

Fluctuation Theorems

Shiling Liang

*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL),
1015 Lausanne, Switzerland*

This report is based on the review article by E.M Sevick et al.[1] on the fluctuation theorem. I also refer to the original papers of Jarzynski [2–4] and adopt notation from a book written by Evans et al. [5]. I discuss the motivation for proposing the fluctuation theorem and provide a detailed derivation of the Evans-Searles Fluctuation Theorem, which shows how macroscopic irreversibility emerged from the time-symmetrical equation of motion and addresses the importance of the initial condition of the system. I also recap the derivation the Crooks Fluctuation Theorem to relate work and free energy of a non-quasistatic process between two states of equilibrium, and I then take the Jarzynski equality as a consequence of the Crooks Fluctuation Theorem. In section III, it is shown that a simple physical example with an analytical solution can be used to verify these relations.

I. INTRODUCTION

Fluctuation theorems present an analytical description of violations of the second law of thermodynamics; however, they also provide rigorous proof to this law from a statistical perspective. Classical thermodynamics can only deal with equilibrium systems and quasistatic processes. The behaviours of non-equilibrium systems are unclear within this scheme. Moreover, reversible equations of motion conflict with the monotonically increasing entropy predicted by the second law. Boltzmann's statistical thermodynamics provides the groundwork for solving this problem by using the second law as a statistical law. However, he was not able to determine a quantitative relation. The most recent progress toward this end was the development of the fluctuation theorem in the 1990s.

In 1994, Evans and Searles[6] proposed the fluctuation theorem, which calculates the probability of negative and positive entropy production. It is valid for systems of any size and with arbitrary distances from equilibrium. It illustrates how irreversible macroscopic phenomena emerge from time-reversible equations of motion. In 1997, Jarzynski[2] developed a more practical relation, Jarzynski equality, which allows us to establish the relationships between work and free energy changes for a non-quasistatic process between two states of equilibrium. This relation reveals more information than do classical thermodynamics, which only reveals discrepancies between work and free energy differences. The next year, Crooks[7] developed a more general form of the Jarzynski equality: the Crooks Fluctuation Theorem. This theorem calculates the relative probability of given forward and backward trajectories.

In this report, I recap the derivations of these relations based on a deterministic Hamiltonian system. I include a short introduction to phase spaces at the beginning and then follow the logical order rather than the chronological order of theorem development (which regards the Jarzynski equality as a consequence of the Crooks Fluctuation Theorem). It is also demonstrated how to view

the second law of thermodynamics is a consequence of the fluctuation theorem.

II. DERIVATION OF FLUCTUATION THEOREMS

A. Phase space and Liouville's Equation

In Hamiltonian mechanics, all possible states of a system expanded a phase space. A typical system in our scale has more than $\sim 10^{23}$ particles. Thus, it is not possible to know their exact configurations, so we use a probability distribution function to characterise such systems. By denoting the probability distribution function as $f(\Gamma, t)$, the probability of finding the system in an infinitesimal phase space volume $d\Gamma$ is

$$P(d\Gamma_t, t) \equiv f(\Gamma_t, t)d\Gamma_t. \quad (1)$$

For a Hamiltonian system, the total time derivative of f is

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial p} \dot{p} + \frac{\partial f}{\partial q} \dot{q} \\ &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \Gamma} \cdot \dot{\Gamma}. \end{aligned} \quad (2)$$

The continuity equation must hold for the probability flow in the phase space

$$\frac{\partial(f\dot{\Gamma})}{\partial \Gamma} + \frac{\partial f}{\partial t} = 0. \quad (3)$$

Using these two equations to eliminate the partial derivative of t , one obtains

$$\frac{df}{dt} = -f \frac{\partial \dot{\Gamma}}{\partial \Gamma}. \quad (4)$$

We now denote $\Lambda \equiv \frac{\partial \dot{\Gamma}}{\partial \Gamma}$, and integrate from the initial time $t = 0$ to $t = \tau$

$$f(\Gamma_\tau, \tau) = f(\Gamma_0, 0) \exp\left(-\int_0^\tau \Lambda dt\right). \quad (5)$$

As the probability of a specific phase space volume must be conserved, we can use equation 2 and equation 5 to obtain the compression factor for the infinitesimal phase space volume:

$$\frac{d\Gamma_\tau}{d\Gamma_0} = \exp\left(\int_0^\tau \Lambda ds\right). \quad (6)$$

B. The Evans-Searles Fluctuation Theorem (1994)

The Evans-Searles fluctuation theorem is the most general one, and it can be applied to a system of any size and any distance from equilibrium.

Using time evolution operator S^t to denote the evolution of a phase space point for a period t and time-reversal mapping operator M^T to reverse the trajectory. For a Hamiltonian system, the reverse mapping is noting but reversing the momentum $p \rightarrow -p$. Since the dynamics are deterministic, a trajectory can be determined by its initial state. For a trajectory origin at phase point Γ_0 , the corresponding anti-trajectory can be easily represented by its initial phase point $\Gamma_0^* \equiv M^T \Gamma_t = M^T S^T \Gamma_0$. Note that the time-reversal mapping does not change the size of phase space volume. With equation (3), we can obtain

$$d\Gamma_0^* = d\Gamma_\tau = d\Gamma_0 \exp\left(\int_0^\tau \Lambda dt\right). \quad (7)$$

Alongside the trajectory of a single-phase space point, we also consider the evolution of an infinitesimal phase space volume $d\Gamma_0$; that is, a bundle of trajectories. The probability of observing this volume is

$$P(d\Gamma_0, 0) \equiv d\Gamma_0 f(\Gamma_0, 0). \quad (8)$$

Macroscopic reversible systems indicate that the probability of observing the forward bundle of trajectories and the corresponding anti-trajectories should be the same:

$$d\Gamma_0 f(\Gamma_0, 0) = d\Gamma_0^* f(\Gamma_0^*, 0). \quad (9)$$

For more general cases, we define a dissipation function to characterize the reversibility of the bundle of trajectories

$$\begin{aligned} \Omega_\tau(\Gamma_0) &\equiv \ln\left(\frac{P(d\Gamma_0, 0)}{P(d\Gamma_0^*, 0)}\right) \\ &= \ln\left(\frac{f(d\Gamma_0, 0)}{f(d\Gamma_0^*, 0)}\right) - \int_0^\tau \Lambda dt. \end{aligned} \quad (10)$$

The positive dissipation function means the anti-trajectories are less probable, and the negative dissipation function gives the opposite case. It takes the value 0 only in the state of equilibrium, where no macroscopic time evolution can be observed. Moreover, the dissipation function has odd parity

$$\Omega_\tau(\Gamma_0) = -\Omega_\tau(\Gamma_0^*). \quad (11)$$

Now it is important to calculate the probability of an anti-trajectory occurring within the entire phase space. We can do the following integral to calculate the relative probability of the dissipation function taking the opposite value.

$$\begin{aligned} &\frac{p(\Omega_\tau = A)}{p(\Omega_\tau = -A)} \\ &= \frac{\int d\Gamma_0 \delta(\Omega_\tau(\Gamma_0) - A) f(\Gamma_0, 0)}{\int d\Gamma_0^* \delta(\Omega_\tau(\Gamma_0^*) + A) f(\Gamma_0^*, 0)} \\ &= \frac{\int d\Gamma_0 \delta(\Omega_\tau(\Gamma_0) - A) f(\Gamma_0, 0)}{\int d\Gamma_0^* \delta(-\Omega_\tau(\Gamma_0) + A) f(\Gamma_0^*, 0)} \\ &= \frac{\int d\Gamma_0 \delta(\Omega_\tau(\Gamma_0) - A) f(\Gamma_0, 0)}{\int d\Gamma_0 \delta(-\Omega_\tau(\Gamma_0) + A) f(\Gamma_0, 0) \exp(-\Omega_\tau(\Gamma_0))} \\ &= \exp(A), \end{aligned} \quad (12)$$

where we used the fact that Γ_0 is a dummy variable for the integral over the whole phase space. Then we use the equation 11. To calculate the third equality, we used the equation 7 with the definition of the dissipation function. The result,

$$\frac{p(\Omega_\tau = A)}{p(\Omega_\tau = -A)} = \exp(A), \quad (13)$$

implies that for all trajectories in the phase space, irreversible cases are exponentially more probable than are reversible cases. We can further calculate the average of the dissipation function over the whole phase space:

$$\begin{aligned} \langle \exp(-\Omega_\tau) \rangle &= \int_{-\infty}^{\infty} dA \exp(-A) p(\Omega_\tau = A) \\ &= \int_{-\infty}^{\infty} dA \exp(A) p(-\Omega_\tau) \exp(A) \\ &= \int_{-\infty}^{\infty} dA p(\Omega_\tau = -A) \\ &= 1. \end{aligned}$$

Note that $\exp(-\Omega_\tau)$ is convex function, which lead to the Jensen's inequality

$$\langle \Omega_\tau \rangle \geq 0. \quad (14)$$

This is exactly what we expect – the second law inequality.

C. Crooks Fluctuation Theorem (1998)

Crooks Fluctuations Theorem and the Jarzynski equality address the non-equilibrium processes between two equilibria (unlike classical thermodynamics, which can only describe quasistatic work relations). Jarzynski equality directly presents the work relation regarding actions performed at any rate. The Crooks Fluctuation Theorem even measures the probability distribution of trajectories. These relations allow us to describe certain microscopic operations, such as the stretching of a polymer.

The Crooks Fluctuation Theorem was developed later than the Jarzynski equality. However, as it is a more general formula, we view the Jarzynski equality as a direct consequence of the Crooks Fluctuation Theorem in the following discussion. The Crooks Fluctuation Theorem reads

$$\frac{p_f(W = \mathcal{A})}{p_r(W = -\mathcal{A})} = \exp(\beta(\mathcal{A} - \Delta F)), \quad (15)$$

where $p_f(W = \mathcal{A})$ is the probability of work done on the system $W = \mathcal{A}$ between the initial equilibrium state A to the final equilibrium state B . Additionally, $p_B(W = \mathcal{A})$ is the probability distribution of the reverse trajectories. On the right-hand side, ΔF is the free energy difference between the two states, and $\beta \equiv \frac{1}{k_B T}$ is the reverse temperature of the heat bath.

The external work is introduced by a control parameter λ . The Hamiltonian of the system can be written as

$$H(\Gamma_t, t, \lambda_t) = T(\mathbf{p}) + V(\mathbf{q}, \lambda), \quad (16)$$

where T and V are the kinetic energy and the potential, respectively. The potential can be adjusted by the external control parameter λ and varies from the initial value $\lambda = A$ to the final value $\lambda = B$. If it varies slowly enough, it becomes a trivial quasistatic case. In more general cases, an external agent drives the system out of equilibrium before it relaxes into a state of equilibrium once again.

Initially, the system is in contact with a heat bath at a temperature of $T = \frac{1}{k_B \beta}$. Thus, it is a canonical ensemble characterised by the Boltzmann probability distribution:

$$p(H(\Gamma_0, A)) = \frac{1}{Z_A} e^{-\beta H(\Gamma_0, A)}. \quad (17)$$

Then the λ varies from an initial value of A to final value of B in a specific protocol. The final equilibrium state has the probability distribution

$$p(H(\Gamma_\tau, B)) = \frac{1}{Z_B} e^{-\beta H(\Gamma_\tau, B)}, \quad (18)$$

and the change of Hamiltonian can be written as

$$H(\Gamma_\tau, B) - H(\Gamma_0, A) = \int_0^\tau dt \frac{\partial H(\Gamma, \lambda)}{\partial \lambda} \dot{\lambda} + \frac{\partial H(\Gamma, \lambda)}{\partial \Gamma} \cdot \dot{\Gamma} \quad (19)$$

The first term of the integral is the total work done on the system, and the second term is the heat absorbed from the environment[8].

$$Q(\Gamma_0, \tau) \equiv \int_0^\tau dt \frac{\partial H(\Gamma, \lambda)}{\partial \Gamma} \cdot \dot{\Gamma} = k_B T \int_0^\tau \Lambda(\Gamma_t) dt \quad (20)$$

$$W(\Gamma_0, \tau) \equiv \int_0^\tau dt \frac{\partial H(\Gamma, \lambda)}{\partial \lambda} \cdot \dot{\lambda}. \quad (21)$$

Hence, we can express the work done as a function of the energy difference associated with the initial and the final states, and we can express the phase space compression factor as

$$W(\Gamma_0, \tau) = H(\Gamma_\tau, B) - H(\Gamma_0, A) - k_B T \int_0^\tau \Lambda(\Gamma_t) dt. \quad (22)$$

As time evolution is deterministic for a Hamiltonian system, the probability distribution of trajectories with $W = \mathcal{A}$ is directly determined by the initial distribution function associated with the work parameter $\lambda = A$

$$p_f(W = \mathcal{A}) = \int d\Gamma_0 \delta(W(\Gamma_0, \tau) - \mathcal{A}) p(H(\Gamma_0, A)). \quad (23)$$

In the same way, we can write the probability of the anti-trajectories during which the system does work \mathcal{A} to the external agent:

$$p_b(W = -\mathcal{A}) = \int d\Gamma_0^* \delta(W(\Gamma_0^*, A) + \mathcal{A}) p(H(\Gamma_0^*, B)). \quad (24)$$

Based on the results obtained, we can write the relative probability as

$$\begin{aligned} & \frac{p_f(W = \mathcal{A})}{p_b(W = -\mathcal{A})} \\ &= \frac{\int d\Gamma_0 \delta(W(\Gamma_0, \tau) - \mathcal{A}) p(H(\Gamma_0, A))}{\int d\Gamma_0^* \delta(W(\Gamma_0^*, \tau) + \mathcal{A}) p(H(\Gamma_0^*, B))} \\ &= \frac{Z_B \int d\Gamma_0 \delta(W(\Gamma_0, \tau) - \mathcal{A}) e^{-\beta H(\Gamma_0, A)}}{Z_A \int d\Gamma_0^* \delta(W(\Gamma_0^*, \tau) + \mathcal{A}) e^{-\beta H(\Gamma_0^*, B)}} \\ &= \frac{Z_B \int d\Gamma_0 \delta(W(\Gamma_0, \tau) - \mathcal{A}) e^{-\beta H(\Gamma_0, A)}}{Z_A \int d\Gamma_0 \delta(W(\Gamma_0, \tau) - \mathcal{A}) e^{-\beta H(\Gamma_0, A)} e^{-\beta W(\Gamma_0, \tau)}} \\ &= \frac{Z_B}{Z_A} e^{\beta \mathcal{A}} \\ &= \exp(-\beta \Delta F + \beta \mathcal{A}), \end{aligned} \quad (25)$$

where the first equality is direct substitution of equation 23 and 24. Then writing explicitly the equilibrium distribution leads to the second equality. To get the third

equality, we used equation 22 to rewrite the denominator. With some simplification, we reach the final expression which is so-called Crooks fluctuation relation

$$\frac{p_f(W = \mathcal{A})}{p_b(W = -\mathcal{A})} = \exp(-\beta\Delta F + \beta\mathcal{A}). \quad (26)$$

D. Jarzynski equality (1997)

As we mentioned before, Jarzynski equality can be regarded as a direct consequence of Crooks fluctuation theorem. With this probability distribution of trajectories given by Crooks FT, we can calculate the Jarzynski average of the work done on the system between initial and final state associated with $\lambda = A$ and $\lambda = B$

$$\begin{aligned} \langle \exp(-\beta W) \rangle &= \int d\mathcal{A} p_f(W = \mathcal{A}) \exp(-\beta\mathcal{A}) \\ &= \int [\mathcal{A} p_b(W = -\mathcal{A}) \exp(-\beta\Delta F)] \\ &= \exp(-\beta\Delta F). \end{aligned} \quad (27)$$

This equality is highly useful; it allows for the calculation of the free energy difference by conducting the experiment many times. Due to the convexity of the exponential function, the Jarzynski equality implies

$$\Delta F \leq W. \quad (28)$$

Therefore we can see it is consistent with the second law of thermodynamics.

III. EXAMPLE

A. Colloidal in an optical trap

The simplest and most practical example of applying fluctuation theorems is through a single colloidal in an optical trap. This system has the smallest possible degrees of freedom, and the optic trap can be approximated as a harmonical potential, which is the simplest potential providing restoring force. For such a system, we can obtain an analytical solution to verify the fluctuation theorems.

We treat the optical trap as a harmonical potential with the spring constant k and assume the particle confined in the trap is over-damped; in this way, we can neglect the kinetic energy. Thus, the Hamiltonian can be written as

$$H(\Gamma_0, \lambda) = \phi_{ext} = \frac{1}{2}k^2\mathbf{r}^2, \quad (29)$$

where λ is the control parameter. At the beginning, $k = k_0$ and the system is in a canonical distribution

$$p_{k_0}(\mathbf{r}) = \frac{2\pi}{\beta k_0} e^{-\beta \frac{1}{2}k_0\mathbf{r}^2}. \quad (30)$$

At time $t = 0$, we suddenly change the intensity of the laser so that the spring constant jumps to k_1 . In this process, the work done on the system is dependent on the distance from the center at that moment

$$W = \int_0^\tau dt \dot{\lambda} \frac{d\phi_{ext}}{d\lambda} = \frac{1}{2}(k_1 - k_0)\mathbf{r}_0^2. \quad (31)$$

Now the system is in a non-equilibrium state and will relax to a state of equilibrium:

$$p_{k_1}(\mathbf{r}) = \frac{2\pi}{\beta k_1} e^{-\beta \frac{1}{2}k_1\mathbf{r}^2}. \quad (32)$$

Accordingly, the free energy difference can be directly obtained from the initial and final equilibrium distributions:

$$\Delta F = -k_B T \ln Z_{k_1} + k_B T \ln Z_{k_0} = k_B T \ln \frac{k_0}{k_1} \quad (33)$$

To calculate the Jarzynski average of the work, we can first find the probability distribution of the trajectories as a function of work by combining equation 30 and equation 31:

$$p_{k_0 \rightarrow k_1}(W) = p_{k_0}(W) = \frac{k_1 - k_0}{\beta k_0} e^{-\beta \frac{k_0 W}{k_1 - k_0}}. \quad (34)$$

Then, we can easily calculate the Jarzynski average and confirm the Jarzynski equality:

$$\begin{aligned} \langle e^{-\beta W} \rangle &= \int_0^\infty dW e^{-\beta W} \frac{k_1 - k_0}{\beta k_0} e^{-\beta \frac{k_0 W}{k_1 - k_0}} \\ &= \frac{k_1}{k_0} \\ &= e^{-\beta \Delta F}. \end{aligned} \quad (35)$$

Following the same procedure, we can also calculate the probability of the reverse process $p_{k_0 \rightarrow k_1}(W)$ and confirm the Crooks Fluctuation Theorem:

$$\frac{p_{k_0 \rightarrow k_1}(W)}{p_{k_1 \rightarrow k_0}(-W)} = \frac{\frac{k_1 - k_0}{\beta k_0} e^{-\beta \frac{k_0 W}{k_1 - k_0}}}{\frac{k_0 - k_1}{\beta k_1} e^{\beta \frac{k_1 W}{k_0 - k_1}}} = e^{\beta(W - \Delta F)}. \quad (36)$$

This example can also be used to verify the Evans-Searles Fluctuation theorem. Since the Evans-Searles Fluctuation Theorem requires that all processes are time-reversible, we need to change the protocol to meet this condition, which involves changing k_1 back to k_0 at time $t = \tau$.

IV. CONCLUSION

Fluctuation theorems bring us a deeper understanding of the irreversible process. Evans-Searles fluctuations

resolved Lochimids objection on Boltzmann's statistical thermodynamics. The Crooks Fluctuation Theorem and Jarzynski equality are powerful tools for investigating small systems with non-negligible fluctuations. However, due to time restrictions, we can only discuss the deterministic dynamics described by Hamiltonian mechanics. The quantum version of the fluctuation theorem and stochastic dynamics are also worth discussing.

V. APPENDIX

A. Hamiltonian mechanics with dissipative force

See in section 2, *Fluctuation Theorem*, E.M Sevick et al.[1].

B. Heat absorbed and the phase space compression factor

In the derivation of the Crooks Fluctuation Theorem, we used the relation

$$Q(\Gamma_0, \tau) \equiv \int_0^\tau dt \frac{\partial H(\Gamma, \lambda)}{\partial \Gamma} \cdot \dot{\Gamma} = k_B T \int_0^\tau \Lambda(\Gamma_t) dt \quad (37)$$

without proof. For an intuitive understanding, we can consider a specific case in which heat slowly enters in a microcanonical ensemble. We begin by writing the above relation in the instantaneous form:

$$\dot{Q}(\Gamma_t) \equiv \frac{\partial H(\Gamma, \lambda)}{\partial \Gamma} \cdot \dot{\Gamma} = k_B T \Lambda(\Gamma_t) \quad (38)$$

To understand this, let us consider the case for states of equilibrium with quasistatic processes, in which Boltzmann's entropy is defined as

$$S = k_B \log \Omega, \quad (39)$$

where Ω is the number of states. In a quasi-static process, we have

$$dS = \frac{dQ}{T}. \quad (40)$$

Thus,

$$\dot{Q} \equiv \frac{dQ}{dt} = T \frac{dS}{dt} = k_B T \frac{d}{dt} \log \Omega = k_B T \Lambda(\Gamma). \quad (41)$$

-
- [1] Edith M Sevick, R Prabhakar, Stephen R Williams, and Debra J Searles. Fluctuation theorems. *Annu. Rev. Phys. Chem.*, 59:603–633, 2008.
- [2] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690–2693, 1997.
- [3] Denis J. Evans and Debra J. Searles. The fluctuation theorem. *Advances in Physics*, 51(7):1529–1585, 2002.
- [4] Chris Jarzynski. Nonequilibrium work theorem for a system strongly coupled to a thermal environment. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(09):P09005, 2004.
- [5] Denis James Evans, Debra Joy Searles, and Stephen Rodney Williams. *Fundamentals of classical statistical thermodynamics: dissipation, relaxation, and fluctuation theorems*. John Wiley & Sons, 2016.
- [6] Denis J Evans and Debra J Searles. Equilibrium microstates which generate second law violating steady states. *Physical Review E*, 50(2):1645, 1994.
- [7] Gavin E Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5-6):1481–1487, 1998.
- [8] For an intuitive understanding of this relation, see the Appendix. More rigorous proof can be obtained with Nosé - Hoover thermostat.