where $\hat{\rho}\simeq 1/n$, one has from Eq. (91)

$$\delta w^2_{\mathrm{av}} = \frac{1}{n(n+1)}\sum_{i=1}^{n}\langle\varepsilon_i|\hat{\Omega}|\varepsilon_i\rangle^2. \tag{92}$$

Under the same natural condition that we adopted for studying the high-temperature behavior of $\delta w^2_{\mathrm{max}}$, that is, if the $\langle\varepsilon_i|\hat{\Omega}|\varepsilon_i\rangle$ are finite and do not scale with n, we see that $\delta w^2_{\mathrm{av}}\propto 1/n$ for $n\gg 1$, which is a typical behavior for dispersions of fluctuating macroscopic quantities in statistical physics [3]. Note that even in this limit there is a difference between (92) and the high-temperature behavior (77) of the maximal dispersion.

D. The maximal and the average dispersion of work illustrated for a two-level system

Let us give concrete expressions of $\delta w^2_{\mathrm{max}}$ and δw^2_{av} for a two-level system $\mathcal{S}$. Two-level systems are of interest both experimentally (many applications) and theoretically (simplest case). The initial Gibbsian density matrix is now a 2×2 diagonal matrix with eigenvalues p_1 and $p_2\leqslant p_1$ as given by Eq. (19). The most general matrix form of the traceless and Hermitian operator $\hat{\Omega}$ in this two-dimensional situation is

$$\hat{\Omega} = \begin{pmatrix} \omega & \chi \\ \chi^* & -\omega \end{pmatrix}. \tag{93}$$

Equations (71) and (91) produce then the following expressions for $\delta w^2_{\mathrm{max}}$ and δw^2_{av}, respectively:

$$\delta w^2_{\mathrm{max}} = \omega^2(1-x^2)\left(1+\frac{|\chi|^2}{\omega^2}\right), \tag{94}$$

$$\delta w^2_{\mathrm{av}} = \omega^2(1-x^2)\left[1-\frac{1}{x^2}-\frac{1-x^2}{2x^3}\ln\frac{1-x}{1+x}\right], \tag{95}$$

where

$$x\equiv p_1-p_2\geqslant 0,\quad 1\geqslant x\geqslant 0, \tag{96}$$

is a monotonically decreasing function of temperature, as follows from Eq. (22). As seen from Eqs. (94) and (95), both $\delta w^2_{\mathrm{max}}$ and δw^2_{av} are monotonically increasing functions of temperature T. It is obvious that $\delta w^2_{\mathrm{max}}>\delta w^2_{\mathrm{av}}$, except for the zero temperature situation $x=1$, where they are both equal to zero. For very high temperatures, that is, for $x\to 0$, $\delta w^2_{\mathrm{av}}=\omega^2/3$ in agreement with Eq. (92), while $\delta w^2_{\mathrm{max}}\to\omega^2+|\chi|^2$. Note that off-diagonal elements of $\hat{\Omega}$ increase $\delta w^2_{\mathrm{max}}$, while δw^2_{av} does not depend on them at all.

V. THERE IS NO DIRECT ANALOG OF THE CLASSICAL BK EQUALITY IN THE QUANTUM SITUATION

The discussion in Sec. IV B implies already that, in contrast to the classical case, the fluctuations of work in the quantum situation are not controlled by any *direct* analog of the classic BK equality (11). In the present section we give another illustration of this fact.

Assume for concreteness that the Gibbsian density matrix $\hat{\rho}$ in Eq. (21) was separated into pure subensembles by means of the measurement of $\hat{H}$, that is, the subensembles are described by the pure density matrices $\{|\varepsilon_l\rangle\langle\varepsilon_l|\}_{l=1}^n$, where $\{|\varepsilon_l\rangle\}_{l=1}^n$ are eigenvectors of $\hat{\rho}$.

According to Eq. (39) one has for the realizations of the random quantity work

$$w_l = \langle\varepsilon_l|\hat{U}_\tau^\dagger\hat{H}\hat{U}_\tau|\varepsilon_l\rangle - \varepsilon_l \tag{97}$$

$$= \sum_{k=1}^{n} C_{kl}\varepsilon_k - \varepsilon_l, \quad l = 1, \ldots, n, \tag{98}$$

where

$$C_{kl} = |\langle \varepsilon_k | \hat{U}_\tau | \varepsilon_l \rangle|^2, \tag{99}$$

is a double-stochastic matrix:

$$\sum_{k=1}^{n} C_{kl} = \sum_{l=1}^{n} C_{kl} = 1. \tag{100}$$

Each w_l has probability p_l given by Eq. (22). One constructs

$$\langle e^{-\beta w} \rangle \equiv \sum_{l=1}^{n} p_l e^{-\beta w_l} = \frac{1}{Z} \sum_{l=1}^{n} \exp\left(-\beta \sum_{k=1}^{n} C_{kl}\varepsilon_k \right), \tag{101}$$

that is, one averages $e^{-\beta w}$ directly as was done in the classical situation. It is shown in Appendix B that

$$1 - \frac{\beta^2 \Delta}{2Z} e^{-\beta \varepsilon_{\min}} \leqslant \langle e^{-\beta w} \rangle \leqslant 1 - \frac{\beta^2 \Delta}{2Z} e^{-\beta \varepsilon_{\max}}, \tag{102}$$

$$\Delta \equiv \varepsilon^{\mathrm{T}}(1 - CC^{\mathrm{T}})\varepsilon = \sum_{k=1}^{n} [\langle \varepsilon_k | \hat{H} | \varepsilon_k \rangle^2 - \langle \varepsilon_k | \hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau | \varepsilon_k \rangle^2], \tag{103}$$

where $\varepsilon^{\mathrm{T}} = (\varepsilon_1, \ldots, \varepsilon_n)$ is the vector of eigenvalues of $\hat{H}$, Z is defined in Eq. (21), and where $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are the minimal and maximal ones among $(\varepsilon_1, \ldots, \varepsilon_n)$. Since all the eigenvalues ν of the product of a double-stochastic matrix and its transpose satisfy $0 \leqslant \nu \leqslant 1$,[26] one has

$$\Delta = \varepsilon^{\mathrm{T}}(1 - CC^{\mathrm{T}})\varepsilon \geqslant 0. \tag{104}$$

Thus $\langle e^{-\beta w} \rangle$ is strictly smaller than unity. As compared to our discussion of the classical situation in Sec. II E, the result $\langle e^{-\beta w} \rangle < 1$ does not in general permit us to draw quantum analogs of the classical features in Secs. II E 2 (presence of active realizations) and II E 3 (shape of the dispersion at high T).

VI. COMPARISON WITH OTHER APPROACHES

In the present section we study two approaches known in the literature. The purpose is to understand whether they have the proper physical meaning for describing fluctuations of work. Since they both allow us to generalize the classical BK equality (though in different ways), the adoption of either of them will mean that there is no major qualitative difference in the behavior of quantum and classical fluctuations of work. It should perhaps be stressed that our concern is the applicability of these approaches for describing fluctuations of work under conditions formulated in the Introduction; their usefulness for other purposes is not put into question.

A. Observable of work

Recall from definitions (29) and (30) that for any initial ensemble described by $\hat{\rho}$, the average of $\hat{\Omega}$ is equal to the work done on the corresponding ensemble.

The approach goes on by stating [6,16,27,32] that the operator $\hat{\Omega}$ is the "observable of work" in the standard sense of quantum observables,[27] i.e., the quantity $\mathrm{tr}[\hat{\rho}\hat{\Omega}^2] - [\mathrm{tr}(\hat{\rho}\hat{\Omega})]^2$ is to be interpreted as the dispersion of work for any $\hat{\rho}$. However, while $\mathrm{tr}[\hat{\Omega}\hat{\rho}]$ happens to be equal to the average energy lost by the work source $\mathcal{W}$, simply due to conservation of the average energy during the system-work-source interaction, this alone is, of course, not sufficient to regard $\hat{\Omega}$ as an operator of work. In fact, such an interpretation relies on the analogy between the definition (30) of $\hat{\Omega}$ and the classical expression for the energy difference (8). Such analogies are widespread in general, and once it is accepted that $\hat{\Omega}$ represents the proper energy difference operator, the extension of its interpretation toward the operator of work seems natural.

Let us, however, recall from our discussion in the Introduction that we expect for a proper approach to fluctuations of work to apply in arbitrary nonequilibrium situation. It is then possible to argue that in general $\hat{\Omega}$ does not have the proper meaning of energy difference operator, let alone its meaning as the operator of work.

Let the ensemble $\mathcal{E}(\hat{\rho})$ have a density matrix $\hat{\rho}(0) = |0\rangle\langle 0|$, such that $|0\rangle$ is an eigenstate of $\hat{\Omega} \equiv \hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau - \hat{H}$ with eigenvalue zero:

$$\hat{\Omega}|0\rangle = 0. \tag{105}$$

Recall that $\hat{U}_\tau^\dagger \hat{H}(\tau) \hat{U}_\tau = \hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau$ is the Hamiltonian in the Heisenberg representation at time τ, while the Schrödinger picture relation $\hat{H}(\tau) = \hat{H}$ is due to the assumed cyclic feature of the process. In general,

[26] For any double-stochastic matrix C_{ik}, consider the matrix CC^{T}, where C^{T} is the transpose of C, and let a_i be an eigenvector of CC^{T} corresponding to a (necessarily non-negative) eigenvalue ν: $\Sigma_{k,l=1}^n C_{ik} C_{lk} a_l = \nu a_i$. One has $|\Sigma_{k,l=1}^n C_{ik} C_{lk} a_l| = |\nu a_i| = \nu |a_i| \leqslant \Sigma_{k,l=1}^n C_{ik} C_{lk} |a_l|$, and then $\nu \Sigma_{i=1}^n |a_k| \leqslant \Sigma_{k=1}^n |a_k|$, that is, $\nu \leqslant 1$, since by defintion an eigenvector should be diffferent from zero. In the same way one proves that for all eigenvalues λ of a stochastic matrix $S_{ik} \geqslant 0$, $\Sigma_i S_{ik} = 1$, one has $|\lambda| \leqslant 1$. Such a matrix has always an eigenvalue equal to 1, since it has a left eigenvector (1, 1,…, 1) corresponding to this eigenvalue, and the spectrum is the same for both the left and the right eigenresolutions.

[27] Once $\hat{\Omega}$ is given an independent meaning as a quantum observable, there arises a question on its measurability, since the standard theories of quantum measurements, see, e.g., [34,47], operate in the Schrödinger representation. We shall not pursue this problem here, but rather take as working hypothesis that this measurement can be carried out.

$$[\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau, \hat{H}] \neq 0, \tag{106}$$

so that $|0\rangle$ is neither an eigenstate of $\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau$, nor an eigenstate of $\hat{H}$.

According to quantum mechanics, Eq. (105) should be interpreted as follows: the operator $\hat{\Omega}$ has on the ensemble $\mathcal{E}(|0\rangle\langle 0|)$ a definite value equal to zero, that is, if it is interpreted as the operator of energy change, then for *all single systems* from $\mathcal{E}(|0\rangle\langle 0|)$ the energy does not change during this thermally isolated process.

However, there are concrete examples (see Appendix I) showing that Eq. (105) can be consistent with

$$\langle 0|[\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau]^m|0\rangle \neq \langle 0|\hat{H}^m|0\rangle \quad \text{for } m > 2. \tag{107}$$

This shows that the energy *does change*, since some of its moments do. In other words, the interpretation of $\hat{\Omega}$ as the energy difference operator is in general unsupportable. Note that the noncommutativity feature as expressed by Eq. (106) is essential for this conclusion, but it is a fact of life.

1. Restricted interpretation of $\hat{\Omega}$

A more restricted interpretation of $\hat{\Omega}$ can be given in the light of the definition of fluctuations of work discussed in Sec. III. This will also show that if ρ commutes with $\hat{\Omega}$ (a semiclassical assumption), our approach is consistent with that of the observable of work.

Let the eigenresolution of $\hat{\Omega}$ be

$$\hat{\Omega} = \sum_{k=1}^{n} \omega_k |\omega_k\rangle\langle\omega_k|. \tag{108}$$

Note that for $\hat{\Omega}$ to have the meaning of the operator of work it is necessary that (i) its eigenvalues $\{\omega_k\}_{i=1}^n$ have the meaning of work by themselves, i.e., ω_k should have both the meaning of average energy lost by the work source $\mathcal{W}$ and the average energy gained by a quantum ensemble, as we discussed in Sec. III D; (ii) probabilities of these realizations of work done on the initial ensemble $\mathcal{E}(\hat{\rho})$ should be given as $\{\langle\omega_k|\hat{\rho}|\omega_k\rangle\}_{k=1}^n$.

Now, if $\hat{\rho}$ and $\hat{\Omega}$ commute,

$$[\hat{\rho}, \hat{\Omega}] = 0, \tag{109}$$

then their eigenvectors can be chosen the same, and, by measuring $\hat{\Omega}$, $\hat{\rho} = \Sigma_{k=1}^n p_k |\omega_k\rangle\langle\omega_k|$ can be separated into subensembles $\{\mathcal{E}(|\omega_k\rangle\langle\omega_k|)\}_{k=1}^n$ with probabilities $p_k = \langle\omega_k|\hat{\rho}|\omega_k\rangle$. The work done on each subensemble $\mathcal{E}(|\omega_k\rangle\langle\omega_k|)$ equals $\omega_k = \langle\omega_k|\hat{\Omega}|\omega_k\rangle$, and one can admit the restricted interpretation of $\hat{\Omega}$ as an operator of work.

Conversely, if $\hat{\rho}$ can be separated into subensembles,

$$\hat{\rho} = \sum_{k=1}^{n} \lambda_k |\psi_k\rangle\langle\psi_k|, \tag{110}$$

and if each of them is allowed to interact with the work source $\mathcal{W}$ such that

$$\omega_k = \langle\psi_k|\hat{\Omega}|\psi_k\rangle, \quad \lambda_k = \langle\omega_k|\hat{\rho}|\omega_k\rangle, \tag{111}$$

then three conditions (110), (111), and (108) imply (109).

To show this we proceed in an indirect way, which is useful by itself. It can be noted that the dispersion

$$\mathrm{tr}[\hat{\rho}\hat{\Omega}^2] - [\mathrm{tr}(\hat{\rho}\hat{\Omega})]^2 = \sum_{k=1}^{n} \langle\omega_k|\hat{\rho}|\omega_k\rangle(\omega_k - W)^2 \tag{112}$$

of the operator $\hat{\Omega}$ provides an upper bound for the maximal dispersion $\delta w_{\max}^2$ of work given by Eq. (71):

$$\mathrm{tr}(\hat{\rho}\hat{\Omega}^2) - W^2 - \delta w_{\max}^2 = \frac{1}{2}\sum_{i,k=1}^{n} \frac{(p_i - p_k)^2}{p_i + p_k} |\langle p_k|\hat{\Omega}|p_i\rangle|^2 \geqslant 0. \tag{113}$$

The equality in the RHS of Eq. (113) is realized only if $\hat{\rho}$ and $\hat{\Omega}$ commute, that is, either $\langle p_k|\hat{\Omega}|p_i\rangle$ is zero for $i \neq k$, or for some pair $i \neq k$ one has $\langle p_k|\hat{\Omega}|p_i\rangle \neq 0$, but the corresponding eigenvalues of $\hat{\rho}$ are degenerate: $p_i = p_k$. Thus $\delta w_{\max}^2$ can be equal to $\mathrm{tr}(\hat{\rho}\hat{\Omega}^2) - W^2$ only if $[\hat{\rho}, \hat{\Omega}] = 0$.

Now note that if Eqs. (110), (111), and (108) are assumed to be valid, they imply $\mathrm{tr}(\hat{\rho}\hat{\Omega}^2) - W^2 - \delta w_{\max}^2 \leqslant 0$ simply due to the definition of the maximal dispersion. This is consistent with Eq. (113) only for $\mathrm{tr}(\hat{\rho}\hat{\Omega}^2) - W^2 - \delta w_{\max}^2 = 0$, which implies $[\hat{\rho}, \hat{\Omega}] = 0$, as we saw above. We conclude that Eqs. (110), (111), and (108) imply Eq. (109), as was promised.

Thus, when $[\hat{\rho}, \hat{\Omega}] \neq 0$, $\hat{\Omega}$ does not qualify as the operator of work even in the restricted sense. We also conclude that though the approach does predict an upper bound for δw^2, this bound is not reachable.[28]

2. Generalization of the classical BK equality

Though $\hat{\Omega}$ does not have the meaning of the operator of work—except in a restricted sense and under condition (109)—there is an operator generalization of Eqs. (11) and (12) which was proposed by Bochkov and Kuzovlev in [16,18]:

$$\frac{\mathrm{tr}\, e^{-\beta\hat{\Omega} - \beta\hat{H}}}{Z} = \left\langle \vec{\exp}\left[-\int_0^\beta ds\, e^{-s\hat{H}} \hat{\Omega} e^{s\hat{H}} \right] \right\rangle \tag{114}$$

$$\equiv \mathrm{tr}\left(\vec{\exp}\left[-\int_0^\beta ds\, e^{-s\hat{H}} \hat{\Omega} e^{s\hat{H}} \right] \hat{\rho} \right) = 1. \tag{115}$$

We recall its derivation in Appendix E.[29] A similar relation was derived in [27].

Let us work out some consequences of Eq. (115). As compared to the classic case, the matters are complicated by the

[28] Note that the difference $\mathrm{tr}(\hat{\rho}\hat{\Omega}^2) - W^2 - \delta W^2 = \Sigma_{\alpha=1}^N \lambda_\alpha (\langle\psi_\alpha|\hat{\Omega}^2|\psi_\alpha\rangle - \langle\psi_\alpha|\hat{\Omega}|\psi_\alpha\rangle^2) \geqslant 0$ is by itself always nonnegative for any separation of $\hat{\rho}$ into subensembles.

[29] For the equilibrium ensemble (21), the Thomson formulation of the second law can be derived from Eqs. (114) and (115) upon the

presence of anti-time-ordering and the integral $\int_0^\beta$ in Eqs. (114) and (115). If one insists on not having them, then the equality (114) and (115) can still be converted into an inequality. By applying the Thompson-Golden inequality [71],[30] $\mathrm{tr}[e^{\hat{A}}e^{\hat{B}}] \geqslant \mathrm{tr}\, e^{\hat{A}+\hat{B}}$, valid for any Hermitian operators $\hat{A}$ and $\hat{B}$ (the equality sign is realized here if and only if $[\hat{A},\hat{B}]=0$), one gets

$$\langle e^{-\beta\hat{\Omega}}\rangle \equiv \mathrm{tr}[\hat{\rho}e^{-\beta\hat{\Omega}}] = \sum_{k=1}^{n} \langle\omega_k|\hat{\rho}|\omega_k\rangle e^{-\beta\omega_k} \geqslant \frac{1}{Z}\,\mathrm{tr}\, e^{-\beta\hat{\Omega}-\beta\hat{H}} = 1, \tag{116}$$

where $|\omega_k\rangle$ and ω_k are eigenvectors and eigenvalues of $\hat{\Omega}$ as defined by Eq. (108).

If now we could interpret $\hat{\Omega}$ as the operator of work, that is, if the eigenvalues ω_k of $\hat{\Omega}$ have the meaning of work by themselves, we would note that $\langle\omega_k|\hat{\rho}|\omega_k\rangle$ is the probability of observing the eigenvalue ω_k upon the measurement of $\hat{\Omega}$ on the state $\hat{\rho}$, and then Eq. (116) would allow us to study fluctuations of work exactly in the way we did in Sec. II E for the classical situation. We would then draw the same general conclusions, and the fact that (116) is an inequality would only *strengthen* these conclusions as compared to the classical situation. However, as we saw above, it is impossible to identify $\hat{\Omega}$ with the operator work, and thus fluctuations of work cannot be studied on the basis of (116), except for the special case $[\hat{\rho},\hat{\Omega}]=0$, where Eqs. (114)–(116) reduce to the usual (essentially classical) BK equality.

B. Approach based on two-time measurements of energy

Yet another, different approach to fluctuations of work and extension of the classical BK equality was proposed in Refs. [28–30]. We shall present it in a more extended form, since that is necessary for understanding its proper physical meaning. On the other hand, in order not to dwell on unnecessary technical details, we shall assume that the spectrum of the Hamiltonian $\hat{H}$ is nondegenerate [compare with Eq. (23)]

$$\varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_n. \tag{117}$$

At the time $t=0$ one measures energy (corresponding to the operator $\hat{H}$) for the ensemble described by the Gibbsian density matrix (21). The probability to get an eigenvalue ε_l of $\hat{H}$ is seen from Eqs. (21) and (20) to be

$$p(l|\mathcal{M}_0) = \langle\varepsilon_l|\hat{\rho}|\varepsilon_l\rangle \tag{118}$$

$$= p_l. \tag{119}$$

Equation (118) is the general quantum formula (Born's rule), while Eq. (119) follows from the Gibbsian form (21), (19), and (22) of $\hat{\rho}$. The symbol $\mathcal{M}_0$ in Eq. (118) reminds us that the probability is conditional and refers to the measurement of $\hat{H}$ done at $t=0$. The necessity of such explicit notations will be seen below. Formally it is always allowed, since *any* probability is conditional.

According to Wigner's formula for multitime probabilities in quantum mechanics [72], the subsequent measurement of the energy at the time τ—represented by the same Hamiltonian $\hat{H}$ due to the cyclic feature of the considered process—will then produce a result ε_k with the conditional probability

$$p(k|l,\mathcal{M}_\tau,\mathcal{M}_0) = |\langle\varepsilon_k|\hat{U}_\tau|\varepsilon_l\rangle|^2. \tag{120}$$

There are three conditionals for the probability in the LHS of Eq. (120): $\mathcal{M}_0$ and $\mathcal{M}_\tau$ stand for the measurements done at $t=0$ and $t=\tau>0$, while the index l indicates the result ε_l of the first measurement. The meaning of Eq. (120) is that the ensemble of systems which during the first measurement at $t=0$ produced the result ε_l is described for $t>0$ by $|\varepsilon_l\rangle\langle\varepsilon_l|$. The members of this ensemble couple to the work source $\mathcal{W}$, the state evolves to $\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|\hat{U}_\tau^\dagger$ at the time $t=\tau$, and is then subjected to the second measurement.

Thus the total probability for having the result ε_l at the moment $t=0$ *and* the result ε_k at $t=\tau$ is given by

$$p(k,l|\mathcal{M}_0,\mathcal{M}_\tau) = p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)p(l|\mathcal{M}_0,\mathcal{M}_\tau) \tag{121}$$

$$= p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)p(l|\mathcal{M}_0). \tag{122}$$

When passing from Eq. (121) to Eq. (122), we used the obvious relation $p(l|\mathcal{M}_0,\mathcal{M}_\tau)=p(l|\mathcal{M}_0)$ (causality, no dependence on the future). It is to be noted that

$$p(k|\mathcal{M}_0,\mathcal{M}_\tau) = \sum_{l=1}^{n} p(k,l|\mathcal{M}_0,\mathcal{M}_\tau) = \sum_l p_l\langle\varepsilon_k|\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|U_\tau^\dagger|\varepsilon_k\rangle, \tag{123}$$

that is, the probability to have the result ε_k in the second measurement is for a general initial density matrix $\hat{\rho}$ not equal to

$$p(k|\mathcal{M}_\tau) = \langle\varepsilon_k|\hat{U}_\tau\hat{\rho}U_\tau^\dagger|\varepsilon_k\rangle, \tag{124}$$

which is the probability to get the result k in a different context, where no first measurement was done. Such an equality is valid, though, if $\hat{\rho}$ commutes with $\hat{H}$, which is the case with the Gibbsian density matrix (21). Let us first restrict our attention to this case.

1. Another generalization of the classical BK equality

One notes from Eq. (120) the double-stochastic feature of $p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)$:

application of the Peierls-Bogoliubov inequality (recalled in Appendix E): $e^{-\beta\,\mathrm{tr}[\hat{\rho}\hat{\Omega}]} \leqslant (1/Z)\mathrm{tr}\, e^{-\beta\hat{\Omega}-\beta\hat{H}} = 1$. From this it follows once again that $W=\mathrm{tr}[\hat{\Omega}\hat{\rho}] \geqslant 0$.

[30] The Thompson-Golden inequality is a particular consequence of the submajorization relation $\lambda(e^{\hat{A}+\hat{B}}) \prec_w \lambda(e^{\hat{A}/2}e^{\hat{B}}e^{\hat{A}/2})$, where $\lambda(\hat{A})$ is the eigenvalue vector of a Hermitian operator $\hat{A}$; see [69] for the definition of submajorization $\prec_w$ and for more details.

$$\sum_{k=1}^{n} p(k|l,\mathcal{M}_\tau,\mathcal{M}_0) = \sum_{l=1}^{n} p(k|l,\mathcal{M}_\tau,\mathcal{M}_0) = 1, \quad (125)$$

and calculates using Eqs. (19), (22), (119), and (125):

$$\langle e^{-\beta(\varepsilon_k-\varepsilon_l)}\rangle_{0,\tau} \equiv \sum_{k,l=1}^{n} p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)e^{-\beta(\varepsilon_k-\varepsilon_l)}$$
$$= \frac{1}{Z}\sum_{k,l=1}^{n} p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)e^{-\beta\varepsilon_k} = 1. \quad (126)$$

This is the equality obtained in Refs. [28–30] as a generalization of the classical BK equality.

Note that for the density matrix (21) the average

$$\sum_{k,l=1}^{n} p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)(\varepsilon_k-\varepsilon_l) = W \quad (127)$$

is equal to the work as defined by Eq. (29). The statement of the second law, $W \geqslant 0$, can once again be deduced from Eq. (126) by employing convexity of the exponent.

2. Critique of the approach

If we were now able to associate the work with a random variable having realizations $\{\varepsilon_k-\varepsilon_l\}_{k,l=1}^n$ and the corresponding probabilities $\{p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)\}_{k,l=1}^n$, it would be possible to study fluctuations of work on the base of Eq. (126), and to draw essentially the same conclusions as we did in Sec. II E for the classical case. It is, however, not difficult to see that the same criticisms we brought in Sec. VI A with respect to the "observable of work" applies here, too.

Keeping in mind our discussion after Eq. (120), note that if the ensemble initially described by $|\varepsilon_l\rangle\langle\varepsilon_l|$ couples to the work source $\mathcal{W}$, its mechanical degree of freedom loses at the time $t=\tau$ the energy

$$\mathrm{tr}(\hat{\Omega}|\varepsilon_l\rangle\langle\varepsilon_l|) = \mathrm{tr}(\hat{H}\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|\hat{U}_\tau^\dagger) - \varepsilon_l. \quad (128)$$

Since the final density matrix $\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|\hat{U}_\tau^\dagger$ need not commute with $\hat{H}$, the energy need not have any definite value at that time, and Eq. (128) does in general not reduce to $\varepsilon_k-\varepsilon_l$ with any fixed k. Such a reduction takes place, however, when

$$\mathrm{tr}(\hat{H}\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|\hat{U}_\tau^\dagger) = \sum_{k=1}^{n} C_{kl}\varepsilon_k = \varepsilon_{\pi(l)} \quad \text{for } l=1,\ldots,n, \quad (129)$$

where C_{kl} is defined via Eq. (99), and where $(\pi(1),\ldots,\pi(n))$ is some permutation of the sequence $(1,\ldots,n)$. Equation (129) can then be rewritten as

$$\sum_{k=1}^{n} \tilde{C}_{kl}(\varepsilon_k-\varepsilon_l) = 0, \quad (130)$$

where the matrix $\tilde{C}=C\Pi$ the product of C and the corresponding permutation matrix Π, and where we noted that once the matrices C and Π are double stochastic [see Eq. (100) for definition), so is $\tilde{C}$. Note, with help of Eq. (117), that for $l=n$ all terms with $k\neq n$ in Eq. (130) are negative unless $\tilde{C}_{k\neq n}=0$, which via the double-stochastic feature of $\tilde{C}$ implies that $\tilde{C}_{n\neq k}=0$ and $\tilde{C}_{nn}=1$. Continuing along the same lines for $l<n$, one gets that Eq. (129) can take place only when $\tilde{C}$ reduces to unity matrix, or, equivalently, C reduces to a permutation matrix

$$\hat{U}_\tau|\varepsilon_l\rangle\langle\varepsilon_l|\hat{U}_\tau^\dagger = |\varepsilon_{\pi(l)}\rangle\langle\varepsilon_{\pi(l)}|. \quad (131)$$

Thus, in general it is the expression (128) and not $\varepsilon_k-\varepsilon_l$ itself that can be interpreted as the work occurring with probability p_l, and this is precisely the point from which we departed in Sec. III.

It is also straightforward to see that the approach does not apply out of equilibrium. The reasons for this are even more straightforward than for the previous approach.

Recall from the Introduction that the proper definition of fluctuations of work is expected to apply to any nonequilibrium initial ensemble $\mathcal{E}(\hat{\sigma})$ with density matrix $\hat{\sigma}$ not commuting with $\hat{H}$:

$$[\hat{\sigma},\hat{H}] \neq 0. \quad (132)$$

In particular, the work averaged over those fluctuations should be equal to the one done on the ensemble.

The present approach is generalized uniquely for arbitrary initial state: the definitions of $p(l|\mathcal{M}_0)$ and $p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)$ in Eqs. (120) and (118) remain unaltered: one substitutes there $\hat{\sigma}$ instead of $\hat{\rho}$.

It is now straightforward to see from Eq. (132) that, due to nondiagonal terms in $\hat{\sigma}$, the average $\Sigma_{k.l=1}^n p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)(\varepsilon_k-\varepsilon_l)$ is not equal to the work $\mathrm{tr}(\hat{\sigma}\hat{\Omega})$ done on the overall ensemble:

$$\mathrm{tr}(\hat{\sigma}\hat{\Omega}) - \sum_{k.l=1}^{n} p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)(\varepsilon_k-\varepsilon_l)$$
$$= \mathrm{tr}[\hat{U}_\tau^\dagger\hat{H}\hat{U}(\hat{\sigma}-|\varepsilon_l\rangle\langle\varepsilon_l|\hat{\sigma}|\varepsilon_l\rangle\langle\varepsilon_l|)] \neq 0. \quad (133)$$

This agument shows again that this approach does not provide a good definition for fluctuations of work.

C. The approaches based on the "observable of work" and on two-time measurements of energy are different

This difference is seen already by comparing Eq. (116) with (126). Still, we want to understand this difference in more detail. More precisely, even though for the initial density matrix commuting with $\hat{H}$, the first and the second moments generated by the two approaches are equal:

$$\mathrm{tr}[\hat{\rho}\hat{\Omega}^p] = \sum_{k,l=1}^{n} p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)(\varepsilon_k-\varepsilon_l)^p, \quad p=0,1,2, \quad (134)$$

already the third moments are in general different, *even though* $[\hat{\rho},\hat{H}]=0$. Indeed, assuming validity of the latter condition, one gets

$$\mathrm{tr}[\hat{\rho}\hat{\Omega}^3] - \sum_{k,l=1}^{n} p(l|\mathcal{M}_0)p(k|l,\mathcal{M}_\tau,\mathcal{M}_0)(\varepsilon_k-\varepsilon_l)^3$$

$$= \mathrm{tr}(\hat{\Omega}[\hat{\Omega},\hat{\rho}]\hat{H}) = \mathrm{tr}(\hat{\rho}[\hat{H},\hat{\Omega}]\hat{\Omega}). \qquad (135)$$

The RHS of Eq. (135) vanishes only if $[\hat{\rho},\hat{\Omega}]=0$, or equivalently $[\hat{H},\hat{\Omega}]=0$, in addition to $[\hat{\rho},\hat{H}]=0$, so that (135) does not vanish in general.

For the two-level example of Sec. IV D (with nondiagonal $\hat{\Omega}$) the RHS of Eq. (135) reads

$$\mathrm{tr}(\hat{\rho}[\hat{H},\hat{\Omega}]\hat{\Omega}) = (p_1-p_2)(\varepsilon_1-\varepsilon_2)|\chi|^2, \qquad (136)$$

where $\hat{\Omega}$ is given by Eq. (93), and where p_k and ε_k are eigenvalues of $\hat{\rho}$ and $\hat{H}$, respectively. For the Gibbsian density matrix $\hat{\rho}$, the RHS of Eq. (136) has negative sign.

Finally, let us point out that differences between the two approaches were recently studied in [32] in a different context.

D. Summary of the discussion of the two approaches

We have discussed two approaches known in the literature, and argued that in the proper quantum domain they do not describe fluctuations of work. Work is a rather particular form of energy having several specific features we discussed in Sec. II B. The approaches miss those features, since, as we argued, they do not ensure that realizations of the claimed random quantity or operator of work have themselves the physical meaning of work. They still allow different generalizations of the classical BK equality which makes them operationally close to the classical situation. These generalizations might be useful for their own sake, but not for discussing fluctuations of work.

VII. CONCLUSION

The second law has a statistical character as it is both formulated and valid for ensembles of identically prepared systems. It is therefore of interest to investigate this statistical aspect in more detail. For the entropic formulation of the second law, this analysis is by now a standard chapter of statistical thermodynamics [1–3].

In the present paper we studied how Thomson's formulation of the second law—no work from an equilibrium ensemble by a cyclic process—emerges through averaging over fluctuations of work in the quantum situation. It will be useful at this moment to recall the special role of Thomson's formulation, and then to proceed with concluding remarks on our results. Recall that a detailed discussion of our approach, including several pertinent questions on its physical meaning, was given by us in Sec. III H.

A. The main features of Thomson's formulation of the second law

(a) The formulation uses the concept of work which is unambiguously defined both conceptually and operationally, both in and out of equilibrium. In this respect work is contrasting to entropy, which is well defined only in (nearly) equilibrium states of macroscopic systems.

(b) Thomson's formulation is valid for any *finite*[31] or *infinite* [20], quantum or classical system interacting with macroscopic sources of work. Not all formulations of the second law have such a universal regime of validity. While all formulations are equivalent in the standard thermodynamical domain, that is, for (nearly) equilibrium states of macroscopic systems, some of them have definite limits when considered for finite systems [23] or for low temperatures (quantum domain) [14,15].

(c) In its literal form Thomson's formulation does not imply any irreversibility, since the dynamics of the system coupled to work source is unitary and thus formally reversible: if some work was put into the initially equilibrium system it can in principle be extracted back. The irreversibility with respect to work transfer comes into existence when one takes into account that in practice no work source can interact with all possible degrees of freedom. In particular, if the system was subjected to a thermal bath after it had interacted with the work source, the system relaxes back to its Gibbsian state and the work which had been put into it cannot be recovered by *any* work source acting on the system only (a similar argument is presented in [6], Chap. 5).

(d) It should perhaps be stressed that Thomson's formulation does not refer to all aspects usually associated with the second law, e.g., by itself it does not explain how a subsystem of a proper macroscopic system (thermal bath) relaxes toward a Gibbsian equilibrium state On the basis of Thomson's formulation it is only possible to argue that—under several assumptions the main of which is the additivity[32]—the Gibbsian state is the only one which forbids work extraction via *any* cyclic thermally isolated process [20,21]. The property of relaxation toward a Gibbsian state is to be viewed as an independent issue of statistical physics; its standard classical understanding was reshaped in literature various times; see, e.g., [14,15,73,74].

B. What appeared to be problematic in defining fluctuations of work in the quantum situation

As we saw in Sec. VI, due to noncommutativity of various quantum observables, there are different quantities which, in the classical limit, coincide with the random quan-

[31] In this context one sometimes hears that the second law must refer to macroscopic systems, and there is no sense in applying it for finite systems. This opinion is not correct, as instanced by Thomson's formulation. If it were not valid for a finite system coupled to work sources, its very application to macroscopic systems would be endangered, because the initial Gibbsian ensemble (21) is prepared under weak interaction with an equilibrium thermal bath; see, e.g., [4]: any cycle violating the formulation for a finite system can be repeated to achieve a violation for the bath.

[32] Additivity means that for two noninteracting systems with Hamiltonians $\hat{H}_1$ and $\hat{H}_2$, the corresponding density matrices factorizes: $\hat{\rho}(\hat{H}_1+\hat{H}_2)=\hat{\rho}(\hat{H}_1)\otimes\hat{\rho}(\hat{H}_2)$. This feature is satisfied for the Gibbsian case: $\hat{\rho}(\hat{H}_1)\propto e^{-\beta\hat{H}_1}$.

tity work. As often, classical reasoning alone is of no help for defining fluctuations of work.

One therefore has first to state what basic physical features the fluctuations of work are expected to have, as we did in Sec. I. Once these features are recognized, the definition of fluctuations of work presented in Sec. III follows naturally.

C. What is similar and what is different in classical and quantum definitions of fluctuations of work?

In both situations the definition of work as a random quantity employs the same idea: the initial homogeneous ensemble of identically prepared systems is separated into irreducible (and homogeneous) subensembles. Both in quantum and classical situations these irreducible subensembles are described maximally completely. In classics they correspond to a trivial subensemble of identical copies of the same system (so that within a subensemble no fluctuations are present), and they are described via phase-space points and trajectories. In quantum mechanics these subensembles, described by pure density matrices (wave functions), provide definite (nonfluctuating) values for the largest possible, *but nonexhaustive*, set of observables, since the pure-state subensemble is not trivial.

In classics the irreducible subensembles of the initial ensemble obviously exist *a priori*, that is, without need of any measurement. In the quantum situation the very structure of subensembles does depend on the measurement applied for the actual separation, or, in other words, for the preparation of an inhomogeneous ensemble. Thus, we need this initial *preparational* measurement, a step which is absent in classics. The above context dependence goes hand in hand with the impossibility of achieving a complete description of individual systems in any subensemble. As the main consequence, the separation of a mixed ensemble is not unique, and thus the random quantity work is *contextual* in the quantum situation.

In the second step, systems from each irreducible subensemble interact with the same macroscopic source. Realizations of the random quantity work are then defined as the average energy increase of the work source when interacting with each subensemble, while the probability of each realization is given by the weight of the corresponding subensemble in the initial mixed ensemble.

In this way the full physical meaning of work is kept, and the approach can be applied to any nonequilibrium initial state of a system interacting with its environment. Indeed, there are experiments in quantum optics which realize observation of fluctuations of work; see Sec. III G.

It remains to mention that the idea of the presented definition of quantum fluctuations of work agrees with the general strategy of studying classical fluctuations for quantities (e.g., temperature) which do not have a direct mechanical meaning [3]. Recall that realizations of the random quantity temperature as proposed in [3]—and checked experimentally in [63]—refer to finite subensembles, such that the physical meaning of temperature is kept.

D. Dispersion of work

The most direct quantity that characterizes fluctuations of work is the dispersion of work we studied in Sec. IV. Although the work is a contextual random quantity and depends on the measurement that was done to separate the initial mixed ensemble into pure subensembles, one can define two reasonable quantities—maximal dispersion and the dispersion for a randomly chosen separation on the initial ensemble—that depend solely on the internal features of the system, that is, on its initial state and its time-dependent Hamiltonian. These quantities were calculated explicitly for any finite quantum system and studied in Sec. IV.

E. Nonexistence of the direct generalization of classical BK equality

In the classical situation fluctuations of work in an initially equilibrium state are controlled by the BK equality [16,24]. This equality allows one to draw a number of model-independent statements on fluctuations of work which we summarized in Sec. II E. In contrast, the *direct* generalization of the BK equality to the quantum domain—which would allow one to draw similar qualitative conclusion on fluctuations of work—does not exist; see Sec. V. As we discussed in detail in Sec. VI, there are quantum generalizations of the BK equality, but they refer to quantities that describe fluctuations of work only if some classical features are present, e.g., those implied by Eq. (109). As the main consequence, fluctuations of work in the quantum situation can have features which are impossible in classics, e.g., (intersubensemble) fluctuations can be absent completely.

ACKNOWLEDGMENTS

It is a pleasure to thank Roger Balian for inspiring discussions. We thank Claudia Pombo for critical remarks and for stressing the importance of understanding the concept of work. The work of A.E.A is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM, financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO.)

APPENDIX A: DERIVATION OF EQ. (15)

Here we recall from [75] a generalization of the Cauchy inequality used in Eq. (15).

Denote by $\Gamma=(x,p)$ the phase space point. Assume that all the integrals over the phase space used below are finite. The desired inequality reads: if $a(\Gamma)$, $b(\Gamma)$, $x(\Gamma)$ are some functions satisfying

$$\int d\Gamma\, a(\Gamma)x(\Gamma)=0, \qquad \int d\Gamma\, b(\Gamma)x(\Gamma)=1, \tag{A1}$$

then

$$\int d\Gamma\, x^2(\Gamma) \geqslant \frac{\int d\Gamma\, a^2(\Gamma)}{\int d\Gamma\, a^2(\Gamma)\int d\Gamma\, b^2(\Gamma)-\left[\int d\Gamma\, a(\Gamma)b(\Gamma)\right]^2}. \tag{A2}$$

To prove this, define

$$A = \int d\Gamma\, a^2(\Gamma), \tag{A3}$$

$$B = \int d\Gamma\, b^2(\Gamma), \tag{A4}$$

$$C = \int d\Gamma\, a(\Gamma) b(\Gamma), \tag{A5}$$

$$y(\Gamma) = \frac{Ab(\Gamma) - Ca(\Gamma)}{AB - C^2}, \tag{A6}$$

and note that

$$\int d\Gamma\, x^2(\Gamma) \geqslant \int d\Gamma\, y^2(\Gamma), \tag{A7}$$

due to

$$\int d\Gamma\, x(\Gamma) y(\Gamma) = \int d\Gamma\, y^2(\Gamma), \tag{A8}$$

which is valid by constructions (A3)–(A6). Equation (A2) follows from Eq. (A7). To get from here Eq. (15) one identifies $x(\Gamma) = \sqrt{\mathcal{P}(\Gamma)}$, $b = \sqrt{\mathcal{P}(\Gamma)} e^{-\beta w(\Gamma)}$, $a = \sqrt{\mathcal{P}(\Gamma)}(f(\Gamma) - \langle f \rangle)$.

APPENDIX B: DERIVATION OF EQ. (102)

Let $f(x)$ be a smooth function, $\{x_i\}_{i=1}^n$ be n points, and

$$\overline{x} = \sum_{k=1}^{n} \lambda_k x_k, \quad \lambda_k \geqslant 0, \quad \sum_{k=1}^{n} \lambda_k = 1. \tag{B1}$$

Apply the incomplete Taylor expansion to $f(x_i)$:

$$f(x_i) = f(\overline{x}) + f'(\overline{x})(x_i - \overline{x}) + \frac{f''(\xi_i)}{2}(x_i - \overline{x})^2, \tag{B2}$$

where ξ_i lies between x_i and $\overline{x}$. Denote by $x_{\max}$ and $x_{\min}$ the maximal and the minimal numbers among x_i. This implies $x_{\max} \geqslant \xi_k \geqslant x_{\min}$. Now assume that $f''(x)$ is monotonically decaying:

$$f''(x_{\max}) \geqslant f''(\xi_i) \geqslant f''(x_{\min}). \tag{B3}$$

Then using Eqs. (B2) and (B3) one has

$$\sum_{k=1}^{n} \lambda_k f(x_k) - f(\overline{x}) = \frac{1}{2} \sum_{k=1}^{n} f''(\xi_k) \lambda_k (x_k - \overline{x})^2, \tag{B4}$$

$$\frac{f''(x_{\min})}{2} \sum_{k=1}^{n} \lambda_k (x_k - \overline{x})^2 \geqslant \sum_{k=1}^{n} \lambda_k f(x_k) - f(\overline{x}) \geqslant \frac{f''(x_{\max})}{2} \sum_{k=1}^{n} \lambda_k (x_k - \overline{x})^2. \tag{B5}$$

To derive Eq. (102), start from Eq. (101), and take in Eq. (B5) the convex function $f = e^{-\beta\varepsilon}$, $\beta = 1/T > 0$, and identify $x_i = \varepsilon_i$, $\lambda_k = C_{kl}$ for each fixed l. The desired Eq. (102) is then recovered upon the summation over l.

APPENDIX C: MAXIMAL DIMENSION OF DISPERSIONLESS OBSERVABLES

Let $\mathrm{tr}(\hat{A}^2\hat{\rho}) = [\mathrm{tr}(\hat{A}\hat{\rho})]^2$ be valid for some Hermitian operator $\hat{A}$ and density matrix $\hat{\rho}$. In the main text we called such operators dispersionless with respect to the ensemble described by the density matrix $\hat{\rho}$.

In the Cauchy inequality $|\mathrm{tr}(\hat{A}\hat{B})|^2 \leqslant \mathrm{tr}(\hat{A}\hat{A}^\dagger)\mathrm{tr}(\hat{B}\hat{B}^\dagger)$, which is valid for any operators $\hat{A}$ and $\hat{B}$, while the equality is realized for $\hat{A} = \alpha\hat{B}^\dagger$, where α is a number. Thus the equality

$$[\mathrm{tr}(\hat{A}\sqrt{\hat{\rho}}\sqrt{\hat{\rho}})]^2 = \mathrm{tr}(\hat{A}^2\hat{\rho})\mathrm{tr}(\hat{\rho}) \tag{C1}$$

implies

$$\hat{A}\sqrt{\hat{\rho}} = \alpha\sqrt{\hat{\rho}}. \tag{C2}$$

Now insert the eigenresolution $\sqrt{\hat{\rho}} = \Sigma_{k=1}^{n} \sqrt{p_k} |\varepsilon_k\rangle\langle\varepsilon_k|$ into Eq. (C2) and multiply it from the right by $|p_m\rangle$, to obtain

$$\sqrt{p_m}\hat{A}|p_m\rangle = \alpha\sqrt{p_m}|p_m\rangle. \tag{C3}$$

It is seen either that only one among the eigenvalues p_k's is nonzero and then the corresponding eigenvector is also an eigenvector for $\hat{A}$, or, more generally, that $\hat{A}$ acts as $\propto \hat{1}$ in the Hilbert space formed by eigenvectors of $\hat{\rho}$ corresponding to its nonzero eigenvalues. In both cases the measurement of $\hat{A}$ on the state $\hat{\rho}$ always produces definite results.

Thus any operator $\hat{A}$ that is dispersionless on the density matrix $\hat{\rho}$ has to have the following block-diagonal matrix representation:

$$\hat{A} = \begin{pmatrix} \alpha\hat{1}_{k\times k} & 0 \\ 0 & \hat{B} \end{pmatrix}, \tag{C4}$$

where α is a real number, $\hat{1}_{k\times k}$ is the $k \times k$ unity matrix in the k-dimensional Hilbert space formed by eigenvectors corresponding to nonzero eigenvalues of $\hat{\rho}$, and finally $\hat{B}$ is an arbitrary $(n-k)\times(n-k)$ Hermitian matrix on the space orthogonal to the zero eigenvalues. It has $(n-k)^2$ free parameters, and another free parameter of $\hat{A}$ is coming with the real number α. Thus, $\hat{A}$ has

$$(n-k)^2 + 1$$

free parameters.

Note finally that various operators that are dispersionless on a pure density matrix need not be mutually commuting. As one of the simplest examples consider

$$\hat{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{D} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \epsilon \end{pmatrix}, \quad |\psi\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

where ϵ is real. It is seen that $\hat{C}|\psi\rangle = |\psi\rangle$ and $\hat{D}|\psi\rangle = \epsilon|\psi\rangle$, but $[\hat{C}, \hat{D}] \neq 0$.

APPENDIX D: RELATION BETWEEN POVMS AND PROJECTIVE MEASUREMENTS

We outline how a POVM given by Eqs. (44)–(46) can be connected with the usual projective measurements.

The general strategy is to couple the system $\mathcal{S}$ with another auxiliary system $\mathcal{G}$. The initial states of $\mathcal{S}$ and $\mathcal{G}$ are, respectively, $\hat{\rho}$ (living in an n-dimensional Hilbert space $\mathcal{H}$) and $\hat{\rho}_{\mathcal{G}}$ existing in an N-dimensional Hilbert space $\mathcal{H}_{\mathcal{G}}$. The initial state of the overall system $\mathcal{S}+\mathcal{G}$ is thus $\hat{\rho}\otimes\hat{\rho}_{\mathcal{G}}$.

Let now the composite system evolve in time under some interaction, and let $\hat{\mathcal{U}}$ be the corresponding evolution operator. The final state is thus

$$\hat{\mathcal{U}}\hat{\rho}\otimes\hat{\rho}_{\mathcal{G}}\hat{\mathcal{U}}^{\dagger}. \tag{D1}$$

Let also

$$\hat{\rho}_{\mathcal{G}}=\sum_{\alpha=1}^{N} h_{\alpha}|h_{\alpha}\rangle\langle h_{\alpha}|, \quad \langle h_{\alpha}|h_{\beta}\rangle=\delta_{\alpha\beta}, \tag{D2}$$

$$\sum_{\alpha=1}^{N} h_{\alpha}=1 \tag{D3}$$

be the eigenresolution of the density matrix $\hat{\rho}_{\mathcal{G}}$.

One now measures for the system $\mathcal{G}$ any Hermitian operator with a nondegenerate spectrum having an orthonormal set of vectors

$$\{|g_{\alpha}\rangle\}_{\alpha=1}^{N} \tag{D4}$$

as its eigenbase. The probability for obtaining the result α, and the postmeasured state of $\mathcal{S}$ obtained upon conditioning on the result α read, respectively,

$$\lambda_{\alpha}=\mathrm{tr}\langle g_{\alpha}|\hat{\mathcal{U}}\hat{\rho}\otimes\hat{\rho}_{\mathcal{G}}\hat{\mathcal{U}}^{\dagger}|g_{\alpha}\rangle=\sum_{\beta=1}^{N} h_{\beta}\mathrm{tr}(\hat{\rho}\hat{G}_{\alpha}^{(\beta)}\hat{G}_{\alpha}^{(\beta)\dagger}), \tag{D5}$$

$$\hat{\rho}'_{\alpha}=\frac{1}{\lambda_{\alpha}}\langle g_{\alpha}|\hat{\mathcal{U}}\hat{\rho}\otimes\hat{\rho}_{\mathcal{G}}\hat{\mathcal{U}}^{\dagger}|g_{\alpha}\rangle=\frac{1}{\lambda_{\alpha}}\sum_{\beta=1}^{N} h_{\beta}\hat{G}_{\alpha}^{(\beta)}\hat{\rho}\hat{G}_{\alpha}^{(\beta)\dagger}, \tag{D6}$$

where by definition

$$\hat{G}_{\alpha}^{(\beta)}=\langle g_{\alpha}|\hat{\mathcal{U}}|h_{\beta}\rangle, \tag{D7}$$

$$\sum_{\alpha=1}^{N}\hat{G}_{\alpha}^{(\beta)\dagger}\hat{G}_{\alpha}^{(\beta)}=\langle h_{\beta}|\hat{\mathcal{U}}^{\dagger}\hat{\mathcal{U}}|h_{\beta}\rangle=1. \tag{D8}$$

If now the initial state $\hat{\rho}_{\mathcal{G}}$ is pure, then only one term survives in the summations over β in Eqs. (D5) and (D6), and we return to POVM measurements as given by Eqs. (44)–(46). The situation is only slightly different for the general case when the initial state $\hat{\rho}_{\mathcal{G}}$ of $\mathcal{G}$ is mixed. Here we get a convex sum of ordinary POVMs which corresponds to a noisy (nonideal) measurement, since the postmeasurement ensembles of the system $\mathcal{S}$ are now described by mixed density matrices $\hat{\rho}'_{\alpha}$, even if the premeasurement ensemble was described by a pure density matrix.

We have shown above how to generate some POVM measurement. Here we discuss (more or less following [34,47,56,57]) how to generate a specific (given) POVM, that is, given operators $\hat{G}_{\alpha}$ existing in the Hilbert space $\mathcal{H}$ and satisfying Eq. (44), one should construct (1) an initial state $\hat{\rho}_{\mathcal{G}}$; (2) a projective measurement base (D4) for $\mathcal{G}$; (3) an evolution operator $\hat{\mathcal{U}}$; such that one gets for an arbitrary $\hat{\rho}$

$$\hat{\mathcal{U}}\hat{\rho}\otimes\hat{\rho}_{\mathcal{G}}\hat{\mathcal{U}}^{\dagger}=\sum_{\alpha=1}^{N}\hat{G}_{\alpha}\hat{\rho}\hat{G}_{\alpha}^{\dagger}\otimes|g_{\alpha}\rangle\langle g_{\alpha}|. \tag{D9}$$

Then the POVM (44)–(46) accounts for what is happening—after the interaction and after the selective measurement in the base (D4)—with the initial ensemble described by $\hat{\rho}$.

To this end let us select two arbitrary orthonormal bases

$$\{|u_k\rangle\}_{k=1}^{n}, \quad \{|g_{\alpha}\rangle\}_{\alpha=1}^{N} \tag{D10}$$

in $\mathcal{H}$ and in $\mathcal{H}_{\mathcal{G}}$, respectively. Let the initial state of $\mathcal{G}$ be chosen as the pure state

$$\hat{\rho}_{\mathcal{G}}=|g_1\rangle\langle g_1|. \tag{D11}$$

Assume that the interaction between $\mathcal{S}$ and $\mathcal{G}$ is chosen such that the corresponding unitary evolution operator $\hat{\mathcal{U}}$ in the composite Hilbert space $\mathcal{H}\otimes\mathcal{H}_{\mathcal{G}}$ results in

$$\hat{\mathcal{U}}|u_k\rangle\otimes|g_1\rangle=\sum_{\alpha=1}^{N}\hat{G}_{\alpha}|u_k\rangle\otimes|g_{\alpha}\rangle. \tag{D12}$$

Note that due to the completeness relation $\Sigma_{\alpha=1}^{N}\hat{G}_{\alpha}^{\dagger}\hat{G}_{\alpha}=\hat{1}$, as given by Eq. (44), one has from Eq. (D12)

$$\langle g_1|\otimes\langle u_k|\hat{\mathcal{U}}^{\dagger}\hat{\mathcal{U}}|u_l\rangle\otimes|g_1\rangle=\sum_{\alpha,\beta=1}^{N}\langle g_{\alpha}|g_{\beta}\rangle\langle u_k|\hat{G}_{\alpha}^{\dagger}\hat{G}_{\beta}|u_l\rangle=\delta_{kl}, \tag{D13}$$

because $\langle g_{\alpha}|g_{\beta}\rangle=\delta_{\alpha\beta}$.

The specification of $\hat{\mathcal{U}}$ is not yet complete. To complete the definition of $\hat{\mathcal{U}}$ in the composite Hilbert space $\mathcal{H}\otimes\mathcal{H}_{\mathcal{G}}$ one should define its action on

$$|\theta_{k\alpha}\rangle=|u_k\rangle\otimes|g_{\alpha}\rangle, \tag{D14}$$

for $\alpha=2,\ldots,N$ in addition to Eq. (D12). This will suffice, since $\{|\theta_{k,\alpha}\rangle\}_{k=1,\alpha=1}^{n,\ N}$ is an orthonormal base in the composite Hilbert $\mathcal{H}\otimes\mathcal{H}_{\mathcal{G}}$.

This completion is possible and one can do that in many different ways, because it amounts to completing the set of n orthonormal vectors

$$\vec{\Theta}_k=\{\{\langle u_l|\hat{G}_{\alpha}|u_k\rangle\}_{l=1}^{n}\}_{\alpha=1}^{N}, \quad k=1,\ldots,n, \tag{D15}$$

to an orthonormal base (in $\mathcal{H}\otimes\mathcal{H}_{\mathcal{G}}$) containing $N\times n$ vectors. Then the columns (or equivalently the rows) of $\hat{\mathcal{U}}$ in the base $|\theta_{k,\alpha}\rangle$ will be a set of of $N\times n$ orthonormal vectors, which is equivalent to $\hat{\mathcal{U}}$ being a unitary matrix.

On the other hand, for a given unitary matrix $\hat{\mathcal{U}}$ there is a Hermitian operator $\hat{H}_{\mathrm{ov}}$ such that $\hat{\mathcal{U}}=\exp[(i\tau/\hbar)\hat{H}_{\mathrm{ov}}]$ with

some time parameter τ. Thus, $\hat{H}_{\rm ov}$ can serve as a Hamiltonian realizing the needed interaction.

APPENDIX E: DERIVATION OF EQS. (114) and (115)

One notes from Eqs. (21) and (28) that

$$\hat{U}_\tau^\dagger \hat{\rho} \hat{U}_\tau = \frac{\exp[-\beta \hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau]}{Z} = \frac{e^{-\beta\hat{\Omega}-\beta\hat{H}}}{Z}, \tag{E1}$$

where we used the definition $\hat{\Omega} = \hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau - \hat{H}$ of $\hat{\Omega}$.

Note the standard equality

$$e^{-\beta\hat{\Omega}-\beta\hat{H}} = \vec{\exp}\left[-\int_0^\beta ds\, e^{-s\hat{H}} \hat{\Omega} e^{s\hat{H}} \right] e^{-\beta\hat{H}}, \tag{E2}$$

where $\vec{\exp}$ means a time-antiordered exponent. The easiest way to derive Eq. (E2) is to differentiate both sides of it with respect to β, and note that they both satisfy the same first-order differential equation and have the same boundary condition at $\beta=0$.

One now gets

$$\hat{U}_\tau^\dagger \hat{\rho} \hat{U}_\tau = \vec{\exp}\left[-\int_0^\beta ds\, e^{-s\hat{H}} \hat{\Omega} e^{s\hat{H}} \right] \hat{\rho}. \tag{E3}$$

Tracing out both sides, one finally obtains Eqs. (114) and (115).

In footnote 29 we used the Peierls-Bogoliubov inequality. The simplest way to derive this inequality from Eqs. (114) and (115) is to note the well-known extremal feature of the free energy:

$$-T \ln \mathrm{tr} e^{-\beta\hat{H}-\beta\hat{\Omega}} = \min\{\mathrm{tr}[\hat{\rho}(\hat{H}+\hat{\Omega})] + T\mathrm{tr}(\hat{\rho}\ln\hat{\rho})\},$$

where the minimization is taken over all possible density matrices. This can alternatively be written as

$$\mathrm{tr}\, e^{-\beta\hat{H}-\beta\hat{\Omega}} = \max \exp\{-\beta \mathrm{tr}[\hat{\rho}(\hat{H}+\hat{\Omega})] - \mathrm{tr}(\hat{\rho}\ln\hat{\rho})\}. \tag{E4}$$

The desired Peierls-Bogoliubov inequality is then obtained by inserting the particular density matrix $\hat{\rho} = e^{-\beta\hat{H}}/Z$ in the RHS of Eq. (E4).

APPENDIX F: PROOF OF HORN'S THEOREM

We intend to prove that given two vectors $x^{\rm T} = (x_1 \geqslant \cdots \geqslant x_n)$ and $y^{\rm T} = (y_1 \geqslant \cdots \geqslant y_n)$ with the following majorization relation (see Sec. IV B 1 for definitions):

$$x \prec y, \tag{F1}$$

there is a real orthogonal matrix $O = (O_{ij})$ such that

$$x_i = \sum_j O_{ij}^2 y_j \Leftrightarrow x = \mathrm{diag}[O\, \mathrm{diag}[y] O^{\rm T}]. \tag{F2}$$

Here $\mathrm{diag}[y]$ means the $n \times n$ diagonal matrix formed by the vector y, while $\mathrm{diag}[O\, \mathrm{diag}[y] O^{\rm T}]$ is the vector formed by diagonal elements of the matrix $O\, \mathrm{diag}[y] O^{\rm T}$, and $O^{\rm T}$ means transposition: $(O^{\rm T})_{kl} = O_{lk}$. Note that for any orthogonal matrix O_{ij}, the matrix O_{ij}^2 is always double stochastic: $\Sigma_i O_{ij}^2 = \Sigma_j O_{ij}^2 = 1$, though the converse is not true [67].

For a given orthogonal matrix O_{ij} there are many unitary matrices Q_{ij} such that $O_{ij}^2 = |Q_{ij}|^2$; e.g., $Q_{ij} = e^{i\phi_j} O_{ij}$, where ϕ_j are arbitrary phases.

The following proof is adopted from Ref. [70] and will be realized in two steps.

First, one shows that Eq. (F1) implies that

$$x = T_1 T_2 \cdots T_{n-1} y, \tag{F3}$$

where the matrices T are the so-called T transform defined as follows. Any T transform $T(m,l;t)$ has three parameters $m < l$ and t, where m and l are natural numbers between zero and n, and where $0 < t < 1$. Its action on any vector y, $y^{\rm T} = (y_1, \ldots, y_n)$, is defined as

$$z = T(m,l;t) y, \tag{F4}$$

where the elements y_l and y_m are mixed in a linear way:

$$z = (y_1, \ldots, y_{m-1}, ty_m + (1-t)y_l, y_{m+1}, \ldots, y_{l-1}, (1-t)y_m + ty_l, y_{l+1}, \ldots, y_n)^{\rm T}. \tag{F5}$$

To get the matrix of $T(m,l;t)$ starting from the $n \times n$ unity matrix $\hat{1}$, one proceeds as follows:

$$(\hat{1})_{mm} = 1 \to (T(m,l;t))_{mm} = t,$$

$$(\hat{1})_{ll} = 1 \to (T(m,l;t))_{ll} = t,$$

$$(\hat{1})_{ml} = 0 \to (T(m,l;t))_{ml} = 1-t,$$

$$(\hat{1})_{lm} = 0 \to (T(m,l;t))_{lm} = 1-t, \tag{F6}$$

while all other elements of the unity matrix are left unchanged.

Equation (F3) can now be proven by induction. It is obvious for $n=2$. Assume it holds for $n-1$. As seen from Eqs. (F1), (84), and (85), one has $y_n \leqslant x_1 \leqslant y_1$, so there exists an index k such that

$$y_k \leqslant x_1 \leqslant y_1. \tag{F7}$$

This implies

$$x_1 = ty_1 + (1-t)y_k \tag{F8}$$

for some $0 \leqslant t \leqslant 1$. Define a T transform $T(1,k;t)$ via

$$\begin{pmatrix} x_1 \\ \hline \overline{y} \end{pmatrix} = T(1,k;t) y, \tag{F9}$$

where

$$\overline{y}^{\rm T} = (y_2, \ldots, y_{k-1}, (1-t)y_1 + ty_k, y_{k+1}, \ldots, y_n). \tag{F10}$$

It is straighforward to show that

$$\overline{y}^{\rm T} \succ (x_2, \ldots, x_n). \tag{F11}$$

Since we assumed that the implication (F1) $\Rightarrow$ (F3) is valid for $n-1$, there is a product of T transforms such that $(x_2, \ldots, x_n) = T_2 \cdots T_{n-1} \overline{y}^{\rm T}$.

Thus this implication is proven by induction.

Let us finally prove the implication (F1) $\Rightarrow$ (F2). Note that with any T transform $T(m,l;t)$ one can associate an orthogonal matrix $V(m,l;t)$ by reshaping Eq. (F6) as follows:

$$(\hat{1})_{mm} = 1 \rightarrow (V(m,l;t))_{mm} = \sqrt{t},$$

$$(\hat{1})_{ll} = 1 \rightarrow (V(m,l;t))_{ll} = \sqrt{t},$$

$$(\hat{1})_{ml} = 0 \rightarrow (V(m,l;t))_{ml} = -\sqrt{1-t},$$

$$(\hat{1})_{lm} = 0 \rightarrow (V(m,l;t))_{lm} = \sqrt{1-t}. \tag{F12}$$

Then Eq. (F4) is equivalent to

$$z = \mathrm{diag}[V(m,l;t)\mathrm{diag}[y]V^{\mathrm{T}}(m,l;t)]. \tag{F13}$$

To prove the implication (F1) $\Rightarrow$ (F2) one again proceeds by induction. It is obviously valid for $n=2$. One assumes its validity for $n-1$. This means Eq. (F11) can be rewritten as

$$(x_2, \ldots, x_n)^{\mathrm{T}} = \mathrm{diag}[\tilde{V}\,\mathrm{diag}[\overline{y}]\tilde{V}^{\mathrm{T}}], \tag{F14}$$

where $\tilde{V}$ is some orthogonal matrix. To complete the proof, define an orthogonal matrix

$$O = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{V} \end{pmatrix} V, \tag{F15}$$

where the matrix V corresponds to the T transform T defined in Eq. (F9) [via the correspondence described in Eq. (F12)], and rewrite Eqs. (F11) and (F9) as

$$x = \mathrm{diag}[O\,\mathrm{diag}[\overline{y}]O^{\mathrm{T}}]. \tag{F16}$$

This proves the implication (F1) $\Rightarrow$ (F2).

Let us realize explicitly the construction given by Eqs. (F1) and (F2) for an example with $n=3$:

$$x = (0,0,0), \quad y = (2,1,-3). \tag{F17}$$

It is obvious that $x \prec y$. For the index k and for the parameter t mentioned before Eq. (F7) one has

$$k = 3, \quad t = \frac{3}{5}. \tag{F18}$$

Equation (F3) reads explicitly

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{3}{5} & 0 & \frac{2}{5} \\ 0 & 1 & 0 \\ \frac{2}{5} & 0 & \frac{3}{5} \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}. \tag{F19}$$

Finally the orthogonal matrix O in Eq. (F16) reads for the present example:

$$O = \begin{pmatrix} \sqrt{\frac{3}{5}} & 0 & -\sqrt{\frac{2}{5}} \\ -\sqrt{\frac{1}{5}} & \sqrt{\frac{1}{2}} & -\sqrt{\frac{3}{10}} \\ \sqrt{\frac{1}{5}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{3}{10}} \end{pmatrix}. \tag{F20}$$

APPENDIX G: DERIVATION OF EQ. (71)

Here we find the maximum of

$$\langle w^2 \rangle = \sum_{\alpha=1}^{N} \lambda_\alpha w_\alpha^2 = \sum_{\alpha=1}^{N} \frac{\langle \widetilde{\psi_\alpha} | \hat{\Omega} | \widetilde{\psi_\alpha} \rangle^2}{\langle \widetilde{\psi_\alpha} | \widetilde{\psi_\alpha} \rangle}, \tag{G1}$$

where the maximization is taken over all possible decompositions (34) of the mixed state $\hat{\rho}$ into pure states. Using Eqs. (55) and (57) one writes equivalently

$$\langle w^2 \rangle = \sum_{\alpha=1}^{N} \frac{[\mathrm{tr}(\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}\hat{\Pi}_\alpha)]^2}{\mathrm{tr}\,\hat{\rho}\hat{\Pi}_\alpha}, \tag{G2}$$

where

$$\hat{\Pi}_\alpha = |\pi_\alpha\rangle\langle\pi_\alpha|. \tag{G3}$$

The maximization in Eq. (G2) is taken over all decompositions of unity

$$\sum_{\alpha=1}^{N} \hat{\Pi}_\alpha = \hat{1}, \tag{G4}$$

where operators $\hat{\Pi}_\alpha$ exist in the n-dimensional Hilbert space $\mathcal{H}$.

The general idea of the following method was adopted from [76]. Introduce an operator $\hat{X}$ via

$$\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2} = \mathrm{Re}(\hat{\rho}\hat{X}) \equiv \frac{1}{2}(\hat{\rho}\hat{X} + \hat{X}\hat{\rho}); \tag{G5}$$

then

$$\mathrm{tr}(\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}\hat{\Pi}_\alpha) = \mathrm{Re}\,\mathrm{tr}(\hat{\Pi}_\alpha\hat{\rho}\hat{X}). \tag{G6}$$

Recall the Cauchy inequality

$$|\mathrm{tr}(\hat{A}\hat{B})|^2 \leqslant \mathrm{tr}(\hat{A}\hat{A}^\dagger)\mathrm{tr}(\hat{B}\hat{B}^\dagger), \tag{G7}$$

which is valid for any operators $\hat{A}$ and $\hat{B}$, with the equality being realized for

$$\hat{A} = \alpha\hat{B}^\dagger, \tag{G8}$$

where α is a number.

Applying first Eq. (G5) and then Eq. (G7) with $\hat{A} = \hat{\Pi}_\alpha^{1/2}\hat{\rho}^{1/2}$, $\hat{B} = \hat{\rho}^{1/2}\hat{X}\hat{\Pi}_\alpha^{1/2}$, one gets

$$[\mathrm{tr}(\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}\hat{\Pi}_\alpha)]^2 \equiv [\mathrm{Re}\,\mathrm{tr}(\hat{\Pi}_\alpha\hat{\rho}\hat{X})]^2 \leqslant |\mathrm{tr}(\hat{\Pi}_\alpha\hat{\rho}\hat{X})|^2 \tag{G9}$$

$$\equiv |\mathrm{tr}(\hat{\Pi}_\alpha^{1/2}\hat{\rho}^{1/2}\hat{\rho}^{1/2}\hat{X}\hat{\Pi}_\alpha^{1/2})|^2 \leqslant \mathrm{tr}(\hat{\Pi}_\alpha\hat{\rho})\mathrm{tr}(\hat{\rho}\hat{X}\hat{\Pi}_\alpha\hat{X}), \tag{G10}$$

one gets for Eqs. (G2) and (G3)

$$\sum_{\alpha=1}^{N} \frac{[\mathrm{tr}(\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}\hat{\Pi}_\alpha)]^2}{\mathrm{tr}\,\hat{\rho}\hat{\Pi}_\alpha} \leqslant \sum_{\alpha=1}^{N} \mathrm{tr}(\hat{\rho}\hat{X}\hat{\Pi}_\alpha\hat{X}) \tag{G11}$$

$$= \mathrm{tr}(\hat{\rho}\hat{X}^2) = \mathrm{tr}(\hat{\rho}^{1/2}\hat{\Omega}\hat{\rho}^{1/2}\hat{X}). \tag{G12}$$

Equation (G9) is realized as an equality for

$$\mathrm{Im\,tr}(\hat{\Pi}_\alpha \hat{\rho} \hat{X}) = 0, \tag{G13}$$

while the Cauchy inequality (G11) becomes an equality for

$$\hat{\rho}^{1/2} \hat{X} |\pi_\alpha\rangle = a_\alpha \hat{\rho}^{1/2} |\pi_\alpha\rangle, \tag{G14}$$

where a_α are some numbers.

Both conditions (G13) and (G14) are realized simultaneously by taking $\{|\pi_\alpha\rangle\}_{\alpha=1}^n$ and $\{a_\alpha\}_{\alpha=1}^n$ as, respectively, eigenvectors and eigenvalues of the Hermitian operator $\hat{X}$. The representation (73) for $\hat{X}$ follows from Eq. (G5). The desired equation (71) for the maximal work dispersion is obtained from Eqs. (G11) and (G12) (the substracted term W^2 is trivial).

APPENDIX H: DERIVATION OF EQ. (91)

Here we calculate the average $\{\langle w^2\rangle\}_{\mathrm{av}}$ of $\langle w^2\rangle$, given by Eq. (G1), over the measure (91). Using Eq. (50) it is straightforward to see that all the terms in the summation in the RHS of Eq. (G1) produce the same average. Thus,

$$\frac{\{\langle w^2\rangle\}_{\mathrm{av}}}{N} = \frac{\int \mathcal{D}M\ \delta\left[\sum_{i=1}^N |M_i|^2 - 1\right] \phi\{M_i\}}{\int \mathcal{D}M\ \delta\left[\sum_{i=1}^N |M_i|^2 - 1\right]}, \tag{H1}$$

where we denoted

$$\mathcal{D}M = \prod_{i=1}^N d\,\mathrm{Re}\,M_i d\,\mathrm{Im}\,M_i \tag{H2}$$

and where one notes from Eq. (50)

$$\phi\{M_i\} = \left| \sum_{j,k=1}^n M_j M_k^* \sqrt{p_j p_k} \langle \varepsilon_k | \hat{\Omega} | \varepsilon_j \rangle \right|^2. \tag{H3}$$

Passing to polar coordinates

$$\int \mathcal{D}M = \int_0^{2\pi} \prod_{i=1}^N d\varphi_i \int_0^\infty \prod_{i=1}^N |M_i| d|M_i| \tag{H4}$$

one gets

$$\frac{\{\langle w^2\rangle\}_{\mathrm{av}}}{N} = \sum_{j,k=1}^n p_j p_k \langle \varepsilon_j | \hat{\Omega} | \varepsilon_j \rangle \langle \varepsilon_k | \hat{\Omega} | \varepsilon_k \rangle \frac{I_{ij}}{I_0}, \tag{H5}$$

where

$$I_{jk} = \int_0^\infty \prod_{i=1}^N dz_i \delta\left[\sum_{i=1}^N z_i - 1\right] \frac{z_j z_k}{\sum_{l=1}^n p_l z_l}, \tag{H6}$$

$$I_0 = \int_0^\infty \prod_{i=1}^N dz_i \delta\left[\sum_{i=1}^N z_i - 1\right]. \tag{H7}$$

These integrals are calculated for $j,k=1,\ldots,n$ by the same method. For example,

$$e^{-r} r^N I_{jj} = \int_0^\infty \prod_{i=1}^N dy_i \delta\left[\sum_{i=1}^N y_i - r\right] \frac{y_j^2 e^{-r}}{\sum_{l=1}^n p_l y_l},$$

$$\Gamma(N+1) I_{jj} = \int_0^\infty \prod_{i=1}^N dy_i y_j^2 \frac{\exp\left(-\sum_{i=1}^N y_i\right)}{\sum_{l=1}^n p_l y_l}, \tag{H8}$$

$$= \int_0^\infty \prod_{l=1}^n dy_l y_j^2 \frac{\exp\left(-\sum_{l=1}^n y_l\right)}{\sum_{l=1}^n p_l y_l},$$

$$= \int_0^\infty ds \int_0^\infty \prod_{i=1}^N dy_i y_j^2\ e^{-\sum_{l=1}^n y_l (s p_l + 1)} \tag{H9}$$

where when passing from Eq. (H8) to Eq. (H9) we changed the integration variable $z_i = y_i / r$ and integrated over r from 0 to ∞.

Further calculations are straightforward and lead to Eq. (91). For dealing with this equation the following formula is useful:

$$\int_0^\infty ds \prod_{k=1}^n \frac{1}{\theta_m + s} = \sum_{k=1}^n \ln\,\theta_k \prod_{l\neq k}^n \frac{1}{\theta_l - \theta_k}, \tag{H10}$$

where the θ_k's are some positive numbers.

APPENDIX I: DERIVATION OF EQ. (107)

Here we give an example of the situation discussed around Eqs. (105)–(107). The effect announced there exists neither for two-level systems—simply because for a 2×2 traceless matrix $\hat{\Omega}$ a zero eigenvalue implies $\hat{\Omega}=0$—nor for three-level systems. The proof of this last fact requires some calculations which will be omitted.

The simplest situation that supports the effect is thus a four-level system. The following example was inspired by [77]. Consider a four-level system with Hamiltonian

$$\hat{H} = \begin{pmatrix} \hat{A} & 0 \\ 0 & \hat{B} \end{pmatrix}, \tag{I1}$$

where $\hat{A}$ and $\hat{B}$ are 2×2 matrices:

$$\hat{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \tag{I2}$$

with a, b, c, and d being some real numbers.

Assume that the unitary operator $\hat{U}_\tau$ is given as an exchange interaction:

$$\hat{U}_\tau = \begin{pmatrix} 0 & \hat{1} \\ \hat{1} & 0 \end{pmatrix}, \tag{I3}$$

where $\hat{1}$ is the 2×2 unit matrix.

The Hamiltonian $\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau$ in the Heisenberg representation at time τ then reads

$$\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau = \begin{pmatrix} \hat{B} & 0 \\ 0 & \hat{A} \end{pmatrix}. \tag{I4}$$

As follows from Eqs. (I1), (I2), and (I4), the matrix $\hat{\Omega}$,

$$\hat{\Omega} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & c-d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d-c \end{pmatrix}, \tag{I5}$$

has a doubly degenerate eigenvalue equal to zero, and the corresponding eigenvectors can be taken as

$$|0_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |0_2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}. \tag{I6}$$

It is now obvious that though

$$\langle 0_1|[\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau]^m|0_1\rangle - \langle 0_1|\hat{H}^m|0_1\rangle = 0 \quad \text{for } m = 1,2,$$

one still has

$$\langle 0_1|[\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau]^3|0_1\rangle - \langle 0_1|\hat{H}^3|0_1\rangle = b^2(c-d) \neq 0,$$

$$\langle 0_1|[\hat{U}_\tau^\dagger \hat{H} \hat{U}_\tau]^4|0_1\rangle - \langle 0_1|\hat{H}^4|0_1\rangle = 2ab^2(c-d) + b^2(c^2-d^2) \neq 0. \tag{I7}$$

These relations were used in Eq. (107).

[1] P. S. Epstein, *Textbook of Thermodynamics* (Wiley & Sons, New York, 1937).
[2] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, 1938).
[3] L. D. Landau and E. M. Lifshitz, *Statistical Physics I* (Pergamon Press, Oxford, 1978).
[4] R. Balian, *From Microphysics to Macrophysics* (Springer, Berlin, 1992), Vol. 1.
[5] F. Schlögl, *Probability and Heat* (Vieweg, Braunschweig, 1989).
[6] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (D. Reidel, Dordrecht, 1983).
[7] J. C. Maxwell, Nature (London) **17**, 257 (1878).
[8] H. S. Leff and A. F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Princeton University Press, Princeton, NJ, 1990).
[9] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, New Haven, 1914).
[10] G. M. Wang *et al.*, Phys. Rev. Lett. **89**, 050601 (2002).
[11] F. Ritort, Ann. Henri Poincare **2**, 63 (2003).
[12] A. E. Allahverdyan and Th. M. Nieuwenhuizen, e-print cond-mat/0207587.
[13] K. Sekimoto, F. Takagi, and T. Hondou, Phys. Rev. E **62**, 7759 (2000).
[14] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. Lett. **85**, 1799 (2000); J. Phys. A **36**, 875 (2003); Phys. Rev. E **64**, 056117 (2001).
[15] Th. M. Nieuwenhuizen and A. E. Allahverdyan, Phys. Rev. E **66**, 036102 (2002).
[16] G. N. Bochkov and Yu. E. Kuzovlev, Sov. Phys. JETP **45**, 125 (1977).
[17] G. N. Bochkov and Yu. E. Kuzovlev, Sov. Phys. JETP **49**, 543 (1979).
[18] G. N. Bochkov and Yu. E. Kuzovlev, Physica A **106**, 443 (1981); **106**, 480 (1981).
[19] I. M. Bassett, Phys. Rev. A **18**, 2356 (1978).
[20] W. Pusz and S. L. Woronowicz, Commun. Math. Phys. **58**, 273 (1978).
[21] A. Lenard, J. Stat. Phys. **19**, 575 (1978).
[22] W. Thirring, *A Course in Mathematical Physics 4: Quantum Mechanics of Large Systems* (Springer, Vienna, 1983).
[23] A. E. Allahverdyan and Th. M. Nieuwenhuizen, e-print cond-mat/0401548; e-print cond-mat/0408537.
[24] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[25] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000); **60**, 2721 (1999).
[26] C. Maes, Ann. Henri Poincare **2**, 29 (2003).
[27] S. Yukawa, J. Phys. Soc. Jpn. **69**, 2370 (2000).
[28] J. Kurchan, e-print cond-mat/0007360.
[29] H. Tasaki, e-print cond-mat/0009244.
[30] S. Mukamel, Phys. Rev. Lett. **90**, 170604 (2003).
[31] W. De Roeck and C. Maes, e-print cond-mat/0309498.
[32] T. Monnai and S. Tasaki, e-print cond-mat/0308337.
[33] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, Europhys. Lett. **66**, 419 (2004).
[34] W. M. de Muynck, *Foundations of Quantum Mechanics: An Empirist Approach* (Kluwer Academic, Dordrecht, 2002).
[35] A. Khrennikov, J. Math. Phys. **45**, 902 (2004); **44**, 2471 (2003); **43**, 789 (2002).
[36] P. Perrot, *A to Z of Thermodynamics* (Oxford University Press, London, 1998).
[37] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen (unpublished).
[38] G. Laufer, Am. J. Phys. **51**, 42 (1983).
[39] A. J. Mallinckrodt and H. S. Leff, Am. J. Phys. **60**, 356 (1992).
[40] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Physica A **305**, 542 (2002).
[41] E. G. D. Cohen and D. Mauzerall, J. Stat. Mech.: Theory Exp.

2004, P07006.
[42] C. Jarzynski, J. Stat. Mech.: Theory Exp. **2004**, P09005.
[43] A. Abragam and M. Goldman, *Nuclear Magnetism* (Clarendon, Oxford, 1982); A. Abragam, *Principles of Nuclear Magnetism* (Clarendon, Oxford, 1961).
[44] W. M. Elsasser, Phys. Rev. **52**, 987 (1937).
[45] J. L. Park, Am. J. Phys. **36**, 211 (1968); Philos. Sci. **35**, 205 (1968).
[46] B. D'Espagnat, *Veiled Reality* (Addison-Wesley, Reading, MA, 1995).
[47] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1995).
[48] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).
[49] O. Cohen, Phys. Rev. A **63**, 016102 (2001).
[50] D. R. Terno, Phys. Rev. A **63**, 016101 (2001).
[51] K. A. Kirkpatrick, e-print quant-ph/0405058; B. d'Espagnat, e-print quant-ph/0111081.
[52] W. N. Itano *et al.*, Phys. Rev. A **47**, 3554 (1993).
[53] G. C. Hegerfeldt, e-print quant-ph/9710027.
[54] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, e-print cond-mat/0408316.
[55] P. Kok and S. L. Braunstein, Phys. Rev. A **61**, 042304 (2000).
[56] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
[57] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
[58] E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936).
[59] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[60] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).
[61] A. Bassi and G. Ghirardi, Phys. Lett. A **309**, 24 (2003).
[62] S. E. Ahnert and M. C. Payne, e-print quant-ph/0408011.
[63] T. C. P. Chui *et al.*, Phys. Rev. Lett. **69**, 3005 (1992).
[64] B. Mandelbrot, Phys. Today **42**(1), 71 (1989).
[65] L. de la Peña and A. M. Cetto, *The Quantum Dice: An Introduction to Stochastic Electrodynamics* (Kluwer, Dordrecht, 1996).
[66] K. Molmer, Y. Castin, and J. Dalibard, J. Opt. Soc. Am. B **10**, 524 (1993).
[67] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic, New York, 1979).
[68] P. M. Alberti and A. Uhlmann, *Stochasticity and Partial Order* (Reidel, Dordrecht, 1982).
[69] R. Bhatia, *Matrix Analysis* (Springer, Berlin, 1997).
[70] T. Ando, Linear Algebr. Appl. **118**, 163 (1989).
[71] S. Golden, Phys. Rev. **137**, B1127 (1965); C. J. Thompson, J. Math. Phys. **6**, 1812 (1965).
[72] E. P. Wigner, Am. J. Phys. **31**, 6 (1963); C. W. Gardiner, *Quantum Noise* (Springer, Berlin, 1996).
[73] P. Bocchieri and A. Loinger, Phys. Rev. **114**, 948 (1959).
[74] H. Tasaki, Phys. Rev. Lett. **80**, 1373 (1998).
[75] D. S. Mitrinović, *Analytical Inequalities* (Springer, Berlin, 1970).
[76] C. A. Fuchs and C. M. Caves, Open Syst. Inf. Dyn. **3**, 1 (1995); e-print quant-ph/9604001.
[77] M. Ozawa, e-print quant-ph/0310072.
[78] A. E. Seigman, *Introduction to Lasers and Masers* (McGraw-Hill, New York, 1971).