

Introduction to quantum statistical thermodynamics

Armen E. Allahverdyan

Yerevan Physics Institute, Alikhanian Brothers Street 2, Yerevan 375036, Armenia

These lectures are inteneded for graduate students with a basic knowledge of quantum mechanics. No deep knowledge of statistical mechanics is assumed. The first purpose is to show how several basic results of statistical mechanics can be derived from within quantum mechanics. The second purpose is dual to the first one: I want to demonstrate how several conceptual problems of quantum mechanics—the measurement problem and the adiabatic dynamics—are elucidated via ideas of statistical mechanics.

Section I recalls basic facts about the Hilbert space formalism. Here I work out several issues that are frequently not displayed in textbooks, e.g. the non-existence of joint probability for non-commuting quantum observables. Section II recalls the basic aspects of the statistical interpretation of quantum mechanics. Here the reader is advised against several common misinterpretations.

One of the main formulations of the second law of thermodynamics is posed and proved in section III. This formulation has definite validity limits, which are studied as well. The next section discusses the generalized Carnot cycle for a model heat engine and the problem of approaching the maximal efficiency while keeping the work-extraction power finite. Here we also discuss how the power problem for heat engines relates to the computational complexity. This problem has relations with protein folding, which are explored in some detail.

Section V discusses the measurement problem: Why and how can we make well-defined statements about an individual quantum system at the end of an ideal measurement in spite of the special, irreducibly probabilistic nature of quantum theory? A resolution of this problem is reviewed.

The last section studies the adiabatic set-up for quantum dynamics and provides substantial generalizations for the adiabatic theorem and the adiabatic (Berry's) phase.

Contents

I. Hilbert space, vectors, operators: short reminder	3
Problem 1	4
Problem 2	4
Problem 3	4
II. Density matrix. Mixed state. Quantum Ensembles.	6
A. Statistical interpretation of quantum mechanics.	6
B. Homogeneous ensembles.	6
C. Pure-state ensembles.	7
D. Inhomogeneous ensembles.	8
E. Prescribed ensemble fallacy.	8
Problem 1	8
Problem 2	9
III. The minimum work principle	10
A. The setup	10
B. The principle	10
C. The minimal work principle for macroscopic systems	11
D. Finite systems	11
E. Level crossing	12
F. Exactly solvable model with level crossing	13
G. Cyclic processes and reversibility	14
H. Summary	14
Problem 1	15
Problem 2	15
Problem 3	15
IV. Carnot cycle and attainability of the maximal efficiency	16
A. Introduction	16

B. Carnot cycle and its generalization	16
C. Maximization of work	17
D. The cycle time	18
E. Sub-optimal engine	19
Problem 1	20
Problem 2	20
1. Problem 3	23
V. Statistical theory of ideal quantum measurement processes	24
Problem 1	29
Problem 2	29
VI. Adiabatic quantum systems with (and without) feedback	30
A. Basic equations	30
B. Linear feedback	31
C. Closed-loop adiabatics	32
D. Examples	34
E. Slow motion of the controlling parameter	35
F. Mixed states	35
Problem 1	36
Problem 2	36
References	37

I. HILBERT SPACE, VECTORS, OPERATORS: SHORT REMINDER

We shall assume that the reader is aware of the basic theory of Hilbert spaces, as presented in theoretical physics (and in quantum mechanics); see the first chapter of [1] for a well-written introduction. Below some facts from this theory will be recalled, thereby fixing the notations. In particular, I assume that the reader knows Dirac's notations for vectors, operators and scalar products.

Quantum states (vectors) and observables (operators or matrices) live in a linear, complex (Hilbert) space \mathbb{H} . The dimension $\dim[\mathbb{H}]$ of this space will be normally assumed to be finite. Most of results directly generalize to separable Hilbert spaces¹.

The scalar product $\langle \psi | \phi \rangle$ between the vectors $|\phi\rangle$ and $\langle \psi | = (\langle \psi |)^{\dagger}$ is defined via 3 conditions [1]:

$$\langle \psi | \phi \rangle = \langle \phi | \psi \rangle^*, \quad (1)$$

$$\langle \psi | (|\phi\rangle + |\chi\rangle) = \langle \psi | \phi \rangle + \langle \psi | \chi \rangle, \quad (2)$$

$$\langle \psi | \psi \rangle \geq 0, \quad \langle \psi | \psi \rangle = 0 \quad \text{iff} \quad |\psi\rangle = 0. \quad (3)$$

For two hermitean ($Y = Y^{\dagger}$, $Z = Z^{\dagger}$) operators Y and Z , $Y \geq Z$ means that all eigenvalues of $Y - Z$ are non-negative, i.e. $\langle \psi | (Y - Z)\psi \rangle \geq 0$ for any $|\psi\rangle \in \mathbb{H}$. The direct sum $Y \oplus Z$ of two operators refers to the following block-diagonal matrix:

$$Y \oplus Z = \begin{pmatrix} Y & 0 \\ 0 & Z \end{pmatrix}. \quad (4)$$

$\text{ran}(Y)$ is the range of Y (set of vectors $Y|\psi\rangle$, where $|\psi\rangle \in \mathbb{H}$). I is the unity operator of \mathbb{H} . $\ker(Y)$ is the subspace of vectors $|\phi\rangle$ with $Y|\phi\rangle = 0$. I_n and 0_n are the $n \times n$ unity and zero matrices, respectively.

A subset of \mathbb{H} that is closed with respect to addition of vectors defines a subspace of \mathbb{H} . In the direct sum of two sub-spaces, $\mathbb{H} \oplus \mathbb{G}$ it is always understood that \mathbb{H} and \mathbb{G} are orthogonal. The vector sum of (not necessarily orthogonal) sub-spaces \mathbb{A} and \mathbb{B} will be denoted as $\mathbb{A} + \mathbb{B}$. This space is formed by all vectors $|\psi\rangle + |\phi\rangle$, where $|\psi\rangle \in \mathbb{A}$ and $|\phi\rangle \in \mathbb{B}$.

A hermitean ($\Pi^{\dagger} = \Pi$) operator Π is called projection if

$$\Pi = \Pi^2. \quad (5)$$

The simplest projections are I and 0 . The eigenvalues of a projection operator are equal to zero or one. The dimension of a projection Π is defined via $\text{tr}(\Pi_k)$.

Quantum observables are represented by hermitean operators. They are observed via their eigenvalues. Consider the eigenresolution of an hermitean operator $Y = Y^{\dagger}$:

$$Y = \sum_k y_k \Pi_k, \quad (6)$$

where y_k are eigenvalues and Π_k are eigenprojectors. The trace $\text{tr}(\Pi_k)$ is equal to the degeneracy of y_k .

Now Π_k can be regarded as the simplest, yes-no observables: the property expressed by $Y = y_k$ is either present (the eigenvalue of Π_k is one), or absent (the eigenvalue of Π_k is zero). This is naturally consistent with

$$I = \sum_k \Pi_k, \quad (7)$$

which means that some feature will be anyhow present, and with the orthonormality

$$\Pi_k \Pi_l = \delta_{kl} \Pi_k, \quad (8)$$

which means that the simultaneous presence of two different features is excluded. Co-existence of two different features is naturally defined via $\Pi_k + \Pi_l$, which is a projector due to (8). Another aspect of (7) is that the features expressed by, respectively, Π_k and $I - \Pi_k$ are complementary.

¹ Separable means that there is at least one countable orthonormal basis. One then can prove that orthonormal bases are countable [1].

Provided that $\dim[\mathbb{H}] \geq 3$, the Gleason's theorem tells that there is a unique probability $p(\cdot) \geq 0$ (Born's rule) given over (7) that holds

$$p(\Pi_k + \Pi_l) = p(\Pi_k) + p(\Pi_l), \quad \text{for any } k, l, \quad (9)$$

$$p(I) = 1. \quad (10)$$

This probability is defined via a density matrix ρ : positive (hermitean) operator that holds $\text{tr}(\rho) = 1$:

$$p(\Pi_k) = \text{tr}(\rho \Pi_k). \quad (11)$$

For a recent clear discussion of the Gleason's theorem (and numerous references on this subject) see [2]; see also [4] for a text-book presentation.

Thus each quantum state is determined by a certain density matrix ρ . In particular, pure states are defined by projector density matrices

$$\rho = |\psi\rangle\langle\psi|. \quad (12)$$

More general (mixed) states hold $\rho^2 \leq \rho$.

There is a hidden assumption in deriving (11) that was articulated in [3]: it is assumed that $p(\Pi_k)$ does not depend on other Π_l 's that make up (7). Indeed, Π_k may enter into different resolutions of unity, e.g. (7) and $\Pi_k + \sum_l Q_l = I$. This is an important aspect of non-contextuality of quantum probability, i.e. the context of other projections is not important in determining $p(\Pi_k)$.

Finally, note that the reasoning around (7, 8) concerning projections was formalized within the logico-algebraic approach to quantum mechanics; see e.g. [4] for further references.

Problem 1

Given two projections P and Q , with (generally) $PQ - QP \equiv [P, Q] \neq 0$, show that

$$[P, (P - Q)^2] = [Q, (P - Q)^2] = 0. \quad (13)$$

Define

$$U' \equiv QP + (I - Q)(I - P), \quad V' \equiv PQ + (I - P)(I - Q). \quad (14)$$

Show that

$$U'V' = V'U' = (I - P - Q)^2. \quad (15)$$

Finally, show that if $|I - P - Q|$ is invertible, then P and Q are unitary equivalent, i.e. there exists a unitary operator U such that

$$P = UQU^\dagger. \quad (16)$$

Problem 2

Show that any positive operator Y —i.e. an operator that holds $\langle\psi|Y|\psi\rangle \geq 0$ for any $|\psi\rangle \in \mathbb{H}$ —is hermitean.

Problem 3

Show that there is no joint probability for non-commuting quantum observables [26].

Solution. Given two sets of non-commuting hermitean projectors:

$$\sum_{k=1}^{n_P} P_k = I, \quad P_k P_i = \delta_{ik} P_k, \quad n_P \leq n, \quad (17)$$

$$\sum_{k=1}^{n_Q} Q_k = I, \quad Q_k Q_i = \delta_{ik} Q_k, \quad n_Q \leq n, \quad (18)$$

we are looking for non-negative operators $\Pi_{ik} \geq 0$ such that for an arbitrary density matrix ρ

$$\sum_{ik} \text{tr}(\rho \Pi_{ik}) = 1, \quad \sum_i \text{tr}(\rho \Pi_{ik}) = \text{tr}(\rho P_k), \quad \sum_k \text{tr}(\rho \Pi_{ik}) = \text{tr}(\rho Q_i). \quad (19)$$

These relations imply

$$\sum_{ik} \Pi_{ik} = I, \quad \Pi_{ik} \leq Q_i, \quad \Pi_{ik} \leq P_k. \quad (20)$$

Now the second (third) relation in (20) implies $\text{ran}(\Pi_{ik}) \subset \text{ran}(Q_i)$ ($\text{ran}(\Pi_{ik}) \subset \text{ran}(P_k)$). Hence $\text{ran}(\Pi_{ik}) \subset \text{ran}(Q_i) \cap \text{ran}(P_k)$.

Thus, if $\text{ran}(Q_i) \cap \text{ran}(P_k) = 0$, then $\Pi_{ik} = 0$, which means that the sought joint probability does not exist.

If $\text{ran}(Q_i) \cap \text{ran}(P_k) \neq 0$, then the largest Π_{ik} that holds the second and third relation in (20) is the projection $g(P_k, Q_i)$ on $\text{ran}(Q_i) \cap \text{ran}(P_k) = 0$. However, the first relation in (20) is still impossible to satisfy (for $[P_i, Q_k] \neq 0$), as seen from the sub-additivity feature:

$$\sum_{ik} g(P_i, Q_k) \leq \sum_k g\left(\sum_i P_i, Q_k\right) = \sum_k g(1, Q_k) = \sum_k Q_k = 1. \quad (21)$$

II. DENSITY MATRIX. MIXED STATE. QUANTUM ENSEMBLES.

For our further purposes we need an account of various features of quantum ensembles and their differences with respect to the classical ones. Though the theory of quantum ensembles is almost as old as quantum mechanics itself, it still attracts lively discussions; see e.g. [5].

A. Statistical interpretation of quantum mechanics.

Within this interpretation of quantum mechanics a quantum ‘state’ is described by a density matrix $\hat{\rho}$. 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 wavefunction $|\uparrow\rangle$ or the pure density matrix $|\uparrow\rangle\langle\uparrow|$. The description is optimal, in the sense that all particles have $\sigma_z = +1$, but incomplete in the sense of 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 [6]. As compared to other interpretations of quantum mechanics, the statistical interpretation is dealing more successfully with a number of conceptual problems, including the quantum measurement problem [6].

B. 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 \S from that ensemble. A typical example is an ensemble prepared by thermalization, that is, by letting each single system \S to interact weakly with an equilibrium thermal bath, and waiting sufficiently long till the equilibrium state of \S is established.

Let us study the feature 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 living in the Hilbert space \mathcal{H}) \hat{A} that are dispersionless on $\mathcal{E}(\hat{\rho})$,

$$\left[\text{tr} \left(\hat{A} \hat{\rho} \right) \right]^2 = \text{tr} \left(\hat{A}^2 \hat{\rho} \right), \quad (22)$$

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

$$\hat{A} \hat{\rho} = \alpha \hat{\rho}, \quad (23)$$

where α is a c-number. This implies $\text{tr} (\hat{A}^m \hat{\rho}) = \left[\text{tr} \hat{A} \hat{\rho} \right]^m$, with $m = 0, 1, 2, 3, \dots$, and the above statement follows.

For a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$, we return from (23) to the standard notion of $|\psi\rangle$ being an eigenstate of \hat{A} .

Any other, non-dispersionless observable \hat{B} — even if it commutes with the density matrix $\hat{\rho}$ — does not have a definite value in a single system \S from $\mathcal{E}(\hat{\rho})$. It is true that for $[\hat{\rho}, \hat{B}] = 0$, $\mathcal{E}(\hat{\rho})$ can be prepared by mixing² 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 $\mathcal{E}(\hat{\rho})$ 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 $\{ |\psi_\alpha\rangle\langle\psi_\alpha| \}_{\alpha=1}^N$

² 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 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., [5], and some confusion arose in literature for not taking it into account.

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 [6]

$$\hat{\rho} = \sum_{\alpha=1}^N \lambda_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|, \quad \lambda_\alpha \geq 0, \quad \sum_{\alpha=1}^N \lambda_\alpha = 1, \quad (24)$$

where $|\psi_\alpha\rangle$ are some normalized — but in general not orthogonal — vectors living in the same n -dimensional Hilbert space \mathcal{H} ³, and where $|\psi_\alpha\rangle\langle\psi_\alpha|$ are distinct.

The eigenresolution

$$\hat{\rho} = \sum_{k=1}^n p_k |p_k\rangle\langle p_k|, \quad (25)$$

$$\hat{H} = \sum_{k=1}^n \varepsilon_k |\varepsilon_k\rangle\langle\varepsilon_k|, \quad (26)$$

$$\langle\varepsilon_k|\varepsilon_l\rangle = \langle p_k|p_l\rangle = \delta_{kl}, \quad (27)$$

is only a particular case of (24), and if now the ensemble $\mathcal{E}(\hat{\rho})$ was prepared by one of the ways corresponding to (24) with non-orthogonal $|\psi_\alpha\rangle$, 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 not any definite value for these subensembles.

We conclude with three related features of a homogeneous ensemble:

- The ensemble cannot be thought to consist of definite subensembles.
- A single system from such an ensemble does not by itself define a subensemble.
- There are thus no homogeneous ensembles in classical statistical physics, since a single system is known to have definite values of all its variables.

C. 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 (24) that for a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$ in the RHS of representation (24) only one term shows up: $|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|^4$. Thus, pure-state ensembles cannot be prepared via mixing of other ensembles of the system §, or, put differently,

- pure-state ensembles are irreducible.
- 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 (22, 23) 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.

Note from (23) that if the mixed density matrix $\hat{\rho}$ has k , $1 \leq k \leq n$, non-zero 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

$$N_k = (n - k)^2 + 1. \quad (28)$$

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 matrices⁵. For a mixed state all dispersionless observables have to be degenerate.

³ Normalization and belonging to \mathcal{H} are necessary for $|\psi_\alpha\rangle\langle\psi_\alpha|$ to describe some ensemble of the systems §.

⁴ This can also be deduced from a more general result: any $|\psi_\alpha\rangle$ that can appear in (24) 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 $\langle 0|\hat{\rho}|0\rangle = \sum_\alpha \lambda_\alpha |\langle 0|\psi_\alpha\rangle|^2 = 0$; thus $\langle 0|\psi_\alpha\rangle = 0$ for $\lambda_\alpha > 0$.

⁵ 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.

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 [6].

D. 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 , $\alpha = 1, \dots, N$. 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 measurements⁶. In that case the above classical quantity is the corresponding record of the macroscopic apparatus by which this measurement was done. 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 $\hat{\rho} = \sum_{\alpha=1}^N \lambda_\alpha \hat{\rho}_\alpha$, but in contrast to the homogeneous situation this is not the full description. The latter is provided by the list

$$\{\lambda_\alpha, \hat{\rho}_\alpha\}_{\alpha=1}^N. \quad (29)$$

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.

E. Prescribed ensemble fallacy.

This fallacy rests on forgetting the difference between homogeneous and inhomogeneous ensembles [7], that is, it rests on neglecting that the overall [post-measurement] density matrix $\hat{\rho} = \sum_{\alpha=1}^N \lambda_\alpha \hat{\rho}_\alpha$ 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 in 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 in the splitting, instead, the splitting is unique, physical and contextual.

In spite of explicit warnings, 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. 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 of 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 (34) is employed.

Problem 1

Give a full description of dispersionless observables.

⁶ These measurements need not be done on the systems § directly, they can be indirect as well. Imagine an ensemble of two spin- $\frac{1}{2}$ particles described by pure density matrix $|\psi\rangle\langle\psi|$, where $|\psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1 \otimes |+\rangle_2 + |-\rangle_1 \otimes |-\rangle_2)$, 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 $\{\frac{1}{2}, |+\rangle_2 \langle +| ; \frac{1}{2}, |-\rangle_2 \langle -|\}$.

Solution.

Let $\text{tr}(\hat{A}^2 \hat{\rho}) = [\text{tr}(\hat{A} \hat{\rho})]^2$ 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 Cauchy inequality $|\text{tr}(\hat{A}\hat{B})|^2 \leq \text{tr}(\hat{A}\hat{A}^\dagger) \text{tr}(\hat{B}\hat{B}^\dagger)$, 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

$$\left[\text{tr}(\hat{A} \sqrt{\hat{\rho}} \sqrt{\hat{\rho}}) \right]^2 = \text{tr}(\hat{A}^2 \hat{\rho}) \text{tr}(\hat{\rho}) \quad (30)$$

implies

$$\hat{A} \sqrt{\hat{\rho}} = \alpha \sqrt{\hat{\rho}}. \quad (31)$$

Now insert the eigenresolution $\sqrt{\hat{\rho}} = \sum_{k=1}^n \sqrt{p_k} |\varepsilon_k\rangle \langle \varepsilon_k|$ into (31) and multiply it from the right by $|p_m\rangle$, to obtain:

$$\sqrt{p_m} \hat{A} |p_m\rangle = \alpha \sqrt{p_m} |p_m\rangle. \quad (32)$$

It is seen that either only one among the eigenvalues p_k 's is non-zero 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 non-zero 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

$$\hat{A} = \begin{pmatrix} \alpha \hat{1}_{k \times k} & 0 \\ 0 & \hat{B} \end{pmatrix}, \quad (33)$$

where α is a real number, $\hat{1}_{k \times k}$ is $k \times k$ unity matrix in the k -dimensional Hilbert space formed by eigenvectors corresponding to non-zero 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.

Problem 2

Provide a simple example for prescribed ensemble fallacy.

III. THE MINIMUM WORK PRINCIPLE

The second law of thermodynamics [8–10], formulated nearly one and half century ago, continues to be under scrutiny [11–13]. While its status within equilibrium thermodynamics and statistical physics is by now well-settled [8–10], its fate in various border situations is far from being clear. In the macroscopic realm the second law is a set of equivalent statements concerning quantities such as entropy, heat, work, etc. However, in more general situations these statements need not be equivalent and some, e.g. those involving entropy, may have only a limited applicability[12, 13]. In contrast to entropy, the concept of work has a well-defined operational meaning for finite systems interacting with macroscopic work sources [9]. It is, perhaps, not accidental that Thomson’s formulation of the second law [8–10] — no work can be extracted from an equilibrium system by means of a cyclical process — was proven [14, 15] both in quantum and classical situation.

Here we study the minimum work principle which extends the Thomson’s formulation to non-cyclic processes [8–10, 15], and provides a recipe for reducing energy costs. After formulating the principle and discussing it for macroscopic systems, we investigate it for finite systems coupled to macroscopic sources of work. Its domain of validity there is found to be large but definitely limited. These limits are illustrated via counterexamples.

A. The setup

Consider a quantum system S which is thermally isolated [8–10]: 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 $R(t) = \{R_1(t), R_2(t), \dots\}$ 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 $H_i = H\{R_i\}$ and $H_f = H\{R_f\}$, respectively. Initially S is assumed to be in equilibrium at temperature $T = 1/\beta \geq 0$, that is, S is described by a gibbsian density operator:

$$\rho(t_i) = \exp(-\beta H_i)/Z_i, \quad Z_i = \text{tr } e^{-\beta H_i}. \quad (34)$$

As usual, this equilibrium state is prepared by a weak interaction between S and a macroscopic thermal bath at temperature T [8, 9, 13], and then decoupling S from the bath in order to achieve a thermally isolated process [8–10].

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

$$i\hbar \frac{d}{dt} \rho(t) = [H(t), \rho(t)], \quad \rho(t) = U(t) \rho(t_i) U^\dagger(t), \quad (35)$$

with time-ordered $U(t) = \overleftarrow{\exp}[-\frac{i}{\hbar} \int_{t_i}^t ds H(s)]$. The work W done on S reads [8–10]

$$W = \int_{t_i}^{t_f} dt \text{tr} [\rho(t) \dot{H}(t)] = \text{tr}[H_f \rho(t_f)] - \text{tr}[H_i \rho(t_i)], \quad (36)$$

where we performed partial integration and inserted (35). This is the energy increase of S , which coincides with the energy decrease of the source.

B. The principle

Let S start in the state (34), and let R move between R_i and R_f along a trajectory $R(t)$. The work done on S 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 velocity, thereby taking a very long time $t_f - t_i$, at the cost of an amount work \widetilde{W} . The minimum-work principle then asserts [8–10]

$$W \geq \widetilde{W}. \quad (37)$$

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 very slow velocity. If during some operation work has to be added ($W > 0$) to S , one wishes to minimize its amount, and operates slowly. For thermally isolated systems, adiabatically slow processes are reversible. This is standard if S is macroscopic [8–10], and below it is shown to hold for a finite S as well, where the definition of reversibility extends unambiguously (i.e., without invoking entropy) [10].

In macroscopic thermodynamics the minimum work principle is derived [8, 10] 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 [15–17]. We discuss one of them now.

C. The minimal work principle for macroscopic systems

It is proven in two steps: first one considers the relative entropy $\text{tr}[\rho(t_f) \ln \rho(t_f) - \rho(t_f) \ln \rho_{\text{eq}}(H_f)]$ between the final state $\rho(t_f)$ given by (35) and an equilibrium state $\rho_{\text{eq}}(H_f) = \exp(-\beta H_f)/Z_f$, $Z_f = \text{tr} e^{-\beta H_f}$, a state corresponding to the final Hamiltonian H_f and the same temperature $T = 1/\beta$. As follows from (35), $\text{tr}[\rho(t_f) \ln \rho(t_f)] = \text{tr}[\rho(t_i) \ln \rho(t_i)]$. This combined with (34, 36) and the non-negativity of relative entropy [9] yields:

$$W \geq F(H_f) - F(H_i) \equiv T \ln \text{tr} e^{-\beta H_i} - T \ln \text{tr} e^{-\beta H_f}, \quad (38)$$

where $F(H_i)$ and $F(H_f)$ are the gibbsian free energies corresponding to $\rho(t_i)$ and $\rho_{\text{eq}}(H_f)$, respectively.

There are several classes of macroscopic systems for which one can show that the free energy difference in the RHS of (38) indeed coincides with the adiabatic work [13, 16, 17].

D. Finite systems

For an arbitrary N -level quantum system S, Eq. (38) does not have the needed physical meaning, since in general $F(H_f) - F(H_i)$ does not coincide with the the adiabatic work. It is known[18] that for finite systems the final density matrix $\rho(t_f)$ given by (35) need not coincide with $\rho_{\text{eq}}(H_f) = \exp(-\beta H_f)/Z_f$. This fact was recently applied for certain irreversible processes[19].

Thus we need an independent derivation of (37). Let the spectral resolution of $H(t)$ and $\rho(t_i)$ be

$$H(t) = \sum_{k=1}^N \varepsilon_k(t) |k, t\rangle \langle k, t|, \quad \langle k, t|n, t\rangle = \delta_{kn}, \quad (39)$$

$$\rho(t_i) = \sum_{k=1}^N p_k |k, t_i\rangle \langle k, t_i|, \quad p_k = \frac{e^{-\beta \varepsilon_k(t_i)}}{\sum_n e^{-\beta \varepsilon_n(t_i)}}. \quad (40)$$

At $t = t_i$ we order the spectrum as

$$\varepsilon_1(t_i) \leq \dots \leq \varepsilon_N(t_i) \implies p_1 \geq \dots \geq p_N. \quad (41)$$

For $t_i \leq t \leq t_f$ we expand on the complete set $|n, t\rangle$:

$$U(t)|k, t_i\rangle = \sum_{n=1}^N a_{kn}(t) e^{-\frac{i}{\hbar} \int_{t_i}^t dt' \varepsilon_n(t')} |n, t\rangle, \quad (42)$$

and use (36) to obtain:

$$W = \sum_{k,n=1}^N |a_{kn}(t_f)|^2 p_k \varepsilon_n(t_f) - \sum_{k=1}^N p_k \varepsilon_k(t_i). \quad (43)$$

A similar formula can be derived to express the adiabatic work \widetilde{W} in coefficients $\tilde{a}_{kn}(t_f)$. From the definition $|a_{kn}(t_f)|^2 = |\langle n, t_f | U | k, t_i \rangle|^2$ it follows that

$$\sum_{k=1}^N |a_{kn}(t_f)|^2 = \sum_{n=1}^N |a_{kn}(t_f)|^2 = 1. \quad (44)$$

With help of the identity: $\sum_{n=1}^N \varepsilon_n x_n = \varepsilon_N \sum_{n=1}^N x_n - \sum_{m=1}^{N-1} [\varepsilon_{m+1} - \varepsilon_m] \sum_{n=1}^m x_n$, we obtain using (43, 44) the general formula for the difference between non-adiabatic and adiabatic work:

$$W - \widetilde{W} = \sum_{m=1}^{N-1} [\varepsilon_{m+1}(t_f) - \varepsilon_m(t_f)] \Theta_m, \quad (45)$$

$$\Theta_m \equiv \sum_{n=1}^m \sum_{k=1}^N p_k (|\tilde{a}_{kn}(t_f)|^2 - |a_{kn}(t_f)|^2). \quad (46)$$

To understand the meaning of this formula, let us first assume that the ordering (41) is kept at $t = t_f$:

$$\varepsilon_1(t_f) \leq \dots \leq \varepsilon_N(t_f). \quad (47)$$

If different energy levels did not cross each other (and equal ones do not become different), Eq. (47) is implied by Eq. (41). According to non-crossing rule [20], if $H\{R\}$ is real and only one of its parameters is varied with time, (47) 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 [20]). No level-crossings and natural conditions of smoothness of $H(t)$ are sufficient for the standard quantum adiabatic theorem [21] to ensure

$$\tilde{a}_{kn}(t_f) = \delta_{kn}. \quad (48)$$

Combined with (41, 48), Eq.(46) brings

$$\begin{aligned} \Theta_m &= \sum_{k=1}^m p_k \left(1 - \sum_{n=1}^m |a_{kn}(t_f)|^2 \right) - \sum_{n=1}^m \sum_{k=m+1}^N p_k |a_{kn}(t_f)|^2 \\ &\geq p_m \left[m - \sum_{k=1}^m \sum_{n=1}^m |a_{kn}(t_f)|^2 - \sum_{n=1}^m \sum_{k=m+1}^N |a_{kn}(t_f)|^2 \right] = 0. \end{aligned}$$

Eqs. (43, 45, 47) together with $\Theta_m \geq 0$ extend the minimum work principle (37) to cases where the adiabatic work is not equal to the difference in free energies.

E. Level crossing

The above non-crossing condition raises the question: Is the minimum 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: 1) 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 [20]. Two parameters can produce robust level-crossing for such Hamiltonians. 2) It is believed that once levels can cross, $\Delta\varepsilon \rightarrow 0$, the very point of the adiabatic theorem 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 [22] in this context. 3) It is sometimes believed that crossing is automatically followed by a population inversion. We shall find no support for that.

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

$$H(t) = h_1(s)\sigma_1 - h_3(s)\sigma_3, \quad s = t/\tau, \quad (49)$$

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. Assume that *i*) for $s \rightarrow s_i < 0$ and for $s \rightarrow s_f > 0$, $h_1(s)$ and $h_3(s)$ go to constant values sufficiently fast; *ii*) for $s \rightarrow 0$ one has: $h_1(s) \simeq \alpha_1 s^2$, $h_3(s) \simeq -\alpha_3 s$, where α_1 and α_3 are positive constants. *iii*) $h_1(s)$ and $h_3(s)$ are non-zero for all s , $s_i \leq s \leq s_f$, except $s = 0$. Not all these points are needed, but we choose them for clarity. Generalizations are indicated below. One writes (49) as

$$H = \begin{pmatrix} -h_3(s) & h_1(s) \\ h_1(s) & h_3(s) \end{pmatrix} = \varepsilon_1(s) \begin{pmatrix} \cos \theta(s) & \sin \theta(s) \\ \sin \theta(s) & -\cos \theta(s) \end{pmatrix},$$

where $\theta(s) = -\arctan[h_1(s)/h_3(s)]$ is a parameter in the interval $-\pi/2 < \theta < \pi/2$, and where

$$\varepsilon_1(s) = \text{sg}(s) \sqrt{h_3^2(s) + h_1^2(s)}, \quad \varepsilon_2(s) = -\varepsilon_1(s) \quad (50)$$

are the adiabatic energy levels which cross at $s = \theta(s) = 0$ ($\sqrt{\dots}$ is defined to be always positive). The adiabatic eigenvectors are, $H(s)|k, s\rangle = \varepsilon_k(s)|k, s\rangle$, $k = 1, 2$,

$$|1, s\rangle = \begin{pmatrix} \cos \frac{1}{2}\theta(s) \\ \sin \frac{1}{2}\theta(s) \end{pmatrix}, \quad |2, s\rangle = \begin{pmatrix} -\sin \frac{1}{2}\theta(s) \\ \cos \frac{1}{2}\theta(s) \end{pmatrix}. \quad (51)$$

Both the eigenvalues and the eigenvectors are smooth functions of s . Eq. (47) is not valid, and (43–46) imply:

$$W - \widetilde{W} = -2\sqrt{h_1^2(s_f) + h_3^2(s_f)} \Theta_1, \quad \tau s_f = t_f. \quad (52)$$

Naively this already proves the violation. More carefully, our strategy is now to confirm (48) in the adiabatic limit $\tau \rightarrow \infty$ and thus to confirm that $\Theta_1 > 0$, implying that the minimum work principle is indeed violated. To this end we apply the standard adiabatic perturbation theory [21]. Substituting (42) into (35) one has:

$$\dot{a}_{kn} = - \sum_{m=1}^N a_{km}(t) e^{\frac{i}{\hbar} \int_{t_i}^t dt' [\varepsilon_n(t') - \varepsilon_m(t')]} \langle n | \partial_t | m \rangle. \quad (53)$$

As $|1\rangle$ and $|2\rangle$ in (51) are real, $\langle n | n \rangle = 1$ implies $\langle n | \partial_t | n \rangle = 0$. Since $\langle n | \partial_t | m \rangle = \frac{1}{\tau} \langle n | \partial_s | m \rangle$ the RHS of (53) contains a small parameter $1/\tau$. It is 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 in the RHS of (53), yielding for $a_{k \neq n}(t_f) = b_{k \neq n}(t_f)$:

$$|a_{k \neq n}(t_f)|^2 = \left| \int_{s_i}^{s_f} ds e^{\frac{i\tau}{\hbar} \int_{s_i}^s du [\varepsilon_k(u) - \varepsilon_n(u)]} \langle n | \partial_s | k \rangle \right|^2, \quad (54)$$

while $|a_{kk}(t_f)|^2 = 1 - \sum_{n \neq k} |a_{kn}(t_f)|^2$. In (54) we put $s\tau = t$, $s'\tau = t'$. For our model (49–51), $\int_{s_i}^s du [\varepsilon_1(u) - \varepsilon_2(u)] = 2 \int_{s_i}^s du \varepsilon_1(u)$ has only one extremal point, at $s = 0$. We also have from (51)

$$\langle 2 | \partial_s | 1 \rangle = \frac{\theta'}{2} = \frac{1}{2} \frac{h_1 h'_3 - h_3 h'_1}{h_1^2 + h_3^2}, \quad \theta' \equiv \frac{d\theta}{ds}. \quad (55)$$

For large τ the integral in (54) can be calculated with use of the saddle-point method:

$$|a_{12}(t_f)|^2 = \frac{\pi\hbar}{\tau} \left[\frac{\langle 2 | \partial_s | 1 \rangle^2 \sqrt{h_1^2 + h_3^2}}{h_1 h'_1 + h_3 h'_3} \right] \Big|_{s=0} = \frac{\pi\hbar\alpha_1^2}{4\tau\alpha_3^3}. \quad (56)$$

Eqs. (54, 56) extend the adiabatic theorem (48) for the level-crossing situation. More general statements of similar adiabatic theorems can be found in Ref. [22]. Inserting $\Theta_1 = (p_1 - p_2) |a_{12}(t_f)|^2 > 0$ in Eq. (52) confirms the violation of the minimum work principle. Eq. (56) shows that the role of the proper internal characteristic time is played by $\hbar\alpha_1^2/\alpha_3^3$ rather than by $\hbar/(\varepsilon_1 - \varepsilon_2)$.

More generally, if $\sqrt{h_3^2(s) + h_1^2(s)}$ is a smooth function for all real s (e.g., it is not $\propto |s|$), there are no crossings of eigenvalues and (37) is valid. If both $h_1(s)$ and $h_3(s)$ are linear for $s \rightarrow 0$, the leading term presented in (56) vanishes due to $\langle 2 | \partial_s | 1 \rangle^2|_{s=0} = 0$, and one needs the second-order in the saddle-point expansion, to be compared with the second-order term of the adiabatic perturbation theory. This leads to the same physical conclusions as (56) did, but with $|a_{12}(t_f)|^2 \sim \tau^{-3}$.

One can calculate $|a_{kn}(t_f)|$ yet in another limiting case, where the characteristic time τ is very short. It is well-known [21] that in this limit energy changes can be calculated with frozen initial state of S. For the present situation this leads from (51) to $|a_{12}(t_f)|^2 = |a_{21}(t_f)|^2 = |\langle 1, t_f | 2, t_i \rangle|^2 = \sin^2 \frac{1}{2}[\theta(t_f) - \theta(t_i)]$, and thus to $\Theta_1 = (p_1 - p_2) \sin^2 \frac{1}{2}[\theta(t_f) - \theta(t_i)]$, again positive.

F. Exactly solvable model with level crossing

Consider a two-level system with Hamiltonian

$$H(t) = \hbar\omega \begin{pmatrix} s \cos^2 s & \frac{1}{2}s \sin 2s \\ \frac{1}{2}s \sin 2s & s \sin^2 s \end{pmatrix}, \quad s = \frac{t}{\tau}, \quad (57)$$

where τ is the characteristic time-scale, and ω is a constant. For $s_i < 0$ denote the adiabatic energy levels as $\varepsilon_1(s_i) = \hbar\omega s_i < \varepsilon_2(s_i) = 0$. They cross when s passes through zero. Eq. (53) for this model can be solved exactly in terms of hypergeometric functions. Postponing the detailed discussion, we present in Fig. 1 the behavior of $|a_{12}(s_f)|^2$ as a function of τ . Since from (45, 57) one has for the present model $W - \widetilde{W} = -\hbar\omega s_f |a_{12}(s_f)|^2 \tanh(\frac{1}{2}\beta\hbar\omega |s_i|)$, violations of the minimum work principle exist for $s_f > 0$, and they are maximal for $|a_{12}(s_f)|^2 \rightarrow 1$. This is seen to be

the case in Fig. 1 for some τ near $\tau = 1$. Notice that both the first-order perturbation theory and the saddle-point approximation are adequately reproducing $|a_{12}(s_f)|^2$ for $\tau \gtrsim 10$.

Let S has a finite amount of levels, and two of them cross. For quasi-adiabatic processes (τ is large but finite) the transition probability between non-crossing levels is exponentially small [21, 23], while as we saw it has power-law smallness for two crossing levels. Then one neglects in (43) the factors $|a_{k \neq n}(t_f)|^2$ coming from any non-crossed levels k and n , and the problem is reduced to the two-level situation. Thus already one crossing suffices to detect limits of the minimum work principle. The reduction to the two-level situation takes place also in a macroscopic system which has few discrete levels at the bottom of a continuous spectrum, since for low temperatures these levels can decouple from the rest of the spectrum.

G. Cyclic processes and reversibility

The above results do not imply any violation of the second law in Thomson's formulation[14]: no work is extracted from S during a cyclic process, $W_c \geq 0$. We illustrate its general proof in the context of the level crossing model given by (49–51). 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 levels backwards, then at the final time of R' Eq. (47) is valid, and (43–48) imply:

$$W_c = |a_{12}|^2(p_1 - p_2)[\varepsilon_2(t_i) - \varepsilon_1(t_i)] \geq \widetilde{W}_c = 0, \quad (58)$$

where $|a_{12}|^2 \leq 1$ now corresponds to the full cyclic process $R + R'$. Eq. (58) confirms the intuitive expectation that non-adiabatic process are less optimal. In particular, this is valid if R' is exactly the same process R moved backwards with the same speed. Then $\widetilde{W}_c = 0$ means that R is a reversible process in the standard thermodynamical sense [8–10]. If R' does not induce another level crossing, i.e., $h_1(s)$ and $h_2(s)$ in Eq. (49) return to their initial values without simultaneously crossing zero, then $\varepsilon_1(t_f) = \varepsilon_2(t_i)$, $\varepsilon_2(t_f) = \varepsilon_1(t_i)$ and Eqs. (43, 48) imply

$$\widetilde{W}_c = (p_1 - p_2)[\varepsilon_2(t_i) - \varepsilon_1(t_i)] \geq W_c = |a_{11}|^2 \widetilde{W}_c > 0.$$

In contrast to (58), 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 minimum work principle for a cyclic process.

H. Summary

We have studied the minimum work principle for finite systems coupled to external sources of work. 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 and have 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 suffices to note violations of the principle. If this is the case, the optimal process occurs at some finite speed.

Level-crossing was observed, e.g., in molecular and chemical physics [25]. It is not a rare effect[24]: if the number of externally varied parameters is larger than two, then for typical spectra level crossings are even much more frequent than avoided crossings[24]. It is possible that the presented limits of the minimum work principle may serve as a test for level crossings.

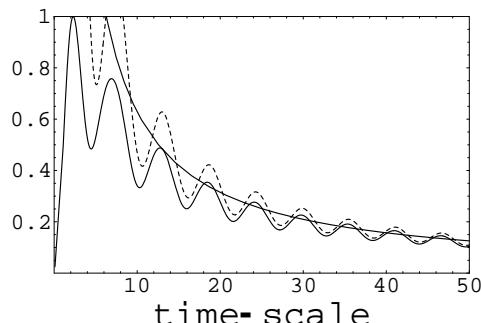


FIG. 1: Amplitude $|a_{12}(s_f)|^2$ versus the time-scale τ for $s_i = -1.5$, $s_f = 1.5$ and $\omega = 1$. Full oscillating curve: the exact value which can reach unity. Dotted curve: result from a first-order adiabatic perturbation theory. The smooth curve presents a saddle-point approximation analogous to (56).

Together with the universal validity of Thomson's formulation of the second law [14, 15], 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 non-equivalence of various formulations of the second law were found in Ref. [12, 13], where for a quantum particle coupled to a macroscopic thermal bath, it was shown that some formulations, e.g., 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.

Problem 1

Derive (58).

Problem 2

Explain why the standard adiabatic theorem is (almost always) derived by assuming no level crossing.

Problem 3

Show in details how the Thomson's formulation of the second law relates to the minimal work principle.

IV. CARNOT CYCLE AND ATTAINABILITY OF THE MAXIMAL EFFICIENCY

A. Introduction

Reciprocating heat engines extract work operating cyclically between two thermal baths at temperatures T_1 and T_2 ($T_1 > T_2$) [8–10]. They have two basic characteristics: (i) efficiency, $\eta = W/Q_1$, is the work W extracted per cycle divided by the heat input Q_1 from the high-temperature bath. (ii) Power W/τ , where τ is the cycle duration. Both these quantities have to be large for a good engine: if η is small, lot of energy is wasted; if the power is small, no sizable work is delivered over a reasonable time [8–10].

The second law establishes the Carnot efficiency

$$\eta_C = 1 - \frac{T_2}{T_1}$$

as an upper bound for η [8–10]. The Carnot cycle reaches the bounding value η_C in the (useless) limit, where the power goes to zero [8–10]. Conversely, realistic engines are not efficient, since they have to be powerful, e.g. the efficiency of Diesel engines amounts to 35–40 % of the maximal value. This *power-efficiency dilemma* motivated a search for the efficiency that would generally characterize the maximal power regime. One candidate for this is the Curzon-Ahlborn efficiency

$$\eta_{CA} = 1 - \sqrt{T_2/T_1}$$

[27], which is however crucially tied to the linear regime $T_1 \approx T_2$ [28]. Beyond this regime η_{CA} is a lower bound of η for a class of model engines [29]. Several recent models for the efficiency at the maximal power overcome η_{CA} with $\eta^* = \frac{\eta_C}{2-\eta_C}$ [30].

As argued in [29, 31], the maximal power regime allows for the Carnot efficiency, at least for certain models. But it is currently an open question whether the maximal efficiency is attained under realistic conditions (see e.g. [31]), and how to characterize the very *realism* of those conditions. Even more generally: what is the origin of the power-efficiency dilemma? We answer these questions by analyzing a generalized Carnot cycle, which in contrast to the original Carnot cycle is not restricted to slow processes. We now summarize our answers.

(1) When the N -particle engine operates at the maximal work extracted per cycle, its efficiency reaches the Carnot bound η_C for $N \gg 1$, while the cycle time is given by the relaxation time of the engine. The maximal work and the Carnot efficiency are achieved due to the flat energy landscape of the engine. For realistic engine-bath interactions this energy landscape leads to a very long [$\mathcal{O}(e^N)$] relaxation time nullifying the power. By realistic we mean interactions that are independent from the engine Hamiltonian. If we assume a proper tuning between engine-bath interaction and the engine Hamiltonian, the relaxation time scales as $\mathcal{O}(\sqrt{N})$, and the maximal efficiency is achievable in the limit $N \gg 1$ at a large power $\mathcal{O}(\sqrt{N})$.

(2) The relaxation of the optimal engine under realistic interactions relates to an important problem of searching an unstructured database for a marked item, where each energy level refers to a database item. This task is computationally complex, i.e. even the most powerful quantum algorithms resolve it in $\mathcal{O}(e^{N/2})$ time-steps [32]. Hence the power-efficiency dilemma relates to computational complexity. The same effect can be reformulated as the Levinthal's paradox of the protein folding problem: if the majority of unfolded states of a protein are assumed to have the same (free) energy, the folding time is very long [33].

(3) A scenario of resolving the Levinthal's paradox proposed in protein science shows the way of constructing sub-optimal engines that operate at a reasonably large values of work, power and efficiency. These sub-optimal engines function as model proteins, but they are restricted to a mesoscopic scales $N \sim 100$; otherwise the relaxation time is again large. Sacrificing some 50–60% of the maximal possible work leads to a reasonable cycle times with the efficiency that achieves some 90 % of the maximal (Carnot) efficiency.

B. Carnot cycle and its generalization

Recall that the Carnot cycle consists of four slow, consecutive pieces [8]: thermally-isolated – isothermal – thermally-isolated – isothermal. Four times slow brings in the vanishing power stressed above; see additionally section I of [36]. Since the overall process is a quasi-equilibrium one, the external fields that extract work from the engine act on it during all the four pieces. One deduces for the isothermal parts: $Q_1 = T_1\Delta S$ and $Q_2 = T_2\Delta S$, where Q_1 (Q_2) is the heat taken from (put into) the T_1 -bath (T_2 -bath), and $\Delta S > 0$ is the entropy change. Since the work extracted is $W = Q_1 - Q_2$, η equals to its maximal value $\eta_C = 1 - \frac{T_2}{T_1}$ [8].

We keep the two isothermal and two thermally isolated pieces of the original Carnot cycle, but do not force them to be slow. In addition, the external fields will act only during the thermally isolated stages. Isothermal pieces amount to free relaxation. Due to these points, we can analyze the engine functioning from the energy conservation. We study the engine via quantum mechanics on a finite Hilbert space, because this reduces the problem to a combinatorial optimization. The final results are interpreted classically and can be also obtained by discretizing the Hamiltonian classical dynamics over phase-space cells.

0. The engine E with the Hamiltonian H_1 starts in an equilibrium state at temperature T_1 described by the density matrix

$$\rho(0) = \rho_1 = e^{-\beta_1 H_1} / (\text{tr } e^{-\beta_1 H_1}), \quad \beta_1 = 1/T_1. \quad (59)$$

1. Between times 0 and τ , E undergoes a thermally isolated process with a time-dependent Hamiltonian $H_{12}(t)$ and the unitary evolution $\rho(\tau) = U_{12}\rho(0)U_{12}^\dagger$:

$$H_{12}(0) = H_1, \quad H_{12}(\tau) = H_2, \quad U_{12} = \mathcal{T}e^{-i \int_0^\tau ds H_{12}(s)}, \quad (60)$$

where \mathcal{T} means chronological ordering. The work taken out of E is determined by energy conservation (see [8] and section II of [36])

$$W_1 = \text{tr}[H_1\rho_1 - H_2U_{12}\rho_1U_{12}^\dagger]. \quad (61)$$

2. Then E is attached to the T_2 -bath and after relaxation time τ_r its density matrix becomes

$$\rho(\tau + \tau_r) = \rho_2 = e^{-\beta_2 H_2} / (\text{tr } e^{-\beta_2 H_2}). \quad (62)$$

The heat that came to E from the T_2 -bath is

$$Q_2 = \text{tr}[H_2\rho_2 - H_2U_{12}\rho_1U_{12}^\dagger]. \quad (63)$$

3. E undergoes another thermally isolated process

$$H_{21}(0) = H_2, \quad H_{21}(\tau) = H_1, \quad U_{21} = \mathcal{T}e^{-i \int_0^\tau ds H_{21}(s)}, \quad (64)$$

completing the cycle with respect to the Hamiltonian. The work taken out of E reads

$$W_2 = \text{tr}[H_2\rho_2 - H_1U_{21}\rho_2U_{21}^\dagger]. \quad (65)$$

4. Finally, E is attached to the T_1 -bath ($T_1 > T_2$) and relaxes to ρ_1 thereby completing the cycle; see (59). The heat that came to E from the T_1 -bath is

$$Q_1 = \text{tr}[H_1\rho_1 - H_1U_{21}\rho_2U_{21}^\dagger]. \quad (66)$$

To stress the differences with the original Carnot cycle: (i) the cycle time $2(\tau + \tau_r)$ need not be much larger than the relaxation time τ_r . (ii) The cycle is out of equilibrium. (iii) The work source and the bath never act simultaneously; either one acts or another. Hence heat and work are deduced from the energy conservation.

The essential assumption of the cycle is that the system-bath coupling is weak; thus we did not count the work necessary for switching the coupling(s) on and off.

C. Maximization of work

We maximize the full extracted work $W = W_1 + W_2$ over H_1, H_2, U_{12}, U_{21} for fixed $T_1 > T_2$ and a fixed number $n+1$ of energy levels of E. The lowest energies of H_1 and H_2 can be set to zero. Introduce the eigen-resolution of H_α

$$H_\alpha = \sum_{k=2}^{n+1} \epsilon_k^{[\alpha]} |k^{[\alpha]}\rangle\langle k^{[\alpha]}|, \quad \alpha = 1, 2. \quad (67)$$

The full work $W = W_1 + W_2$ reads from (61, 65)

$$W = \sum_{\alpha=1}^2 \sum_{k=2}^{n+1} p_k^{[\alpha]} \epsilon_k^{[\alpha]} \quad (68)$$

$$- \sum_{k,l=2}^{n+1} \left[p_k^{[2]} \epsilon_l^{[1]} C_{kl}^{[21]} + p_k^{[1]} \epsilon_l^{[2]} C_{kl}^{[12]} \right], \quad (69)$$

where $\{p_k^{[\alpha]}\}_{k=1}^{n+1}$ are eigenvalues of ρ_α given by (59, 62, 67), and where

$$C_{kl}^{[\alpha\gamma]} = |\langle k^{[\alpha]} | U_{\alpha\gamma} | l^{[\gamma]} \rangle|^2, \quad (\alpha, \gamma) = (1, 2), (2, 1). \quad (70)$$

$C_{kl}^{[\alpha\gamma]}$ are doubly stochastic matrices: $\sum_{k=1}^{n+1} C_{kl}^{[\alpha\gamma]} = \sum_{l=1}^{n+1} C_{kl}^{[\alpha\gamma]} = 1$. Such a matrix C_{kl} can be represented as a convex sum of permutation matrices (Birkhoff's theorem) [35]: $C_{kl} = \sum_\delta \lambda_\delta \Pi_{kl}^{[\delta]}$, where $\lambda_\delta \geq 0$, $\sum_\delta \lambda_\delta = 1$, and where $\Pi^{[\delta]}$ permutes the components of any vector on which it acts. Hence we can maximize W in (68, 69) over $\{\lambda_\delta\}$. The optimal $C_{kl}^{[12]}$ and $C_{kl}^{[21]}$ amount to permutation matrices, since $\{\lambda_\delta\}$ enter linearly into W . Without loss of generality we can assume $\epsilon_1^{[\alpha]} \leq \dots \leq \epsilon_{n+1}^{[\alpha]}$ and hence $p_1^{[\alpha]} \geq \dots \geq p_{n+1}^{[\alpha]}$. Then the optimal permutations $C_{kl}^{[12]}$ and $C_{kl}^{[21]}$ are unit matrices; see (69). In contrast to the original Carnot cycle, the optimal thermally isolated processes can be realized as sudden (quick) changes of the Hamiltonian eigenvalues without changing the eigenvectors. A prominent example of such a process is the Mössbauer effect [37]. It is an advantage that thermally isolated processes can be fast; otherwise it is difficult to maintain thermal isolation, which is a known problem of the original Carnot cycle [8, 9].

The work

$$W = W_1 + W_2 = \sum_{k=2}^{n+1} (p_k^{[1]} - p_k^{[2]})(\epsilon_k^{[1]} - \epsilon_k^{[2]}), \quad (71)$$

is to be still maximized over $\{\epsilon_k^{[1]}\}_{k=2}^{n+1}$ and $\{\epsilon_k^{[2]}\}_{k=2}^{n+1}$; see (70). W is symmetric with respect to permutations within $\{\epsilon_k^{[1]}\}_{k=2}^{n+1}$ and within $\{\epsilon_k^{[2]}\}_{k=2}^{n+1}$. We checked numerically that this symmetry is not broken and hence the maximum of W is reached for

$$\epsilon^{[\alpha]} \equiv \epsilon_2^{[\alpha]} = \dots = \epsilon_{n+1}^{[\alpha]}, \quad \alpha = 1, 2, \quad (72)$$

i.e. all excited levels have the same energy. Thus the thermally isolated pieces of the cycle consist, respectively, of sudden changes $\epsilon^{[1]} \rightarrow \epsilon^{[2]}$ and $\epsilon^{[2]} \rightarrow \epsilon^{[1]}$.

With new variables $e^{-\beta_\alpha \epsilon^{[\alpha]}} \equiv u_\alpha$ we write the maximal work as

$$W_{\max}[u_1, u_2] = \frac{(T_1 \ln \frac{1}{u_1} - T_2 \ln \frac{1}{u_2})(u_1 - u_2)n}{[1 + nu_1][1 + nu_2]}, \quad (73)$$

where u_1 and u_2 are found from

$$\partial_{u_1} W_{\max}[u_1, u_2] = \partial_{u_2} W_{\max}[u_1, u_2] = 0. \quad (74)$$

u_1 and u_2 depend on T_2/T_1 and on n . Noting (66) and the result before (71) we obtain $Q_1 = \text{tr}(H_1(\rho_1 - \rho_2))$ for the heat obtained from the high-temperature bath. Using (72) we get from $\eta = W/Q_1$ and from (73):

$$\eta = 1 - [T_2 \ln u_2] / [T_1 \ln u_1]. \quad (75)$$

Note from (73) that $W_{\max}[u_1, u_2] > 0$ and $T_2 < T_1$ imply $1 > \epsilon^{[2]}/\epsilon^{[1]} > T_2/T_1$. Hence (75) implies $\eta \leq \eta_C = 1 - T_2/T_1$, as expected.

Both $W_{\max}[u_1, u_2]$ and η increase with n . For $\ln[n] \gg 1$ we get asymptotically from (74):

$$u_1 = \frac{(1 - \theta) \ln[n]}{n}, \quad u_2 = \frac{\theta}{n \ln[n](1 - \theta)}, \quad (76)$$

where $\theta \equiv T_2/T_1$. This produces

$$W_{\max}[u_1, u_2] = (T_2 - T_1) \ln[n] - \mathcal{O}(1/\ln[n]), \quad (77)$$

$$\eta = \eta_C - \mathcal{O}(1/\ln[n]), \quad \eta_C \equiv 1 - T_2/T_1. \quad (78)$$

The maximal work $W_{\max}[u_1, u_2]$ scales as $\ln[n]$, since this is the “effective number of particles” for the engine. In the macroscopic limit $\ln[n] \gg 1$, the efficiency converges to its maximal value $\eta_C = 1 - T_2/T_1$; see (78).

D. The cycle time

It amounts to two times the relaxation time τ_r of the system with spectrum (72) and energy gap $\epsilon \sim \ln[n]$; see (72, 76). (Recall that the thermally isolated stages of the cycle are very quick.) The magnitude of τ_r essentially depends on the scenario of relaxation.

First (specific) scenario. We can assume that the Hamiltonian (67, 72) of the heat engine is known. Then there exist system-bath interaction scenarios that lead to $\tau_r = \mathcal{O}(\sqrt{\ln[n]})$; see section VI of [36]. Hence for this type of relaxation the Carnot efficiency is achievable at a large power $\mathcal{O}(\sqrt{\ln[n]}) \gg 1$; see (77). However, in these scenarios the system-bath interaction Hamiltonian (that governs the relaxation) is special: it depends on the engine Hamiltonian (67, 72).

Second (realistic) scenario. Assuming that the system-bath interaction does not depend on the Hamiltonian (67, 72), we can estimate τ_r within the weak-coupling, Markov master equation approach that leads to $\tau_r = \mathcal{O}(n)$; see section III of [36]. For a qualitative understanding of this situation, consider the relaxation as a random walk in the energy space, e.g. in the second step of the cycle, where the engine starts with almost unpopulated ground state, and it has to achieve ground state probability ≈ 1 after relaxation; see (76). So, if every transition from one excited energy level to another takes a finite time, one will need to perform in average $n/2$ transitions before jumping to the ground state. Now note from (78) that the convergence of η to η_C is controlled by $\mathcal{O}(1/\ln[n])$: a small step towards η_C will lead to a large increase in τ_r nullifying the power $\mathcal{O}(\ln[n]/n)$ for $n \gg 1$; see (77). Hence for this type of relaxation the Carnot efficiency is not achievable at a finite power.

The second relaxation scenario of the system with Hamiltonian (67, 72) is similar to the known combinatorial optimization problem: finding a marked item in an unstructured database [32] of $n + 1$ items. This problem is mapped to physics by associating each item to an eigenvector of a Hamiltonian [32]. The marked item relates to the lowest energy level 0, while all other (excited) eigenvalues of the Hamiltonian ϵ are equal. The resulting system has unknown eigenvectors of the Hamiltonian, but known eigenenergies. Now the searching process can be organized as a relaxation of the system from an excited state to a low-temperature equilibrium state. This state is dominated by the ground level due to a large ϵ . Once the relaxation is over, the unknown item (eigenvector) can be revealed by measuring the energy.

For classical algorithms the search time of this problem scales as $\mathcal{O}(n)$ for $n \gg 1$ [32]. It is thus not much better than going over all possible candidates for the solution, a typical situation of a computationally complex problem. For quantum algorithms (Grover's search) the search time scales as $\mathcal{O}(\sqrt{n})$ [32]. This is still not suitable for our purposes, since it nullifies the power for $\ln[n] \gg 1$.

E. Sub-optimal engine

Within the second (realistic) relaxation scenario, we shall modify the optimal engine so that the power is finite, but both the work and efficiency are still large. We are guided by the analogy between the relaxation of the Hamiltonian (67, 72) under the second scenario and the Levinthal's paradox from protein physics [33]. In fact, (67, 72) is the simplest model employed for illustrating the paradox; see [33, 34] and section V of [36]. Here the ground state refers to the unique folded (native) state. To ensure its stability, it is separated by a large gap from excited (free) energy levels. The essence of the paradox is that assuming many *equivalent* unfolded (excited) states, the relaxation time to the native state is unrealistically long. Recall that the states ρ_1 and ρ_2 of the optimal engine refer respectively to unfolded and folded states of the protein model. Indeed $nu_\alpha/(1+nu_\alpha)$ ($\alpha = 1, 2$) is the overall probability of the excited levels; see (72). Hence for $\ln[n] \gg 1$ the ground state (excited levels) dominates in ρ_2 (ρ_1); see (76).

The resolution of the paradox is to be sought via resolving the degeneracy of excited levels: if there are energy differences, some (unfavorable) transitions will not be made shortening the relaxation time [33, 34]. In resolving the energy degeneracy we follow the simplest model proposed in [33].

The model has $N \gg 1$ degrees of freedom $\{\sigma_i\}_{i=1}^n$; each one can be in $\zeta + 1$ states: $\sigma_i = 0, \dots, \zeta$. Whenever $\sigma_i = 0$ for all i 's, the model protein is in the folded (ground) state with energy zero [33, 34]. The ground state has zero energy. Excited states with $s \geq 1$ have energy $\epsilon^{[\alpha]} + s\delta^{[\alpha]}$, where $\epsilon > 0$ and s is the number of (misfolded) degrees of freedom with $\sigma_i \neq 0$. $\delta > 0$ is the parameter that (partially) resolves the degeneracy of excited states; we revert to the previous, work-optimal, model for $\delta \rightarrow 0$. For *different* eigenvalues of the Hamiltonian H_α we have

$$\left\{ (1 - \delta_{\text{Kr}}[s, 0]) (\epsilon^{[\alpha]} + s\delta^{[\alpha]}) \right\}_{s=0}^N, \quad \alpha = 1, 2, \quad (79)$$

where $\delta_{\text{Kr}}[s, 0]$ is the Kronecker delta, and where each energy $\epsilon^{[\alpha]} + s\delta^{[\alpha]}$ is degenerate $\frac{\zeta^s N!}{s!(N-s)!}$ times; thus the total number of energy levels is $(1 + \zeta)^N$.

Given (79), the cycle consists of two isothermal and two thermally isolated pieces with sudden changes $(\delta^{[1]}, \epsilon^{[1]}) \rightarrow (\delta^{[2]}, \epsilon^{[2]}) \rightarrow (\delta^{[1]}, \epsilon^{[1]})$; see (59–71). Below we shall also assume

$$\beta_1 \delta^{[1]} = \beta_2 \delta^{[2]}, \quad (80)$$

TABLE I: Parameters of the sub-optimal engine: work W , efficiency η and the cycle time $2\tau_r$; see (81–85). W_{\max} is the maximal work extracted for the optimal engine at a vanishing power; see (73, 74). For the sub-optimal engine: $K = \zeta e^{-\beta_1 \delta^{[1]}}$, $N = 140$, $\zeta = 4$, $T_1 = 1$, $T_2 = 1/2$. Carnot and Curzon-Ahlborn efficiencies are, respectively, $\eta_C = 1/2$ and $0.5858 \eta_C$. Also, $p_1^{[\alpha]} = [1 + (1 + K)^N e^{-\beta_\alpha \epsilon^{[\alpha]}}]^{-1}$ ($\alpha = 1, 2$) are the ground-state probabilities of $\rho_\alpha \propto e^{-\beta_\alpha H_\alpha}$; see (79).

K	τ_r	W/W_{\max}	W	η/η_C	$p_1^{[1]}$	$p_1^{[2]}$
0.1	4.45×10^{-5} s	0.2267	23.52	0.8751	0.0392	0.9808
0.2	4.35 s	0.3884	40.3	0.9110	0.0237	0.9883
0.24	357 s	0.4393	45.58	0.9181	0.0210	0.9896

because this makes the sub-optimal engine structurally very similar to the optimal one. Now the work $W = W_1 + W_2$ is calculated from (61, 65, 66, 79, 80):

$$W[v_1, v_2; K] = \frac{m(\Delta\epsilon + \frac{KN\Delta\delta}{1+K})(v_1 - v_2)}{(1 + mv_1)(1 + mv_2)}, \quad (81)$$

$$\Delta\epsilon = \epsilon^{[1]} - \epsilon^{[2]} = T_1 \ln[1/v_1] - T_2 \ln[1/v_2], \quad (82)$$

$$\Delta\delta = \delta^{[1]} - \delta^{[2]} = (T_1 - T_2) \ln[\zeta/K]. \quad (83)$$

where $K = \zeta e^{-\beta_1 \delta^{[1]}}$, $m = (1 + K)^N$, and where $v_\alpha \equiv e^{-\beta_\alpha \epsilon^{[\alpha]}}$ ($\alpha = 1, 2$) are determined from maximizing (81). Note the analogy between (73) and (81), with m being an analogue of n ; they get equal for $\delta \rightarrow 0$. Note that in (81) we neglected factor $\mathcal{O}(\frac{1}{m})$ assuming that $m \gg 1$.

Likewise, we get for the efficiency [cf. (75)]:

$$\eta = 1 - \frac{T_2}{T_1} \times \frac{\ln \frac{1}{v_2} + \frac{NK \ln(\zeta/K)}{1+K}}{\ln \frac{1}{v_1} + \frac{NK \ln(\zeta/K)}{1+K}}. \quad (84)$$

For this model [33] assumes a local Markov relaxation dynamics, where each degree of freedom makes a transition $\sigma_i \rightarrow \sigma_i \pm 1$ in 10^{-9} seconds; this value is chosen conventionally to fit experimental magnitudes for the elementary dynamic step [33]. The model has a single relaxation time [33] that is easily reproduced in the general master-equation framework (see section IV of [36]):

$$\tau_r = 10^{-9}(1 + K)^N / (NK) \text{ seconds}, \quad (85)$$

where the factor N is due to the N -times degenerate first excited level.

For $\delta^{[\alpha]} \rightarrow 0$ ($\alpha = 1, 2$), where the excited energy levels become degenerate, $\tau_r \propto (1 + \zeta)^N$ scales linearly over the number of energy levels, as expected. When $\delta^{[\alpha]}$ are not zero, τ_r can be of order of 1 second for $N \sim 100$, because $1 + K$ is close to 1. However, for the macroscopic situation ($N \sim 10^{23}$) τ_r is still huge. In this sense, the model is incomplete, but still useful for analyzing the mesoscopic situation $N \sim 100$ that is relevant for the protein folding problem [34].

Table I illustrates the characteristics of the sub-optimal engine and compares them with those of the optimal one. Reasonable cycle times can coexist with a finite fraction ($\sim 40\%$) of the maximal work and with sizable efficiencies ($\sim 90\%$ of the maximal value) that are larger than the Curzon-Ahlborn value. Hence, albeit within the second (realistic) scenario it is impossible to approach the maximal efficiency as close as desired, reasonably large efficiencies at a finite (or even large) power are possible.

Problem 1

Discuss the standard thermodynamic Carnot cycle and argue that it reaches the maximal efficiency at zero power.

Problem 2

Estimate the relaxation time of the optimal engine via master equation.

Solution.

1. Consider a system with $n \gg 1$ degenerate (excited) levels with energy $\varepsilon > 0$, and a single ground state with energy 0. In this section we shall estimate the relaxation time of this system within the Markov master-equation framework. We start with simplifying assumptions for showing the origin of a long relaxation time in this system. We then demonstrate the result at more general level.

Let $\{p_i\}_{i=0}^n$ be the probability of energy levels. The master equation reads

$$\dot{p}_0 = \sum_{i=1}^n w_{0i} p_i - p_0 \sum_{i=1}^n w_{i0}, \quad (86)$$

where w_{0i} is rate of the transition $i \rightarrow 0$. Since all energy levels besides the lowest one have the same energy ε , the detailed balance condition reads (which reflects the fact that the bath is in equilibrium at temperature $1/\beta$):

$$w_{0i} e^{-\beta\varepsilon} = w_{i0}. \quad (87)$$

2. Let us now assume that in (86) all the excited energy levels are equivalent, and hence

$$w_{0i} = w_{01} \quad \text{and} \quad w_{i0} = w_{10}, \quad (88)$$

do not depend on i' (this assumption is relaxed at the end of the section):

$$\dot{p}_0 = -w_{01} p_0 [ne^{-\beta\varepsilon} + 1] + w_{01}, \quad (89)$$

meaning that the relaxation time τ_r for $p_0(t)$ to converge exponentially to its equilibrium value

$$p_0^{[\text{eq}]} = 1/(1 + ne^{-\beta\varepsilon}) \quad (90)$$

reads

$$1/\tau_r = w_{01}[ne^{-\beta\varepsilon} + 1]. \quad (91)$$

A popular choice for the rate is given by the transition state theory [77]:

$$w_{01} = \frac{\kappa}{n+1} e^{-\beta(\varepsilon^* - \varepsilon)}, \quad (92)$$

where $\varepsilon^* > \varepsilon$ is the transition state energy: once the system gets at that state, it has equal probability to move to any state; hence the factor $\frac{1}{n+1}$ in (92). Here κ does not depend on n ; it is determined by the the energy landscape in the vicinity of the transition state and the excited state.

In our situation

$$e^{-\beta\varepsilon} = \mathcal{O}(\ln[n]/n) \quad \text{or} \quad e^{-\beta\varepsilon} = \mathcal{O}\left(\frac{1}{n \ln[n]}\right). \quad (93)$$

Eqs. (92, 91) then imply that the relaxation time τ_r is roughly (neglecting logarithmic factors)

$$\tau_r = \mathcal{O}(n) \gg 1. \quad (94)$$

This conclusion is then based on two physical aspects: first that there are many states with the same energy [the factor $\frac{\kappa}{n+1}$ in (92)]. Second is that the energy gap is large; see (93).

3. For more general (than the transition state theory) choices of w_{01} we still shall obtain the same result if we require that for any excited state energy $\varepsilon > 0$, (89) produces a well-defined and finite limit for $n \rightarrow \infty$. This request is based on two hidden assumptions: *i*) the master equation is derived within the weak-coupling assumption, hence it cannot contain very fast (in the limit $n \rightarrow \infty$) characteristic times; *ii*) the limit $n \rightarrow \infty$ can be taken independently from ε . Now we naturally get that w_{01} and w_{10} have an overall dependence $\mathcal{O}(\frac{1}{n})$:

$$w_{01} = \hat{w}_{01}(\varepsilon)/n, \quad w_{10} = \hat{w}_{10}(\varepsilon)/n, \quad (95)$$

where $\hat{w}_{10}(\varepsilon)$ and $\hat{w}_{01}(\varepsilon)$ can depend on energy ε , but they do not depend on n directly [cf. (92)]. Then (91) assumes a finite limit for $n \gg 1$. We now get from (95, 93):

$$1/\tau_r = \hat{w}_{01}(\varepsilon) \times \mathcal{O}\left(\frac{1}{n}\right). \quad (96)$$

It is natural to assume that for $\varepsilon \rightarrow \infty$, $\hat{w}_{01}(\varepsilon)$ stays at least bounded [cf. (92, 93)], and then we are back to (94). There is an important relaxation scenario (going back to Arrhenius and improved by Kramers) where \hat{w}_{01} does not depend on ε (provided that $\varepsilon > 0$ is sufficiently large); it is given by $\hat{w}_{01} \propto e^{-\beta V}$, where $V > 0$ is the barrier height [77].

Thus we note that the conclusion (94) on long characteristic times is not completely straightforward and—if taken out of the usual relaxation theories, e.g. the transition state theory or the Arrhenius theory—it requires several hidden assumptions. In section V of this supplementary material we show that (94) is violated, and the relaxation time can be much shorter, if allow the system-bath interaction to depend on the system features.

4. Finally, let us return to (88) and show that this assumption can be relaxed without changing our main conclusions. For a sufficiently large energy gap $\varepsilon > 0$ between the ground state and the excited state, we can apply the adiabatic approximation meaning that the excited levels probability equilibrate between themselves much quicker than the ground state level probability. Hence they all get into the same value:

$$p_i(t) = (1 - p_0(t))/n, \quad (97)$$

before $p_0(t)$ start to change appreciably. Employing (97) in (86) we revert to (86), where now instead of w_{01} we should employ $\frac{1}{n} \sum_{i=1}^n w_{0i}$. Note that (97) is especially plausible in our situation, since we also start the relaxation process from equilibrium states at a temperature different from the bath temperature. At such an initial state the probabilities of the excited levels are equal.

5. To go beyond the adiabatic approximation we will now consider the whole master equation. Let us first note that the detailed balance conditions (87) for the transition rates w_{ij} between excited levels lead to $w_{i \neq j} = w_{ji}$; and write down the rest of the master equation (86)

$$\dot{p}_i = \sum_{j \neq i} w_{ij} p_j - \sum_{j \neq i} w_{ji} p_i - w_{0i} p_i + w_{i0} p_0 \quad (98)$$

in the following form:

$$\dot{p}_i = \sum_j \bar{w}_{ij} p_j - w_{0i} p_i + e^{-\beta\varepsilon} w_{0i} p_0. \quad (99)$$

Where the matrix \bar{w} is defined as: $\bar{w}_{ii} = -\sum_{j \neq i} w_{ij}$, $\bar{w}_{i \neq j} = w_{ij}$; and is a symmetric matrix satisfying $\sum_i w_{ij} = 0$ for $\forall j$ and can be shown to be non-positive.

We now form the quantities $Q_i = p_i - e^{-\beta\varepsilon} p_0$, for which the master equation (86, 99) produces

$$\dot{Q}_i = \sum_j \bar{w}_{ij} Q_j - w_{0i} Q_i - e^{-\beta\varepsilon} \sum_j w_{0j} Q_j. \quad (100)$$

Introducing the positive and symmetric matrix $W_{ij} = -\bar{w}_{ij} + \delta_{ij} w_{0i}$ and bra-ket notation for vectors, we rewrite (100) as

$$\frac{d}{dt} |Q\rangle = - (W + e^{-\beta\varepsilon} |1\rangle\langle w|) |Q\rangle, \quad (101)$$

where $|1\rangle = (1, \dots, 1)$ and $\langle w| = (w_{01}, \dots, w_{0n})$. In the equilibrium state $Q_i = 0$.

Now the relaxation time of the system can be estimated via the minimal eigenvalue ω_{\min} of the matrix $W + e^{-\beta\varepsilon} |1\rangle\langle w|$. Below, we show numerically that in the asymptotic limit of $n \rightarrow \infty$,

$$\omega_{\min} = (1 + ne^{-\beta\varepsilon}) \frac{\sum_i w_{0i}}{n} + \frac{f_n}{n^2}, \quad (102)$$

$$f_n = \mathcal{O}(1), \quad (103)$$

where f_n does not depend on n for a sufficiently large n . Hence for $n \gg 1$ we neglect the last term in (102) and obtain for the relaxation time:

$$\frac{1}{\tau_r} = (1 + ne^{-\beta\varepsilon}) \frac{\sum_i w_{0i}}{n}, \quad (104)$$

confirming that conditions (95, 93) imply the relaxation time to grow proportionally with n . Eq. (104) is the same result as was obtained above via the adiabatic approximation. Hence this approximation holds up to the second-order term of the asymptotic expansion of ω_{\min} in terms of n .

1. Problem 3

What is the protein folding problem?

Solution.

1. Early experiments have shown that proteins can fold (i.e. reach the native, functional state) in a reasonably short time; see [78]. Moreover, they do so spontaneously (without external guidance) and starting from different initial conditions. These experiments created what is known to be the modern thermodynamic paradigm on the protein folding: the native state corresponds to the (relatively) unique global minimum of free energy [78].

2. It however still remained unclear how specifically proteins fold, i.e. what is their kinetics. Levinthal assumed that all unfolded states (conformations) are more or less equivalent [79]. Hence during the relaxation to the folded state all possible conformations are tried out to find the energetically most favorable one. This will take an enormous amount of time, because for a (hypothetical) small protein with 100 residues, the number of possible unfolded states would be about 3^{100} [79]. Since this conclusion is clearly unsupportable (hence the Levinthal's paradox), there should be some structure in the set of unfolded states that makes them non-equivalent.

3. One (by now classical) view suggests that in the course of its relaxation the protein passes through a unique path of partially unfolded intermediate states [80]. They allow stepwise folding, drastically reducing the scale of conformational search. In that view—which was supported by experiments on sufficiently long proteins [80]—the protein folding problem is reduced to relaxation in a finite number of states [78]. Still it was unclear how the protein reaches one of those intermediates, since now the Levinthal's paradox can be reformulated with respect to partial relaxation.

4. In addition to the latter objection, experiments on short proteins have shown that folding intermediates are absent. Hence a new view emerged that explains the protein folding kinetics as taking place on a funnel (free) energy landscape, where different unfolded states have different (free) energies and it is this difference that drives the protein towards the minimum (free) energy state [33, 81]. An example of the above scenario is provided by the Zwanzig's model [33], as reviewed above.

Still for realistic proteins the Zwanzig's model is clearly oversimplified, e.g., it does not include conformational entropy, disorder, residue sequence, *etc.* Thus for a deeper understanding of the protein folding one should go to more realistic models [81], which however share the two main points of the the Zwanzig's model: *i)* unfolded states have different (free) energies; *ii)* the folding time can be made finite via fine-tuning only for sufficiently short proteins having 100 – 200 coarse-grained degrees of freedom. It is argued that longer proteins will fold hierarchically, i.e. first certain domains will fold independently from each other (these domains thus play the role of folding intermediates), and only after that the protein will relax globally. The existence of fine-tuning is explained via evolution, a notorious solver of difficult problems in biology [78].

V. STATISTICAL THEORY OF IDEAL QUANTUM MEASUREMENT PROCESSES

One of the main foundational challenges of quantum theory is the so-called measurement problem: Why does each individual run of an ideal measurement yield a well-defined outcome, in spite of the existence of quantum coherences? Does measurement theory require a specific principle or interpretation of quantum mechanics? Already raised by the founding fathers, this crucial question has witnessed a revival [38–44]. Many authors explore models, others propose alternative interpretations of quantum mechanics or go beyond it.

As in [45, 46], we will approach the problem through a standard minimalist quantum approach, by analysing the dynamics of the isolated system constituted by the tested system S coupled to the measuring apparatus A. A key point is the *macroscopic size* of the apparatus, which forces us to rely on non-equilibrium quantum statistical mechanics. Moreover, being irreducibly probabilistic, quantum physics does not describe individual objects; we must deal with *statistical ensembles* (§1), then attempt to infer the properties of a single run of the measurement from those of the subensembles of runs in which it may embedded.

We deal with ideal, non demolishing measurements. Their purpose is to test a single observable $\hat{s} = \sum_i s_i \hat{\Pi}_i$ of S characterised by its eigenvalues s_i and the associated eigenprojectors $\hat{\Pi}_i$, while perturbing S minimally. Although ideal measurements are not currently performed in real experiments, their theoretical elucidation is a necessary step towards a full understanding of actual measurement processes, since any general quantum measurement (POVM) can be represented as a partial trace over an ideal measurement [6]. We denote by $\hat{\mathcal{D}}(t)$ the joint density operator of S+A for a large ensemble of runs, and by $\hat{r}(t) = \text{tr}_A \hat{\mathcal{D}}(t)$ and $\hat{\mathcal{R}}(t) = \text{tr}_S \hat{\mathcal{D}}(t)$ the marginal density operators of S and A, respectively. At the initial time $t = 0$, S and A are uncorrelated, S lies in some state $\hat{r}(0)$ and A in a metastable state $\hat{\mathcal{R}}(0)$, so that $\hat{\mathcal{D}}(0) = \hat{r}(0) \otimes \hat{\mathcal{R}}(0)$. Triggered by its coupling with S, A may relax towards one or another among its stable states $\hat{\mathcal{R}}_i$, which should have equal entropies and energies so as to avoid bias in the measurement. These states can be distinguished from one another through observation or registration of the pointer variable A_i , identified with the expectation value $\text{tr}_A \hat{\mathcal{R}}_i \hat{A}$ of some collective observable \hat{A} of A. The final indication A_i of the pointer must be fully correlated with the eigenvalue s_i of the tested observable \hat{s} and with the production of the final state \hat{r}_i for S. The analysis of the process should therefore explain how S+A, starting from the initial state $\hat{\mathcal{D}}(0)$, reaches at a final time t_f , *for each run of the measurement*, one among the states [47]

$$\hat{\mathcal{D}}_i = \hat{r}_i \otimes \hat{\mathcal{R}}_i, \quad p_i \hat{r}_i = \hat{\Pi}_i \hat{r}(0) \hat{\Pi}_i, \quad (105)$$

with Born's probability $p_i = \langle \hat{\Pi}_i \rangle = \text{tr}_S \hat{r}(0) \hat{\Pi}_i$.

Some parts of this task have been achieved for various models. Here we deal with arbitrary ideal measurement processes, gathering the key ideas that underlie their full dynamical solution. Due to the generality of our scope, we can describe only qualitatively this solution, but we will demonstrate the feasibility of the program by recalling in footnotes the outcomes of the detailed dynamical solution [43, 48] of the Curie–Weiss (CW) model⁷.

1. Statistical formulation of the principles of quantum mechanics. We tackle the measurement problem within a statistical formulation of quantum mechanics, also called “statistical interpretation” or “ensemble interpretation” [38]. It is a *minimalist*, formal description suited to both microscopic and macroscopic systems. Its spirit is the same as in the C^* -algebraic approach [44], although we deal with finite non relativistic systems. It does not prejudge any specific interpretation of quantum oddities [6]. Physical interpretations should emerge at the macroscopic scale, in experimental contexts, as will be discussed in §5.

Physical quantities pertaining to a system (or rather to an ensemble of similar systems) are described as “observables” represented by Hermitean matrices in a Hilbert space. Observables behave as random objects, but, unlike ordinary random variables, their randomness arises from their non-commutative nature and is inescapable.

A “quantum state”, whether pure or not, encompasses the probabilistic predictions that one may make about the various observables. It is characterised by a correspondence between observables \hat{O} and real numbers, implemented as $\hat{O} \mapsto \langle \hat{O} \rangle = \text{tr} \hat{\mathcal{D}} \hat{O}$ by means of a Hermitean, normalised and nonnegative density operator $\hat{\mathcal{D}}$. Such a definition looks analogous to that of a state in classical statistical mechanics, encoded in a density in phase space. However, “quantum expectation values” $\langle \hat{O} \rangle$, “quantum probabilities” such as $\langle \hat{\Pi}_i \rangle$ and “quantum correlations” such as $\langle \hat{s} \hat{A} \rangle$

⁷ In the CW model [6, sect. 3], S is a spin $\frac{1}{2}$, the measured observable being its z-component \hat{s}_z , with outcomes $i = \uparrow$ or \downarrow . The apparatus simulates a magnetic dot, including $N \gg 1$ spins $\hat{\sigma}^{(n)}$, which interact through the Ising coupling J , and a phonon thermal bath at temperature $T < J$; these spins and the phonons are coupled through a dimensionless weak coupling γ . Initially prepared in its metastable paramagnetic state, A may switch to one or the other stable ferromagnetic state. The pointer observable $\hat{A} = N \hat{m} = \sum_{n=1}^N \hat{\sigma}_z^{(n)}$ is the total magnetisation in the z-direction of the N Ising spins. The coupling between S and A is $\hat{H}_{SA} = -\sum_{n=1}^N g \hat{s}_z \hat{\sigma}_z^{(n)}$, while $\hat{H}_S = 0$.

present a priori only a formal similarity with standard expectation values, probabilities and correlations; fluctuations arise not only from some ignorance but also from the operator character of physical quantities.

As a standard probability distribution, a quantum state gathers information and refers, implicitly or not, to a statistical ensemble \mathcal{E} : We should regard an individual system as embedded in a large, real or virtual, ensemble \mathcal{E} of systems produced under the same conditions. So a state does not “belong to a system”, it is not an intrinsic property. Information may be updated as for ordinary probabilities by acknowledging and sorting the outcomes of measurements so as to split \mathcal{E} into subensembles, to each of which a new state will be assigned (§5).

2. System plus apparatus in thermodynamic equilibrium. We suppose that the compound system S+A is isolated, including in A a thermal bath or an environment if present. The properties of S+A are governed by the Hamiltonian $\hat{H} = \hat{H}_S + \hat{H}_A + \hat{H}_{SA}$, which must have specific features. If A is decoupled from S, the part \hat{H}_A governing the macroscopic apparatus should produce an initial metastable state $\hat{\mathcal{R}}(0)$ with lifetime longer than the duration of the measurement and several thermodynamic equilibrium states $\hat{\mathcal{R}}_i$, the possible final states. A typical example¹ is given by spontaneously broken discrete invariance, the pointer variable A_i being the order parameter.

As we wish to deal with ideal measurements, the process should perturb S as little as possible: any observable compatible with \hat{s} , i.e., commuting with its eigenprojectors $\hat{\Pi}_i$, should remain unaffected. The conservation of all these observables [47] is expressed by the fact that \hat{H} depends on S only through the projectors $\hat{\Pi}_i$. In particular the coupling must have the form $\hat{H}_{SA} = \sum_i \hat{\Pi}_i \otimes \hat{h}_i$, where \hat{h}_i are operators of A. Moreover, if \hat{s} takes the value s_i , that is, $\hat{\Pi}_i$ the value 1, A should end up near its stable state $\hat{\mathcal{R}}_i$, the pointer variable being close to A_i . This can be achieved if each \hat{h}_i behaves as a source that breaks explicitly the invariance⁸ by energetically favouring A_i .

Before analysing the dynamics of the measurement process (§3 and §4), we determine for orientation the general form $\hat{\mathcal{D}}_{eq}$ of the thermodynamic equilibrium states associated with the Hamiltonian \hat{H} of S+A. We rely on the maximum von Neumann entropy criterion for assigning a state to a system in equilibrium [49, 50]. We thus maximise $-\text{tr}\hat{\mathcal{D}} \ln \hat{\mathcal{D}}$ under constraints accounting for the conserved quantities. As usual for isolated macroscopic systems, thermal equilibrium is implemented by specifying the average energy $\langle \hat{H} \rangle$. The other constants of the motion, to wit, the expectation values of all observables of S that commute with the projections $\hat{\Pi}_i$, are accounted for by associating a Lagrange multiplier with each of them.

This yields for the equilibrium states a Gibbsian form where the exponent involves an arbitrary linear combination of all the conserved observables, i.e., $\hat{\mathcal{D}}_{eq} \propto \exp(-\beta \hat{H} + \sum_i \hat{y}_i)$ where \hat{y}_i is any operator of S acting inside the diagonal block i (so that $\hat{y}_i = \hat{\Pi}_i \hat{y}_i \hat{\Pi}_i$). Noting that the exponent, which commutes with the projections $\hat{\Pi}_i$, has a block diagonal structure, we find for these equilibrium states the related block diagonal structure

$$\hat{\mathcal{D}}_{eq} = \sum_i q_i \hat{x}_i \otimes \hat{\mathcal{R}}_i^h, \quad \sum_i q_i = 1. \quad (106)$$

Each factor $q_i \hat{x}_i$, which arises from $\exp(-\beta \hat{H}_S + \hat{y}_i)$, is an arbitrary block diagonal operator of S, where $\hat{x}_i = \hat{\Pi}_i \hat{x}_i \hat{\Pi}_i$, $\text{tr}_S \hat{x}_i = 1$ behaves as a density operator of S. (If the eigenvalue s_i is non degenerate, \hat{x}_i reduces to $\hat{\Pi}_i$.) Each factor $\hat{\mathcal{R}}_i^h \propto \exp[-\beta(\hat{H}_A + \hat{h}_i)]$ in (106) can be interpreted as a canonical equilibrium density operator in the space of A. Here the invariance is *explicitly broken* by adding to the Hamiltonian of A the source term \hat{h}_i arising from \hat{H}_{SA} .

This term should be sufficiently large so that the distribution $\text{tr}_A \hat{\mathcal{R}}_i^h \delta(A - \hat{A})$ of \hat{A} has a single narrow peak, and sufficiently small so that this peak lies close to A_i , so as to let \hat{A} behave as a pointer observable. These properties are easy to satisfy for a macroscopic apparatus. Thermodynamic equilibrium (106) thus entails a complete correlation between the eigenvalue s_i of \hat{s} and the value A_i of the pointer variable. The moderate size of \hat{h}_i also ensures that the state $\hat{\mathcal{R}}_i^h$ lies in the basin of attraction of the state $\hat{\mathcal{R}}_i$ with broken symmetry, so that $\hat{\mathcal{R}}_i^h$ will relax smoothly to $\hat{\mathcal{R}}_i$ at the end of the measurement process when the coupling \hat{H}_{SA} is switched off.

Let us return to measurements. We wish to explain how a final state $\hat{\mathcal{D}}_i$ of the form (1) should be assigned with probability p_i to each run. It is thus necessary (but not sufficient) to prove, by studying the dynamics of a large

⁸ The interaction Hamiltonian \hat{H}_{SA} allows to describe not only ideal measurements involving well separated eigenvalues s_i of \hat{s} , but also more general measurements for which the projectors $\hat{\Pi}_i$, still associated through \hat{h}_i with the pointer indications A_i , are no longer in one-to-one correspondence with the eigenvalues of \hat{s} . For instance, if some $\hat{\Pi}_i$ encompasses the eigenspaces of several different neighbouring eigenvalues, selecting the outcome A_i will not discriminate them, and the final state $\hat{r}_i = \hat{\Pi}_i \hat{r}(0) \hat{\Pi}_i / p_i$ of S will not be associated with a single eigenvalue of \hat{s} as in an ideal measurement. As another example, consider two orthogonal rank-one projectors $\hat{\Pi}_1$ and $\hat{\Pi}_2$, coupled with sources \hat{h}_1 and \hat{h}_2 that produce different outcomes A_1 and A_2 , and assume that $\hat{\Pi}_1 + \hat{\Pi}_2$ spans the two-dimensional eigenspace associated with a degenerate eigenvalue of \hat{s} ; reading the outcome A_1 (or A_2) then provides more information than this eigenvalue.

statistical ensemble \mathcal{E} of runs for which S+A lies initially in the state $\hat{\mathcal{D}}(0) = \hat{r}(0) \otimes \hat{\mathcal{R}}(0)$, that its final state is

$$\hat{\mathcal{D}}(t_f) = \sum_i p_i \hat{\mathcal{D}}_i = \sum_i p_i \hat{r}_i \otimes \hat{\mathcal{R}}_i. \quad (107)$$

We can identify (107) with a thermodynamic equilibrium state (106), for which relaxation of $\hat{\mathcal{R}}_i^h$ to $\hat{\mathcal{R}}_i$ has taken place after switching off \hat{H}_{SA} . Dynamics and conservation laws will determine the free parameters of $\hat{\mathcal{D}}_{eq}$ from the initial state $\hat{\mathcal{D}}(0)$ as $q_i \hat{x}_i = \hat{\Pi}_i \hat{r}(0) \hat{\Pi}_i \equiv p_i \hat{r}_i$. i.e., $q_i = p_i$, $\hat{x}_i = \hat{r}_i$.

We can also identify the state $\hat{\mathcal{D}}_i$ defined in (105), expected to describe the subensemble \mathcal{E}_i obtained by selecting within \mathcal{E} the runs tagged by the value A_i of the pointer, with an equilibrium state for which all q_j with $j \neq i$ vanish. More generally, for an *arbitrary subset* \mathcal{E}_{sub} of runs⁹ having yielded a proportion q_i of individual runs with outcomes A_i , the expected final state

$$\hat{\mathcal{D}}_{sub}(t_f) = \sum_i q_i \hat{\mathcal{D}}_i \quad (108)$$

is a thermodynamic equilibrium state (106) with $\hat{x}_i = \hat{r}_i$.

Thus, an ideal measurement process appears as a mere relaxation of S+A to equilibrium, for the full ensemble \mathcal{E} of runs and for arbitrary subensembles \mathcal{E}_{sub} . In quantum mechanics, relaxation of $\hat{\mathcal{D}}(t)$ and $\hat{\mathcal{D}}_{sub}(t)$ towards Gibbsian thermodynamic equilibrium states (3) and (4) is not granted [51]. We must therefore justify these properties within the quantum statistical dynamics framework. We sketch the main steps of such a proof in §3 and §4.

3. Dynamics of S+A for the full set of runs. Our first task [45] consists in deriving the final state (107) by solving the Liouville–von Neumann equation $i\hbar d\hat{\mathcal{D}}(t)/dt = [\hat{H}, \hat{\mathcal{D}}(t)]$ with initial condition $\hat{\mathcal{D}}(0) = \hat{r}(0) \otimes \hat{\mathcal{R}}(0)$. Taking into account the above form of \hat{H} and the commutation $[\hat{H}_S, \hat{r}(0)] = 0$ which ensures that the marginal state $\hat{r}(t)$ of S is perturbed only by the interaction \hat{H}_{SA} during the process, we check that $\hat{\mathcal{D}}(t)$ can be parameterised as

$$\hat{\mathcal{D}}(t) = \sum_{i,j} \hat{\Pi}_i \hat{r}(0) \hat{\Pi}_j \otimes \hat{\mathcal{R}}_{ij}(t) \quad (109)$$

in terms of a set $\hat{\mathcal{R}}_{ij}(t) = \hat{\mathcal{R}}_{ji}^\dagger(t)$ of operators in the Hilbert space of A, to be determined by the equations

$$i\hbar \frac{d\hat{\mathcal{R}}_{ij}(t)}{dt} = (\hat{H}_A + \hat{h}_i) \hat{\mathcal{R}}_{ij}(t) - \hat{\mathcal{R}}_{ij}(t) (\hat{H}_A + \hat{h}_j), \quad (110)$$

with the initial conditions $\hat{\mathcal{R}}_{ij}(0) = \hat{\mathcal{R}}(0)$. The dynamics thus involves *solely the apparatus*, irrespective of the tested system, a specific property of ideal measurements.

Though macroscopic, A should be treated as a finite system so as to ensure a better control of the dynamics and to discuss under which conditions the process can be regarded as an ideal measurement. We must then explain how the expected irreversible relaxation from $\hat{\mathcal{D}}(0)$ to $\hat{\mathcal{D}}(t_f)$ can be governed by the reversible equations (110), so that we run into the old paradox of irreversibility. As usual in statistical mechanics, it is legitimate for finite but large systems to disregard events having an extremely small probability, to forget about recurrences that would occur after large, unattainable times, and to neglect physically irrelevant correlations between a macroscopic number of degrees of freedom. Such approximations, although not exact mathematically, are fully justified when their outcome is physically indistinguishable from the exact solution. A large apparatus, and a suitable choice of parameters in \hat{H}_A and \hat{H}_{SA} will therefore be needed, for each model, to explain the required relaxations and to estimate their time scales, as will be illustrated by the CW model treated extensively in [43].

Two types of relaxation arise independently from the dynamical equations (110). (i) For $i \neq j$, the coherent contributions $\hat{\mathcal{R}}_{ij}(t)$ decay owing to the difference between \hat{h}_i and \hat{h}_j and eventually vanish. The off-diagonal blocks of the density matrix $\hat{\mathcal{D}}(t)$ are thus truncated as regards the physically attainable observables. Depending on the model, this decay may be governed by different mechanisms.

⁹ Subsets obtained by extracting runs at random from \mathcal{E} would be described by the same state $\hat{\mathcal{D}}(t)$ as the full set \mathcal{E} . If \hat{r}_i is a mixed state, the runs described by (4) are picked up at random within \mathcal{E}_i .

For $i = j$, the evolution of $\hat{\mathcal{R}}_{ii}(t)$ governed by (110) is a mere relaxation from the metastable state $\hat{\mathcal{R}}(0)$ to the equilibrium state $\hat{\mathcal{R}}_i^h$ in the presence of the source \hat{h}_i , and then to $\hat{\mathcal{R}}_i$ after \hat{H}_{SA} is switched off. The correlation between s_i and A_i needed to register the outcome is thereby established¹⁰. Thus, microscopic dynamics confirms the thermodynamic surmise (107) for the final state of S+A in the ensemble \mathcal{E} .

4. Final states for arbitrary subensembles. We have shown how $\hat{\mathcal{D}}(t)$ evolves from $\hat{\mathcal{D}}(0)$ to $\hat{\mathcal{D}}(t_f) = \sum_i p_i \hat{\mathcal{D}}_i$ for the large set \mathcal{E} of runs. Our next task consists in proving, again dynamically, that S+A ends up in a state of the form (108), $\hat{\mathcal{D}}_{\text{sub}}(t) = \sum_i q_i \hat{\mathcal{D}}_i$, for all possible subsets \mathcal{E}_{sub} of \mathcal{E} . If the density operator $\hat{\mathcal{D}}(t_f)$ did behave as an ordinary probability distribution, this result would be obvious, as the form $\sum_i p_i \hat{\mathcal{D}}_i$ of $\hat{\mathcal{D}}(t_f)$ would mean that \mathcal{E} contains a proportion p_i of individual runs having ended up in the state $\hat{\mathcal{D}}_i$; the form (108) of $\hat{\mathcal{D}}_{\text{sub}}$ would follow for any subset gathering a proportion q_i of runs i . However, in quantum mechanics, contrary to standard probability theory, specifying the state $\hat{\mathcal{D}}$ for an ensemble \mathcal{E} (with \mathcal{N} elements), that is, the correspondence $\hat{O} \mapsto \langle \hat{O} \rangle = \text{tr } \hat{\mathcal{D}} \hat{O}$, does not imply the existence of probabilities for individual systems (§1) so that the above reasoning is fallacious [52].

Indeed, when an ensemble \mathcal{E} (with \mathcal{N} elements described by $\hat{\mathcal{D}}$) gathers a subensemble \mathcal{E}_{sub} (with $0 < \mathcal{N}_{\text{sub}} < \mathcal{N}$ elements described by $\hat{\mathcal{D}}_{\text{sub}}$) and its complement, one has

$$\hat{\mathcal{D}}(t) = k \hat{\mathcal{D}}_{\text{sub}}(t) + (1 - k) \hat{\mathcal{D}}_{\text{Csub}}(t), \quad (111)$$

involving the weight $k = \mathcal{N}_{\text{sub}}/\mathcal{N}$ and the nonnegative density operator $\hat{\mathcal{D}}_{\text{Csub}}(t)$ of the complement of \mathcal{E}_{sub} .

However, if we conversely consider some decomposition of a known mixed state $\hat{\mathcal{D}}$ having the form (7), we run into a severe difficulty. Due to the matrix nature of quantum states, nothing tells us that the set \mathcal{E} described by $\hat{\mathcal{D}}$ may be split into two subsets that would be described by the two terms $\hat{\mathcal{D}}_{\text{sub}}$ and $\hat{\mathcal{D}}_{\text{Csub}}$. We have to face a *quantum ambiguity*: A mixed state possesses $\hat{\mathcal{D}}$ of S+A many *different mathematical decompositions* (111) which are physically incompatible [52]. We cannot infer from its sole knowledge whether a term $\hat{\mathcal{D}}_{\text{sub}}$ issued from some decomposition of $\hat{\mathcal{D}}$ is a density operator describing a real subset of runs or not. (We keep the notation $\hat{\mathcal{D}}_{\text{sub}}$ also in the latter case.) In particular, the form (107) of $\hat{\mathcal{D}}$, though suggestive, is not sufficient to imply the existence of subensembles of runs that would be described by $\hat{\mathcal{D}}_i$.

In order to overcome this quantum ambiguity, we adopt the following strategy. We start from the state $\hat{\mathcal{D}}(t_{\text{split}})$, taken at a time t_{split} earlier than t_f but sufficiently late so that $\hat{\mathcal{D}}(t_{\text{split}})$ has already reached the form $\sum_i p_i \hat{\mathcal{D}}_i$, after \hat{H}_{SA} has been switched off. We consider *all mathematically allowed decompositions* of $\hat{\mathcal{D}}(t_{\text{split}})$ of the form (111), involving two nonnegative Hermitean operators. Although nothing ensures that the operators $\hat{\mathcal{D}}_{\text{sub}}(t_{\text{split}})$ thus defined are physically meaningful, we are ascertained that *their class includes all physical states* associated with real subsets of runs. Our purpose is then to show, by taking $\hat{\mathcal{D}}(t_{\text{split}})$ as initial condition and solving for $t > t_{\text{split}}$ the equations of motion for $\hat{\mathcal{D}}(t)$ governed by the Hamiltonian $\hat{H}_S + \hat{H}_A$, that *any admissible candidate* $\hat{\mathcal{D}}_{\text{sub}}(t)$ *for a physical state ends up in the expected form* $\hat{\mathcal{D}}_{\text{sub}}(t_f) = \sum_i q_i \hat{\mathcal{D}}_i$. We shall thereby have proven the relaxation towards the required equilibrium form (108) for all physical subsets of runs, although it will be impossible before the reasoning of §5 to know which among the operators $\hat{\mathcal{D}}_{\text{sub}}(t_f)$ thereby constructed are the physical ones.

We begin with the determination of the general form, issued from (111), of the initial operators $\hat{\mathcal{D}}_{\text{sub}}(t_{\text{split}})$. To simplify the discussion, we assume here the eigenvalues of \hat{s} to be non degenerate¹¹ so that $\hat{r}_i = \hat{\Pi}_i = |s_i\rangle\langle s_i|$. As A is macroscopic, the fluctuations of \hat{H}_A around $\langle \hat{H}_A \rangle$ and of the pointer observable around the macroscopic value A_i are relatively small, and we can replace in $\hat{\mathcal{D}}(t_{\text{split}})$ the canonical equilibrium states $\hat{\mathcal{R}}_i$ by microcanonical ones, $\hat{\mathcal{R}}_i^\mu$.

¹⁰ Authors do not always give the same meaning to the various words used. We term as *truncation* the disappearance of the off-diagonal blocks of the density matrix of S+A under the effect of an arbitrary mechanism, and specialise *decoherence* to the production of this effect by interaction with an environment or a thermal bath. We term as *registration* the process which leads each diagonal block to the correlated state $\hat{r}_i \otimes \hat{\mathcal{R}}_i$, and as *reduction* the transition from $\hat{r}(0)$ to some \hat{r}_i for an individual run. While much attention has been paid to the vanishing of the off-diagonal blocks, the relaxation of the diagonal blocks is too often disregarded, although it produces the correlations that ensure registration. In the CW model [6, sect. 7], this process is triggered by \hat{h}_i which makes $\hat{\mathcal{R}}(0)$ unstable and should be sufficiently large to exclude false registrations ($g \gg J/\sqrt{N}$). Later on, the relaxation of $\hat{\mathcal{R}}_{ii}(t)$ to $\hat{\mathcal{R}}_i^h$, and finally to $\hat{\mathcal{R}}_i$ after \hat{H}_{SA} is switched off, is governed by the dumping of free energy from the magnet to the phonon bath; its characteristic duration is the registration time $\hbar/\gamma(J - T)$.

¹¹ For degenerate eigenvalues s_i , the only change in the forthcoming derivation, if the states $\hat{r}_i \equiv |i\rangle\langle i|$ are pure, is the replacement of $|s_i\rangle$ by $|i\rangle$. If \hat{r}_i is mixed, we note that this operator of S is not modified by the process, while remaining fully coupled with A_i for $t > t_{\text{split}}$. We should therefore preserve this property when we split $\hat{\mathcal{D}}$ so as to build the candidates $\hat{\mathcal{D}}_{\text{sub}}$ for states of physical subensembles. The microcanonical relaxation of A then produces again the final state (10), at least for all physical subensembles.

Within the Hilbert space of A, we denote as $|A_i, \eta\rangle$ a basis of kets characterised by a value of \hat{A} close to A_i (within a small margin δA_i), and by a value of the energy corresponding to the microcanonical equilibrium state $\hat{\mathcal{R}}_i^\mu$. As the spectrum is dense, the index η may take a very large number G_i of values, and $\hat{\mathcal{R}}_i^\mu$ is expressed by

$$\hat{\mathcal{R}}_i^\mu = \frac{1}{G_i} \sum_{\eta} |A_i, \eta\rangle \langle A_i, \eta|. \quad (112)$$

The state $\hat{\mathcal{D}}(t_{\text{split}}) \simeq \sum_i p_i \hat{r}_i \otimes \hat{\mathcal{R}}_i^\mu$ thus involves only diagonal matrix elements within the Hilbert subspace $\mathcal{H}_{\text{corr}}$ of S+A spanned by the correlated kets $|s_i\rangle |A_i, \eta\rangle$. Hence, since both operators on the right side of (111) must be nonnegative, their matrix elements must lie within this Hilbert subspace $\mathcal{H}_{\text{corr}}$, so that any initial (normalised) operator constrained by (111) has the form

$$\hat{\mathcal{D}}_{\text{sub}}(t_{\text{split}}) = \sum_{i,j,\eta,\eta'} |s_i\rangle |A_i, \eta\rangle K(i, \eta; j, \eta') \langle s_j | \langle A_j, \eta'|. \quad (113)$$

The evolution for $t > t_{\text{split}}$ of the operator $\hat{\mathcal{D}}_{\text{sub}}(t)$ issued from (113) is governed by $\hat{H}_{\text{S}} + \hat{H}_{\text{A}}$, and, as in section 3, \hat{H}_{S} is ineffective. The Hamiltonian \hat{H}_{A} of the apparatus is chosen so as to involve interactions inducing among the kets $|A_i, \eta\rangle$ rapid transitions¹² that modify η without affecting the macroscopic value of A_i . Apart from the existence here of *several thermodynamic equilibrium states*¹³ labelled by i , the mechanism is the same as in the standard microcanonical relaxation [8, 53–55] which leads to the equalisation of the populations of the microstates for a single microcanonical equilibrium. Such a “quantum collisional process” is irreversible for a large apparatus. Acting on both $|A_i, \eta\rangle$ and $\langle A_j, \eta'|$ in (113), it produces over the same time scale τ_{sub} two different effects. (i) For $i \neq j$, all contributions to (113) fade out. (ii) For $i = j$, all terms such that $\eta \neq \eta'$ disappear, while the coefficients $K(i, \eta; i, \eta)$ of the various terms η all tend to one another, their sum remaining constant. The duration τ_{sub} of these relaxations being much shorter than t_f , the mechanism is already effective before t_{split} , so that anyway $\hat{\mathcal{D}}_{\text{sub}}$ reaches at the final time $t_f > t_{\text{split}} + \tau_{\text{sub}}$ the microcanonical equilibrium

$$\hat{\mathcal{D}}_{\text{sub}}(t_f) = \sum_i q_i \hat{r}_i \otimes \hat{\mathcal{R}}_i^\mu, \quad q_i = \sum_{\eta} K(i, \eta; i, \eta). \quad (114)$$

Since the above derivation holds for arbitrary operators $\hat{\mathcal{D}}_{\text{sub}}$ issued from a mathematical decomposition (111) of $\hat{\mathcal{D}}$, it encompasses *all the physical subsets* \mathcal{E}_{sub} of runs, which therefore end up in the required form (108). The coefficients q_i of the various physical subensembles are related to one another by a *hierarchic structure*: If two disjoint subensembles $\mathcal{E}_{\text{sub}}^{(1)}$ and $\mathcal{E}_{\text{sub}}^{(2)}$ of \mathcal{E} , described by $\hat{\mathcal{D}}_{\text{sub}}^{(1)}$ and $\hat{\mathcal{D}}_{\text{sub}}^{(2)}$, and having $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ elements, respectively, merge into a new subensemble \mathcal{E}_{sub} , the above proof implies for the coefficients the standard addition rule

$$[\mathcal{N}^{(1)} + \mathcal{N}^{(2)}] q_i = \mathcal{N}^{(1)} q_i^{(1)} + \mathcal{N}^{(2)} q_i^{(2)}. \quad (115)$$

5. *Emergence of classical probabilistic interpretation.* In order to elucidate the measurement problem for ideal quantum measurements, it remains to show that the operators $\hat{\mathcal{D}}_i$ are not only the building blocks of the final density operators $\hat{\mathcal{D}}(t_f)$ of S+A (associated with the full ensemble \mathcal{E} of runs) and $\hat{\mathcal{D}}_{\text{sub}}(t_f)$ (associated with its subensembles

¹² Two different mechanisms achieving such a process have been studied for the CW model [6, §11.2], and it has been shown that they produce the result (114). In the more realistic one [6, Appendices H and I], the transitions that modify η are produced by an interaction \hat{V} between the magnet and the bath which has a variance $v^2 = \text{tr } \hat{V}^2$; an average delay θ separates successive transitions. Microcanonical relaxation may take place even if \hat{V} is not macroscopic, with a variance scaling as $v \propto N^a$ ($a < 1$) for large N . For a short θ , scaling as $\theta \propto 1/N^b$ ($a < b < 2a$), the characteristic time $\tau_{\text{sub}} = \hbar^2/v^2\theta$ scales as $1/N^c$ where $c = 2a - b$, $0 < c < a < 1$; it is short compared to the registration time, which dominates t_f because registration involves a macroscopic dumping of energy from the magnet to the bath, in contrast to the present relaxation.

¹³ This type of relaxation also occurs in the dynamics of finite-temperature quantum phase transitions with *spontaneously broken invariance*, explaining within quantum theory why the order parameter may take a well-defined value in a single experiment. This analogue of the measurement problem is solved along the same lines for macroscopic systems, so that the community has rightfully not been bothered about it. But the problem remains for phase transitions in finite systems, which require an analysis of time scales.

\mathcal{E}_{sub}), but also that they describe the outcomes (1) of individual runs. However, we have stressed (§1) that quantum mechanics, in its statistical formulation, does not deal with individual systems but only provides information on statistical ensembles – possibly gedanken but physically consistent. In the strict framework of quantum statistical mechanics, the most detailed result about ideal measurements that can be derived is the structure (108) of the final density operators of S+A for all possible subensembles \mathcal{E}_{sub} of \mathcal{E} . An essential feature of this result will allow extrapolation to individual runs, to wit, the elimination owing to dynamics of the quantum ambiguity, yielding the hierarchic structure (115) of the states of the subensembles \mathcal{E}_{sub} .

Indeed, the latter structure is just the same as the one that lies at the core of the definition of ordinary probabilities when they are interpreted as relative frequencies of occurrence of individual events [56]. Once the quantum ambiguity has been removed and the property (115) established, it is natural to complement the inherently probabilistic principles of quantum mechanics with the following interpretation of the final mixed states (108) associated with the various subensembles \mathcal{E}_{sub} : A coefficient q_i that arose abstractly from the above relaxation process is identified¹⁴ as the proportion of runs having produced the outcome A_i in the subensemble \mathcal{E}_{sub} ; each \mathcal{E}_i contains Born's proportion p_i of runs. An individual run then belongs to some subset \mathcal{E}_i and can be assigned the state $\hat{\mathcal{D}}_i$, so that a *solution of the quantum measurement problem is achieved*.

The runs are tagged by the value A_i of the pointer variable, which characterises the factor $\hat{\mathcal{R}}_i$ of $\hat{\mathcal{D}}_i$ and which can macroscopically be observed or registered. By picking out the runs having yielded A_i , one extracts from the whole ensemble \mathcal{E} the subensemble \mathcal{E}_i that ends up in the state $\hat{\mathcal{D}}_i$. Two steps are thus necessary to go from the initial state $\hat{\mathcal{D}}(0)$ to $\hat{\mathcal{D}}_i$. First, the Hamiltonian evolution (irreversible since A is large) of the coupled system S+A for the full ensemble \mathcal{E} leads to the state $\hat{\mathcal{D}}(t_f)$ expressed by (107); correlations are created, and information is transferred from S to A with some loss. However, the subsequent reduction of the state from $\hat{\mathcal{D}}(t_f)$ to one of its components $\hat{\mathcal{D}}_i$ is not a consequence of some evolution, but the mere result of selecting the particular outcome A_i . This change of state, associated with the extraction from \mathcal{E} of the subset \mathcal{E}_i is merely an updating of information¹⁵, in the same way as the change of an ordinary probability distribution resulting from some gain of information. A further step is involved in the *reduction of the state of S* from $\hat{r}(0)$ to \hat{r}_i , the disregard of A after selection of A_i .

A “state” defined by the mapping $\hat{O} \mapsto \langle \hat{O} \rangle = \text{tr} \hat{\mathcal{D}} \hat{O}$ has only a mathematical nature; the *c*-numbers $\langle \hat{O} \rangle$ present only a formal analogy with ordinary expectation values. However, at the end of a measurement, the coefficients $p_i = \langle \hat{\Pi}_i \rangle = \text{trs} \hat{r}(0) \hat{\Pi}_i$ can be identified with ordinary probabilities in the frequency interpretation. Accordingly, $\langle \hat{s} \rangle = \text{tr} \hat{r}(0) \hat{s} = \sum_i p_i s_i$ and $A_i = \text{tr}_A \hat{\mathcal{R}}_i \hat{A}$ appear as ordinary statistical averages, and $\langle \hat{s}^2 \rangle - \langle \hat{s} \rangle^2$ as an ordinary variance; the correlation between \hat{s} and \hat{A} in $\hat{\mathcal{D}}(t_f)$ has a classical nature. As usual in statistical mechanics, the macroscopic behaviour of the apparatus, in particular the irreversibility of the process, emerges from the underlying reversible microscopic theory. But moreover another remarkable type of emergence occurs at the end of the measurement process: *Classical probabilities* emerge from quantum theory, although the underlying “quantum probabilities” were non-commutative and could not a priori be regarded as frequencies in the absence of an experimental context. Certain properties of S, encoded statistically in the initial state $\hat{r}(0)$ of the ensemble \mathcal{E} , are selected by the measurement context and revealed by the outcomes of the individual runs of the process.

Problem 1

Provide a simple example of quantum ambiguity.

Problem 2

Derive (106).

¹⁴ In other words, there exist subensembles \mathcal{E}_i for which all but one of the coefficients q_i vanish. This property, together with (115), was included in the definition of probabilities as frequencies [56].

¹⁵ Measurements involve both a physical process of interaction between S and A and a selection of outcomes for repeated experiments. If we do not select the indications of A, knowledge about S is updated by replacing $\hat{r}(0)$ by $\sum_i p_i \hat{r}_i$. If the tested observable is not fully specified, the least biased subsequent predictions should rely on a state obtained by averaging over all possible interaction processes. If for instance, one is aware that an ensemble of spins initially prepared in the state $\hat{r}(0)$ have been measured in some direction, but if one knows neither in which direction nor the results obtained, one should assign to the final state the density operator $\frac{1}{3}[\hat{I} + \hat{r}(0)]$ as being the best (but imperfect) description. Indeed, a quantum state is not an intrinsic property of a physical system but it reflects our knowledge about the ensemble in which it is embedded.

VI. ADIABATIC QUANTUM SYSTEMS WITH (AND WITHOUT) FEEDBACK

The adiabatic theorem of quantum mechanics governs the evolution of a quantum system subject to slowly varying external fields. Its applications span a vast array of fields, such as two-level systems (nuclei undergoing magnetic resonance or atoms interacting with a laser field), quantum phase transitions, quantum field theory (where a low-energy theory is derived by tracing out fast, high-energy degrees of freedom), and Berry's phase [37, 57–59]. This phase and the adiabatic theorem also find applications in quantum information processing [60, 61]. For a recent discussion on the validity of the adiabatic approach see [62].

A general perspective of the quantum adiabatic physics is that it studies a system subject to a slow, *open loop* (i.e., no feedback) control, where the evolution of the external fields is given *a priori* via time-dependent parameters of the system Hamiltonian. In view of numerous application of this setup, it is natural to wonder about the quantum adiabatic *closed-loop* control, where the external controlling fields evolve under feedback from the controlled quantum system. We expect to find in this case a much richer dynamic behaviour, keeping simultaneously the system-independent features of the adiabatic approach. Now any feedback needs information on the evolution of the system. This information is to be gained via measurements, which in the quantum situation are normally related with unpredictable disturbances and irreversibility. Thus quantum control has so far been restricted to open-system dynamics [63].

However, also quantum measurements can be non-disturbing, if N non-interacting quantum particles (spins, *etc*) couple to the proper measuring apparatus. For $N \gg 1$ [analog of the classical limit] one can measure single-particle observables (almost) without disturbing the single-particle density matrix, since the disturbance caused by such measurements scales as $\frac{1}{N\sigma^2}$, where σ is the measurement precision [64]. The knowledge of these observables allows to implement feedback [65]. Non-disturbing measurements on ensembles of few-level systems are routinely employed in NMR physics (e.g., in ensemble computation) and quantum optics [66–68]. An experimental example of realizing feedback in NMR physics is presented in [68].

Here we develop an adiabatic theory under feedback obtained via such non-disturbing measurements. We intend to show that the inclusion of the feedback in quantum adiabatic dynamics leads to much richer dynamic behavior (non-linear replicator equation for the adiabatic probabilities, new gauge-invariant adiabatic phase), while still keeping the main attractive point of the adiabatic approach that is the ability to gain important information without actually solving equations of motion.

A. Basic equations

Consider a d -level quantum system described by a pure state $|\psi\rangle$ (generalization to mixed states is indicated below). The system evolves according to the Schrödinger equation with Hamiltonian $H[R(t)]$, where $R(t)$ is a classical controlling parameter ($\hbar = 1$):

$$i|\dot{\psi}(t)\rangle = H[R(t)]|\psi(t)\rangle. \quad (116)$$

By means of a continuous non-disturbing measurement performed on an ensemble of identical, non-interacting systems (each one described by $|\psi(t)\rangle$) one finds the average $\langle\psi(t)|A|\psi(t)\rangle$ of a monitoring observable A (in NMR physics A typically corresponds to the magnetization). This average enters the feedback dynamics of R

$$\dot{R} = \varepsilon F(R, \langle\psi(t)|A|\psi(t)\rangle), \quad (117)$$

where $\varepsilon \ll 1$ is a small dimensionless parameter. We assume that $|F(., .)|$ is bounded from above, which means that R is a slow variable: its derivative is bounded by a small number. For $F = F(R)$ (no feedback) we recover the standard adiabatic setup. The dynamics (116) conserves the purity of $|\psi(t)\rangle$, but the overlap $\langle\phi(t)|\psi(t)\rangle$ between two different wave-functions is not conserved in time, since H depends on $|\psi(t)\rangle\langle\psi(t)|$ via (117).

We assume that R is a scalar variable, though most of our results can be generalized directly to the case of several variables (vector-valued R).

The way we introduced (117) referred to the situation, where the feedback is engineered artificially. However, there is an important case, where the above scheme occurs naturally. For a particular case

$$F = -\langle\psi|\partial_R H|\psi\rangle, \quad (118)$$

Eqs. (116, 117) can be viewed as a *hybrid* dynamics, where a classical particle with coordinate R performs an overdamped motion and couples to the quantum system. Then ε in (117) corresponds to an inverse damping constant, while $\langle\psi|\partial_R H|\psi\rangle$ is the average (mean-field) force acting on the particle. (Overdamping means that the acceleration is irrelevant and the friction force is directly equated to the average potential force.) Hybrid theories are frequently

employed in optics, atomics physics and chemical physics; see [69] for the current state of art. In this context it is useful to recall that also the non-feedback [adiabatic] situation, where a quantum system interacts with a given external classical parameter, refers to an interaction of the quantum system with a classical one (the parameter is then the coordinate of this classical system). However, in the this standard case the feedback effect is neglected.

Let us now introduce the adiabatic eigensolution of the Hamiltonian for a fixed value of R ($n = 1, \dots, d$):

$$H[R] |n[R]\rangle = E_n[R] |n[R]\rangle, \quad \langle n[R] | m[R]\rangle = \delta_{mn}. \quad (119)$$

For simplicity we assume that the adiabatic energy levels are non-degenerate. The representation (119) has a gauge-freedom:

$$|n[R]\rangle \rightarrow e^{i\alpha_n[R]} |n[R]\rangle, \quad (120)$$

where $\alpha_n[R]$ is an arbitrary single-valued function depending on n and R . All observables should be gauge-invariant. Now let us expand $|\psi(t)\rangle$ as

$$\begin{aligned} |\psi(t)\rangle &= \sum_n c_n(t) e^{i\gamma_n(t)} |n[R(t)]\rangle, \\ \gamma_n(t) &\equiv - \int_0^t dt' E_n[R(t')], \quad c_n(t) = \langle n[R(t)] | \psi \rangle e^{-i\gamma_n(t)}, \end{aligned} \quad (121)$$

where $\gamma_n(t)$ are the dynamical phases, c_n are the adiabatic amplitudes, and where all summation indices run from 1 to d (the total number of energy levels), if not stated otherwise.

One gets from (116, 117, 121):

$$\dot{c}_n = -\varepsilon \sum_k c_k \langle n | k' \rangle F(R, \{c\}, \{e^{i\Delta\gamma(t)}\}) e^{i(\gamma_k(t) - \gamma_n(t))}, \quad (122)$$

where prime denotes differentiation over R ,

$$|k'\rangle = \partial_R |k[R]\rangle, \quad (123)$$

and where $F(R, c, e^{i\Delta\gamma(t)})$ is already reparametrized in terms of all c_l and all phase-differences $e^{\gamma_l(t) - \gamma_m(t)}$. To make this point clear we indicated this dependence via curly brackets in $F(R, \{c\}, \{e^{i\Delta\gamma(t)}\})$. The amplitudes c_n and R are slow variables, since, e.g., $|\dot{c}_n|$ is bounded from above by $\mathcal{O}(\varepsilon)$ provided that the feedback $F(R)$ and the Hamiltonian $H(R)$ are bounded functions of R . However, the contribution from the dynamical phases γ_n changes fast, since on the slow time $\tau = \varepsilon t$ it behaves as $\sim e^{i\tau/\varepsilon}$; see (121). If the spacings between the adiabatic energy levels $E_n[R]$ remain large enough, the existence of some intermediate time τ_f is guaranteed, over which the dynamical phase contribution performs many oscillations, but c_n and R do not change appreciably. Note that in particular we assume that the adiabatic energy levels do not cross as functions of R [72] (in addition of these levels being non-degenerate).

The adiabatic approximation divides c_n into the time-averaged (over τ_f) part \bar{c}_n and the small (at least as $\mathcal{O}(\varepsilon)$) oscillating part: $c_n = \bar{c}_n + \delta c_n$ [71]. To leading order we neglect in the RHS of (122) all the oscillating factors and substitute $c \rightarrow \bar{c}$ and $R \rightarrow \bar{R}$:

$$\bar{c}_n^\bullet = - \sum_k \bar{c}_k \langle n | k' \rangle \overline{F(\bar{R}, \{\bar{c}\}, \{e^{i\Delta\gamma}\}) e^{i(\gamma_k - \gamma_n)}}, \quad (124)$$

where $\tau = \varepsilon t$, $\bar{X} \equiv \int_0^{\tau_f} \frac{ds}{\tau_f} X(s)$, and where $X^\bullet \equiv dX/d\tau$. Likewise, (117) produces for the averaged motion of R

$$\bar{R}^\bullet = \overline{F(\bar{R}, \{\bar{c}\}, \{e^{i\Delta\gamma}\})}.$$

B. Linear feedback

The simplest example of feedback is

$$F = \langle \psi | A | \psi \rangle = \sum_{nm} c_n^* c_m A_{nm} e^{i(\gamma_m - \gamma_n)}, \quad (125)$$

where $\langle n | A | m \rangle \equiv A_{nm}$. Eq. (125) can be regarded as the first term of the Taylor expansion assuming that $F(x)$ depends weakly on its argument. Eq. (124) leads to

$$\bar{c}_l^\bullet = - \sum_{knm} \bar{c}_k \langle l | k' \rangle \bar{c}_n^* \bar{c}_m A_{nm} \overline{e^{i(\gamma_m - \gamma_n + \gamma_k - \gamma_l)}}. \quad (126)$$

In working out (126) we shall assume that the time-integrated energy-level differences are distinct:

$$\gamma_m(t) - \gamma_n(t) \neq \gamma_l(t) - \gamma_k(t), \text{ if } m \neq n \text{ and } m \neq l. \quad (127)$$

This condition is generic for few-level systems. It does not hold for cases like harmonic oscillator, which should be separately worked out from (126). Now in the RHS of (126) the non-zero terms are those with $m = n$ and $l = k$, and those with $m = l$ and $k = n$ (but $n \neq l$, not to count twice the term $m = n = k = l$):

$$\bar{c}_l^\bullet = -\bar{c}_l \langle l|l' \rangle \bar{R}^\bullet - \bar{c}_l \sum_{n(\neq l)} |\bar{c}_n|^2 \langle l|n' \rangle A_{nl}, \quad (128)$$

where $\langle l|l' \rangle$ is imaginary, since $\partial_R \langle l|l \rangle = 0$. The nontrivial (second) term in the RHS of (128) is due to non-diagonal elements of A . Defining the phase and modulus of \bar{c}_n ,

$$\bar{c}_n = \sqrt{p_n} e^{i\phi_n}, \quad \sum_n p_n = 1, \quad (129)$$

we get from (128) [and likewise from (117, 125)]

$$p_l^\bullet = -2p_l \sum_{n(\neq l)} p_n \Re\{\langle l|n' \rangle A_{nl}\}, \quad (130)$$

$$\phi_l^\bullet = i\langle l|l' \rangle R^\bullet - \sum_{n(\neq l)} p_n \Im\{\langle l|n' \rangle A_{nl}\}, \quad (131)$$

$$\bar{R}^\bullet = \sum_n p_n A_{nn}. \quad (132)$$

Eqs. (130–132) are our central results. Before exploring them in more detail let us discuss the standard (open-loop, i.e., no feedback) adiabatics, where $A = A(R)$ is a c-number. Now R moves in a prescribed way according to $R^\bullet = A(R)$. Eq. (130) leads to the conservation of the probabilities $p_l^\bullet = 0$ (adiabatic theorem): the system does not get enough energy to move out of the given energy level [37, 57]. The RHS of (131) reduces to Berry's factor $\phi_{B,l}^\bullet = i\langle l|l' \rangle R^\bullet$. As seen from (121), though $\phi_{B,l}$ is by itself not gauge-invariant, it does bring an observable (Berry phase) contribution to the average of an observable (hermitean operator) over the state $|\psi(t)\rangle = \sum_n c_n(0)e^{i\phi_{B,n}(\tau)+i\gamma_n(t)}$, provided that this operator is not diagonal over the adiabatic basis (119).

The Berry phase was observed in numerous experiments; see [58, 59] for review. It is constrained by the following conditions.

1. The Berry phase nullifies, $\langle l|l' \rangle = 0$, if the adiabatic eigenvectors $|l\rangle$ can be made real via a gauge transformation, e.g., a spinless particle without magnetic field. (This statement does not hold if there are level-crossings.)
2. $\phi_{B,l} = 0$ for a cyclic motion of a single slow parameter R , where R is switched on at the initial time and then switched off at the final time. The Berry phase may be different from zero if there is more than one slow parameter $\mathbf{R} = (R_1, R_2, \dots)$ on a closed curve \mathcal{C} : $\mathbf{R}(0) = \mathbf{R}(\tau)$ [58]. Then one gets a gauge-invariant expression $\phi_{B,l} = i \oint_C d\mathbf{R} \langle l|\partial_{\mathbf{R}} l \rangle$ [58, 59].

To our opinion these constraints seriously limit applications of the Berry phase. We shall see below that the adiabatic phases generated by the feedback-driven adiabatic dynamics are free of these limitations.

C. Closed-loop adiabatics

Eq. (130) for p_l arises out of the averaging over the fast dynamic phases under condition (127). Eq. (130) is nonlinear over p_n due to the feedback. The probabilities p_n are no longer conserved [due to the resonance between the oscillations of c_n and those of R , see (126)], and if p_n 's are known, the phases ϕ_l are obtained directly from (131). The matrix

$$a_{ln} \equiv -2\Re\{\langle l|\partial_R n \rangle \langle n|A|l \rangle\}, \quad a_{ln} = -a_{nl}, \quad (133)$$

in (130) is antisymmetric; in particular, $a_{ll} = 0$, which means $\sum_l p_l(\tau) = 1$. The edges of the probability simplex, e.g. $p_l = \delta_{l1}$, are (possibly unstable) stationary solutions of (130), and $p_l(\tau)$ is always non-negative.

It is noteworthy that (130) coincides with the replicator equation for a zero-sum population game [74, 75]. Consider a population of agents that consists of groups $l = 1, \dots, d$. The fraction p_l of each group in the total population changes due to interaction between the groups, so that p_l^\bullet is proportional to p_l itself, while the proportionality coefficient is the average payoff of the group l : $p_l^\bullet = p_l \sum_n a_{ln} p_n$ [74, 75]. Here the payoff matrix a_{ln} determines the fitness increase of the group l in its interaction with the group n . The actual mechanism of this interaction depends on the concrete

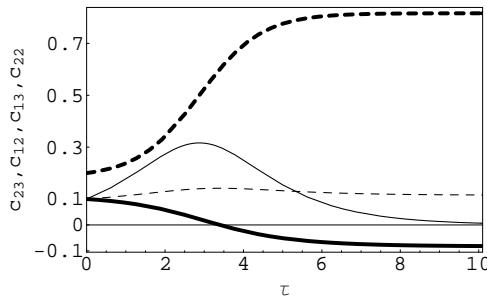


FIG. 2: Adiabatic amplitudes c_{nl} versus τ for a three-level system. c_{nl} are obtained from solving (154, 155) (hybrid dynamics). The external field is not acting on the first energy level: $|\langle 2|1' \rangle| = |\langle 3|1' \rangle| = 0$; thus c_{11} is constant. We put $|\langle 2'|3 \rangle| = 1$. All $|k\rangle$ are real (no Berry phases), and additionally the adiabatic energies are constant: $A_{ll} = -\partial_R E_l = 0$; thus \bar{R} is constant according to (147). Normal curve: c_{23} . Dashed curve: c_{12} . Thick curve: c_{13} . Thick dashed curve: c_{22} . Recall that $c_{33} + c_{22} = \text{const.}$

implementation of the model (inheritance, learning, imitation, infection, etc) [75]. The condition $a_{nl} = -a_{ln}$ means a zero-sum game (e.g., poker): the gain of one group equals to the loss of the other. Thus in (130) the population game, with (in general) τ -dependent payoffs a_{ln} , is now played by the energy levels. Interesting features of the replicator equation can be found without solving it; see (135–138).

We note that this seems to be the first *physical* realization of the replicator dynamics, which so far was limited to the phenomenology of population dynamics. This, in particular, opens up a way for its precise experimental investigation (something that arguably is hardly possible in a field like population dynamics).

For the open-loop control changing of R on the slow time scale is mandatory, otherwise no adiabatic motion occurs at all. The closed-loop situation is different, since now for $\langle n|A|n \rangle \equiv A_{nn} = 0$ the slow motion of R is absent,

$$\bar{R}^\bullet = 0, \quad (134)$$

[see (132)], with still non-trivial adiabatic dynamics. Eq. (134) implies that the fast motion of R averages out on the slow time. Let us focus on this situation, since we cannot study (130–132) in full generality.

Eqs. (130, 133), now with τ -independent a_{ln} , is conveniently studied via the time-averages [75]:

$$\frac{1}{T} \ln \frac{p_l(T)}{p_l(0)} = \sum_n a_{ln} \tilde{p}_n(T), \quad \tilde{p}_n(T) = \int_0^T \frac{d\tau}{T} p_n(\tau), \quad (135)$$

where T defines the time-window of the averaging. We shall normally take $T \rightarrow \infty$; see below.

There are now two different dynamic scenarios depending on the concrete form of τ -independent a_{lp} in (130, 133).

1. If all $p_l(t)$ (which were non-zero at the initial time $\tau = 0$) stay non-zero for all times, $\ln p_l(T)$ in the LHS of (135) is limited, which means that this LHS can be neglected for $T \rightarrow \infty$. We then get from (135) [74, 75]

$$\sum_n a_{ln} \tilde{p}_n(\infty) = 0. \quad (136)$$

Thus all $p_l(t)$ may remain non-zero for all times provided that there is a probability vector $\tilde{p}(\infty)$ that satisfies (136). Clearly, $\tilde{p}(\infty)$ is a stationary state of (130, 133). Recall that the [non-negative] relative entropy is defined as

$$S[\tilde{p}(\infty)|p(\tau)] = \sum_l \tilde{p}_l(\infty) \ln [\tilde{p}_l(\infty)/p_l(\tau)], \quad (137)$$

where $p(\tau)$ is a time-dependent solution of (130). $S[\tilde{p}_n(\infty)|p(\tau)]$ is equal to zero if and only if $\tilde{p}(\infty) = p(\tau)$. Due to (136), $S[\tilde{p}_n(\infty)|p(t)]$ is a constant of motion [thus an adiabatic invariant], since

$$S^\bullet[\tilde{p}(\infty)|p(\tau)] = \sum_l p_l(\tau) a_{ln} \tilde{p}_n(\infty). \quad (138)$$

Eq. (130) can be recast into a Hamiltonian form [74], where the constant of motion $S[\tilde{p}(\infty)|p(\tau)]$ becomes the Hamiltonian. The non-linearity of this dynamics is essential, since it can demonstrate chaos for $d \geq 5$. In some related systems the chaotic behavior was seen in [76].

2. If the matrix a_{ln} is such that (136) does not have any probability vector solution, $\frac{1}{T} \ln \frac{p_l(T)}{p_l(0)}$ in (135) is necessarily finite for at least one l . The corresponding probability $p_l(T)$ goes to zero (for a large T): $p_l(T) \rightarrow p_l(\infty) = 0$, so that

for all k one has $\sum_n a_{kn} \tilde{p}_n(\infty) \leq 0$. This inequality is strict at least for $k = l$. Eq. (138) shows that $S[\tilde{p}(\infty)|p(\tau)]$ now decays to zero meaning that $p(\tau)$ relaxes to $\tilde{p}(\infty)$. This relaxation is due to the non-linearity of (130); it is impossible without feedback.

Eq. (131) for the phases integrates as

$$\phi_l(\tau) = -\tau \sum_{n(\neq l)} \tilde{p}_n(\tau) b_{ln}, \quad b_{ln} \equiv \Im\{\langle l|n' \rangle \langle n|A|l \rangle\}, \quad (139)$$

where $\tilde{p}_n(\tau)$ satisfies the algebraic equation (135), and b_{ln} is symmetric: $b_{ln} = b_{nl}$. Eq. (139) gives the phases of the adiabatic feedback control. Clearly, $\phi_l(\tau)$ is free of the constraints for the open-loop (Berry) phase $\phi_{B,l}$:

- i) it is gauge-invariant together with b_{ln} , see (120, 139);
- ii) its existence does not require complex adiabatic eigenvectors $|l\rangle$, provided that the monitoring observable A has at least some complex elements $\langle n|A|l \rangle$;
- iii) it does not require several control parameters for cyclic processes;
- iv) even if a_{nl} , defined via (133), is zero, i.e., if the probabilities p_n are conserved, the feedback-driven phases ϕ_l in (139) can be non-zero. Here we have an important situation, where the adiabatic theorem holds, but the new adiabatic phases are non-trivial.

Note that $\phi_l = 0$ if the evolution starts from one adiabatic eigenvector $p_n(0) = \delta_{nk}$, i.e., in the expansion (121) only one adiabatic eigenfunction with index k is present: $c_n(0) = \delta_{nk}$ (however this stationary state of (130) need not be stable, as we saw above). In contrast, the Berry phase may be non-zero even for this case, although its observation does require interference with another eigenstate.

D. Examples

We now apply our findings to two simple examples. For a two-level system (130, 133) reduce to (recall that $\bar{R}^\bullet = 0$, and thus a_{12} does not depend on time τ)

$$p_1(\tau) = \frac{p_1(0)e^{a_{12}\tau}}{1 + p_1(0)[e^{a_{12}\tau} - 1]}, \quad (140)$$

which means that independent of the initial value $p_1(0)$, $p_1 \rightarrow 1$ ($p_1 \rightarrow 0$) if $a_{12} > 0$ ($a_{12} < 0$). Properly choosing the time τ and a_{12} , and knowing $p_1(0)$, we can reach any value $0 \leq p_1(\tau) \leq 1$. For this simplest two-level example the oscillatory behavior (described after (138)) is absent.

Eq. (139) produces for the phases

$$\phi_{1,2}(\tau) = \pm \frac{b_{12}}{a_{12}} \ln [p_1(0)(e^{\mp a_{12}\tau} - 1) + 1]. \quad (141)$$

Two basic examples of two-level systems are the spin- $\frac{1}{2}$ and the polarization states of light. The standard Berry phase was observed in both these cases [58].

For the three-level situation the internal stationary vector is obtained from (136) (up to normalization)

$$\tilde{p}_1(\infty) = \frac{a_{23}}{a_{12} + a_{23} - a_{13}}, \quad \tilde{p}_2(\infty) = -\frac{a_{13}}{a_{12} + a_{23} - a_{13}}, \quad (142)$$

$$\tilde{p}_3(\infty) = \frac{a_{12}}{a_{12} + a_{23} - a_{13}}, \quad (143)$$

provided all these probabilities are positive, as e.g. for $a_{12} > 0$, $a_{23} > 0$, $a_{13} < 0$. In game-theoretic terms this means that 1 beats 2, 2 beats 3, but 3 beats 1 (cyclic dominance or rock-scissor-paper game) [75]. Now the τ -dependent solution $p(\tau)$ of (130) oscillates around the stable point given by (142, 143) above [75].

The relaxation regime is realized if one of the probabilities in (142, 143) are not positive. For example, if 1 beats both others ($a_{12} > 0$, $a_{13} > 0$), the only attractor of (130) is $\tilde{p}(\infty) = (1, 0, 0)$. The latter conclusion holds also for a τ -dependent \bar{R} , if the conditions $a_{12}(\tau) > 0$ and $a_{13}(\tau) > 0$ are satisfied for all τ 's. However, the general arguments (135–138) do not hold for τ -dependent a_{ln} .

In the last example [as well as in the two-level example given by (140)] we meet a situation, where for an *arbitrary* initial pure state, the dynamics leads to relaxation to a *definite* state, i.e., to the state $p_1 = 1$ with the example given by (140). The reason for this relaxation behavior, which is clearly absent for the usual linear Schrödinger dynamics, is the underlying non-linearity of the feedback-driven adiabatic dynamics, as expressed by the dependence of F on $|\psi\rangle\langle\psi|$ in (117). To some extent the above relaxation effect resembles the collapse of the wave-function known from

the phenomenology of quantum measurement. Recall from our discussion around (118) that this non-linearity may be caused by a mean-field interaction with a classical [i.e., macroscopic] system. However, the studied situation does not correspond to any measurement, since the "collapse" is not probabilistic and does not correspond to a macroscopic change in any measuring apparatus.

E. Slow motion of the controlling parameter

Now R moves slowly due to $\langle n|A|n\rangle \neq 0$. Now (130) becomes a driven replicator equation, since a_{ln} are τ -dependent. The general theory for a time-dependent replicator is (to our knowledge) lacking. There are two cases, where the above results suffice for analyzing the driven situation. In the above three-level example assume that conditions $\frac{a_{23}(\tau)}{a_{12}(\tau)} < 0$ or $\frac{a_{13}(\tau)}{a_{12}(\tau)} > 0$, are satisfied for all τ . Then the same argument on the relaxation to a single state applies. If the opposite conditions $\frac{a_{23}(\tau)}{a_{12}(\tau)} > 0$ and $\frac{a_{13}(\tau)}{a_{12}(\tau)} < 0$ hold for all τ , all $p_l(\tau)$ are non-zero for all times.

F. Mixed states

So far we focused on pure states of the quantum system. Now we assume that the quantum state ρ is mixed and the feedback goes via the average $\text{tr}(A\rho)$; compare with closed loop equation (117). Since the closed loop equations (116) is not linear, the mixed-state dynamics (in general) does not reduce to the pure case. Starting from the feedback-driven von Neumann equation [compare with (116)]

$$i\dot{\rho}(t) = [H[R(t)], \rho], \quad \dot{R} = \varepsilon F(R, \text{tr}\{A\rho\}), \quad (144)$$

and defining for the adiabatic amplitude

$$c_{nm} \equiv \langle n|\rho|m\rangle e^{i\gamma_m - i\gamma_n}, \quad (145)$$

[compare with (121)], and proceeding along the lines of (119–127) we obtain

$$\begin{aligned} \bar{c}_{nm}^\bullet + \bar{R}^\bullet \bar{c}_{nm} (\langle n|n'\rangle + \langle m'|m\rangle) \\ = - \sum_{l(\neq n)} \langle n|l'\rangle A_{ln} \bar{c}_{nl} \bar{c}_{lm} - \sum_{l(\neq m)} \langle l'|m\rangle \bar{c}_{nl} \bar{c}_{lm} A_{ml} \\ \bar{R}^\bullet = \sum_l \bar{c}_{ll} A_{ll}. \end{aligned} \quad (146) \quad (147)$$

There is a case where the pure-state analysis applies directly: Pseudo-pure states in NMR are important for ensemble computation and are given as

$$\rho = (1 - \eta) \hat{1} + \eta |\psi\rangle\langle\psi|, \quad (148)$$

where $\hat{1}$ is the unit matrix, and where $0 < \eta < 1$ is a parameter [67]. Since $\hat{1}$ is an invariant of (146), Eq. (146) reduces to (128), but with $A_{nl} \rightarrow \eta^2 A_{nl}$. Thus for the pseudo-pure case we get the same (though rescaled in time with the factor η^2) adiabatic dynamics.

In general the phases of c_{nm} do not decouple from $|c_{nm}|$, and we do not have a general theory for mixed states.

Hybrid dynamics. Let us study in more detail the hybrid (quantum-classical) dynamics $A = -\partial_R H$; see our discussion after (118). Let us first of all address the pure-state dynamics. Eq. (119) implies for $n \neq l$

$$A_{nl} \equiv \langle n|A|m\rangle = (E_n - E_l)\langle n|l'\rangle, \quad (149)$$

$$\langle l|n'\rangle A_{nl} = (E_l - E_n)|\langle l|n'\rangle|^2. \quad (150)$$

Substituting (150) into (139) we see that the new phases ϕ_l nullify. Eqs. (150, 130) predict for the probabilities

$$p_l^\bullet = 2p_l \sum_{n(\neq l)} p_n (E_n - E_l) |\langle l|n'\rangle|^2. \quad (151)$$

We shall order the energies as

$$E_1 < \dots < E_d. \quad (152)$$

Since we assumed that there are no level-crossings (see our discussion after (123)), this ordering is conserved at all times, once it was imposed initially.

Considering (151) for $l = d$ we see that p_d goes to zero for large times $\tau \rightarrow \infty$. Continuing this logic for $l = d-1, \dots, 1$ we see that all p_l with $l > 1$ go to zero for large times, while the probability p_1 for the lowest level goes to one. The origin of this phenomenon of relaxation to the ground-state (or cooling) is related to (118), which implies that the external parameter R interacts with a zero-temperature thermal bath.

Let us now turn to the mixed-state dynamics, which will be seen to be more interesting. Eq. (146) implies

$$\bar{c}_{nn}^\bullet = 2 \sum_{l(\neq n)} (E_l - E_n) |\langle l' | n \rangle|^2 |\bar{c}_{nl}|^2. \quad (153)$$

Let $n = 1$ be the lowest energy level. If all $|\langle l' | 1 \rangle|$ differ from zero, the non-diagonal elements $\bar{c}_{l \neq 1}$ have to nullify for large τ , since \bar{c}_{11} should be limited. Continuing this reasoning for $n > 1$, we get that all non-diagonal elements $\bar{c}_{n \neq l}$ nullify for large times, if all $|\langle l' | n \rangle|$ are positive.

Once the non-diagonal elements $\bar{c}_{n \neq m}$ nullify, the large-time behavior of the diagonal elements \bar{c}_{nn} is determined by the fact that the dynamics (144) conserves the eigenvalues of the density matrix ρ . Inspecting (146) and using (150, 152) one can see that the long-time diagonal values of $\bar{c}_{nn}(\infty)$ are equal to the eigenvalues $\lambda_n(0)$ of the initial matrix $\{\bar{c}_{nm}(0)\}_{n,m=1}^d$: the lowest energy gets the largest eigenvalue, $\bar{c}_{11}(\infty) = \max[\lambda_1(0), \dots, \lambda_d(0)]$, the second energy level gets the second largest eigenvalues among $\{\lambda_n(0)\}$, and so on till finally the highest energy level gets the smallest eigenvalue: $\bar{c}_{dd}(\infty) = \min[\lambda_1(0), \dots, \lambda_d(0)]$. The resulting long-time state with the density matrix $\delta_{nm} c_{nn}(\infty)$ has the lowest energy among all density matrices with energies (152) and eigenvalues $\lambda_n(0)$. Thus what we have here is an effect of controlled decoherence.

The above analysis required that all $|\langle l' | n \rangle|$ are non-zero. If, however, $|\langle l' | n \rangle| = 0$ for some $n \neq l$, the element c_{nl} survives at long-times and undergoes a non-trivial evolution. An example of such evolution is presented in Fig. 2 under the following simplifying conditions (which are in a sense minimal for the existence of this effect): (1) The number of energy levels is equal to three: $d = 3$. (2) The field R acts only on energy levels 2 and 3; the level 1 does not feel R . Thus $|\langle 2 | 1' \rangle| = |\langle 3 | 1' \rangle| = 0$ and $|\langle 2' | 3 \rangle| > 0$. (3) The adiabatic energies do not depend on R . Thus $A_{ll} = -\partial_R E_l = 0$ and \bar{R} is also a constant of the averaged dynamics; (147). The R -dependence is restricted to the eigenvectors of the Hamiltonian $H[R]$ in (119). (4) These adiabatic eigenvectors $|k\rangle$ are real (no Berry phases). Eqs. (146, 150) now read

$$\bar{c}_{22}^\bullet = 2\beta |\bar{c}_{23}|^2, \quad |\bar{c}_{23}|^\bullet = \beta |\bar{c}_{23}|(1 - \bar{c}_{11} - 2\bar{c}_{22}), \quad (154)$$

$$\bar{c}_{12}^\bullet = \beta \bar{c}_{13} |\bar{c}_{23}| e^{-i\psi_{23}}, \quad \bar{c}_{13}^\bullet = -\beta \bar{c}_{12} |\bar{c}_{23}| e^{i\psi_{23}}, \quad (155)$$

where $\bar{c}_{11}^\bullet = 0$, $\psi_{23}^\bullet = 0$, and where $\beta \equiv (E_3 - E_2)|\langle 2' | 3 \rangle|^2$ and $\bar{c}_{23} \equiv |\bar{c}_{23}|e^{i\psi_{23}}$. The numerical solution of (146, 150) under the above conditions (1)-(4) is presented in Fig. 2. The fact that \bar{c}_{11} is constant is natural, since the external field R does not act on the first energy level due to condition (2). However, we see that c_{12} and c_{13} do change in time, an effect that is impossible for pure states; see (130, 131), which show that under $|\langle 2 | 1' \rangle| = |\langle 3 | 1' \rangle| = 0$ both p_1 and ϕ_1 are constant (in the slow time). Another interesting aspect seen in Fig. 2, where all $c_{l \neq n}$ are real, is that c_{13} changes its sign. This is an example of the adiabatic phase for the considered case. Note that though c_{23} decays to zero [according to the argument presented after (153)], it increases in the intermediate times.

In summary, we studied how the feedback generated by non-disturbing (ensemble) measurements affects the adiabatic (i.e., slowly driven) quantum dynamics. For the simplest linear feedback we have found that *i*) the populations are no longer constant. Instead, they satisfy the canonical [replicator] equation of the population game theory, allowing us to visualize the corresponding dynamics as a zero-sum game played by the adiabatic energy levels. The [non-linear] replicator equation generates a non-trivial (possibly chaotic) Hamiltonian motion, or alternatively, relaxation toward a certain state. *ii*) In addition to the Berry phase, the feedback generates a new, explicitly gauge-invariant phase, which exists under a wider range of conditions. There are scenarios of feedback, where the probabilities are constant (resembling the ordinary situation), but the new phases are still non-trivial. All these results extend directly to pseudo-pure quantum states. For the properly mixed-state situation we analyzed in detail the hybrid (quantum-classical) case.

Problem 1

Derive (154, 155).

Problem 2

Give an evolutionary interpretation of the replicator dynamics (consult [75]).

-
- [1] R. Shankar, *Principles of Quantum Mechanics* (Plenum Press, NY, 1980).
- [2] F. Logiurato and A. Smerzi, Journal of Modern Physics, **3**, 1802 (2012).
- [3] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).
- [4] R. I. G. Hughes, *The Structure and Interpretation of Quantum Mechanics* (Harvard University Press, Cambridge, 1992).
- [5] O. Cohen, Phys. Rev. A **63**, 016102 (2001).
D. R. Terno, Phys. Rev. A **63**, 016101 (2001).
K.A. Kirkpatrick, quant-ph/0405058.
B. d'Espagnat, quant-ph/0111081.
- [6] W.M. de Muynck, *Foundations of Quantum Mechanics, an Empiricist Approach*, (Kluwer Academic Publishers, 2002).
- [7] P. Kok and S. L. Braunstein, Phys. Rev. A **61**, 042304 (2000)
- [8] L.D. Landau and E.M. Lifshitz, *Statistical Physics, I*, (Pergamon Press Oxford, 1978).
- [9] R. Balian, *From Microphysics to Macrophysics*, volume I, (Springer, 1992).
- [10] P. Perrot, *A to Z of Thermodynamics*, (Oxford University Press, 1998).
- [11] *Quantum Limits to the Second Law*, ed. D.P. Sheehan, (AIP Conf. Proc. 643, 2002)
- [12] 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).
- [13] Th.M. Nieuwenhuizen and A.E. Allahverdyan, Phys. Rev. E **66**, 036102 (2002)
- [14] G.N. Bochkov and Yu.E. Kuzovlev, Sov. Phys. JETP, **45**, 125 (1977). W. Pusz and S.L. Woronowicz, Comm. Math. Phys., **58**, 273 (1978). A. Lenard, J. Stat. Phys., **19**, 575 (1978). J. Kurchan, cond-mat/0007360. A.E. Allahverdyan and Th.M. Nieuwenhuizen, Physica A **305**, 542 (2002).
- [15] I.M. Bassett, Phys. Rev. A **18**, 2356 (1978).
H. Tasaki, cond-mat/0009206.
- [16] H. Narnhofer and W. Thirring, Phys. Rev. A, **26**, 3646 (1982).
- [17] K. Sekimoto and S. Sasa, J. Phys. Soc. Jpn. **66**, 3326 (1997). C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997). G.E. Crooks, Phys. Rev. E **60**, 2721 (1999). C. Maes, J. Stat. Phys. **95**, 367 (1999). H. Tasaki, cond-mat/0009244.
- [18] L. J. Broer, Physica **17**, 531 (1951). R. D. Mountain, Physica **30**, 808 (1964). M.J. Klein, Phys. Rev. **86**, 807 (1952). R.M. Wilcox, Phys. Rev. **174**, 624 (1968).
- [19] K. Sato, K. Sekimoto, T. Hondou, and F. Takagi, Phys. Rev. E **66**, 016119 (2002). H. Tasaki, cond-mat/0008420.
- [20] H.C. Longuet-Higgins, Proc. R. Soc. A, **344**, 147 (1975). C.A. Mead, J. Chem. Phys., **70**, 2276 (1979).
- [21] A. Galindo and P. Pascual, *Quantum Mechanics*, (Springer-Verlag, 1991).
- [22] M. Born and V. Fock, Z. Phys, **51**, 165 (1928). G.A. Hagedorn, Ann. Phys., **196**, 278 (1989). J.E. Avron and A. Elgart, Comm. Math. Phys., **203**, 445 (1999).
- [23] J. Hwang and P. Pechukas, J. Chem. Phys., **67**, 4640 (1977). Y.N. Demkov, et al., Phys. Rev. A, **18**, 2089 (1978).
- [24] D.G. Truhlar and A.C. Mead, Phys. Rev. A **68**, 032501 (2003).
- [25] D.R. Yarkony, Rev. Mod. Phys., **68** (1996). 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).
- [26] E.P. Wigner, in *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. Van der Merwe (MIT Press, Cambridge, MA, 1971).
M.D. Srinivas and E. Wolf, Phys. Rev. D **11**, 1477 (1975).
P. Kruszynski and W. M. de Muynck, J. Math. Phys. **28**, 1761 (1987).
L. Cohen. Phil. Sci. **33**, 317 (1966).
H. Margenau and L. Cohen, in *Quantum Theory and Reality*, ed. M. Bunge (Springer, Berlin, 1968)
- [27] J. Yvon, *The Saclay reactor*, Proceedings of the International Conference on Peaceful uses of Atomic Energy. p. 387, Geneva (1955).
P. Chambadal, *Les Centrales Nucléaires* (Armand Colin, Paris, France, 1957).
I. I. Novikov, J. Nuclear Energy II **7**, 125 (1958).
F. Curzon and B. Ahlborn, Am. J. Phys. **43**, 22 (1975).
- [28] C. Van den Broeck, Phys. Rev. Lett. **95**, 190602 (2005).
M. Esposito, K. Lindenberg and C. Van den Broeck, Phys. Rev. Lett. **102**, 130602 (2009).
- [29] A.E. Allahverdyan, R.S. Johal, and G. Mahler, Phys. Rev. E **77**, 041118 (2008)
- [30] See, e.g., M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Phys. Rev. Lett. **105**, 150603 (2010).
- [31] G. Benenti, K. Saito, and G. Casati, Phys. Rev. Lett. **106**, 230602 (2011).
U. Seifert, Phys. Rev. Lett. **106**, 020601 (2011).
- [32] E. Farhi and S. Gutmann, Phys. Rev. A **57**, 2403 (1998).
- [33] R. Zwanzig, PNAS, **92**, 9804 (1995).
- [34] K. Sneppen and G. Zocchi, *Physics in Molecular Biology* (Cambridge University Press, Cambridge, 2005).
- [35] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic Press, NY, 1979).
- [36] A. E. Allahverdyan, K. V. Hovhannisyan, A. V. Melkikh and S. G. Gevorkian, Phys. Rev. Lett. **109**, 248903 (2013).
- [37] A. B. Migdal, *Qualitative methods in quantum theory* (Westview Press, 2000).
- [38] D. Home and M. A. B. Whitaker, Physics Reports **210**, 223 (1992).
- [39] M. Namiki and S. Pascazio, Phys. Rep. **232**, 301 (1993).

- [40] W. H. Zurek, Rev. Mod. Phys. **75**, 715 (2003).
- [41] G. L. Sewell, AIP Conf. Proc. 962, 215 (2007).
- [42] F. Laloë, *Do we really understand quantum mechanics?* (Cambridge Univ. Press, 2012).
- [43] A. E. Allahverdyan, R. Balian and Th. M. Nieuwenhuizen, Phys. Rep. **525**, 1 (2013).
- [44] H. Narnhofer and W. F. Wreszinski, arXiv:1309.2550v1, (2013).
- [45] A. Daneri, A. Loinger and G. M. Prosperi, Nucl. Phys. **33**, 297 (1962).
- [46] K. Hepp, Helv. Phys. Acta, **45**, 237 (1972).
- [47] The form of \hat{r}_i obeys the rule proposed by G. Lüders, Ann. Phys. (Leipzig) **6** (8), 322 (1951). English translation and discussion: K. A. Kirkpatrick, Ann. Phys. (Leipzig) **15** (9), 663 (2006).
- [48] A. E. Allahverdyan, R. Balian and Th. M. Nieuwenhuizen, Europhys. Lett. **61**, 452 (2003).
- [49] E. T. Jaynes, Phys. Rev. **106**, 620 and **108**, 171 (1957).
- [50] R. Balian, *From microphysics to macrophysics, I, II*, (Springer, Berlin, 2007).
- [51] J. S. Bell, Helv. Phys. Acta, 48:93, 1975.
- [52] J. L. Park, Foundations of Physics, 18, 225, 1988.
- [53] L. van Hove, in Fundamental Problems in Statistical Mechanics, edited by E. D. G. Cohen (North-Holland, Amsterdam) 1962, p. 157.
- [54] N. G. van Kampen, in Fundamental Problems in Statistical Mechanics, edited by Cohen E. D. G. (North-Holland, Amsterdam) 1962, ibid., p. 173.
- [55] M. Bander, arXiv:cond-mat/9609218.
- [56] R. von Mises, *Probability, Statistics and Truth* (Macmillan, London, 1957).
- [57] A. Messiah, *Quantum mechanics* (North-Holland, Amsterdam, 1962), Vol. 2.
- [58] M.V. Berry, Proc. R. Soc. Lond. A, **392**, 45 (1984).
A. Shapere and F. Wilczek, *Geometric Phases in Physics* (World Scientific, Singapore, 1989).
J. Anandan, J. Christian and K. Wanelik, Am. J. Phys. **65**, 180 (1997).
- [59] G.G. de Polavieja and E. Sjöqvist, Am J Phys **66**, 431 (1998). Y. Ben-Aryeh, arXiv:quant-ph/0402003.
- [60] J. Pachos, P. Zanardi, and M. Rasetti, Phys. Rev. A **61**, 010305 (2000).
J.A. Jones, V. Vedral, A. Ekert, and G. Castagnoli, Nature **403**, 869 (2000).
- [61] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).
- [62] S. Duki, H. Mathur, Onuttom Narayan, Phys. Rev. Lett. **97**, 128901 (2006).
J. Ma, Yongping Zhang, Enge Wang, Biao Wu, Phys. Phys. Rev. Lett. **97**, 128902 (2006).
K.-P. Marzlin and B.C. Sanders, Phys. Rev. Lett. **97**, 128903 (2006).
D.M. Tong, K. Singh, L. C. Kwek, C. H. Oh, Phys. Rev. Lett. **98**, 150402 (2007).
R. MacKenzie et al., Phys. Rev. A **73**, 042104 (2006).
- [63] Special issue on quantum control, J. Opt B, **7** (2005).
- [64] D. Poulin, Phys. Rev. A, **62** 012307 (2005).
- [65] S. Lloyd and J.-J. Slotine, Phys. Rev. A, **62** 012307 (2000). C. Altafini, arXiv:quant-ph/0601016.
- [66] O.R. Ernst, G. Bodenhausen and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford University Press, Oxford, 1987).
- [67] J.A. Jones, Prog. NMR Spectr. **38**, 325 (2001).
- [68] P. Broekart and J. Jeener, J. Magn. Reson. A **113**, 60 (1995).
- [69] M.J. Hall and M. Reginatto, Phys. Rev. A, **72**, 062109 (2005). L. Diosi et al., Phys. Rev. A **61**, 022108 (2000).
- [70] This way of applying the adiabatic approximation coincides with the general method proposed in [71]. The method also allows to find the higher-order corrections and determine the applicability region of the adiabatic approach.
- [71] N.N. Bogoliubov and Yu.A. Mitropolskii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon & Breach, New York, 1961); see chapter V.
- [72] The ordinary (non-feedback) adiabatic approach can be generalized to the case, where some level-crossings are allowed; see [73] for detailed references. However, for the present feedback driven adiabatic dynamics allowing level-crossings requires more involved analysis.
- [73] A.E. Allahverdyan and Th.M. Nieuwenhuizen, Phys. Rev. E **71**, 046107 (2005).
- [74] E. Akin and V. Losert, J. Math. Biology, **20**, 231 (1984).
- [75] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).
J. Hofbauer and K. Sigmund, Bull. Am. Math. Soc. **40**, 479 (2003).
- [76] G. Picard and T.W. Johnston, Phys. Rev. Lett. **48**, 1610 (1982).
Y. Sato, E. Akiyama and J.D. Farmer PNAS **99**, 4748 (2002).
- [77] H-X. Zhou, Quarterly Reviews of Biophysics **43**, 219 (2010).
- [78] P. Echenique, *Introduction to protein folding for physicists*, arXiv:0705.1845.
- [79] C. Levinthal, *How to fold graciously*, in Mossbauer Spectroscopy in Biological Systems, edited by J. T. P. DeBrunner and E. Munck, pp. 22-24 (Allerton House, Monticello, Illinois, 1969, University of Illinois Press).
- [80] R.L. Baldwin, Protein Sci. **6**, 2031 (1997).
- [81] P.G. Wolynes, Phil. Trans. R. Soc. A **363**, 453 (2005).