Minimal work principle: Proof and counterexamples

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*

The minimal work principle states that work done on a thermally isolated equilibrium system is minimal for adiabatically slow (reversible) realization of a given process. This principle, one of the formulations of the second law, is studied here for finite (possibly large) quantum systems interacting with macroscopic sources of work. It is shown to be valid as long as the adiabatic energy levels do not cross. If level crossing does occur, counterexamples are discussed, showing that the minimal work principle can be violated and that optimal processes are neither adiabatically slow nor reversible. The results are corroborated by an exactly solvable model.

I. INTRODUCTION

The second law of thermodynamics [1–4], formulated nearly one and a half century ago, continues to be under scrutiny [5–9]. While its status within equilibrium thermodynamics and statistical physics is by now well settled [1–4], its fate in various border situations is far from being clear [5,7–9]. In the macroscopic realm the second law is a set of equivalent statements concerning quantities such as entropy, heat, work, etc. In more general situations these statements need not be equivalent and some, e.g., those involving entropy, may have only a limited applicability [7,9]. In contrast to entropy, the concept of work has a well-defined operational meaning for an arbitrary system interacting with macroscopic work sources [1–3]. Moreover, the definition of work is not confined to (nearly) equilibrium situations: it is defined equally well for processes that start and end in arbitrary nonequilibrium states [1–3].

In this perspective the second law can be regarded as a set of several statements concerning the work. They were first derived from observations and have direct counterparts in everyday experience. Two formulations of the second law are at the focus of our present interest [1–4,10–19]:

(1) Thomson's formulation: no work can be extracted from an equilibrium system by means of a cyclic process generated by an external work source.

(2) The minimal work principle: when varying the speed of a given process done on an (initially) equilibrium system, the work is minimal for the slowest realization of the process (more details of the principle are given below).

These two statements are well defined—both operationally and conceptually—for a finite system coupled to macroscopic sources of work. It is therefore interesting to consider whether they are valid in this domain. This is an inquiry towards one of the most pertinent questions in the foundations of statistical physics: whether the second law is valid for finite systems, or, put differently, whether the thermodynamical limit is really necessary for the validity of the second law. Some historical developments of this question are recalled in the Appendix.

Thomson's formulation was derived with great rigor from the principles of quantum mechanics [10–13]. It is valid not only for a macroscopic, but also for a finite system coupled to sources of work. This fact is not occasional: would Thomson's formulation be violated for a finite (e.g., few-level) system, it would immediately imply violations for large collections of such systems, i.e., for some macroscopic systems.

There are also several pertinent derivations of the minimal work principle. They cover some important but still particular cases, such as the linear response limit (weak coupling to the work source) [16,17], quasi-slow processes [18], and constant temperature processes for macroscopic systems [12,15,19,18]. So far all these derivations were confirming the validity of the minimal work principle.

Our present purpose is to study the minimal work principle for a finite (possibly large) quantum system coupled to sources of work. We derive the principle for finite systems and then discuss its limits. It is found that—in contrast to Thomson's formulation—the domain of the validity of the minimal work principle is large but definitely limited. These limits are connected with crossing of adiabatic energy levels, and they are illustrated via counterexamples which include an exactly solvable model. A more detailed discussion on what, to our opinion, qualifies as a limit of the second law, is given in the Appendix.

The phenomenon of level-crossing (conical intersections), which was once little more than a theoretical curiosity, recently attracted much attention in chemical physics, quantum chemistry, and biophysics; see Refs. [20,21] for a review. Its importance was recognized for such essential processes as charge transfer reactions, light harvesting, and ultrafast decay of excited states. We thus show here that the same phenomenon of level crossing is crucial for the proper understanding of the second law. It is important to note that for the typical spectra level crossings are much more frequent than the avoided crossings, if the number of parameters varied under interaction with the work sources is larger than two [22].

The paper is organized as follows. In Sec. II we recall the basic setup of the problem. Section III discusses the minimal work principle and its various implications. Then the principle is derived for finite systems. Section V studies the limits of the principle, while Sec. VI discusses relations of the principle to Thomson's formulation of the second law and to

cyclic processes. The last section presents our conclusions.

II. THE SETUP

A. Thermally isolated processes

Consider a quantum system S which is thermally isolated [1,2,4]: it moves according to its own dynamics and interacts with an external macroscopic work source. This interaction is realized via time dependence of some parameters $R(t) = \{R_1(t), R_2(t), \ldots\}$ of the system's Hamiltonian $H(t) = H\{R(t)\}$. They move along a certain trajectory $R(t)$ which at some initial time t_i starts from $R_i = R(t_i)$, and ends at $R_f = R(t_f)$. The initial and final values of the Hamiltonian are

$$H_i = H\{R_i\} \quad \text{and} \quad H_f = H\{R_f\}, \tag{1}$$

respectively.

The Hamiltonian $H(t)$ generates a unitary evolution:

$$i\hbar \frac{d}{dt}\rho(t) = [H(t),\rho(t)], \quad \rho(t) = U(t)\rho(t_i)U^\dagger(t), \tag{2}$$

with time-ordered

$$U(t) = \widetilde{\exp}\left(-\frac{i}{\hbar}\int_{t_i}^t ds\, H(s)\right). \tag{3}$$

The interaction of S with the source is related with flows of work which qualifies as a high-graded (ordered or mechanical) type of energy. The work W done on S reads [1–4]

$$W = \int_{t_i}^{t_f} dt\, \text{tr}[\rho(t)\dot{H}(t)] = \text{tr}[H_f\rho(t_f)] - \text{tr}[H_i\rho(t_i)], \tag{4}$$

where we performed partial integration and inserted (2). This is the average energy increase of S, which, due to energy conservation, coincides with the average energy decrease of the source [1,2]. Thus there are two ways to measure work. One can let the work source interact with an ensemble of systems S and then to measure the energy change of the work-source. Due to the macroscopic size of the latter, its energy practically coincides with its average [1,2]. Or, alternatively, one can measure directly the initial and final average[1] energies of S, e.g., via measuring the Hamiltonians H_i and H_f, respectively, on the ensemble of systems S.

B. Initial state

Initially the system S is assumed to be in equilibrium at temperature $T = 1/\beta \geq 0$, that is, S is described by a Gibbsian density operator:

$$\rho(t_i) = \frac{1}{Z_i}\exp(-\beta H_i), \quad Z_i = \text{tr}\, e^{-\beta H_i}. \tag{5}$$

This equilibrium state can be prepared by a weak interaction between S and a macroscopic thermal bath at tempera-

---

[1]Normally this averaging is done either by letting many identically prepared systems interact with the work source, or operating with a single system but repreparing its state after each interaction period.

ture T [1–3,9], and then decoupling S from the bath in order to achieve a thermally isolated process [1–4]. The fact of preparing S via a thermal bath has the following implications:

(1) Reproducibility: The preparation process leading to $\rho(t_i)$ can, in principle, be repeated the needed amount of times, and thus various effects displayed by S and described by density matrix ρ can be amplified to the needed extent. Moreover, this reproduction is in principle not connected with energy costs: due to the weak coupling to the bath no (or little) energy should be spent for switching the coupling on and off.

(2) According to quantum mechanics, $\rho(t_i)$ allows us to predict probabilities for various results obtained via measurements done on S. In this sense $\rho(t_i)$ can be thought to describe an ensemble of identically prepared systems S, rather than one preparation of a single system. Since it was prepared via a thermal bath only—i.e., without additional measurements employed to separate the ensemble into subensembles—no physical interpretation should be based on a choice of subensembles for $\rho(t_i)$. In particular, it will be incorrect to interpret $\rho(t_i)$ as "S is with some probability in a state with a definite but unknown energy (eigenvalue of H_i)." This interpretation is allowed for a classical Gibbs distribution, but in the quantum situation it may lead to inconsistencies with experiment; see, e.g., in Ref. [23]. One of the reasons prohibiting such an interpretation is that separation of a mixed-state ensemble into pure-state components (subensembles) depends on the concrete measurement done for this purpose and therefore is not unique.

III. THE MINIMAL WORK PRINCIPLE

A. Formulation of the principle

Let S start in the state (5), and let the parameters $R(t)$ move between R_i and R_f along a trajectory $R(t)$. The work done on S during this process is W. Consider the adiabatically slow realization of this process: R proceeds between the same values R_i and R_f and along the same trajectory, but now with a homogeneously vanishing speed, thereby taking a very long time $t_f - t_i$, at the cost of an amount of work $\widetilde{W}$. The minimal work principle then asserts [1,2,4]

$$W \geq \widetilde{W}. \tag{6}$$

This is a statement on optimality: if work has to be extracted from S, W is negative, and to make it as negative as possible one proceeds with a very slow speed. If during some operation work has to be added ($W > 0$) to S, one wishes to minimize its amount, and again operates slowly.

The following remarks are intended to clarify the physical meaning of the principle and to prevent its improper use.

(a) For thermally isolated systems, adiabatically slow processes are reversible. This is standard if S is macroscopic [1,2,4], and below in Sec. VI it is shown to hold for a finite S as well, where the definition of reversibility extends unambiguously (i.e., without invoking entropy) [4].

(b) The formulation of the minimal work principle does not by itself give any detailed information on the precise

meaning of being "slow." Only its derivation from first principles can be informative in this respect, and this is one of the reasons why such derivations are really needed. It suffices to say at the present moment that "slow" means "slower than the characteristic times of the system relevant for calculating work," and certainly does not mean "slower than characteristic times of the system S without any interaction with external sources of work." Indeed, we shall see below that, in general, the relevant characteristic times become determined only once the very process $R(t)$—with its initial and final points R_i and R_f—is given. The operational definition of slow processes in the context of the minimal work principle is straightforward: one increases the characteristic time scale τ of the process until the work as a function of τ saturates, i.e., it no longer depends on τ. It is conceivable that for some systems there will be several types of slow processes, that is, the work as a function of τ will display several plateaus. If so, then each type of slow process can be studied for its own sake [24–26].

(c) Note that as far as the work—i.e., average energy lost by the work source—is concerned, any processes can be considered as a part of a thermally isolated one, provided S is defined to be the whole system which interacts with the source of work and which together with the latter forms a closed system[2] [see additionally in this context Eq. (18) and the remark after it]. Thus, the statement of the minimal work principle is more general than it may seem.

(d) Finally, it may be useful to comment on the used nomenclature. Following Ref. [1] we call a slow thermally isolated process adiabatic. Now and then for sharpening the needed context we use equivalently the term "adiabatically slow." Note that this by no means presupposes validity of the quantum mechanical adiabatic theorem.

B. The minimal work principle for macroscopic systems

In macroscopic thermodynamics the minimal work principle is derived [1,4] from certain axioms which ensure that, within the domain of their applicability, this principle is equivalent to other formulations of the second law. Derivations in the context of statistical thermodynamics are presented in Refs. [9,15–19].

In the following discussion we will reproduce a proof of the minimal work principle for a class of thermally isolated processes realized on a macroscopic system. Our purpose is to understand why precisely the principle holds in this situation, and thereby to motivate its investigation for finite systems.

The derivation proceeds in two steps. First one considers the relative entropy,

$$S[\rho(t_f)\|\rho_{eq}(H_f)] = \mathrm{tr}[\rho(t_f)\ln \rho(t_f) - \rho(t_f)\ln \rho_{eq}(H_f)], \quad (7)$$

between the final state $\rho(t_f)$ given by (2) and an equilibrium state

$$\rho_{eq}(H_f) = \exp(-\beta H_f)/Z_f, \quad Z_f = \mathrm{tr}\, e^{-\beta H_f}, \quad (8)$$

which corresponds to the final Hamiltonian H_f and to the same temperature $T = 1/\beta$ as in the initial state $\rho(t_i)$.

The relative entropy $S[\rho\|\sigma]$ is known to be non-negative for any density matrices ρ and σ. Among other useful features, it can serve as a (nonsymmetric) "distance" between ρ and σ, since $S[\rho\|\sigma] = 0$ implies $\rho = \sigma$. Applications of relative entropy in statistical physics and quantum information theory are reviewed in Refs. [2,27].

As follows from the unitarity of the evolution operator (2),

$$S_f = -\mathrm{tr}[\rho(t_f)\ln \rho(t_f)] = -\mathrm{tr}[\rho(t_i)\ln \rho(t_i)] = S_i. \quad (9)$$

This is the well-known conservation of Von Neumann entropy during a thermally isolated process.

The definition of relative entropy (7), combined with Eqs. (9), (8), and (5) and with the definition of work (4) brings

$$S[\rho(t_f)\|\rho_{eq}(H_f)] = \mathrm{tr}[\rho(t_i)\ln \rho(t_i) - \rho(t_f)\ln \rho_{eq}(H_f)]$$
$$= \beta W + \ln Z_f - \ln Z_i. \quad (10)$$

Using the non-negativity of the relative entropy (7) one gets

$$W = F(H_f) - F(H_i) + TS[\rho(t_f)\|\rho_{eq}(H_f)] \quad (11)$$

$$\geq F(H_f) - F(H_i), \quad (12)$$

where $F(H)$ is the free energies corresponding to Hamiltonian H and temperature T:

$$F(H) \equiv -T \ln \mathrm{tr}\, e^{-\beta H}. \quad (13)$$

This inequality is well known [1,2] (though it is derived and formulated less explicitly) and sometimes is viewed as a proof of the minimal work principle.[3] However, we still have to show that the work $\widetilde{W}$ in the slow limit coincides with the difference of free energies given by the right-hand side (RHS) of Eq. (12). Recall that the latter quantity was so far defined only formally.

Equation (12) gets the needed physical meaning when one *assumes* that for a macroscopic system S, the final state $\widetilde{\rho}(t_f)$—reached from $\rho(t_i) = \exp(-\beta H_i)/Z_i$ by the adiabatically slow process—can be approximated by $\rho_{eq}(H_f)$ defined in Eq. (8):

$$S[\widetilde{\rho}(t_f)\|\rho_{eq}(H_f)] \approx 0, \quad (14)$$

where $\approx$ means that the equality is supposed to recover in the macroscopic (thermodynamic) limit for S. This then leads from Eq. (11) to the needed relation

$$\widetilde{W} \approx F(H_f) - F(H_i). \quad (15)$$

The statement (6) of the minimal work principle then follows from Eqs. (9) and (15).

---

[2] This is akin to the known statement that any dynamical evolution in quantum mechanics can be viewed as a part of a unitary evolution, provided all the environment of the system was included into the description. Such statements are sometimes regarded as too general and therefore useless. However, they can be very useful as instanced by the formalism of completely positive operations for open quantum systems [3].

---

[3] Recently inequality (12) was generalized within so-called work-fluctuations theorems [13,28]. These theorems account for fluctuations which appear when measuring the work via the system's energy.

In the second step of the derivation one should prove Eq. (15). There are several classes of macroscopic systems for which one can show that the free energy difference $F(H_f) - F(H_i)$ in (12) indeed coincides with the adiabatic work [9,15,19]. We recall below one of them.

Assume that S consists of two parts: macroscopic thermal bath B and a subsystem (particle) P coupled to it:

$$H(t) = H_P\{R(t)\} + H_B + gH_I, \qquad (16)$$

where $H_P(t)$ and H_B stand for the Hamiltonians of P and B, respectively, and where H_I is the interaction Hamiltonian with g being the corresponding coupling constant. The source of work interacts with the particle only, thus only $H_P\{R(t)\}$ is time dependent.

The following conditions are usually considered to be sufficient for the validity of Eq. (15); see [1,2]. Along the lines of Ref. [9], we present them in a slightly more formalized and particular way, and we recall that they were checked in models [9,19].

(i) The thermal bath B is composed of a macroscopic number of harmonic oscillators with a proper (e.g., ohmic) spectrum of their interaction with P.

(ii) The evolution generated by $H(t)$ in (16) starts from an equilibrium state (5) for the total system $S = B + P$.

(iii) The characteristic time of the external process $R(t)$ is assumed to be much larger than the relevant times of the particle P. These times include those generated by the Hamiltonian $H_P\{R(t)\}$, as well as the relaxation times of P induced by the bath. The latter times are controlled by the interaction Hamiltonian H_I and they become very long for $g \to 0$. Note that since the thermal bath B is assumed to be macroscopic (dense spectrum), there are characteristic times of B (so-called Heisenberg times) which are proportional to inverse level-spacing and are thus very large. They, however, do not enter into the definition of adiabatically slow processes. All the relevant characteristic times are finite in the thermo-dynamical limit for the bath, at least for the type of models considered in [9].

The above three conditions are sufficient for the density operator of P to be given as [9]

$$\rho_P(t) = \frac{1}{Z(t)} \text{tr}_B e^{-\beta H(t)}, \quad Z(t) = \text{tr}\, e^{-\beta H(t)}, \qquad (17)$$

where $\text{tr}_{B(P)}$ means trace over the bath (particle) degrees of freedom. The physical meaning of (17) is obvious: in the slow limit P is in the local equilibrium (or local stationary) state.[4]

Note that the state of P at time t need not at all coincide with the local Gibbsian $e^{-\beta H_P(t)}/Z_P(t)$, since no weak coupling assumption on the system-bath interaction was made.

Equations (4), (16), and (17) suffice to derive relation (15) between the work $\widetilde{W}$ done for the adiabatically slow process and the difference $F(H_f) - F(H_i)$ of free energies. Indeed, note from Eq. (16) that $\partial_t H(t) = \partial_t H_P(t)$ and (4) can be written as

$$W = \int_{t_i}^{t_f} dt\ \text{tr}_P[\rho_P(t)\partial_t H_P(t)]. \qquad (18)$$

This relation shows that the work defined globally via the energy difference of the overall system of the particle and the bath can also be calculated via integration of the expression on the RHS of Eq. (18), which contains only quantities referring to the particle P.

Now proceed with application of Eq. (17):

$$W = \int_{t_i}^{t_f} \frac{dt}{Z(t)} \text{tr}_P \text{tr}_B[e^{-\beta H(t)}\partial_t H_P(t)] = \int_{t_i}^{t_f} \frac{dt}{Z(t)} \text{tr}[e^{-\beta H(t)}\partial_t H(t)]$$

$$= -T\int_{t_i}^{t_f} dt \frac{d}{dt} \ln Z(t) = F(H_f) - F(H_i). \qquad (19)$$

Thus the minimal work principle (6) is proved for this class of processes realized on macroscopic systems. One may note once again that although the quantities $F(H_f)$ and $F(H_i)$ are very large in the macroscopic limit for the bath, their difference is finite and is of the order of the particle's energy. This is also seen from (18) which contains only quantities referring to the particle.

IV. FINITE SYSTEMS

Let us now turn to a finite N-level quantum system S.[5] The first thing to do is to apply here the reasoning developed above for macroscopic systems. In fact, Thomson's formulation of the second law, $W \geq 0$, is valid for finite systems for precisely the same reasons as it is valid for macroscopic ones. Indeed, since this formulation refers to cyclic processes, one puts $H_f = H_i$ in Eqs. (7) and (12)—this can be done without altering the validity of (12), since ρ_{eq} is in general an auxiliary density matrix—and gets $W \geq 0$ for cyclic processes, which is the statement of Thomson's formulation.

However, when trying to apply the above reasoning for finite systems, one immediately sees that there are no reasons why the work in the slow limit should be equal to the difference in free energies. Nor are there reasons to expect the validity of Eq. (14), that is, to expect that in general the final

---

[4]For the actual calculation in Eq. (17), one may need to keep the bath large but finite, to carry out the trace tr_B, and only then to go to the macroscopic limit for the bath. Then the density matrix $\rho_P(t)$ is finite and well-behaved although the quantities like $Z(t) = \text{tr}\, e^{-\beta H(t)}$ may not be well-defined in the macroscopic limit for the bath. This is a standard procedure in statistical physics and it is legitimate for the present situation, since the thermodynamical limit for the bath commutes with the limit of show processes, as we already recalled above. The validity of such commutation was also seen in Ref. [18].

[5]Note that an attempt was made recently to study the minimal work principle for finite systems [14]. The author of Ref. [14] obtained that the principle has the same range of validity as Thomson's formulation of the second law. To our opinion this is incorrect. It is, however, to be mentioned that the author of Ref. [14] has stressed the preliminary character of his results, and that his results concerning cyclic processes are correct.

state $\tilde{\rho}(t_f)$ reached from $\rho(t_i) = \exp(-\beta H_i)/Z_i$ during the adiabatically slow process should be equal to $\rho_{eq}(H_f)$ defined by (8). The reason for this can be seen by noting that due to the unitarity of the evolution operator U in (3), the spectrum of $\tilde{\rho}(t_f)$ coincides with that of the initial density matrix $\rho(t_i)$, and in general does not have the Gibbsian shape of the spectrum of $\rho_{eq}(H_f)$.

The facts that $\tilde{\rho}(t_f) \neq \rho_{eq}(H_f)$ and $\tilde{W} \neq F(H_f) - F(H_i)$ for finite and certain macroscopic systems are known and were studied especially in the linear response regime [29–37]. Here are some results obtained within that activity:

(1) A class of finite systems was determined for which the adiabatic work coincides with the difference in free energies [29,30].

(2) Conditions were determined for macroscopic systems, under which the adiabatic work converges to the free energy difference in the thermodynamical limit [30,31,35]. Concrete estimates for the rate of convergence were given.

(3) It was found that for certain nontrivial macroscopic systems the above convergence can be absent even in the thermodynamical limit [30,32]. It was argued that quantum systems are more vulnerable in this respect than classical ones [30].

(4) The fact that $\tilde{\rho}(t_f) \neq \rho_{eq}(H_f)$ was recently applied for studying certain processes which are reversible in the macroscopic limit, but become irreversible for finite system [36].

Derivation of the minimal work principle

Thus we cannot rely on macroscopic analogies and we need an independent derivation of the minimal work principle (6) for finite systems. Some ideas of the following derivation were adopted from Ref. [11].

Let the spectral resolution of $H(t)$ and $\rho(t_i)$ be[6]

$$H(t) = \sum_{k=1}^{N} \varepsilon_k(t)|k,t\rangle\langle k,t|, \quad \langle k,t|n,t\rangle = \delta_{kn}, \qquad (20)$$

$$\rho(t_i) = \sum_{k=1}^{N} p_k|k,t_i\rangle\langle k,t_i|, \quad p_k = \frac{e^{-\beta\varepsilon_k(t_i)}}{\sum_n e^{-\beta\varepsilon_n(t_i)}}. \qquad (21)$$

At $t=t_i$ we order the spectrum as

$$\varepsilon_1(t_i) \leq \cdots \leq \varepsilon_N(t_i) \qquad (22)$$

$$\Rightarrow p_1 \geq \cdots \geq p_N. \qquad (23)$$

For any t in the interval $t_i \leq t \leq t_f$ we expand over the complete set $|n,t\rangle$:

$$U(t)|k,t_i\rangle = \sum_{n=1}^{N} a_{kn}(t)\exp\left(-\frac{i}{\hbar}\int_{t_i}^{t} dt'\,\varepsilon_n(t')\right)|n,t\rangle, \quad (24)$$

where

---

[6]The eigenstates of $\rho(t_i)$ are used below as a calculational tool. It is by no means implied that "what really happens" is that the system S is—with some probability and before doing any measurement—in one of those states. See also the second remark in Sec. II B.

$$a_{kn}(t) \equiv a_{kn}(t;t_i) = \langle n,t|U(t)|k,t_i\rangle\exp\left(\frac{i}{\hbar}\int_{t_i}^{t} dt'\,\varepsilon_n(t')\right) \qquad (25)$$

are the expansion coefficients, and where $|a_{kn}(t_f)|^2$ is the probability to measure energy $\varepsilon_n(t_f)$ at the final time t_f, provided the initial state was $|k,t_i\rangle\langle k,t_i|$.

Now we use Eqs. (21) and (24) to obtain for the work (4):

$$W = \sum_{k,n=1}^{N} |a_{kn}(t_f)|^2 p_k \varepsilon_n(t_f) - \sum_{k=1}^{N} p_k \varepsilon_k(t_i). \qquad (26)$$

A similar formula can be derived to express the adiabatic work $\tilde{W}$ in terms of coefficients $\tilde{a}_{kn}(t_f)$:

$$\tilde{W} = \sum_{k,n=1}^{N} |\tilde{a}_{kn}(t_f)|^2 p_k \varepsilon_n(t_f) - \sum_{k=1}^{N} p_k \varepsilon_k(t_i). \qquad (27)$$

From the definition (25),

$$|a_{kn}(t_f)|^2 = |\langle n,t_f|U|k,t_i\rangle|^2, \qquad (28)$$

it follows that[7]

$$\sum_{k=1}^{N} |a_{kn}(t_f)|^2 = \sum_{n=1}^{N} |a_{kn}(t_f)|^2 = 1. \qquad (29)$$

Employing the identity (summation by parts)

$$\sum_{n=1}^{N} \varepsilon_n x_n = \varepsilon_N \sum_{n=1}^{N} x_n - \sum_{m=1}^{N-1} [\varepsilon_{m+1} - \varepsilon_m] \sum_{n=1}^{m} x_n, \qquad (30)$$

with $x_n = \sum_{k=1}^{N} |a_{kn}(t_f)|^2 p_k$ and $x_n = \sum_{k=1}^{N} |\tilde{a}_{kn}(t_f)|^2 p_k$, we obtain from Eqs. (26) and (27) and, using Eqs. (29), we obtain a general formula for the difference between the non-adiabatic and adiabatic work:

$$W - \tilde{W} = \sum_{m=1}^{N-1} [\varepsilon_{m+1}(t_f) - \varepsilon_m(t_f)]\Theta_m, \qquad (31)$$

$$\Theta_m \equiv \sum_{n=1}^{m} \sum_{k=1}^{N} p_k[|\tilde{a}_{kn}(t_f)|^2 - |a_{kn}(t_f)|^2]. \qquad (32)$$

Let us now assume that the ordering (22) is kept at the final time $t=t_f$:

$$\varepsilon_1(t_f) \leq \cdots \leq \varepsilon_N(t_f). \qquad (33)$$

If different energy levels did not cross each other—i.e., different ones do not become equal and equal ones do not become different—Eq. (33) is implied by the initial ordering (22).

---

[7]As possible physical interpretation of feature (29), note that for the uniform distribution of the initial states, $p(k,t_i) = 1/N$, the prediction probability $p(n;t_f|k,t_i) = |a_{kn}(t_f)|^2$ is equal to the retrodiction probability $p(k,t_i|n;t_f) = [p(k,t_i)p(n;t_f|k,t_i)]/[\sum_{k=1}^{N} p(k,t_i)p(n;t_f|k,t_i)] = [(1/N)p(n;t_f|k,t_i)]/[\sum_{k=1}^{N} 1/Np(n;t_f|k,t_i)] = p(n;t_f|k,t_i).$

The behavior of energy levels with respect to level-crossing is governed by the noncrossing rule, which was frequently discussed in the literature and derived in a rather general situation [38–41]. We shall need it in the following particular formulation:

If $H\{R\}$ is real and only one of its parameters is varied with time, (33) is satisfied for any discrete-level quantum system: level crossing, even if it happens in model-dependent calculations or due to approximate symmetry, does not survive arbitrary small perturbation where it is substituted by avoided crossing (for a more general $H\{R\}$ the conditions prohibiting level-crossing are more restrictive; see [40]). To get a stable point of level crossing, one needs at least two independently varying parameters for a real $H\{R\}$.

The rule is known since the early days of quantum mechanics [38] and is presented in textbooks; see, e.g., Ref. [39]. It is an important tool in atomic and molecular spectroscopy, and development of these fields led people to reconsider its derivation. In particular, Ref. [42] correctly criticizes the standard proof of the no-crossing rule for being imprecise and not general enough. In response to this and several related criticisms, (at least) two complete and general derivations appeared which settled the issue [40,41].

No level crossings and natural conditions of smoothness of $H(t)$ are *sufficient* for the standard quantum adiabatic theorem [43] to ensure

$$\tilde{a}_{kn}(t_{\mathrm{f}}) = \delta_{kn}. \tag{34}$$

This is the known statement on the absence of transitions for slow variations.[8]

Combined with Eqs. (23) and (34), the definition (32) brings

$$\Theta_m = \sum_{k=1}^{m} p_k \left(1 - \sum_{n=1}^{m} |a_{kn}(t_{\mathrm{f}})|^2 \right) - \sum_{n=1}^{m} \sum_{k=m+1}^{N} p_k |a_{kn}(t_{\mathrm{f}})|^2$$

$$\geq -p_m \left(\sum_{k=1}^{m} \sum_{n=1}^{m} |a_{kn}(t_{\mathrm{f}})|^2 + \sum_{n=1}^{m} \sum_{k=m+1}^{N} |a_{kn}(t_{\mathrm{f}})|^2 \right) + p_m m$$

$$= p_m(m - m) = 0. \tag{35}$$

----

[8]This standard statement of the adiabatic theorem was elaborated in literature several times so as to provide information on the internal characteristic times $\mathcal{T}$ [44]. To be slow then will mean $\tau \gg \mathcal{T}$. Dealing with a discrete spectrum and assuming, as we did above, that there are no level-crossings and that adiabatic energy levels are smooth functions of time, $\mathcal{T}$ can roughly be estimated via the inverse of the minimal spacing between the involved energy levels [44]. In this context, the levels which provide a small spacing are said to define avoided crossing. One should, however, keep in mind that this estimation cannot be extrapolated to situations with level crossing, where the minimal level spacing is zero. In those cases the whole situation changes and the characteristic internal times $\mathcal{T}$ can well be finite; see below and in Refs. [45,46] (compare also with our discussion in Sec. III A). The same is the case for some situations in a continuous spectrum, where again the minimal level-spacing is zero, but the relevant characteristic times can be finite; see, e.g., [47,48]. In short, the inverse of the minimal level-spacing defines the relevant characteristic time only for some particular cases, not in general.

Once Θ_m are non-negative, the statement $W \geq \tilde{W}$ of the minimal work principle follows from Eqs. (26), (31), and (33).

Recall once again the basic ingredients of the proof:

(1) The condition (29) which came from the unitarity of the evolution during a thermally isolated process,

(2) The no-crossing assumption which led to Eq. (33) and simultaneously—via the standard adiabatic theorem—to Eq. (34).

(3) We did not assume that the adiabatic work is either related or equal to the difference in free energies.

(4) Only two features of the initial state $\rho(t_{\mathrm{i}})$ were used: $[H_{\mathrm{i}}, \rho(t_{\mathrm{i}})] = 0$ determined the specific form (26) and (27) of the work, while the order relation (23) for the eigenvalues of the initial density matrix was used in proving $\Theta_m \geq 0$ in Eq. (35). This means that the Gibbsian property (5) is not necessary for the validity of the derivation.

As an immediate application of the obtained results note that if only one parameter is varied, we are ensured of the absence of level crossings and the minimal work principle is valid. Among many examples of this situation there is the case of a gas—which may consist of any number of particles—in a rectangular container interacting with one of its walls moving in time. This is the standard setup for a one-parameter thermally isolated process. Recently it was studied for some simple quantum systems [6,34].

V. LEVEL CROSSING

The above noncrossing condition raises the question: Is the minimal work principle also valid if the adiabatic energy levels cross? Before addressing this question in detail, let us mention some popular misconceptions which surround the level crossing problem:

The no-crossing rule is said to exclude all crossings. This is incorrect as the exclusion concerns situations where, in particular, only one independent parameter of a real Hamiltonian $H\{R\}$ is varied [40]. Two parameters can produce robust level crossing for such Hamiltonians.

One accepts that level crossing can happen, but believes it to be very rare and thus irrelevant for any sensible physical situation. This opinion is invalidated by whole chapters of chemical physics [20,21]: level crossing is not only observed in many- (at least two) atom molecules, where internuclear distances play the role of classical time-dependent parameters, but is necessary for the proper description of some known chemical reactions, as well as for predicting new ones. Over the years several methods were developed for identifying and locating the points of level crossing.[9] In particular, the method based on the geometrical phase ("Berry phase") allows one to deduce the existence of level crossing from the behavior of the system at points remote from the crossing; see Refs. [20] for more information.

Note as well that level crossing is a more frequent phenomenon than avoided crossing [22], if the number of inde-

----

[9]This is an important issue, since some numerical or approximate analytical methods may easily miss points of level crossing or may produce spurious ones.

pendently varying parameters is larger than two.

It is believed that once levels can cross, $\Delta\varepsilon\to 0$, the very point of defining slow processes disappears as the internal characteristic time $\hbar/\Delta\varepsilon$ of S is infinite. This view misidentifies the proper internal time as seen below (see also the discussion in Sec. III A and footnote 8). The absence of level crossings is indeed a sufficient condition for the validity of Eq. (34) (no transitions for slow changes). It is, however, by no means necessary [45,46,49].

It is sometimes believed that crossing is automatically followed by a population inversion. We shall find no support for that.

A. A two-level example within adiabatic perturbation theory

As a first example we consider a spin-1/2 particle with Hamiltonian

$$H(s) = h_1(s)\sigma_1 - h_3(s)\sigma_3, \tag{36}$$

$$= \begin{pmatrix} -h_3(s) & h_1(s) \\ h_1(s) & h_3(s) \end{pmatrix}, \quad s = t/\tau, \tag{37}$$

where σ_1, σ_3, and $\sigma_2 = i\sigma_1\sigma_3$ are Pauli matrices, and where s is the reduced time with τ being the characteristic time scale. The magnetic fields h_1 and h_3 smoothly vary in time.

Let us now make the following assumptions: (*i*) for $s \to s_i < 0$ and for $s \to s_f > 0$, $h_1(s)$ and $h_3(s)$ go to constant values sufficiently fast: (*ii*) At $s = t = 0$ both $h_1(s)$ and $h_3(s)$ turn to zero. This condition indicates a level crossing, since crossed eigenvalues of the traceless 2×2 matrix $H(s)$ mean that this matrix is equal to zero at the point of the crossing. Besides these two basic conditions, we shall assume few auxiliary ones, whose purpose is to make the discussion below more transparent. (*iii*) $h_1(s)$ and $h_3(s)$ are nonzero for all s, $s_i \leq s \leq s_f$, except $s = 0$. (*iv*) Due to the latter condition $h_3(s)/h_1(s)$ is finite for the involved s, except possibly $s \to 0$. If it is infinite in this limit, then obviously its inverse $h_1(s)/h_3(s)$ is finite; now for all s, $s_i \leq s \leq s_f$, and goes to zero for $s \to 0$. Since one of these ratios has to be finite, we shall assume that this finite ratio is $h_1(s)/h_3(s)$. (*v*) $h_3(s) > 0$ for $s < 0$ and $h_3(s) < 0$ for $s > 0$.

Here is a concrete example realizing the last two assumptions. For $s \to 0$,

$$h_1(s) \simeq \alpha_1 s^2, \quad h_3(s) \simeq -\alpha_3 s, \tag{38}$$

where α_1 and α_3 are positive constants.

With the above conditions on $h_1(s)$ and $h_3(s)$, one can propose a useful parametrization of the Hamiltonian (37). Recalling that $h_3(s)$ changes its sign at $s = 0$, Eq. (37) is written as

$$H(s) = \mathrm{sgn}(s)\sqrt{h_3^2(s) + h_1^2(s)}\begin{pmatrix} \cos\theta(s) & \sin\theta(s) \\ \sin\theta(s) & -\cos\theta(s) \end{pmatrix}, \tag{39}$$

where

$$\sin\theta(s) \equiv \mathrm{sgn}(s)\frac{h_1(s)}{\sqrt{h_3^2(s) + h_1^2(s)}}, \tag{40}$$

$$\cos\theta(s) \equiv -\mathrm{sgn}(s)\frac{h_3(s)}{\sqrt{h_3^2(s) + h_1^2(s)}}. \tag{41}$$

Here $\sqrt{\cdots}$ is defined to be always positive, $\mathrm{sgn}(s)$ is the sign function and

$$\theta(s) = -\arctan\left(\frac{h_1(s)}{h_3(s)}\right) \tag{42}$$

is a parameter in the interval $-\pi/2 < \theta(s) < \pi/2$. Note that the presence of $\mathrm{sgn}(s)$ in the above expressions is necessary for having a smooth parametrization. Otherwise, e.g., $h_3(s)/\sqrt{h_3^2(s) + h_1^2(s)}$ is discontinuous at the crossing point $s = 0$.

The eigenvalues and eigenvectors of $H(s)$,

$$H(s)|k,s\rangle = \varepsilon_k(s)|k,s\rangle, \quad k = 1,2, \tag{43}$$

are now read off by inspecting Eq. (40):

$$\varepsilon_1(s) = \mathrm{sgn}(s)\sqrt{h_3^2(s) + h_1^2(s)}, \quad \varepsilon_2(s) = -\varepsilon_1(s), \tag{44}$$

$$|1,s\rangle = \begin{pmatrix} \cos\frac{1}{2}\theta(s) \\ \sin\frac{1}{2}\theta(s) \end{pmatrix}, \quad |2,s\rangle = \begin{pmatrix} -\sin\frac{1}{2}\theta(s) \\ \cos\frac{1}{2}\theta(s) \end{pmatrix}. \tag{45}$$

It is seen that the eigenvalues $\varepsilon_1(s)$ and $\varepsilon_2(s)$ of $H(s)$ (adiabatic energy levels) cross at $s = \theta(s) = 0$. Note from the above conditions on $h_1(s)$ and $h_2(s)$ and from Eqs. (42) and (44) that both the eigenvalues and the eigenvectors of $H(s)$ are smooth functions of s. This fact is important for subsequent calculations, but first of all it is necessary for the very definition of level crossing.

Now Eq. (22) for the ordering of initial energy levels is valid, but Eq. (33) for the same ordering of final energy levels is not valid due to the level crossing at $s = 0$. Equations (26)–(32) imply

$$W - \tilde{W} = -2\sqrt{h_1^2(s_f) + h_3^2(s_f)}\,\Theta_1, \quad \tau s_f = t_f, \tag{46}$$

where Θ_1 is defined by (32).

Our strategy is now to confirm relation (34) in the slow limit $\tau\to\infty$ and thus to confirm that $\Theta_1 > 0$, implying that the minimal work principle is indeed violated.

To this end we apply the standard quantum mechanical adiabatic perturbation theory [43]. Substituting Eq. (24) into Eq. (2) one has

$$\dot{a}_{kn} = -\sum_{m=1}^{N} a_{km}(t)\exp\left(\frac{i}{\hbar}\int_{t_i}^{t}dt'[\varepsilon_n(t') - \varepsilon_m(t')]\right)\langle n,s|\partial_t|m,s\rangle. \tag{47}$$

As $|1,s\rangle$ and $|2,s\rangle$ in Eq. (45) are real, the normalization $\langle n,s|n,s\rangle = 1$ implies $\langle n,s|\partial_t|n,s\rangle = 0$. Since $\langle n,s|\partial_t|m,s\rangle = (1/\tau)\langle n,s|\partial_s|m,s\rangle$ the RHS of (47) contains a small parameter $1/\tau$. It is therefore more convenient to introduce new variables: $a_{kn}(t) = \delta_{kn} + b_{kn}(t)$, $b_{kn}(t_i) = 0$. To leading order in

$1/\tau$, b_{kn} can be neglected on the RHS of (47), and, after changing variables as $s\tau=t$, $s'\tau=t'$, one gets for $a_{k\neq n}(t_f)$ $=b_{k\neq n}(t_f)$

$$|a_{k\neq n}(t_f)|^2 = \left| \int_{s_i}^{s_f} ds \exp\left(\frac{i\tau}{\hbar} \int_{s_i}^{s} du[\varepsilon_k(u) - \varepsilon_n(u)] \right) \langle n|\partial_s|k\rangle \right|^2,$$

(48)

while, due to normalization,

$$|a_{kk}(t_f)|^2 = 1 - \sum_{n\neq k} |a_{kn}(t_f)|^2.$$

(49)

For our model described by Eqs. (36)–(45), the quantity

$$\int_{s_i}^{s} du[\varepsilon_1(u) - \varepsilon_2(u)]$$

(50)

$$= 2 \int_{s_i}^{s} du\varepsilon_1(u)$$

(51)

has only one extremal point, at $s=0$. We also have from (45)

$$\langle 2,s|\partial_s|1,s\rangle = \frac{\theta'(s)}{2} = \frac{1}{2} \frac{h_1(s)h_3'(s) - h_3(s)h_1'(s)}{h_3^2(s) + h_1^2(s)}, \quad (52)$$

where $\theta'(s) \equiv d\theta/ds$.

For large τ the integral in Eq. (48) can be calculated with use of the saddle-point method:

$$|a_{12}(t_f)|^2 = \left. \frac{\pi\hbar}{\tau} \left[\frac{\langle 2,s|\partial_s|1,s\rangle^2 \sqrt{h_1^2(s) + h_3^2(s)}}{h_1(s)h_1'(s) + h_3(s)h_3'(s)} \right] \right|_{s=0}.$$

(53)

Substituting Eq. (38) into Eq. (53), one gets

$$|a_{12}(t_f)|^2 = \frac{\pi\hbar\alpha_1^2}{4\tau\alpha_3^3}.$$

(54)

Equations (48), (53), and (54) extend the statement of the adiabatic theorem (34) for the level crossing situation. More general versions of similar adiabatic theorems can be found in Refs. [45,46]. Inserting

$$\Theta_1 = (p_1 - p_2)|a_{12}(t_f)|^2 > 0$$

(55)

in Eq. (46) confirms the violation of the minimal work principle. Equation (54) also shows that for the considered process the role of the proper internal characteristic time is played by $\hbar\alpha_1^2/\alpha_3^3$ rather than by $\hbar/(\varepsilon_1-\varepsilon_2)$.

Let us recall once again that the violation of the minimal work principle is due to common influence of the following factors:

(i) There is a level crossing: a more populated state goes to higher energies, while a less populated one goes to lower energies.

(ii) For slow processes there are no transitions between various energy levels: $\tilde{a}_{kn}(t_f) = \delta_{kn}$.

(iii) For not very slow processes there do occur transitions: $a_{kn}(t_f) \neq \delta_{kn}$. They cost less work.

Note as well that when the external fields $h_1(s)$ and $h_3(s)$ in Eq. (36) are such that $\sqrt{h_3^2(s) + h_1^2(s)}$ is a smooth function

for all real s, there are no crossings of eigenvalues and (6) is valid. In the example of level crossing given above, one has $\sqrt{h_3^2(s) + h_1^2(s)} \propto |s|$ for small s. Note that if level crossing is absent, the transition probability $|a_{12}(t_f)|^2$ is small as $O(e^{-\tau})$, since the integral in Eq. (48) is determined by the extremal point of (50) which is now complex; see [44] for more details.

B. Fast processes

One can calculate $|a_{kn}(t_f)|^2$ in yet another limiting case, where the characteristic time τ is very short, while the change in $h_1(t)$ and $h_3(t)$ is finite and there is the level crossing at $t=0$.

It is well known [43] that in this limit energy changes can be calculated with the help of the frozen initial state of S. For the present situation this leads from Eq. (45) to

$$|a_{12}(t_f)|^2 = |a_{21}(t_f)|^2 = |\langle 1,t_f|2,t_i\rangle|^2 = \sin^2 \tfrac{1}{2}[\theta(t_f) - \theta(t_i)],$$

(56)

and thus to

$$\Theta_1 = (p_1 - p_2)\sin^2 \tfrac{1}{2}[\theta(t_f) - \theta(t_i)],$$

(57)

which is again positive. This demonstrates that even very fast processes (i.e., $\tau \to 0$) can be more optimal than slow ones. It is conceivable that violations of the minimal work principle are maximal for some finite τ. This expectation is confirmed below.

C. Exactly solvable model with level crossing

The above results obtained by perturbation theory will now be corroborated on an exactly solvable displaying level crossing.

Consider a two-level system with Hamiltonian

$$H(s) = \hbar\omega \begin{pmatrix} s\cos^2 s & \frac{1}{2}s\sin 2s \\ \frac{1}{2}s\sin 2s & s\sin^2 s \end{pmatrix}, \quad s = \frac{t}{\tau}, \quad (58)$$

where τ is the characteristic time scale, s is the reduced time, and where ω is a coupling constant. The model belongs to the pool of exactly solvable driven two-level systems. We learned about it from Ref. [45], where some asymptotic features of its solution were studied.

The eigenvectors

$$|1,s\rangle = \begin{pmatrix} \cos s \\ \sin s \end{pmatrix}, \quad |2,s\rangle = \begin{pmatrix} -\sin s \\ \cos s \end{pmatrix}$$

(59)

of $H(s)$ correspond to the eigenvalues

$$\varepsilon_1(s) = \hbar\omega s, \quad \varepsilon_2(s) = 0,$$

(60)

respectively. It is seen that the energy levels cross when $s = t/\tau$ passes through zero. Equations (47) read for the present case

$$\frac{da_{11}(s)}{ds} = e^{i\omega\tau s^2/2}a_{12},$$

(61)

$$\frac{\mathrm{d}a_{12}(s)}{\mathrm{d}s} = -e^{-i\omega\tau s^2/2}a_{11}. \tag{62}$$

These equations can be solved exactly in terms of hypergeometric functions

$$a_{11}(s) = c_1 H_e\left(-\frac{i}{\omega\tau}, s\chi\right) + c_2 F_1\left(\frac{i}{2\omega\tau}, \frac{1}{2}, s^2\chi^2\right), \tag{63}$$

$$a_{12}(s) = \frac{c_1}{\chi^2}e^{i\pi/4 - s^2\chi^2}H_e\left(-1 - \frac{i}{\omega\tau}, s\chi\right) - c_2 s e^{-s^2\chi^2}F_1$$
$$\times\left(1 + \frac{i}{2\omega\tau}, \frac{3}{2}, s^2\chi^2\right), \tag{64}$$

where H_e and $F_1 \equiv {}_1F_1$ are, respectively, Hermite and hypergeometrical functions [50], and where

$$\chi \equiv \exp\left(\frac{i\pi}{4}\right)\sqrt{\frac{\omega\tau}{2}}. \tag{65}$$

The integration constants c_1 and c_2 are determined from the initial conditions $a_{11}(s_i)=1$ and $a_{12}(s_i)=0$.

In the slow limit $\omega\tau \gg 1$, the transition probability can be calculated via the first-order adiabatic perturbation theory result (48):

$$a_{12}(s_f) = -\int_{s_i}^{s_f} \mathrm{d}s\, e^{-i\omega\tau s^2/2}. \tag{66}$$

With help of the saddle-point method one gets for (66)

$$|a_{12}(s_f)|^2 = \frac{2\pi}{\omega\tau}, \tag{67}$$

in agreement with (54).[10] The characteristic time for the slow process is seen to be $1/\omega$.

Recalling Eq. (29), one deduces from Eqs. (31) and (32) for the present model:

$$W - \widetilde{W} = \hbar\omega s_f |a_{12}(s_f)|^2 \tanh\left(\tfrac{1}{2}\beta\hbar\omega s_i\right) \tag{68}$$

To have level crossing we take $s_i < 0$, $s_f > 0$. Equation (68) then indeed predicts

$$W - \widetilde{W} < 0. \tag{69}$$

It is seen that violations of the minimal work principle exist for $s_f > 0$, and they are maximal for $|a_{12}(s_f)|^2 \to 1$. This is seen to be the case in Figs. 1 and 2 for some τ near $\tau = 1$. Note from Fig. 1 that both the first-order perturbation theory result (66) and the saddle-point approximation to it given by Eq. (67) are adequately reproducing $|a_{12}(s_f)|^2$ for $\tau \gtrsim 10$. Moreover, the first-order perturbation theory is seen to provide an upper bound for the exact expression and predicts the appearance of large oscillations around $\tau \sim 1$.

---

[10]To make a detailed comparison between Eqs. (67) and (54), substract from $H(s)$ [defined by (58)] an irrelevant factor $\frac{1}{2}\mathrm{tr}H(s)$, and note from Eqs. (36) and (38) that the following relations are valid for the present model: $\alpha_3 = \frac{1}{2}\hbar\omega$, $\alpha_1 = \hbar\omega$. Substituting them into Eq. (54) one gets Eq. (67).
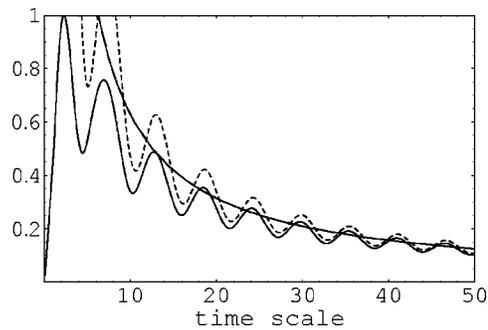

FIG. 1. Amplitude $|a_{12}(s_f)|^2$ versus the time scale τ for $s_i = -1.5$, $s_f = 1.5$, and $\omega = 1$. Full oscillating curve: the exact value which can reach unity. Dotted curve: result from a first-order adiabatic perturbation theory. The smooth curve presents a saddle-point approximation (67).

For $\tau \to 0$, $|a_{12}(s_f)|^2$ goes to its value $\sin^2(s_f - s_i)$ predicted by the sudden perturbation theory. This amounts to $|a_{12}(s_f)|^2 = 0.019\,91$ for the situations in Figs. 1 and 2.

Thus, all the basic conclusions drawn from the adiabatic perturbation theory are confirmed by this exactly solvable model.

D. Many-level systems

Let S have a finite number of levels, and assume that two of them cross. For quasi-adiabatic processes (τ is large but finite) and analytically varying Hamiltonian $H(t)$, the transition probability between noncrossing levels is exponentially small [44], while, as we saw, it has power-law smallness for two crossing levels. Then one neglects in (26) the factors $|a_{k\neq n}(t_f)|^2$ coming from any noncrossed levels k and n, and the problem is reduced to the two-level situation. Thus already one crossing suffices to detect limits of the minimal work principle (provided, of course, that the crossed levels are sufficiently populated initially).

The reduction to the two-level situation takes place also in a macroscopic system which has few discrete levels located below a continuous spectrum and separated from the latter by a finite gap. It is known, see, e.g., in Ref. [46], that the transitions between these discrete levels and the continuous part of the spectrum vanish exponentially for quasi-adiabatic
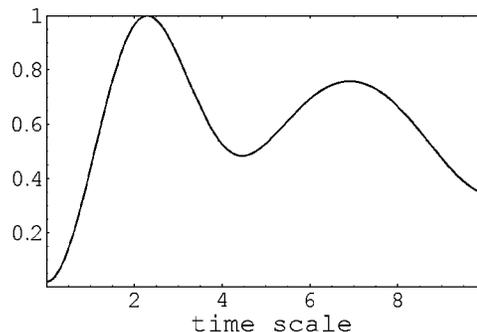

FIG. 2. The magnified version of Fig. 1: The exact value of $|a_{12}(s_f)|^2$ versus the characteristic time scale τ for $s_i = -1.5$, $s_f = 1.5$, and $\omega = 1$.

processes. These discrete levels thus decouple from the rest of the spectrum and the problem returns to an effectively two-level situation.

VI. CYCLIC PROCESSES AND REVERSIBILITY

The above results on the limits of the minimal work principle do not imply any violation of the second law in Thomson's formulation [10–12]: no work is extracted from S during a cyclic process, $W_c \geq 0$. We illustrate its general proof obtained in Refs. [10–13] in the context of the level crossing model given by Eqs. (36)–(45).

Assume that the trajectory $R(t) = (h_1(t), h_2(t))$ described there is supplemented by another trajectory $R'(t)$ which brings the parameters back to their initial values $(h_1(t_i), h_3(t_i))$ so that the overall process $R + R'$ is cyclic. If R' crosses the adiabatic energy levels backwards, then at the final time t'_f of the full cyclic process $R + R'$ one has

$$\varepsilon_1(t'_f) = \varepsilon_1(t_i), \quad \varepsilon_2(t'_f) = \varepsilon_2(t_i). \quad (70)$$

Together with

$$|a_{12}(t'_f)|^2 = |a_{21}(t'_f)|^2 = 1 - |a_{11}(t'_f)|^2 = 1 - |a_{12}(t'_f)|^2,$$

and Eqs. (26) and (34) as applied to the full cyclic process (i.e., changing in them $t_f \to t'_f$), Eq. (70) implies

$$W_c = |a_{12}(t'_f)|^2 (p_1 - p_2)[\varepsilon_2(t_i) - \varepsilon_1(t_i)] \geq \widetilde{W}_c = 0. \quad (71)$$

Equation (71) confirms the intuitive expectation that non-adiabatic processes are less optimal. In particular, this is valid if $R' = R_{mr}$ is exactly the same process R moved backwards with the same speed (mirror reflection). Then $\widetilde{W}_c = 0$ means that the original process R is reversible: According to the standard thermodynamical definition [1,2,4], a process R is reversible, if, after supplementing it with its mirror reflection process R_{mr}, the work done for the total cyclic process $R + R_{mr}$ is zero.

If R' does not induce another level crossing, i.e., if $h_1(s)$ and $h_2(s)$ in Eq. (36) do return to their initial values without simultaneously crossing zero, then the levels are interchanged,

$$\varepsilon_1(t'_f) = \varepsilon_2(t_i), \quad \varepsilon_2(t'_f) = \varepsilon_1(t_i), \quad (72)$$

and this time Eqs. (26) and (34) imply

$$\widetilde{W}_c = (p_1 - p_2)[\varepsilon_2(t_i) - \varepsilon_1(t_i)], \quad (73)$$

$$\widetilde{W}_c \geq W_c = [1 - |a_{12}(t'_f)|^2]\widetilde{W}_c > 0. \quad (74)$$

It is seen that in contrast to the situation described by Eq. (71), non-adiabatic processes are more optimal if $R + R'$ contains one level-crossing (or an odd number of them). We thus have found here a violation of the minimal work principle for a cyclic process.

VII. CONCLUSIONS

This paper was devoted to one of the persistent questions in statistical physics: whether the second law is valid for finite systems, and, if it is not valid, what are the possible scenarios of its violation. The proper way to answer this question is to take a formulation of the second law which has a clear meaning for finite systems, and to study it from the first principles of quantum mechanics, without invoking any postulate.

Along these lines we have studied the minimal work principle for finite systems coupled to external sources of work. The principle states that the work done on an (initially) equilibrium system during a thermally isolated process is minimal for the smallest speed of the process. As compared to other formulations of the second law, this principle has a direct practical meaning as it provides a recipe for reducing energy costs of various processes. We gave its general proof for finite systems which starts from first principles of quantum mechanics and avoids the usual lore associated with the second law (chaos, thermodynamic limit, coarse graining, various definitions of entropy). We have also shown that it may become limited if there are crossings of adiabatic energy levels: optimal processes need to be neither slow nor reversible. Already one crossing can suffice to note violations of the principle. If this is the case, the optimal process occurs at some finite, system-dependent speed.

Level crossing was observed, e.g., in molecular and chemical physics [20,21]. It is not a rare effect [22]: if the number of externally varied parameters is larger than two, then for typical spectra level crossings are much more frequent than avoided crossings [22].

Together with the universal validity of Thomson's formulation of the second law [10–13], the limits of the principle imply that the very equivalence between various formulations of the second law may be broken for a finite system coupled to macroscopic sources of work: different formulations are based on different physical mechanisms and have different ranges of validity. Similar results on nonequivalence of various formulations of the second law were found in Refs. [7,9], where for a quantum particle coupled to a macroscopic thermal bath, it was shown that some formulations, e.g., the Clausius inequality and positivity of the energy dispersion rate, are satisfied at sufficiently high temperatures of the bath, but can be invalid at low temperatures, that is, in the quantum regime. The physical mechanism responsible for this is the formation of a cloud of bath modes around the particle, well known in the cases of Kondo-cloud and polaron-cloud, but more general.

There are still many issues to consider before the minimal work principle and its limits can be said to be properly understood. In particular, we need a better description for the transition from finite to macroscopic systems. It might also be of interest to find an explicit example of a macroscopic system displaying limits of the principle. A separate problem is to study the minimal work principle in the (semi)classical limit of finite quantum systems. This problem is special due to the fact that the limit of slow processes need not commute with the classical limit [51].

We, however, had first to understand that limits of the minimal work principle do exist in principle and sometimes in practice, and this is the main message of the present paper.

APPENDIX: LIMITS TO THE SECOND LAW

As was stressed in the Introduction, one of basic purposes of this paper was to understand the limits of the minimal work principle, which is a particular formulation of the second law. We therefore consider it necessary to clarify what in general we mean by "limits to the second law."

It is first of all important that one is dealing with a concrete formulation of the second law, which, within the studied situation, has a clear conceptual and operational meaning. For example, Thomson's formulation of the second law as applied to finite systems is certainly of that type, and so is the minimal work principle. Both these formulations of the second law operate with the concept of work, which—in contrast to entropy—is defined unambiguously for finite systems at all times.

If within the studied situation we find the statement not satisfied—e.g., the minimal work principle need not be satisfied in the presence of level crossings—then we encountered a limit of that particular formulation. Note that for finite systems it will not be legitimate to speak about limits of any entropic statement of the second law, since entropy for a finite system does not have a clear and well-accepted physical meaning. One can define here various entropies (especially in the quantum situation) and it is not clear with respect to which one the corresponding statements of second law are to be formulated. Here one encounters limits of the concept of entropy rather than limits of the second law. It is only in the macroscopic situation that the coarse-grained entropy becomes a meaningful quantity on sufficiently long time scales.

In the context of limits of the second law there is an opinion that fluctuations of various quantities—e.g., fluctuations of work, provided they can be sensibly defined—provide violations of the second law. This opinion is supported by an observation that for a typical system in the thermodynamical limit the fluctuations vanish and, for example, Thomson's formulation in this limit is a statement on energy difference of a macroscopic system [i.e., in the definition (4) of work one need not take averages over different systems or different realizations]. According to this opinion, the second law is formulated as a statement on the random quantity work, and once the latter can fluctuate for finite systems assuming sometimes negative values, this gives violations of the second law.

It is true that when the second law was first deduced in the 19th century, it was formulated for a single closed system, in a way resembling the laws of ordinary mechanics. However, already in the beginning of 20th century it was clearly understood [52–55] that this law has only 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 when the first robust observations of fluctuations were made [53]. Together with theoretical works of Boltzmann in 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 micro-physics through averaging over fluctuations. A detailed summary of this activity is presented in the 1937 book by Epstein [53], while Tolman in 1938 [54] and Kemble in 1939 [55] discuss theoretical aspects of the situation. Since then, this understanding of the second law entered into several modern books of statistical physics and thermodynamics [1,2].

It is also true that when the statistical character of the second law was not yet widely accepted, several known scientists made statements on violations of the second law by fluctuations [56]. One of these citations is by Maxwell, "the second law is drawn from our experience of bodies consisting of immense number of molecules….it is continually being violated in any sufficiently small group of molecules. As the number…is increased…the probability of a measurable variation may be regarded as practically an impossibility." If one cites this quotation, one should keep in mind that Thomson's formulation of the second law is perfectly valid for any "small group of molecules." As we saw in the present paper, the same—modulo level-crossing—concerns the minimal work principle.

In summary, fluctuations do not provide violations of the second law, since this law is formulated with respect to averages. For a recent discussion on this point, see [57].

---

[1] L. D. Landau and E. M. Lifshitz, *Statistical Physics, I* (Pergamon, Oxford, 1978).

[2] R. Balian, *From Microphysics to Macrophysics* (Springer, New York, 1992) Vol. 1.

[3] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (D. Reidel, Dordrecht, 1983).

[4] P. Perrot, *A to Z of Thermodynamics* (Oxford University Press, Oxford, 1998).

[5] *Quantum Limits to the Second Law*, edited by D. P. Sheehan, AIP Conf. Proc. Vol. 643 (AIP, New York, 2002).

[6] V. V. Mityugov, Usp. Fiz. Nauk **170**, 681 (2000).

[7] 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).

[8] J. Gemmer, A. Otte, and G. Mahler, Phys. Rev. Lett. **86**, 1927 (2001).

[9] Th. M. Nieuwenhuizen and A. E. Allahverdyan, Phys. Rev. E **66**, 036102 (2002).

[10] G. N. Bochkov and Yu. E. Kuzovlev, Sov. Phys. JETP **45**, 125 (1977); W. Pusz and S. L. Woronowicz, Commun. Math. Phys.

58, 273 (1978).

[11] A. Lenard, J. Stat. Phys. **19**, 575 (1978); A. E. Allahverdyan and Th. M. Nieuwenhuizen, Physica A **305**, 542 (2002).

[12] I. M. Bassett, Phys. Rev. A **18**, 2356 (1978).

[13] J. Kurchan, e-print cond-mat/0007360.

[14] H. Tasaki, e-print cond-mat/0009206.

[15] H. Narnhofer and W. Thirring, Phys. Rev. A **26**, 3646 (1982).

[16] L. P. Kadanoff and P. C. Martin, Ann. Phys. (N.Y.) **24**, 419 (1963).

[17] D. Forster, *Hydrodynamics, Broken Symmetry, and Correlation Functions* (Benjamin, New York, 1983), see especially Sec. 2.8.

[18] R. Fukuda, Prog. Theor. Phys. **77**, 825 (1987).

[19] K. Sekimoto and S. Sasa, J. Phys. Soc. Jpn. **66**, 3326 (1997).

[20] D. R. Yarkony, Rev. Mod. Phys. **68**, 985 (1996); J. Phys. Chem. A **105**, 6277 (2001).

[21] H. Koppel, Chem. Phys. **77**, 359 (1983); H. D. Meyer, *ibid.* **82**, 199 (1983); C. A. Mead and D. G. Truhlar, J. Chem. Phys. **70**, 2284 (1979); C. A. Mead, Rev. Mod. Phys. **64**, 51 (1992).

[22] D. G. Truhlar and A. C. Mead, Phys. Rev. A **68**, 032501 (2003).

[23] P. Kok and S. L. Braunstein, Phys. Rev. A **61**, 042304 (2000).

[24] Th. M. Nieuwenhuizen, Phys. Rev. E **61**, 267 (2000).

[25] A. E. Allahverdyan and Th. M. Nieuwenhuizen, Phys. Rev. E **62**, 845 (2000).

[26] M. Kawasaki, Phys. Rev. E **65**, 046145 (2002).

[27] V. Vedral, Rev. Mod. Phys. **74**, 197 (2002).

[28] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[29] M. J. Klein, Phys. Rev. **86**, 807 (1952).

[30] L. Rosenfeld, Physica (Amsterdam) **27**, 67 (1961).

[31] R. D. Mountain, Physica (Amsterdam) **30**, 808 (1964).

[32] R. M. Wilcox, Phys. Rev. **174**, 624 (1968).

[33] A. A. Samokhin, Phys. Lett. **36A**, 372 (1971).

[34] S. N. Andreev, A. A. Rukhadze, and A. A. Samokhin, Condens. Matter Phys. **7**, 451 (2004).

[35] S. V. Mashkevich and V. S. Mashkevich, Phys. Rev. E **51**, 245 (1995).

[36] K. Sato, K. Sekimoto, T. Hondou, and F. Takagi, Phys. Rev. E **66**, 016119 (2002); H. Tasaki, e-print cond-mat/0008420.

[37] A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen, Europhys. Lett. **67**, 565 (2004).

[38] J. von Neumann and E. Wigner, Z. Phys. **30**, 467 (1929).

[39] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1958).

[40] H. C. Longuet-Higgins, Proc. R. Soc. London, Ser. A **344**, 147 (1975).

[41] C. A. Mead, J. Chem. Phys. **70**, 2276 (1979).

[42] G. J. Hatton, Phys. Rev. A **14**, 901 (1976); K. Razi Naqvi, *ibid.* **15**, 1807 (1977); G. J. Hatton, *ibid.* **16**, 1347 (1977).

[43] A. Galindo and P. Pascual, *Quantum Mechanics* (Springer-Verlag, New York, 1991); A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam, 1962).

[44] J. Hwang and P. Pechukas, J. Chem. Phys. **67**, 4640 (1977); Y. N. Demkov *et al.*, Phys. Rev. A **18**, 2089 (1978); M. Wilkinson and M. Morgan, Phys. Rev. A **61**, 062104 (2000).

[45] G. A. Hagedorn, Ann. Phys. (N.Y.) **196**, 278 (1989).

[46] J. E. Avron and A. Elgart, Commun. Math. Phys. **203**, 445 (1999); Phys. Rev. A **58**, 4300 (1998).

[47] J. E. Avron, J. S. Howland, and B. Simon, Commun. Math. Phys. **128**, 497 (1990).

[48] A. Elgart and J. H. Schenker, Rev. Math. Phys. **14**, 569 (2002).

[49] M. Born and V. Fock, Z. Phys. **51**, 165 (1928).

[50] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965); A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953).

[51] M. V. Berry, J. Phys. A **17**, 1225 (1984).

[52] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (New Haven, 1914).

[53] P. S. Epstein, *Textbook of Thermodynamics* (Wiley, New York, 1937).

[54] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, 1938).

[55] E. C. Kemble, Phys. Rev. **56**, 1013 (1939).

[56] We learned on this citation from a recent survey of some historical developments in statistical mechanics: J. L. Lebowitz, Rev. Mod. Phys. **71**, 346 (1999).

[57] A. E. Allahverdyan and Th. M. Nieuwenhuizen, e-print cond-mat/0207587.