World Scientific
www.worldscientific.com

Quantum fluctuation relations

Paolo Facchi* and Giancarlo Garnero

Dipartimento di Fisica and MECENAS
Università di Bari, I-70126 Bari, Italy

INFN, Sezione di Bari, I-70126 Bari, Italy
*paolo.facchi@ba.infn.it

Marilena Ligabò

Dipartimento di Matematica
Università di Bari, I-70125 Bari, Italy

We present here a set of lecture notes on exact fluctuation relations. We prove the
Jarzynski equality and the Crooks fluctuation theorem, two paradigmatic examples of
classical fluctuation relations. Finally, we consider their quantum versions, and analyze
analogies and differences with the classical case.

Keywords: Quantum thermodynamics; fluctuation relations.

0. Introduction

We present here the notes of three lectures given by one of us at the "Fifth
International Workshop on Mathematical Foundations of Quantum Mechanics and
its applications" held in February 2017 in Madrid at the *Instituto de Ciencias
Matemáticas* (ICMAT).

We will consider some results about fluctuation theorems both for classical and
for quantum systems, a research topic that recently has attracted a great deal
of attention. The statistical mechanics of classical and quantum systems driven
far from equilibrium has witnessed quite recently a sudden development with the
discovery of various exact fluctuation theorems which connect equilibrium thermo-
dynamic quantities to non-equilibrium ones. There are excellent reviews on this
topic, which cover both classical [1] and quantum fluctuation relations [2, 3]. Here,
we will follow more closely the exposition by Campisi, Hänggi and Talkner [3], to
which we refer the reader for further information.

In the first lecture, we will recall the derivation of Einstein's fluctuation-
dissipation relation for a Brownian particle, which is the inception of classical fluc-
tuation relations. Moreover, we will identify the fundamental ingredients which are

already present in this early derivation. Then, we will consider the Green–Kubo formula, which represents the first general approach to quantum fluctuation-dissipation relations. The link between the correlation function of a quantum system and its linear response function will be shown and the classical limit of the Green–Kubo formula will be considered.

The second lecture is devoted to exact classical fluctuation relations. In particular, we will give explicit deductions of the Jarzynski equality and the Crooks fluctuation theorem, two paradigmatic examples of classical fluctuation theorems. In the proofs, two ingredients will be crucial: reversibility at the microscopic level and the Gibbs probability distribution on the initial conditions of the system.

Finally, in the third lecture, we will consider the quantum case. After introducing the operational definition of measurement of work as a two-time energy measurement, and properly defining microreversibility for time-dependent unitary evolution, we prove both the Jarzynski equality and the Crooks fluctuation theorem for a quantum system.

1. Lecture 1: Fluctuation-dissipation Relations

We start with some classical results about fluctuation-dissipation relations.

At the microscopic level matter is in a permanent state of agitation and undergoes thermal and quantum fluctuations. Statistical Mechanics is able to provide explanations and quantitative results on those fluctuating quantities.

A paradigmatic example is a rarefied gas at thermal equilibrium, which is classically described by the Maxwell–Boltzmann distribution of velocities. This distribution is derived under the assumption that the classical dynamics of the microscopic constituents is Hamiltonian, and that the atoms of the gas interact via negligible short-range forces. Moreover, the Maxwell–Boltzmann distribution describes a situation of thermal equilibrium.

What happens to other fluctuating quantities?

In this lecture, we will be mainly interested in the work exchanged during out-of-equilibrium transformations and its fluctuations. In this analysis, a crucial role will be played by two ingredients:

- the initial state of a physical system is a thermal state, and, as such, is described by the Gibbs canonical distribution:

$$\rho_\beta = \frac{\mathrm{e}^{-\beta H_0}}{Z_0}, \tag{1}$$

where $\beta^{-1} \propto T$ is the temperature at equilibrium, H_0 is the Hamiltonian at the initial time, and Z_0 is the partition function;
- the dynamics is reversible at the microscopic level.

The first hypothesis is of statistical nature, because we are assuming a well-defined initial probability distribution on the initial state. On the other hand, the second one is only stating the Hamiltonian nature of microscopic dynamics.

Next, we would like to understand what happens after forcing the system out of equilibrium, not necessarily in an adiabatic way.

1.1. *Einstein's relation*

The history of fluctuation relations can be traced back to the work of Einstein [5]. In 1905, he proved that the linear response of a system in thermal equilibrium, driven out of equilibrium by an external force, is determined by the fluctuation properties at equilibrium.

Einstein considers the case of a Brownian particle in a fluid (see Fig. 1) and determines a relation between the mobility μ and the diffusion constant D:

$$\mu = \frac{D}{k_B T}, \tag{2}$$

where k_B is the Boltzmann constant and T is the absolute temperature of the fluid at equilibrium. We recall that in a dissipative fluid, the mobility represents the ratio of the suspended particle's terminal drift velocity v_d to an applied force F:

$$\mu = \frac{v_d}{F}, \tag{3}$$

It is apparent from Eq. (2) that Einstein's relation links a non-equilibrium quantity, say μ, related to the force that drags the system out of its initial state, with the temperature T of the gas at equilibrium.

We briefly recall the derivation of Einstein's relation. Suppose the force F is conservative, say $F(x) = -\nabla U(x)$, where $U : \mathbb{R}^3 \to \mathbb{R}$ is a smooth potential. Then, the drift velocity at x reads

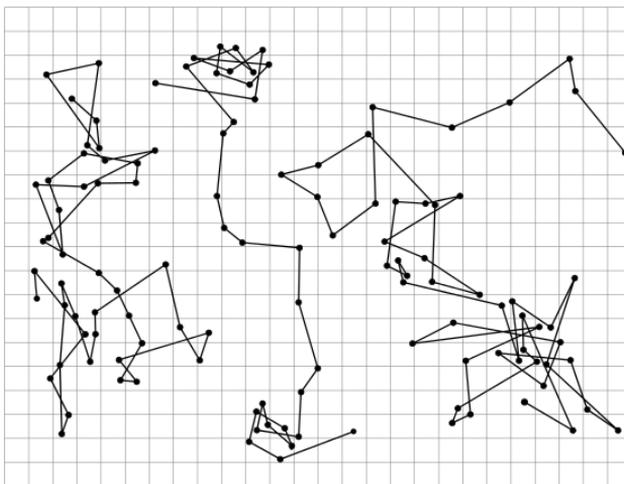$$v_d(x) = \mu(x)F(x) = -\mu(x)\nabla U(x). \tag{4}$$

Fig. 1. Evidence of Brownian motion as depicted for the first time by Jean Perrin in 1908 [4].

Moreover, assume that the concentration is at equilibrium and thus is determined by the Maxwell–Boltzmann statistics,

$$\rho(x) = A\mathrm{e}^{-\frac{U(x)}{k_B T}}, \tag{5}$$

where A is a normalization constant. The current density due to drift reads

$$J_{\mathrm{drift}}(x) = \rho(x)v_d(x) = -\rho(x)\mu(x)\nabla U(x). \tag{6}$$

There is a second contribution to the current density which is due to diffusion and, according to Fick's law, is proportional to the gradient of the concentration:

$$J_{\mathrm{diffusion}}(x) = -D(x)\nabla\rho(x). \tag{7}$$

At equilibrium there is a balance between these currents, namely

$$J_{\mathrm{drift}}(x) + J_{\mathrm{diffusion}}(x) = 0. \tag{8}$$

By deriving (5), we obtain

$$\nabla\rho(x) = -\frac{\nabla U(x)}{k_B T}\rho(x) \tag{9}$$

and plugging it in the balance equation (8), we get

$$-\rho(x)\nabla U(x)\left[\mu(x) - \frac{D(x)}{k_B T}\right] = 0, \tag{10}$$

and Einstein's fluctuation-dissipation relation (2) follows. It is evident from the above derivation that (2) is an approximate relation, since Fick's law (7) is only valid in the linear regime.

Einstein's relation was the first of a series of fluctuation-dissipation relations, which predict the behavior of systems that obey the detailed balance principle and are weakly perturbed from thermal equilibrium: thermal fluctuations of a physical observable are related to the linear response, quantified by the admittance or impedance of the same physical observable. The key idea is that the response of a system, which is at thermodynamic equilibrium, to a small applied force is the same as the response to statistical fluctuations at equilibrium.

A second example of a fluctuation-dissipation relation was provided by the Johnson–Nyquist noise [6, 7]. This phenomenon is due to the thermal agitation of electrons in a conductor at equilibrium. The overall effect is an electrical thermal noise which can be measured and appears as a difference voltage acting at the extrema of an isolated resistor. This time-dependent voltage, known as noise voltage, depends on the conductor's temperature and its mean square value is given by

$$\langle V^2\rangle = 4R\Delta\nu k_B T, \tag{11}$$

where R is the resistance and $\Delta\nu$ is the bandwidth of the observed frequencies.

1.2. *Green–Kubo relations*

Let us now quickly review the general framework of the fluctuation-dissipation relations provided in a quantum-mechanical setting by the Green–Kubo relations [8, 9].

Consider an isolated quantum system, whose Hamiltonian operator is H_0, which is a self-adjoint operator on a Hilbert space $\mathcal{H}$. Suppose that the system is at thermal equilibrium at temperature T, say

$$\rho_\beta = \frac{e^{-\beta H_0}}{Z_0}, \tag{12}$$

where $\beta = 1/k_B T$, and $Z_0 = \mathrm{Tr}(e^{-\beta H_0})$ is the partition function. Assume that the system is perturbed by an external time-dependent force, so that the total Hamiltonian reads

$$H(\Lambda_t) = H_0 - \Lambda_t Q, \tag{13}$$

where $t \in [0, \tau] \mapsto \Lambda_t \in \mathbb{R}$, with $\tau > 0$, and $Q = Q^\dagger$ is the observable coupled to the force Λ_t. For simplicity, we shall assume that Q is bounded.

The motion of the system is perturbed by the force Λ, but the perturbation is small if the force is weak. We will confine ourselves to weak perturbations and look at the response of the system in the linear approximation. The response is observed through the average change $\Delta B(t)$ of a bounded observable B. It is not difficult to prove [9] that at the first order in Λ,

$$\Delta B(t) := \mathrm{Tr}(B\rho_t) - \mathrm{Tr}(B\rho_\beta) = \int_0^t \Phi_{BQ}(t-s)\Lambda_s \mathrm{d}s, \tag{14}$$

where ρ_t denotes the evolution at time t of ρ_β under the action of the perturbed Hamiltonian (13). The kernel Φ_{BQ} is the so called response function and is given by

$$\Phi_{BQ}(t) = \frac{\langle [Q, B(t)] \rangle_\beta}{i\hbar}, \tag{15}$$

where $[A, B] = AB - BA$ is the commutator,

$$B(t) = e^{iH_0 t/\hbar} B e^{-iH_0 t/\hbar} \quad (t \in \mathbb{R}) \tag{16}$$

is the (unperturbed) evolution of the observable B, and $\langle A \rangle_\beta = \mathrm{tr}(A\rho_\beta)$ is the thermal expectation value.

The second ingredient is the correlation function Ψ_{BQ}:

$$\Psi_{BQ}(t) = \frac{\langle \{Q, B(t)\} \rangle_\beta}{2}, \tag{17}$$

where $\{A, B\} = AB + BA$ is the anticommutator.

The quantum fluctuation-dissipation theorem [10] links the above two functions, that is

$$\boxed{\hat{\Psi}_{BQ}(\omega) = \frac{\hbar\omega}{2}\coth\left(\frac{\beta\hbar\omega}{2}\right)\hat{\Phi}_{BQ}(\omega),} \tag{18}$$

where $\hat{f}$ denotes the Fourier transform of the function f,

$$\hat{f}(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{d}t. \tag{19}$$

Notice that the classical limit, $\hbar \to 0$, of the quantum fluctuation-dissipation theorem reads

$$\hat{\Psi}(\omega) = \beta \hat{\Phi}(\omega), \tag{20}$$

since $\coth(x) \sim 1/x$ as $x \to 0$. The classical limit is in accordance with Einstein's relation (2). The Green–Kubo relations started a new trend of research on higher order fluctuation-dissipation relations beyond the linear regime [11–13].

2. Lecture 2: Classical Fluctuation Relations

In this section, we are going to deduce the so-called *Jarzynksi equality* [13] and the *Crooks fluctuation theorem* [14], which are two paradigmatic examples of exact classical fluctuation theorems.

Consider a fluctuating quantity x (for example the number of transported electrons in a resistance, or the heat, or the work in non-equilibrium transformations) and call $p_F(x)$ (the subscript F stands for *forward*) the probability density function of x during a non-equilibrium thermodynamic transformation; call $p_B(x)$ the probability density function of the same quantity x but under the backward transformation (the subscript B stands for *backward*). Then, due to microreversibility, fluctuation relations are usually expressed as a link between $p_F(x)$ and $p_B(x)$ of the form

$$p_F(x) = \mathrm{e}^{\beta(x-a)} p_B(-x), \tag{21}$$

where a is a quantity related to the equilibrium starting points of the forward and backward processes.

Equation (21) relates non-equilibrium quantities, say the probability distributions of x for the forward and backward processes, to equilibrium quantities, say the constants β and a. Thus, microreversibility implies that at the macroscopic level, the forward probability is exponentially more likely than the backward one. For example it could happen that the entropy of a small isolated system might spontaneously decrease, e.g. the water in a glass could spontaneously freeze at room temperature. However, relation (21) maintains that this process is extremely highly unprobable.

As already discussed in Lecture 1, in the analysis of classical fluctuation relations two ingredients are fundamental:

- reversibility at microscopic scales;
- initial condition at thermal equilibrium.

We would like to analyze the work fluctuations for a classical non-autonomous system and deduce, as a consequence, the corresponding exact fluctuation relation. Consider a classical system described by the Hamiltonian function

$$H(z, \lambda) = H_0(z) - \lambda Q(z), \tag{22}$$

where $z = (q, p)$ is a point in the phase space Γ of the physical system, H_0 is the unperturbed Hamiltonian function, and λ is a real parameter representing the external force coupled to the conjugate variable $Q(z)$. We assume that both H_0 and Q are smooth functions on Γ.

The Gibbs canonical state at temperature β^{-1} associated to the Hamiltonian (22) is

$$\rho_\beta^\lambda(z) = \frac{e^{-\beta H(z, \lambda)}}{Z(\lambda)} = \frac{e^{-\beta(H_0(z) - \lambda Q(z))}}{Z(\lambda)}, \tag{23}$$

$$Z(\lambda) = \int_\Gamma e^{-\beta H(z, \lambda)} dz, \quad z \in \Gamma, \ \lambda \in \mathbb{R}. \tag{24}$$

The logarithm of the partition function is related to the Helmholtz free energy F by [15]

$$F(\lambda) = -\beta^{-1} \ln Z(\lambda). \tag{25}$$

Note that for an unbounded phase space, such as $\Gamma = \mathbb{R}^{2d}$, the canonical state (23) is a well-defined probability density if $H(z, \lambda)$ is confining, that is, for all $\lambda \in \mathbb{R}$, $H(z, \lambda) \to +\infty$ (sufficiently fast) as $|z| \to \infty$.

Next, consider the time reversal transformation on the phase space Γ:

$$\theta : \Gamma \to \Gamma, \quad (q, p) \mapsto (q, -p). \tag{26}$$

We assume that:

(1) the unperturbed Hamiltonian H_0 is invariant under time-reversal transformations, say $H_0(\theta(z)) = H_0(q, -p) = H_0(z)$ for every point $z = (q, p)$ in phase space Γ;
(2) the conjugate variable Q has a definite behavior under time-reversal transformations, say $Q(\theta(z)) = \eta_Q Q(z)$, with $\eta_Q = \pm 1$. (For example if $Q(q, p) = q_1$, then $\eta_Q = 1$, while if $Q(q, p) = q_1 p_2 - p_2 q_1$, then $\eta_Q = -1$).

As a consequence, we get that

$$H(\theta(z), \lambda) = H(z, \eta_Q \lambda), \tag{27}$$

for all $z \in \Gamma$ and $\lambda \in \mathbb{R}$. Indeed,

$$H(\theta(z), \lambda) = H_0(\theta(z)) - \lambda Q(\theta(z)) = H_0(z) - \lambda \eta_Q Q(z) = H(z, \eta_Q \lambda). \tag{28}$$

In fact, condition (27) is equivalent to the assumptions on H_0 and Q, as the reader can easily prove.

Suppose now that a given force *protocol*, assumed to be smooth,

$$t \in [0, \tau] \mapsto \Lambda_t \in \mathbb{R} \tag{29}$$

is assigned to the external force λ in (22), so that the Hamilton function $H(z, \Lambda_t)$ is time-dependent and the overall quantity $-\Lambda_t Q(z)$ is the time-dependent perturbation term.

Our intention is to implement a time-reversed protocol, since it is evidently impossible to turn back the physical time (a procedure that could be implemented on a computer simulation). Nevertheless, we need a way to implement a backward protocol in real time in order to deduce classical fluctuation relations.

We define the *backward protocol* as

$$\tilde{\Lambda} : [0, \tau] \to \mathbb{R}, \quad \tilde{\Lambda}_t := \eta_Q \Lambda_{\tau-t}. \tag{30}$$

In this way, modulo a sign η_Q related to the time-reversal parity of Q, the external force traces back its evolution from $\tilde{\Lambda}_0 = \eta_Q \Lambda_\tau$ to $\tilde{\Lambda}_\tau = \eta_Q \Lambda_0$.

We will denote by $\varphi_{t,0}[z_0; \Lambda]$ the solution, assumed to exist and to be unique, of the Hamilton equations at time t under the external protocol Λ,

$$\begin{cases} \dot{q}(t) = \nabla_p H(z(t), \Lambda_t), \\ \dot{p}(t) = -\nabla_q H(z(t), \Lambda_t), \end{cases} \tag{31}$$

with initial conditions $z(0) = z_0 = (q_0, p_0)$, so that

$$z \in \Gamma \mapsto \varphi_{t,0}[z; \Lambda] \in \Gamma \tag{32}$$

is the Hamiltonian flow in $[0, t]$ under the protocol Λ.

It is an instructive exercise on the use of Hamilton equations to prove that under the microreversibility assumption (27), the flow generated by the backward protocol $\tilde{\Lambda}$ and the forward flow are related as follows [16]:

$$\varphi_{t,0}[z_0; \Lambda] = \theta(\varphi_{\tau-t,0}[\theta(z(\tau)); \tilde{\Lambda}]), \quad z(\tau) = \varphi_{\tau,0}[z_0; \Lambda]. \tag{33}$$

Note that $\theta(z(\tau))$ is the initial position under the backward protocol $\tilde{\Lambda}$. The lazy reader can convince himself of the validity of (33) by carefully looking at Fig. 2.
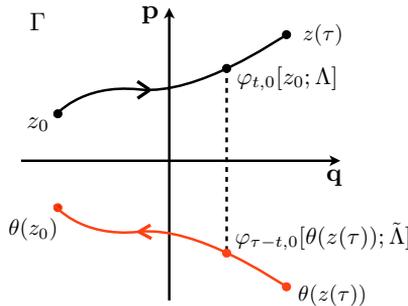

Fig. 2. Microreversibility [3]. The initial point z_0 at time $t = 0$ gets dragged under the protocol Λ along the trajectory $\varphi_{t,0}[z_0; \Lambda]$, until it reaches its final position $z(\tau)$ at time $t = \tau$. Plotted in red is the trajectory under the backward protocol $\tilde{\Lambda}$. It starts at $t = 0$ in $\theta(z(\tau))$, evolves along $\varphi_{\tau-t,0}[\theta(z(\tau)); \tilde{\Lambda}]$ until time $t = \tau$, at $\theta(z_0)$.

The second ingredient is the initial state and its nature is statistical: we assume that the initial conditions z_0 of the system are randomly sampled from the Gibbs canonical distribution (23) at $t = 0$,

$$\rho_\beta^{\Lambda_0}(z) = \frac{\mathrm{e}^{-\beta H(z,\Lambda_0)}}{Z(\Lambda_0)}. \tag{34}$$

Then, we let the system evolve under the protocol Λ until time $t = \tau$.

Since the dynamics is Hamiltonian, the *work* $W[z_0; \Lambda]$ done by the external protocol on the system is given by the difference between its final and its initial energy, namely

$$W[z_0; \Lambda] = H(z(\tau), \Lambda_\tau) - H(z_0, \Lambda_0), \tag{35}$$

where $z(\tau) = \varphi_{\tau,0}[z_0; \Lambda]$. Clearly, $W[z_0; \Lambda]$ depends on both the initial conditions z_0 and the protocol Λ.

The work can be written as

$$W[z_0; \Lambda] = -\int_0^\tau \dot{\Lambda}_t Q(z(t))\mathrm{d}t, \tag{36}$$

where $\dot{\Lambda}_t = \frac{\mathrm{d}}{\mathrm{d}t}\Lambda_t$. Indeed, we get

$$W[z_0; \Lambda] = H(z(\tau), \Lambda_\tau) - H(z_0, \Lambda_0) = \int_0^\tau \frac{\mathrm{d}}{\mathrm{d}t}H(z(t), \Lambda_t)\mathrm{d}t, \tag{37}$$

where $z(t) = \varphi_{t,0}[z_0; \Lambda]$. Next, we expand the total derivative of $H(z(t), \Lambda_t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}H(z(t), \Lambda_t) = \nabla_z H(z(t), \Lambda_t) \cdot \dot{z}(t) + \frac{\partial}{\partial \lambda}H(z(t), \Lambda_t)\dot{\Lambda}_t \tag{38}$$

$$= \frac{\partial}{\partial \lambda}H(z(t), \Lambda_t)\dot{\Lambda}_t = -Q(z(t))\dot{\Lambda}_t. \tag{39}$$

In the second equality, we used the orthogonality condition between the gradient of the Hamiltonian and the velocity vector field,

$$\nabla_z H(z(t), \Lambda_t) \cdot \dot{z}(t) = \nabla_q H(z(t), \Lambda_t) \cdot \dot{q}(t) + \nabla_p H(z(t), \Lambda_t) \cdot \dot{p}(t)$$
$$= \nabla_q H(z(t), \Lambda_t) \cdot \nabla_p H(z(t), \Lambda_t)$$
$$- \nabla_p H(z(t), \Lambda_t) \cdot \nabla_q H(z(t), \Lambda_t)$$
$$= 0, \tag{40}$$

which is a direct consequence of Hamilton equations (31). Thus, we have proved Eq. (36). So far, we have only made use of Hamiltonian dynamics.

2.1. *Jarzynski equality*

We want to compute the average of $\mathrm{e}^{-\beta W[z_0;\Lambda]}$ over all the possible initial conditions drawn from the canonical distribution (34):

$$\langle \mathrm{e}^{-\beta W}\rangle_\Lambda := \int_\Gamma \mathrm{e}^{-\beta W[z_0;\Lambda]}\rho_\beta^{\Lambda_0}(z_0)\mathrm{d}z_0. \tag{41}$$

The Jarzynski equality states that

$$\langle e^{-\beta W} \rangle_\Lambda = e^{-\beta(F(\Lambda_\tau) - F(\Lambda_0))}, \tag{42}$$

where F is the Helmoltz free energy (25)

This is a fluctuation relation since it connects a non-equilibrium quantity (on the left-hand side) with an equilibrium one (on the right-hand side). The right-hand side is an equilibrium quantity since it depends solely on the initial and final equilibrium states. Note however that (42) is exact: it holds no matter how strong is the external force Λ and how far from the initial equilibrium $\rho_\beta^{\Lambda_0}$ the system is driven by Λ.

Let us consider the average (41) and let us perform some simple computations:

$$\langle e^{-\beta W} \rangle_\Lambda = \int_\Gamma e^{-\beta W[z_0; \Lambda]} \rho_\beta^{\Lambda_0}(z_0) dz_0$$

$$= \frac{1}{Z(\Lambda_0)} \int_\Gamma e^{-\beta(H(z(\tau), \Lambda_\tau) - H(z_0, \Lambda_0))} e^{-\beta H(z_0, \Lambda_0)} dz_0,$$

$$= \frac{1}{Z(\Lambda_0)} \int_\Gamma e^{-\beta H(z(\tau), \Lambda_\tau)} dz_0. \tag{43}$$

By the Liouville's theorem, the Hamiltonian flow $\varphi_{\tau,0}[\,\cdot\,; \Lambda]$ is a canonical transformation on the phase space Γ, and thus is volume preserving and has unit Jacobian determinant:

$$z = z(\tau) = \varphi_{\tau,0}[z_0; \Lambda], \quad \left| \frac{\partial z}{\partial z_0} \right| = 1. \tag{44}$$

Therefore, the integration over the initial points z_0 in (43) can be traded for an integration over the final points z, namely

$$\frac{1}{Z(\Lambda_0)} \int_\Gamma e^{-\beta H(z(\tau), \Lambda_\tau)} dz_0 = \frac{1}{Z(\Lambda_0)} \int_\Gamma e^{-\beta H(z, \Lambda_\tau)} dz \overset{(24)}{=} \frac{Z(\Lambda_\tau)}{Z(\Lambda_0)}. \tag{45}$$

Thus, we get

$$\langle e^{-\beta W} \rangle_\Lambda = \frac{Z(\Lambda_\tau)}{Z(\Lambda_0)}, \tag{46}$$

which gives the Jarzynski equality by using the definition of the Helmholtz free energy (25).

A straightforward consequence of (42) follows from the convexity of the exponential function, see Fig. 3. By Jensen's inequality, we get

$$e^{-\beta(F(\Lambda_\tau) - F(\Lambda_0))} = \langle e^{-\beta W} \rangle_\Lambda \geq e^{-\beta \langle W \rangle_\Lambda}. \tag{47}$$

Therefore, one has

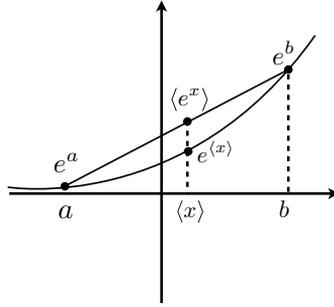$$\langle W \rangle_\Lambda \geq F(\Lambda_\tau) - F(\Lambda_0) =: \Delta F, \tag{48}$$

Fig. 3. Given two points a, b and denoting the average with $\langle x \rangle = (a + b)/2$, from the convexity of the exponential function it follows that $\langle e^x \rangle \geq e^{\langle x \rangle}$, where $\langle e^x \rangle = (e^a + e^b)/2$ is the average of the exponential function over the points a and b.

which is an expression of the second law of thermodynamics. Indeed, if one defines the dissipated work as [15]

$$W_{\text{diss}} = W - \Delta F, \tag{49}$$

then inequality (48) states that the dissipated work on average can only be absorbed:

$$\langle W_{\text{diss}} \rangle_\Lambda = \langle W \rangle_\Lambda - \Delta F \geq 0. \tag{50}$$

In other words, the dissipated work done on a system initially at thermal equilibrium is always non-negative, independently of the protocol Λ.

2.2. *Crooks fluctuation theorem*

Next, we will consider the Crooks fluctuation theorem and deduce from it the Jarzynski equality. In order to do so, we need to introduce probability density functions and make use of the microreversibility assumption (27).

The probability density function (from now on PdF) of the work under the protocol Λ is given by

$$p[w; \Lambda] = \langle \delta(w - W) \rangle_\Lambda, \tag{51}$$

where the average is over the initial Gibbs ensemble $\rho_\beta^{\Lambda_0}$ in (34), namely

$$p[w; \Lambda] = \int_\Gamma \frac{e^{-\beta H(z_0, \Lambda_0)}}{Z(\Lambda_0)} \delta(w - W[z_0; \Lambda]) dz_0, \tag{52}$$

and δ is the Dirac measure. From the PdF of the work, one can get the average of any continuous bounded function f. Indeed,

$$\langle f(W) \rangle_\Lambda = \int_\Gamma f(W[z_0; \Lambda]) \frac{e^{-\beta H(z_0, \Lambda_0)}}{Z(\Lambda_0)} dz_0$$

$$= \int_\Gamma dz_0 \frac{e^{-\beta H(z_0, \Lambda_0)}}{Z(\Lambda_0)} \int_\mathbb{R} dw f(w) \delta(w - W[z_0; \Lambda])$$

$$= \int_{\mathbb{R}} \mathrm{d}w f(w) \int_{\Gamma} \mathrm{d}z_0 \frac{\mathrm{e}^{-\beta H(z_0, \Lambda_0)}}{Z(\Lambda_0)} \delta(w - W[z_0; \Lambda])$$

$$= \int_{\mathbb{R}} f(w) p[w; \Lambda] \mathrm{d}w. \tag{53}$$

Our objective now is to prove the Crooks fluctuation theorem:

$$\boxed{p[w; \Lambda] = \mathrm{e}^{\beta(w - \Delta F)} p[-w; \tilde{\Lambda}],} \tag{54}$$

where

$$\Delta F = F(\Lambda_\tau) - F(\Lambda_0) = -\frac{1}{\beta} \ln\left(\frac{Z(\Lambda_\tau)}{Z(\Lambda_0)}\right). \tag{55}$$

The above equality relates the probability of absorbing work under the protocol Λ to that of releasing work under the backward protocol $\tilde{\Lambda}$. It is worth noticing that the ratio between the two probabilities is exponentially large for $w > \Delta F$.

We start with a simple manipulation of the PdF (52):

$$p[w; \Lambda] = \int_{\Gamma} \frac{\mathrm{e}^{-\beta H(z_0, \Lambda_0)}}{Z(\Lambda_0)} \delta(w - W[z_0; \Lambda]) \mathrm{d}z_0$$

$$= \frac{1}{Z(\Lambda_0)} \mathrm{e}^{\beta w} \int_{\Gamma} \mathrm{e}^{-\beta H(z(\tau), \Lambda_\tau)} \delta(w - H(z(\tau), \Lambda_\tau) + H(z_0, \Lambda_0)) \mathrm{d}z_0$$

$$= \frac{Z(\Lambda_\tau)}{Z(\Lambda_0)} \mathrm{e}^{\beta w} \int_{\Gamma} \frac{\mathrm{e}^{-\beta H(z(\tau), \Lambda_\tau)}}{Z(\Lambda_\tau)} \delta(-w - H(z_0, \Lambda_0) + H(z(\tau), \Lambda_\tau)) \mathrm{d}z_0$$

$$= \mathrm{e}^{\beta(w - \Delta F)} \int_{\Gamma} \rho_\beta^{\Lambda_\tau}(z(\tau)) \delta(-w - H(z_0, \Lambda_0) + H(z(\tau), \Lambda_\tau)) \mathrm{d}z_0, \tag{56}$$

where we used the definition of work (35) and, in the last equality, the definition of $\rho_\beta^{\Lambda_\tau}(z)$ in Eq. (23).

Next, we add the ingredient of microreversibility in order to rewrite the above integral. We recall the relation between the Hamiltonian flow under the protocol Λ and the one generated by the backward protocol $\tilde{\Lambda}$ given in Eq. (33), which is valid for all $t \in [0, \tau]$. Thus, at time $t = 0$, we get an alternative way of writing the initial datum z_0:

$$z_0 = \varphi_{0,0}[z_0; \Lambda] = \theta(\varphi_{\tau,0}[\theta(z(\tau)); \tilde{\Lambda}]). \tag{57}$$

Using the microreversibility assumption (27), we can rewrite the Hamiltonian at $t = 0$ as follows:

$$H(z_0, \Lambda_0) = H(\theta(\varphi_{\tau,0}[\theta(z(\tau)); \tilde{\Lambda}]), \Lambda_0) = H(\varphi_{\tau,0}[\theta(z(\tau)); \tilde{\Lambda}], \eta_Q \Lambda_0)$$

$$= H(\varphi_{\tau,0}[\theta(z(\tau)); \tilde{\Lambda}], \tilde{\Lambda}_\tau), \tag{58}$$

since $\tilde{\Lambda}_\tau = \eta_Q \Lambda_0$. Moreover, the same assumption implies that

$$H(z(\tau), \Lambda_\tau) = H(\theta(z(\tau)), \tilde{\Lambda}_0), \tag{59}$$

because $\tilde{\Lambda}_0 = \eta_Q \Lambda_\tau$. We are now ready to plug Eqs. (57)–(59) into Eq. (56) and the conservation of measure induced by the transformation in Eq. (44), obtaining

$$
\begin{aligned}
\frac{p[w; \Lambda]}{\mathrm{e}^{\beta(w-\Delta F)}} &= \int_\Gamma \rho_\beta^{\Lambda_\tau}(z(\tau))\delta(-w - H(z_0, \Lambda_0) + H(z(\tau), \Lambda_\tau))\mathrm{d}z_0 \\
&= \int_\Gamma \rho_\beta^{\Lambda_\tau}(z(\tau))\delta(-w - H(\varphi_{\tau,0}[\theta(z(\tau)); \tilde{\Lambda}], \tilde{\Lambda}_\tau) + H(\theta(z(\tau)), \tilde{\Lambda}_0))\mathrm{d}z_0 \\
&= \int_\Gamma \rho_\beta^{\Lambda_\tau}(\theta(z))\delta(-w - H(\varphi_{\tau,0}[z; \tilde{\Lambda}], \tilde{\Lambda}_\tau) + H(z, \tilde{\Lambda}_0))\mathrm{d}z \\
&= \int_\Gamma \rho_\beta^{\tilde{\Lambda}_0}(z)\delta(-w - H(\varphi_{\tau,0}[z; \tilde{\Lambda}], \tilde{\Lambda}_\tau) + H(z, \tilde{\Lambda}_0))\mathrm{d}z \\
&= p[-w; \tilde{\Lambda}], \quad\quad\quad (60)
\end{aligned}
$$

where in (60), we performed the change of coordinate induced by the time reversal transformation $z = \theta(z(\tau))$ (see Eq. (38)), whose Jacobian has unit determinant, and we used the relation

$$
\rho_\beta^{\Lambda_\tau}(\theta(z)) = \frac{\mathrm{e}^{-\beta H(\theta(z), \Lambda_\tau)}}{Z(\Lambda_\tau)} = \frac{\mathrm{e}^{-\beta H(z, \eta_Q \Lambda_\tau)}}{Z(\eta_Q \Lambda_\tau)} = \rho_\beta^{\eta_Q \Lambda_\tau}(z) = \rho_\beta^{\tilde{\Lambda}_0}(z). \quad\quad (61)
$$

The Crooks fluctuation relation (54) states that if we consider a positive work $w > \Delta F > 0$, the probability that the work is injected into the system is larger by a factor $\mathrm{e}^{\beta(w-\Delta F)}$ than the probability that it might be absorbed under the backward forcing. In other words, energy consuming processes are exponentially more probable than energy releasing processes. Crooks fluctuation theorem expresses the second law of thermodynamics at a detailed level which quantifies the relative frequency of energy releasing processes.

Moreover, from the Crooks fluctuation theorem, it is possible to recover the Jarzynski equality. Indeed, from (53) and (54), one has

$$
\begin{aligned}
\langle \mathrm{e}^{-\beta W} \rangle_\Lambda &= \int_\mathbb{R} \mathrm{e}^{-\beta w} p[w; \Lambda]\mathrm{d}w \\
&= \int_\mathbb{R} \mathrm{e}^{-\beta w}\mathrm{e}^{\beta(w-\Delta F)} p[-w; \tilde{\Lambda}]\mathrm{d}w \\
&= \mathrm{e}^{-\beta \Delta F} \int_\mathbb{R} p[-w; \tilde{\Lambda}]\mathrm{d}w = \mathrm{e}^{-\beta \Delta F}, \quad\quad (62)
\end{aligned}
$$

since $\int_\mathbb{R} p[-w; \tilde{\Lambda}]\mathrm{d}w = 1$.

A final comment is in order. A straightforward corollary of the Crooks fluctuation relation (54) is the following generalization of the Jarzinski equality, whose proof is left to the reader.

Let $B : \Gamma \to \mathbb{R}$ such that $B(\theta(z)) = \eta_B B(z)$, for all $z \in \Gamma$, with $\eta_B = \pm 1$, then

$$
\left\langle \exp\left(\int_0^\tau u_t B(t)\mathrm{d}t\right) \mathrm{e}^{-\beta W} \right\rangle_\Lambda = \mathrm{e}^{-\beta \Delta F} \left\langle \exp\left(\int_0^\tau \tilde{u}_t B(t)\mathrm{d}t\right) \right\rangle_{\tilde{\Lambda}}, \quad (63)
$$

for all test functions $u : [0, \tau] \to \mathbb{R}$, where $\tilde{u}_t := \eta_B u_{\tau-t}$, for all $t \in [0, \tau]$, is the backward function. The relation (63) is the generating functional of the fluctuation-dissipation relations at all order: take the functional derivatives with respect to Λ and u at $\Lambda = u \equiv 0$.

3. Lecture 3: Quantum Fluctuation Relations

In this lecture, we would like to discuss the quantum version of the fluctuation relations, whose classical version was proved in the previous section.

From the axioms of quantum mechanics [17], the Hamiltonian function on phase space has to be substituted with a self-adjoint operator. For this reason, we are going to consider the following Hamiltonian operator on the Hilbert space $\mathcal{H}$:

$$H(\lambda) = H_0 - \lambda Q, \tag{64}$$

where H_0 and Q are self-adjoint operators, while λ is a real parameter representing the external force. The canonical Gibbs state is described quantum mechanically by a density matrix [18]:

$$\rho_\beta^\lambda = \frac{1}{Z(\lambda)} e^{-\beta H(\lambda)}, \tag{65}$$

where

$$Z(\lambda) = \text{tr}(e^{-\beta H(\lambda)}) \tag{66}$$

is the partition function. The Helmholtz free energy is defined in terms of the partition function as in the classical case (25).

In the following, we will assume that $\{H(\lambda)\}_{\lambda \in \mathbb{R}}$ is a family of (unbounded) self-adjoint operators on a common domain D. Moreover, we assume that ρ_β^λ is trace-class for all $\lambda \in \mathbb{R}$ and $\beta > 0$. This implies that the Hamiltonians (64) have a discrete spectrum with finite multiplicity, namely,

$$H(\lambda) = \sum_m E_m^\lambda \Pi_m^\lambda, \tag{67}$$

where $\{E_m^\lambda\}$ are the distinct eigenvalues of $H(\lambda)$ ($E_m^\lambda \neq E_n^\lambda$ for $m \neq n$), and the eigenprojections $\{\Pi_m^\lambda\}$ are of finite rank, $\text{tr}(\Pi_m^\lambda) < +\infty$ for all m. Moreover, if $\mathcal{H}$ is infinite-dimensional, then $E_m^\lambda \to +\infty$ as $m \to +\infty$.

Assume now that a given protocol $t \in [0, \tau] \mapsto \Lambda_t \in \mathbb{R}$ is assigned to the external force in (64), so that the Hamiltonian is time-dependent, $t \mapsto H(\Lambda_t)$. The quantum evolution in the interval $[s, t]$ is described by the Schrödinger equation, which, from the operator point of view, reads

$$i\frac{\partial}{\partial t} U_{t,s}[\Lambda]\psi = H(\Lambda_t) U_{t,s}[\Lambda]\psi, \quad U_{s,s}[\Lambda] = \mathbb{I}, \quad (t \geq s), \tag{68}$$

for all $\psi \in D$: if the system is initially in the state $\psi \in D$, according to the Schrödinger equation, it will evolve at time t to the state $U_{t,s}[\Lambda]\psi \in D$. Here and in

the following we set $\hbar = 1$. Note that one can instead consider the derivative with respect to the initial time s and get

$$i\frac{\partial}{\partial s}U_{t,s}[\Lambda]\psi = U_{t,s}[\Lambda]H(\Lambda_s)\psi, \quad U_{t,t}[\Lambda] = \mathbb{I}, \quad (t \geq s), \tag{69}$$

for all $\psi \in D$, which is a final value problem. We will explicitly denote the unique solution to (68) or to (69) by

$$U_{t,s}[\Lambda] = \mathcal{T}\exp\left(-i\int_s^t H(\Lambda_u)\mathrm{d}u\right) = \mathcal{T}\exp\left(-i\int_s^t (H_0 - \Lambda_u Q)\mathrm{d}u\right), \tag{70}$$

where $\mathcal{T}$ is the time-ordered product.

In analogy with the classical case (35), let us tentatively define the work as the difference between the Hamiltonian operator at time $t = \tau$ and the Hamiltonian operator at time $t = 0$ in the Heisenberg picture:

$$W[\Lambda] = U_{\tau,0}^\dagger[\Lambda]H(\Lambda_\tau)U_{\tau,0}[\Lambda] - H(\Lambda_0). \tag{71}$$

In order to get Jarzynski equality, then, one could try to follow step by step the derivation used for the classical case (see Eq. (43)), but in this case, the cancellation in Eq. (43) cannot be made. Indeed, in general the Hamiltonians (64) at different times do not commute, unless H_0 commutes with Q:

$$[H(\Lambda_t), H(\Lambda_s)] = (\Lambda_t - \Lambda_s)[H_0, Q]. \tag{72}$$

In fact, it can be shown that with the definition of work given by (71), one has that

$$\langle e^{-\beta W}\rangle_\Lambda := \mathrm{tr}(\rho_\beta^{\Lambda_0} e^{-\beta W[\Lambda]}) = e^{-\beta\Delta F} \tag{73}$$

if and only if $[H(\Lambda_t), H(\Lambda_s)] = 0$ for all $t, s \in [0, \tau]$. The last condition applies either to the case of a constant protocol, which would imply $\Delta F = 0$ or to the commutative case $[H_0, Q] = 0$ that is, morally, to a classical situation.

At first look, it may seem that there could be no genuinely quantum counterpart of the Jarzynski equality. The problem, in fact, relies on the definition (71) of work under the protocol Λ, and one has to think more deeply about the meaning of work.

It is well known that work characterizes a process rather than a state of the system, and indeed it depends on the trajectory followed by the system from its initial to its final state (see Eq. (35)). As such, in quantum mechanics, work cannot be associated to a self-adjoint operator whose eigenvalues are determined by an energy measurement at a given time. Instead, in order to determine the exchanged work, energy must be measured twice, at two instants of time, say at time $t = 0$ and time $t = \tau$. The difference of the outcomes of these two measurements will yield the work performed on the system in that particular instance.

This two-time measurements procedure is the operational description of the measure of work (35) along a single trajectory $z_0 \mapsto z(\tau)$, which in the classical case is a particular instance of all possible trajectories drawn from the initial thermal state (34). This description can be immediately exported to the quantum world, where, however, the measurement process will add quantum fluctuations to

the classical statistical fluctuations due to the choice of a thermal initial state. As a consequence, the difference of the outcomes of a two-time measurement is different from the measurement of the difference between the corresponding measured operators (71), whose operational meaning is unclear.

Let us look step by step at the two-measurement process:

(1) As in the classical case the system is prepared in the Gibbs state

$$\rho_\beta^{\Lambda_0} = \frac{\mathrm{e}^{-\beta H(\Lambda_0)}}{Z(\Lambda_0)} = \frac{1}{Z(\Lambda_0)} \sum_n \mathrm{e}^{-\beta E_n^{\Lambda_0}} \Pi_n^{\Lambda_0}. \tag{74}$$

(2) At time $t = 0$, the energy of the system is measured and the outcome is, say, the eigenvalue $E_n^{\Lambda_0}$ for some n, and the state of the system becomes

$$\rho_n = \frac{\Pi_n^{\Lambda_0} \rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0}}{p_n^{\Lambda_0}}, \quad p_n^{\Lambda_0} = \mathrm{tr}(\rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0}), \tag{75}$$

where $p_n^{\Lambda_0} = \mathrm{tr}(\rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0})$ is the probability of getting the outcome $E_n^{\Lambda_0}$.

(3) Then, one lets the system evolve for a time τ, so that its state becomes

$$\rho_n(\tau) = U_{\tau,0}[\Lambda] \rho_n U_{\tau,0}^\dagger[\Lambda]. \tag{76}$$

(4) Finally, at time $t = \tau$ a second energy measurement is performed and the outcome $E_m^{\Lambda_\tau}$ is obtained with probability

$$p_{m|n} = \mathrm{tr}(\Pi_m^{\Lambda_\tau} \rho_n(\tau)), \tag{77}$$

and the state of the system becomes

$$\rho_{m,n} = \frac{\Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \rho_n U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau}}{p_{m|n}}$$

$$= \frac{\Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0} \rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0} U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau}}{p_{m|n} p_n}. \tag{78}$$

Summarizing, the overall work done on the system in this particular instance is given by the difference of the final and initial measurement outcomes,

$$W = E_m^{\Lambda_\tau} - E_n^{\Lambda_0}, \tag{79}$$

whose probability is

$$p_{m,n} = p_{m|n} p_n^{\Lambda_0}$$

$$= \mathrm{tr}(\Pi_m^{\Lambda_\tau} \rho_n(\tau)) \, \mathrm{tr}(\rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0})$$

$$= \mathrm{tr}\left(\Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \frac{\Pi_n^{\Lambda_0} \rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0}}{p_n^{\Lambda_0}} U_{\tau,0}^\dagger[\Lambda]\right) p_n^{\Lambda_0}$$

$$= \mathrm{tr}(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0} \rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0}), \tag{80}$$

where, in the last equality, the cyclicity of the trace was used. From Eq. (74), it follows that

$$\Pi_n^{\Lambda_0} \rho_\beta^{\Lambda_0} \Pi_n^{\Lambda_0} = \frac{1}{Z(\Lambda_0)} e^{-\beta E_n^{\Lambda_0}} \Pi_n^{\Lambda_0}, \tag{81}$$

so that the probability of obtaining the outcomes $(E_n^{\Lambda_0}, E_m^{\Lambda_\tau})$, and thus the state (78), reads

$$p_{m,n} = \frac{e^{-\beta E_n^{\Lambda_0}}}{Z(\Lambda_0)} \operatorname{tr}(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0}). \tag{82}$$

Now we can compute the average of $e^{-\beta W}$ over all possible outcomes:

$$\begin{aligned}
\langle e^{-\beta W} \rangle_\Lambda &:= \sum_{m,n} e^{-\beta(E_m^{\Lambda_\tau} - E_n^{\Lambda_0})} p_{m,n} \\
&= \sum_{m,n} \frac{e^{-\beta E_m^{\Lambda_\tau}}}{Z(\Lambda_0)} \operatorname{tr}(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0}) \\
&= \sum_m \frac{e^{-\beta E_m^{\Lambda_\tau}}}{Z(\Lambda_0)} \operatorname{tr}\left(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \sum_n \Pi_n^{\Lambda_0} \right) \\
&= \sum_m \frac{e^{-\beta E_m^{\Lambda_\tau}}}{Z(\Lambda_0)} \operatorname{tr}(\Pi_m^{\Lambda_\tau}) = \frac{1}{Z(\Lambda_0)} \operatorname{tr}\left(\sum_m e^{-\beta E_m^{\Lambda_\tau}} \Pi_m^{\Lambda_\tau} \right) \\
&= \frac{1}{Z(\Lambda_0)} \operatorname{tr}(e^{-\beta H(\Lambda_\tau)}) = \frac{Z(\Lambda_\tau)}{Z(\Lambda_0)} = e^{-\beta \Delta F},
\end{aligned} \tag{83}$$

where we used the cyclicity of the trace, the relation

$$\sum_m \operatorname{tr}(\Pi_m^{\Lambda_0} A) = \operatorname{tr}\left(\sum_m \Pi_m^{\Lambda_0} A \right) = \operatorname{tr}(A), \tag{84}$$

valid for all trace-class operators A, and the equality

$$e^{-\beta H(\Lambda_\tau)} = \sum_m e^{-\beta E_m^{\Lambda_\tau}} \Pi_m^{\Lambda_\tau}. \tag{85}$$

Thus, we have proved the Jarzynski equality (42) for a quantum system.

This equality takes into account the presence of thermal fluctuations of work due to the Gibbs initial state, as in the classical case, together with its quantum fluctuations inherent in the two-time measurement process.

3.1. *Microreversibility*

Let $\Theta : \mathcal{H} \to \mathcal{H}$ be the quantum time-reversal (anti-unitary) operator, with $\Theta^2 = \mathbb{I}$. We assume that the Hamiltonian is time-reversal invariant, namely that for

all $\lambda \in \mathbb{R}$:

$$\Theta H(\lambda)\Theta = H(\eta_Q \lambda), \tag{86}$$

which is the quantum version of the assumption (27). As in the classical case, η_Q is the time-reversal parity of the observable Q, namely

$$\Theta Q \Theta = \eta_Q Q, \quad \eta_Q = \pm 1. \tag{87}$$

From the spectral decomposition (67), one gets that the property (86) implies that

$$E_m^{\eta_Q \lambda} = E_m^{\lambda}, \quad \Theta \Pi_m^{\lambda} \Theta = \Pi_m^{\eta_Q \lambda}, \tag{88}$$

for all m and λ.

We will prove that

$$U_{t,0}[\Lambda] = \Theta U_{\tau-t}[\tilde{\Lambda}]\Theta U_{\tau,0}[\Lambda], \tag{89}$$

where $t \in [0, \tau] \mapsto \tilde{\Lambda}_t := \eta_Q \Lambda_{\tau-t}$ is the backward protocol (30). Notice the analogy with the classical case where

$$\varphi_{t,0}[\cdot; \Lambda] = \theta(\varphi_{\tau-t,0}[\theta(\varphi_{\tau,0}[\cdot; \Lambda]); \tilde{\Lambda}]), \tag{90}$$

see Eq. (33). In order to prove (89), we observe that, from (68), $U_{\tau-t,0}[\tilde{\Lambda}]$ satisfies the following integral equation on the common domain D:

$$\begin{aligned}
U_{\tau-t,0}[\tilde{\Lambda}] &= \mathbb{I} - i \int_0^{\tau-t} ds\, H(\tilde{\Lambda}_s)U_{s,0}[\tilde{\Lambda}] \\
&= \mathbb{I} - i \int_0^{\tau-t} ds\, H(\eta_Q \Lambda_{\tau-s})U_{s,0}[\tilde{\Lambda}] \\
&= \mathbb{I} - i \int_t^{\tau} d\sigma\, H(\eta_Q \Lambda_{\sigma})U_{\tau-\sigma,0}[\tilde{\Lambda}].
\end{aligned} \tag{91}$$

Therefore, by using (86), one has

$$\begin{aligned}
\Theta U_{\tau-t,0}[\eta_Q \tilde{\Lambda}]\Theta &= \mathbb{I} + i \int_t^{\tau} d\sigma\, \Theta H(\eta_Q \Lambda_{\sigma})U_{\tau-\sigma,0}[\tilde{\Lambda}]\Theta \\
&= \mathbb{I} + i \int_t^{\tau} d\sigma\, H(\Lambda_{\sigma})\Theta U_{\tau-\sigma,0}[\tilde{\Lambda}]\Theta.
\end{aligned} \tag{92}$$

On the other hand, from (69), one gets

$$U_{\tau,t}[\Lambda] = \mathbb{I} - i \int_t^{\tau} d\sigma\, U_{\tau,\sigma}[\Lambda]H(\Lambda_{\sigma}) \tag{93}$$

and thus

$$U_{\tau,t}^{\dagger}[\Lambda] = \mathbb{I} + i \int_t^{\tau} d\sigma\, H(\Lambda_{\sigma})U_{\tau,\sigma}^{\dagger}[\Lambda]. \tag{94}$$

By comparing (92) with (94), we see that the unitaries $\Theta U_{\tau-t,0}[\tilde{\Lambda}]\Theta$ and $U_{\tau,t}^{\dagger}[\Lambda]$ satisfy the same integral equation, whence by uniqueness they are equal:

$$\Theta U_{\tau-t,0}[\tilde{\Lambda}]\Theta = U_{\tau,t}^{\dagger}[\Lambda] = U_{t,0}[\Lambda]U_{\tau,0}^{\dagger}[\Lambda]. \tag{95}$$

3.2. *Quantum Crooks fluctuation theorem*

The probability $p_{m,n}$ of getting the outcomes $(E_n^{\Lambda_0}, E_m^{\Lambda_\tau})$ in the two-measurement process of the work done on a quantum system was derived in (82). It follows that the probability of getting a value w of the work is

$$P[w; \Lambda] = \sum_{m,n} p_{m,n} \delta(w, E_m^{\Lambda_\tau} - E_n^{\Lambda_0}), \tag{96}$$

where $\delta(x, y)$ is the Kronecker delta. We want to prove that

$$\boxed{P[w; \Lambda] = e^{\beta(w - \Delta F)} P[-w; \tilde{\Lambda}].} \tag{97}$$

We first observe that

$$P[w; \Lambda] = \sum_{m,n} p_{m,n} \delta(w, E_m^{\Lambda_\tau} - E_n^{\Lambda_0})$$

$$= \sum_{m,n} \frac{e^{-\beta E_n^{\Lambda_0}}}{Z(\Lambda_0)} \operatorname{tr}(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0}) \delta(w, E_m^{\Lambda_\tau} - E_n^{\Lambda_0})$$

$$= \frac{e^{\beta w}}{Z(\Lambda_0)} \sum_{m,n} e^{-\beta E_m^{\Lambda_\tau}} \operatorname{tr}(U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\Lambda_\tau} U_{\tau,0}[\Lambda] \Pi_n^{\Lambda_0}) \delta(-w, E_n^{\Lambda_0} - E_m^{\Lambda_\tau})$$

$$= e^{\beta(w - \Delta F)} \sum_{m,n} \frac{e^{-\beta E_m^{\tilde{\Lambda}_0}}}{Z(\tilde{\Lambda}_0)} \operatorname{tr}(U_{\tau,0}[\Lambda] \Pi_n^{\eta_Q \tilde{\Lambda}_\tau} U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\eta_Q \tilde{\Lambda}_0}) \delta(-w, E_n^{\tilde{\Lambda}_\tau} - E_m^{\tilde{\Lambda}_0}), \tag{98}$$

where we used the definition (30) of the backward protocol $\tilde{\Lambda}$, and the assumptions (88). Then, we use microreversibility, namely (89) for $t = 0$ and obtain that

$$U_{\tau,0}^\dagger[\Lambda] = \Theta U_{\tau,0}[\tilde{\Lambda}] \Theta \tag{99}$$

and

$$U_{\tau,0}[\Lambda] = \Theta U_{\tau,0}^\dagger[\tilde{\Lambda}] \Theta, \tag{100}$$

therefore

$$\frac{P[w; \Lambda]}{e^{\beta(w - \Delta F)}} = \sum_{m,n} \frac{e^{-\beta E_m^{\tilde{\Lambda}_0}}}{Z(\tilde{\Lambda}_0)} \operatorname{tr}(U_{\tau,0}[\Lambda] \Pi_n^{\eta_Q \tilde{\Lambda}_\tau} U_{\tau,0}^\dagger[\Lambda] \Pi_m^{\eta_Q \tilde{\Lambda}_0}) \delta(-w, E_n^{\tilde{\Lambda}_\tau} - E_m^{\tilde{\Lambda}_0})$$

$$= \sum_{m,n} \frac{e^{-\beta E_m^{\tilde{\Lambda}_0}}}{Z(\tilde{\Lambda}_0)} \operatorname{tr}(\Theta U_{\tau,0}^\dagger[\tilde{\Lambda}] \Theta \Pi_n^{\eta_Q \tilde{\Lambda}_\tau} \Theta U_{\tau,0}[\tilde{\Lambda}] \Theta \Pi_m^{\eta_Q \tilde{\Lambda}_0}) \delta(-w, E_n^{\tilde{\Lambda}_\tau} - E_m^{\tilde{\Lambda}_0})$$

$$= \sum_{m,n} \frac{e^{-\beta E_m^{\tilde{\Lambda}_0}}}{Z(\tilde{\Lambda}_0)} \operatorname{tr}(U_{\tau,0}^\dagger[\tilde{\Lambda}] \Pi_n^{\tilde{\Lambda}_\tau} U_{\tau,0}[\tilde{\Lambda}] \Pi_m^{\tilde{\Lambda}_0}) \delta(-w, E_n^{\tilde{\Lambda}_\tau} - E_m^{\tilde{\Lambda}_0})$$

$$= P[-w; \tilde{\Lambda}], \tag{101}$$

where we used again (88) and $\Theta^2 = \mathbb{I}$.

Acknowledgments

References

[1] U. Seifert, *Rep. Progr. Phys.* **75** (2012) 126001.
[2] M. Esposito, U. Harbola and S. Mukamel, *Rev. Mod. Phys.* **81** (2009) 1665.
[3] M. Campisi, P.Hänggi and P. Talkner, *Rev. Mod. Phys.* **83** (2011) 771.
[4] Wikipedia, https://en.wikipedia.org/wiki/Brownian-motion.
[5] A. Einstein, *Ann. Phys. (Leipzig)* **322** (1905) 549.
[6] J. Johnson, *Phys. Rev.* **32** (1928) 97.
[7] H. Nyquist, *Phys. Rev.* **32** (1928) 110.
[8] M. S. Green, *J. Chem. Phys.* **22** (1954) 398.
[9] R. Kubo, *J. Phys. Soc. Jpn.* **12** (1957) 570.
[10] H. B. Callen and T. A. Welton, *Phys. Rev.* **83** (1951) 34.
[11] D. J. Evans, E. G. D. Cohen and G. P. Morriss, *Phys. Rev. Lett.* **71** (1993) 2401.
[12] G. Gallavotti and E. G. D. Cohen, *Phys. Rev. Lett.* **74** (1995) 2694.
[13] C. Jarzynski, *Phys. Rev. Lett.* **78** (1997) 2690.
[14] G. E. Crooks, *Phys. Rev. E* **60** (1999) 2721.
[15] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, London, 1958).
[16] R. L. Stratonovich, *Nonlinear Nonequilibrium Thermodynamics II: Advanced Theory*, Springer Series in Synergetics, Vol. 59 (Springer-Verlag, Berlin, 1994).
[17] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
[18] K. Huang, *Statistical Mechanics* (Wiley & Sons, 1987).