Fluctuation Theorems (Jarzynski, Crooks)

Quantum Foundations · Module 9 — Quantum Thermodynamics

Node 9.3

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Thermodynamics **Node:** 9.3

Fluctuation theorems are exact identities on the statistics of work done by a non-equilibrium protocol, valid arbitrarily far from equilibrium. They sharpen the second law from an inequality $\langle W\rangle \geq \Delta F$ to an equality $\langle e^{-\beta W}\rangle = e^{-\beta\Delta F}$ and, in the refined Crooks form, to a detailed-balance statement relating forward and time-reversed protocols. In the quantum setting they are consequences of the TPM scheme of 9.2 together with unitarity and time-reversal symmetry.

Formal Statement

Let the system start in a thermal state $\rho_0 = e^{-\beta H(0)}/Z_0$, evolve under a driving protocol that changes $H(t)$ from $H(0)$ to $H(\tau)$, and let W be the TPM work of 9.2. Let $\Delta F = -\beta^{-1}\log(Z_\tau/Z_0)$ be the equilibrium free-energy difference.

Quantum Jarzynski equality (Tasaki 2000, Kurchan 2000):

$$\boxed{\langle e^{-\beta W}\rangle_F = e^{-\beta\Delta F}}$$

where $\langle\cdot\rangle_F$ is the average over the forward-protocol work distribution $P_F(W)$.

Quantum Crooks fluctuation theorem:

$$\boxed{\frac{P_F(W)}{P_R(-W)} = e^{\beta(W-\Delta F)}}$$

where P_R is the work distribution under the **time-reversed protocol**, run with initial state $e^{-\beta H(\tau)}/Z_\tau$ and Hamiltonian trajectory $\tilde{H}(t) = \Theta H(\tau - t)\Theta^{-1}$, with Θ the antiunitary time-reversal operator.

Jarzynski is obtained from Crooks by integrating over W. The second law $\langle W\rangle \geq \Delta F$ follows from Jarzynski by Jensen's inequality applied to the convex function $e^{-\beta x}$.

Intuition

The second law says "on average, you can't extract more work than the free-energy difference allows." Fluctuation theorems say something stronger: the probability of a rare *violation* (an individual run where $W < \Delta F$) is *exactly* exponentially suppressed, and the suppression ratio is universal. If you see an individual run where the work is $\Delta F - k_BT$, it is $e \approx 2.7$ times less likely than the time-reversed run producing work $\Delta F - k_BT + 2k_BT = \Delta F + k_BT$.

This refines the second law into a *microscopic reversibility* principle. The underlying physics is: unitary quantum dynamics is time-reversal symmetric; the only thing that breaks symmetry is the thermal initial condition, and that asymmetry is exactly Gibbs-weighted. All fluctuation theorems are statements that this Gibbs-weight is the source of thermodynamic irreversibility.

Structural Role

9.3 closes the loop opened by 9.1 and 9.2. It says: the Gibbs state (9.1), plugged into the TPM scheme (9.2), satisfies exact non-equilibrium identities. This has three consequences:

- The second law becomes a *theorem*, not a postulate, once microscopic reversibility and the Gibbs initial condition are granted.
- Free energy differences ΔF — usually equilibrium-only quantities — can be extracted from non-equilibrium measurements. This is the basis of experimental free-energy estimation (Liphardt et al. 2002 for classical; Batalhão et al. 2014 for quantum NMR).
- Landauer's principle (9.4) and engine bounds (9.5) are corollaries in special cases.

The quantum version is structurally identical to the classical one because the derivation only uses unitarity, TPM, and Gibbs weights — no commutativity assumption is needed.

Mathematical Development

Derivation of Jarzynski from TPM

Start from the definition:

$$\langle e^{-\beta W}\rangle_F = \sum_{n,m} p_n^0\, p_{m|n}\, e^{-\beta(E_m(\tau)-E_n(0))}.$$

Substitute $p_n^0 = e^{-\beta E_n(0)}/Z_0$:

$$= \frac{1}{Z_0}\sum_{n,m} e^{-\beta E_m(\tau)}\, p_{m|n}.$$

The quantum transition probabilities $p_{m|n} = |\langle m_\tau|U_\tau|n_0\rangle|^2$ satisfy $\sum_n p_{m|n} = 1$ (rows of a doubly stochastic matrix, by unitarity of U_τ). Therefore

$$\langle e^{-\beta W}\rangle_F = \frac{1}{Z_0}\sum_m e^{-\beta E_m(\tau)} = \frac{Z_\tau}{Z_0} = e^{-\beta\Delta F}. \qquad \blacksquare$$

The derivation uses three ingredients: (i) Gibbs initial condition, (ii) unitary forward evolution (so $p_{m|n}$ is doubly stochastic), (iii) TPM definition of W. No assumption about the protocol being slow, weak, or near-equilibrium enters.

Derivation of Crooks

Compute the joint forward probability:

$$P_F(n\to m) = \frac{e^{-\beta E_n(0)}}{Z_0}\,|\langle m_\tau|U_\tau|n_0\rangle|^2.$$

For the time-reversed protocol, start in $e^{-\beta H(\tau)}/Z_\tau$, apply $\tilde{U}_\tau = \Theta U_\tau^\dagger \Theta^{-1}$, and measure at the endpoints. Microreversibility of unitary dynamics gives

$$|\langle \tilde{n}_0|\tilde{U}_\tau|\tilde{m}_\tau\rangle|^2 = |\langle m_\tau|U_\tau|n_0\rangle|^2.$$

Hence

$$P_R(m \to n) = \frac{e^{-\beta E_m(\tau)}}{Z_\tau}\,|\langle m_\tau|U_\tau|n_0\rangle|^2.$$

Divide the two:

$$\frac{P_F(n \to m)}{P_R(m \to n)} = \frac{Z_\tau}{Z_0}\,e^{-\beta(E_n(0)-E_m(\tau))} = e^{\beta(W-\Delta F)}.$$

The left-hand ratio, summed over pairs with fixed W, gives $P_F(W)/P_R(-W)$. ■

Second law via Jensen

By Jensen's inequality on the convex function e^{-x}:

$$e^{-\beta\langle W\rangle} \le \langle e^{-\beta W}\rangle = e^{-\beta\Delta F} \quad \Longrightarrow \quad \langle W\rangle \ge \Delta F.$$

Jarzynski contains the second law as a one-line corollary. Any protocol for which $\langle W\rangle = \Delta F$ exactly must have a delta-peaked $P(W)$, i.e. be reversible.

Fluctuation-dissipation at linear response

Expanding Crooks around equilibrium $W \approx \Delta F$, the fluctuation theorem reproduces the linear-response fluctuation-dissipation theorem: the dissipated work $W - \Delta F$ has Gaussian statistics with variance $2k_BT\langle W - \Delta F\rangle$. Quantum quenches that stay near-equilibrium obey this Gaussian limit; far-from-equilibrium quenches exhibit the full non-Gaussian Crooks identity.

Worked Example

Sudden quench of a qubit

Take the sudden-quench protocol from 9.2: $H(0) = \frac{\omega}{2}Z$, $H(\tau) = \frac{\omega}{2}(\cos\theta Z + \sin\theta X)$, $U_\tau = I$. The forward work distribution is (from 9.2):

$$P_F(W=0) = \cos^2(\theta/2), \quad P_F(W=+\omega) = p_0^0\sin^2(\theta/2), \quad P_F(W=-\omega) = p_1^0\sin^2(\theta/2).$$

Check Jarzynski. Compute $\langle e^{-\beta W}\rangle_F$:

$$\cos^2(\theta/2) + \frac{\sin^2(\theta/2)}{1+e^{-\beta\omega}}(e^{-\beta\omega} + e^{-\beta\omega}e^{\beta\omega}) = \cos^2(\theta/2) + \sin^2(\theta/2)\cdot\frac{1+e^{-\beta\omega}}{1+e^{-\beta\omega}} = 1.$$

Compute $e^{-\beta\Delta F} = Z_\tau/Z_0$. Both Hamiltonians have eigenvalues $\pm\omega/2$, so $Z_\tau = Z_0$ and $\Delta F = 0$. Consistent: $\langle e^{-\beta W}\rangle_F = 1 = e^{-0}$. ✓

Check Crooks. The reverse protocol starts in $e^{-\beta H(\tau)}/Z_\tau$ and quenches back to $H(0)$. Working it out, one finds

$$P_R(W=0)=\cos^2(\theta/2),\quad P_R(W=+\omega)=p_0^\tau\sin^2(\theta/2),\quad P_R(W=-\omega)=p_1^\tau\sin^2(\theta/2),$$

where $p_0^\tau=1/(1+e^{-\beta\omega})=p_0^0$ because the new Hamiltonian has the same spectrum. The Crooks ratio at $W=+\omega$:

$$\frac{P_F(+\omega)}{P_R(-\omega)}=\frac{p_0^0\sin^2(\theta/2)}{p_1^0\sin^2(\theta/2)}=\frac{p_0^0}{p_1^0}=e^{\beta\omega}=e^{\beta(W-\Delta F)}.\checkmark$$

Experimental verification. This exact protocol (or its analogue) has been run in the lab: Batalhão et al. (2014) verified quantum Jarzynski for a chloroform NMR system, and An et al. (2015) did so for a trapped ion. Measured $P(W)$ agreed with theory within error bars across the full distribution, not just the mean.

Cross-Links

- **9.1 Gibbs state** — the initial condition whose Boltzmann weights drive the whole derivation.
- **9.2 Quantum work and heat** — supplies the TPM definition used in both Jarzynski and Crooks.
- **9.4 Landauer's principle** — a corollary of the second law in the special case of information erasure.
- **9.5 Quantum heat engines** — fluctuation theorems bound the efficiency fluctuations of finite-time cycles.
- **5.1 Density matrices / time reversal** — microreversibility $|\langle m_\tau|U_\tau|n_0\rangle|^2=|\langle\tilde{n}_0|\tilde{U}_\tau|\tilde{m}_\tau\rangle|^2$.
- **Statistics** — Crooks is a detailed-balance relation between forward and reverse path measures.

Structural Compression

Invariant: $\langle e^{-\beta W}\rangle_F=e^{-\beta\Delta F}$ for any protocol taking a quantum system out of thermal equilibrium.

Mechanism: Gibbs initial state $\times$ double stochasticity of unitary transition probabilities $\times$ TPM definition of work.

Cross-System Echo: Classical Jarzynski (1997) and Crooks (1999) — derived from Hamiltonian dynamics + Gibbs. Quantum version derived from unitary dynamics + Gibbs. Same structural skeleton; quantum version is actually *easier* to prove because unitarity gives double stochasticity for free.

Exercises

1. Fill in the three lines of the Jarzynski derivation, identifying exactly where unitarity of U_τ is used and what goes wrong if U_τ is replaced by a non-unital CPTP map.
2. Derive Crooks from the joint ratio $P_F(n\to m)/P_R(m\to n)$ and verify that integrating over W reproduces Jarzynski.
3. Show that for an *isothermal reversible* protocol (infinitely slow quasi-static), $P(W)$ is a delta function and $\langle W\rangle=\Delta F$ exactly.
4. For a sudden quench of a harmonic oscillator, $H(0)=\hbar\omega(a^\dagger a+1/2)\to\hbar\omega'(a^\dagger a+1/2)$, compute the full work distribution and verify Jarzynski using the partition functions $Z=(2\sinh(\beta\hbar\omega/2))^{-1}$.
5. Show that the Kullback–Leibler divergence $D(P_F\|\tilde{P}_R)$, where $\tilde{P}_R(W)=P_R(-W)e^{-\beta W}/\langle e^{-\beta W}\rangle$, equals the dissipated work $\beta(\langle W\rangle-\Delta F)$. Interpret structurally.
6. Suppose an experiment reports $\langle W\rangle-\Delta F=0.5\,k_BT$ with Gaussian statistics. Use the linear-response limit of Crooks to predict the variance, and verify consistency with a direct calculation.

7. **Argue, structurally**, why fluctuation theorems take exactly the same algebraic form in classical and quantum thermodynamics, despite the fact that classical work is a path integral and quantum work is a pair of measurement outcomes. Identify what feature of the derivation is *not* used in either case.

Quantum Foundations · Module 9 — Quantum Thermodynamics · Node 9.3
Concepts: random-variables, expectation-value
Prerequisites: 9.2

hbar.university — own.your.cognition