

PAPER • OPEN ACCESS

Introducing one-shot work into fluctuation relations

To cite this article: Nicole Yunger Halpern *et al* 2015 *New J. Phys.* **17** 095003

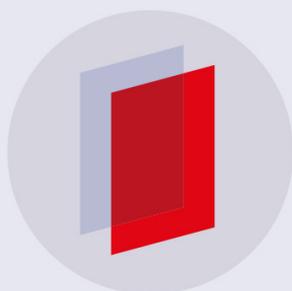
View the [article online](#) for updates and enhancements.

Related content

- [The role of quantum information in thermodynamics—a topical review](#)
John Goold, Marcus Huber, Arnau Riera *et al.*
- [Entropic equality for worst-case work at any protocol speed](#)
Oscar C O Dahlsten, Mahn-Soo Choi, Daniel Braun *et al.*
- [From free energy measurements to thermodynamic inference in nonequilibrium small systems](#)
A Alemany, M Ribezzi-Crivellari and F Ritort

Recent citations

- [Fluctuations of work cost in optimal generation of correlations](#)
Emma McKay *et al*
- [Beyond Number of Bit Erasures](#)
Joshua A. Grochow and David H. Wolpert
- [Phase Transition in Protocols Minimizing Work Fluctuations](#)
Alexandre P. Solon and Jordan M. Horowitz



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.



PAPER

Introducing one-shot work into fluctuation relations

OPEN ACCESS

RECEIVED

23 February 2015

REVISED

19 July 2015

ACCEPTED FOR PUBLICATION

10 August 2015

PUBLISHED

11 September 2015

Nicole Yunger Halpern¹, Andrew J P Garner^{2,3}, Oscar C O Dahlsten² and Vlatko Vedral^{2,3}¹ Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA² Atomic and Laser Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford, OX13PU, UK³ Center for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore

E-mail: nicoleyh@caltech.edu

Keywords: one-shot statistical mechanics, fluctuation theorems, nonequilibrium thermodynamics, small-scale systems, irreversible thermodynamics, resource theory

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

Any further distribution of
this work must maintain
attribution to the
author(s) and the title of
the work, journal citation
and DOI.

**Abstract**

Two approaches to small-scale and quantum thermodynamics are fluctuation relations and one-shot statistical mechanics. Fluctuation relations (such as Crooks' theorem and Jarzynski's equality) relate nonequilibrium behaviors to equilibrium quantities such as free energy.

One-shot statistical mechanics involves statements about every run of an experiment, not just about averages over trials. We investigate the relation between the two approaches. We show that both approaches feature the same notions of work and the same notions of probability distributions over possible work values. The two approaches are alternative toolkits with which to analyze these distributions. To combine the toolkits, we show how one-shot work quantities can be defined and bounded in contexts governed by Crooks' theorem. These bounds provide a new bridge from one-shot theory to experiments originally designed for testing fluctuation theorems.

1. Introduction

The probabilistic nature of thermalization prevents us from deterministically predicting the amount of work performed on a system in any given run of an experiment. This stochasticity necessitates a statistical treatment of work, especially when the deviation from the mean value of work is large. Two popular frameworks employed for this purpose are *fluctuation theorems* [1–15] and *one-shot statistical mechanics* [16–23]. The former framework's purpose is to quantify the behaviors of nonequilibrium classical and quantum systems. The latter framework concerns statements true of every trial (realization) of an experiment.

The relationship between these frameworks has been unclear (though work by Åberg [19] suggests that a connection could be fruitful). We will demonstrate that these approaches are not competitors. Rather, the approaches are mutually compatible tools. Combined, they describe general thermal behaviors of small classical and quantum systems.

We will begin with a technical introduction to fluctuation theorems and one-shot statistical mechanics. We then present our main claim: *that one-shot statistical mechanics can be applied to settings governed by fluctuation theorems* (see figure 1). We substantiate this claim by generalizing the characteristic functions of the work probability distributions for classical and quantum systems. From this generalization, we derive bounds on one-shot work quantities in settings governed by fluctuation theorems. We demonstrate how this generalization can be employed in two mathematical formalisms: a *work-extraction game* [18, 19] and *thermodynamic resource theories* [20, 22–26]. To conclude with two pedagogical examples, we apply the generalization to specific fluctuation settings: *Landauer bit reset* [18, 19, 24, 27–29] and *experimental DNA unzipping* [30–32]. The examples illustrate the opportunity to test one-shot results with experiments devised originally for fluctuation theorems.

2. Preliminaries

2.1. Fluctuation theorems

Consider a classical system that is coupled to a heat bath and driven externally. Due to the probabilistic nature of thermalization, the amount of heat transferred between the system and the bath in any given trial cannot be predicted. Hence the amount of work done by the drive, in any given trial, cannot be predicted. The protocol can be associated with a *work distribution* $P(W)$, the probability density associated with some trial's costing an amount W of work. In equilibrium thermodynamics, $P(W)$ peaks tightly at the average value $W = \langle W \rangle$. This value suffices to describe the work performed in each trial. In more general, nonequilibrium, thermodynamics, the average does not suffice. Yet thermodynamic state variables related to averages (temperature, free energy, etc) are used in nonequilibrium thermodynamics. Fluctuation relations link these variables to probability distributions over work or heat. We will focus mostly on continuous variables W , which have been used in classical and quantum contexts⁴ (e.g., [6, 8, 10, 15]).

One such fluctuation relation is Crooks' theorem [2]. Though originally derived in a classical setting, it has been shown to govern quantum processes [3–12, 14, 15]. Crooks' theorem describes the fluctuations in the work expended on systems subject to a time-changing Hamiltonian $H(\lambda_t)$ in the presence of a heat bath. An experimenter can change the external scalar parameter λ_t during the time interval $t \in [-\tau, \tau]$ by performing work. We denote the bath's inverse temperature by $\beta = 1/k_B T$. The external driving can be performed in either a forward or a reverse direction. The forward process begins at time $-\tau$, when the system occupies the thermal state $e^{-\beta H_{-\tau}}/Z_{-\tau}$ of the initial Hamiltonian $H_{-\tau} \equiv H(\lambda_{-\tau})$. The reverse process begins at time τ , when the system occupies the thermal state $e^{-\beta H_\tau}/Z_\tau$ associated with the final Hamiltonian $H_\tau \equiv H(\lambda_\tau)$ of the forward protocol. ($Z_{\pm\tau}$ denote normalization factors.)

Suppose that an agent implements the protocol in both directions many times, measuring the work invested in each forward trial and the work extracted from each reverse. Two probability distributions encapsulate these measurements: $P_{\text{fwd}}(W)$ denotes the probability that some forward trial will require work W (or the probability per unit work, if P_{fwd} denotes a probability density), and $P_{\text{rev}}(-W)$ denotes the probability that some reverse trial will output work W .

If the system's interactions with the bath are Markovian and *microscopically reversible* [33], and if the initial state is thermal, the work probability distributions satisfy *Crooks' theorem* [2],

$$\frac{P_{\text{fwd}}(W)}{P_{\text{rev}}(-W)} = e^{\beta(W-\Delta F)}. \quad (1)$$

This ΔF denotes the difference $F(e^{-\beta H_\tau}/Z_\tau) - F(e^{-\beta H_{-\tau}}/Z_{-\tau})$ between the Helmholtz free energies of thermal states over the forward process's final and initial Hamiltonians.

Multiplying each side of Crooks' theorem by $P_{\text{rev}}(-W)e^{-\beta W}$ and integrating over W yields *Jarzynski's equality* [1],

$$\langle e^{-\beta W} \rangle_{\text{fwd}} = e^{-\beta \Delta F}, \quad (2)$$

wherein $\langle \cdot \rangle_{\text{fwd}}$ denotes an expectation value calculated from P_{fwd} . Applied to a work distribution $P(W)$ constructed from simulations or experiments, Jarzynski's equality can be used to calculate the equilibrium quantity ΔF . Combined with Jensen's inequality, $\langle e^x \rangle \geq e^{\langle x \rangle}$, Jarzynski's equality implies a lower bound on the average work required to complete a trial. This bound, $\langle W \rangle \geq \Delta F$, has been considered a statement of the Second Law of Thermodynamics [34].

The left-hand side (lhs) of Jarzynski's equality has been recognized as the characteristic function, or Fourier transform, of $P_{\text{fwd}}(W)$ [6, 15]. If $u = i\beta$ denotes the variable conjugate to W , the characteristic function is

$$\chi_{\text{fwd}}(\beta) \equiv \mathcal{F}\{P_{\text{fwd}}(W)\} \equiv \int_{-\infty}^{\infty} dW P_{\text{fwd}}(W) e^{iuW} = \int_{-\infty}^{\infty} dW P_{\text{fwd}}(W) e^{-\beta W}. \quad (3)$$

In terms of the characteristic function, Jarzynski's equality reads

$$\chi_{\text{fwd}}(\beta) = e^{-\beta \Delta F}. \quad (4)$$

The reverse process corresponds to the characteristic function $\chi_{\text{rev}}(\beta) \equiv \int_{-\infty}^{\infty} dW P_{\text{rev}}(W) e^{-\beta W}$, in terms of which $\chi_{\text{rev}}(\beta) = e^{\beta \Delta F}$.

⁴ W can be continuous even if the system has a discrete energy spectrum: consider a system that interacts with the heat bath while two (or more) energy levels shift at different rates. The system can jump from one level to another at any time, due to thermalization. The work cost of a trajectory that jumps at time t can differ infinitesimally from the work cost of a trajectory that jumps at time $t + dt$.

2.2. One-shot statistical mechanics

Mean values do not necessarily reflect a system's *typical* behavior. Consider a system that must output at least some threshold amount of work to trigger another process. One such threshold is the activation energy required to begin a chemical reaction. The system might output below-threshold work usually but far-above-threshold work occasionally. The average work might exceed the threshold, but the second process is usually not triggered.

By spotlighting statistics other than the mean, one-shot information theory extends idealized protocols implemented $n \rightarrow \infty$ times to realistic finite- n protocols that might fail. Conventional statistical mechanics describes the optimal rate at which work can be extracted *asymptotically*. Consider transforming n copies of one equilibrium state into n copies of another quasistatically, in the presence of a temperature- T heat bath. In the asymptotic, or thermodynamic, limit as $n \rightarrow \infty$, the average work required per copy approaches the difference ΔF between the states' free energies. The free energy depends on the Shannon entropy. In reality, states are transformed finitely many times, and realistic processes have probabilities δ of failing to accomplish their purposes. Finite- n work-consumption rates have been quantified with *one-shot entropies* [17, 20–23]. So have the efficiencies of finite- n data compression, randomness extraction, quantum key distribution, and hypothesis testing [16, 35–37].

One one-shot entropy is the order- ∞ Rényi entropy $H_\infty(\mathcal{P})$, known also as the *min-entropy*. For any discrete probability distribution \mathcal{P} whose greatest element is \mathcal{P}^{\max} ,

$$H_\infty(\mathcal{P}) \equiv -\log(\mathcal{P}^{\max}). \quad (5)$$

(All logarithms in this article are base- e .) We will discuss two popular models in which one-shot entropies are applied to thermodynamics: *work-extraction games* and *thermodynamic resource theories*.

2.2.1. Work-extraction game

In the work-extraction game described by Egloff *et al* [18], a player transforms a system in a state ρ , governed by a Hamiltonian H_ρ , into a state σ governed by $H_\sigma : (\rho, H_\rho) \mapsto (\sigma, H_\sigma)$. For simplicity, we take a quasiclassical model such that states are assumed to commute with their Hamiltonians. The agent has access to a temperature- T heat bath.

The player should choose an *optimal strategy* to maximize the transformation's work output (or minimize the transformation's work cost). The strategy consists of a sequence of operations of two types: (1) without investing work, the player can couple the system to the bath in any manner modeled by a stochastic matrix that preserves the Gibbs state $e^{-\beta H_\rho}/Z_\rho$. (Such thermalization models are discussed in appendix A.) (2) By investing or extracting work, the agent can shift the Hamiltonian's levels.

The primary result in [18] implies an upper bound on the work extractable (up to a probability δ of failure) during the transformation $(\rho, H_\rho) \mapsto (\sigma, H_\sigma)$. Egloff *et al* show that the *optimal strategy* has a probability $1 - \delta$ of outputting at least the work

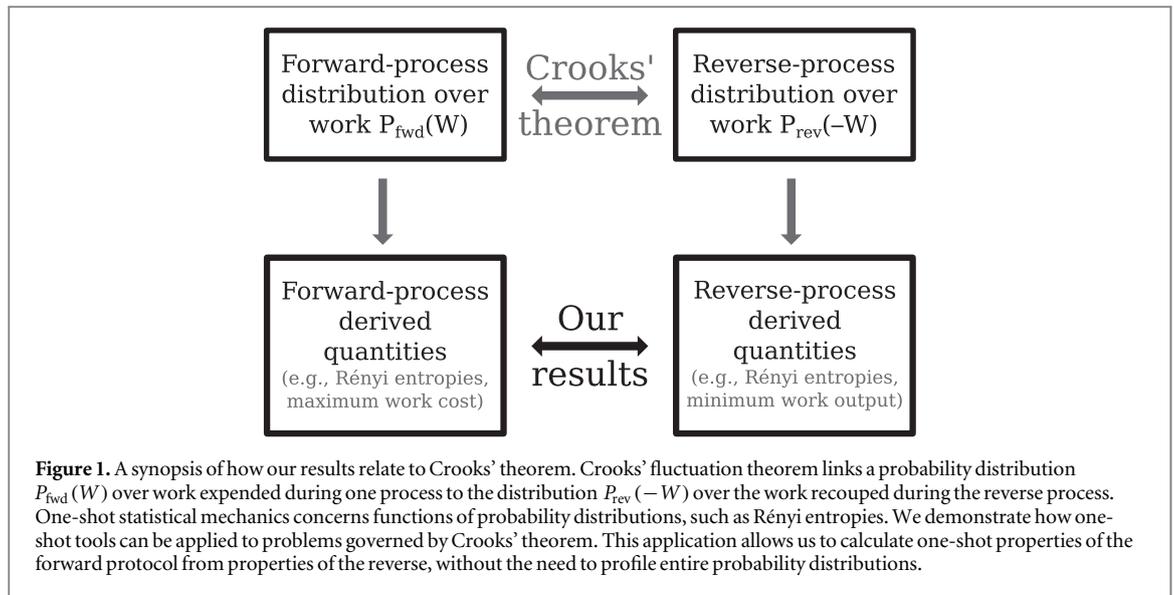
$$w_{\text{best}}^\delta(\rho, H_\rho \mapsto \sigma, H_\sigma) = T \log \left(M \left(\frac{G^T(\rho)}{1 - \delta} \parallel G^T(\sigma) \right) \right) \quad (6)$$

in each trial. G^T denotes *Gibbs-rescaling* relative to the temperature T . Gibbs-rescaling facilitates the comparison of the work values of states governed by different Hamiltonians. A state can have the capacity to perform work due to the state's information content (e.g., because the state is pure) and energy contents (e.g., because the state has weight on high energy levels). Gibbs-rescaling recasts the state's work capacity as entirely informational. This recasting enables us to compare the final and initial states, even though different Hamiltonians govern them. M denotes the *relative mixedness*, a measure of how much more mixed one state is, or how much less information-sourced work capacity a state has. Dissipative processes yield less than the optimal amount w_{best}^δ of work. Hence equation (6) upper-bounds the amount that could be extracted with an arbitrary (possibly suboptimal) strategy.

2.2.2. Thermodynamic resource theories

Resource theories have been used to calculate how efficiently scarce quantities can be distilled and converted into other forms via cheap (or 'free') operations [25]. To an agent able to perform only certain operations for free, each state has some value, or worth. We can quantify this value with resource theories. *Thermodynamic resource theories* model exchanges of heat amongst systems and baths [20, 22–24, 26]. Each resource theory is defined by the inverse temperature β of a heat bath from which the agent can draw Gibbs states for free. More generally, energy-conserving *thermal operations* can be performed for free. Nonequilibrium states have value because work can be extracted from them.

Horodecki and Oppenheim introduced one-shot tools into thermodynamic resource theories [20]. They focused on quasiclassical resource theories, in which states commute with the Hamiltonians that govern them. Horodecki and Oppenheim calculated the minimum work required to create a state within trace distance ε of a



target state ρ . They also analyzed the transfer of work from ρ to a battery defined by a two-level Hamiltonian of gap w . The maximum w such that the battery ends within trace distance δ of its excited state was shown to be related to a one-shot entropy of ρ . One-shot information theory has since been applied to *catalysis* (the facilitation of a transformation by an ancilla) [22, 26], to arbitrary baths such as particle baths [23], and to quantum problems (that involve states that do not commute with the Hamiltonian) [38, 39].

3. Unification of fluctuation theorems and one-shot statistical mechanics

Fluctuation theorems and one-shot statistical mechanics concern properties of work distributions beyond averages. The two frameworks do not compete to describe the same concept in alternative ways. Rather, the frameworks complement each other and can be combined into a general description of small-scale classical and quantum systems. Fluctuation theorems are restricted to systems that satisfy certain physical assumptions and that can undergo forward and reverse protocols. Crooks' theorem [2], for example, relies on the dynamics' Markovianity and microscopic reversibility, and on the system's beginning in a thermal state. The tools of one-shot statistical mechanics (e.g. Rényi entropies, and bounds on work values in every trial of an experiment) can be applied more generally to the statistics produced by any system that consumes work. The formalisms are not incompatible: *the tools of one-shot statistical mechanics can be applied to the work distribution of any process governed by Crooks' theorem* (see figure 1).

We will substantiate this claim by focusing on the one-shot concept of *guaranteed work*: an upper bound (up to some error) on the work required to complete some process that applies not just on average, but in every trial. We will define this quantity in contexts governed by Crooks' theorem and will relate the quantity to the one-shot entropy H_∞ . Our results describe all quantum and classical systems whose thermalization satisfies microscopic reversibility and Markovianity and whose work distribution is continuous. (For details about these assumptions' realizations in two common one-shot frameworks, see appendix A.)

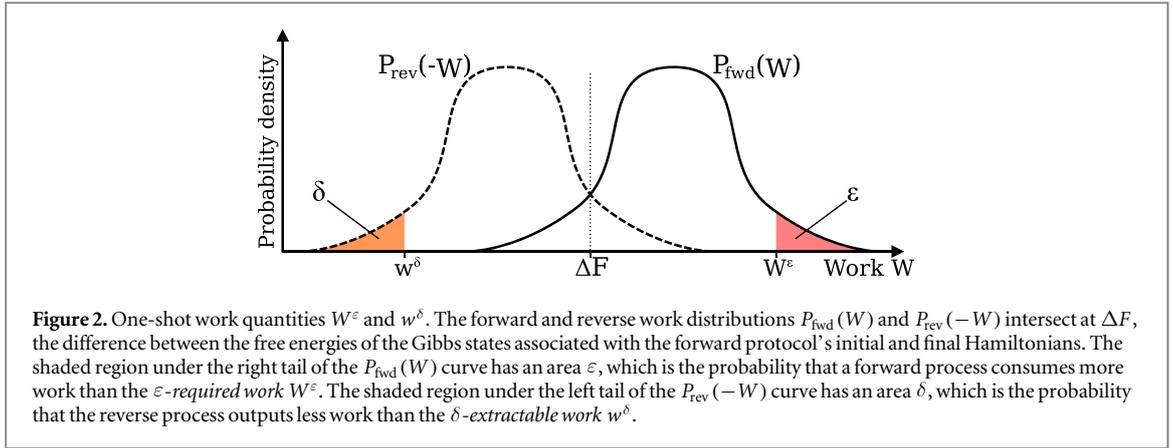
3.1. One-shot work quantities in fluctuation contexts

Suppose we consider the behavior of a system evolving under a process that is driven by a single external parameter, and otherwise satisfies the conditions for Crooks' theorem to hold. For clarity in this article, we shall choose examples where the *forward process* tends to cost work to complete. We will upper-bound the amount of work required to complete a single trial of a process successfully, such that this bound is only exceeded with probability ε (as illustrated on the right-hand side (rhs) of figure 2).

Definition 1. Each implementation of the forward protocol has a probability $1 - \varepsilon$ of requiring no more work than the ε -required work W^ε that satisfies

$$\int_{-\infty}^{W^\varepsilon} dW P_{\text{fwd}}(W) = 1 - \varepsilon. \quad (7)$$

The trial has a probability $\varepsilon \in [0, 1]$ of requiring more work than W^ε .



Similarly, we will lower-bound the amount of work extracted in the reverse process, for all but δ of the trials (illustrated on the lhs of figure 2).

Definition 2. Each implementation of the reverse protocol has a probability $1 - \delta$ of outputting at least the δ -extractable work w^δ that satisfies

$$\int_{w^\delta}^{\infty} dW P_{\text{rev}}(-W) = 1 - \delta. \quad (8)$$

The trial has a probability $\delta \in [0, 1]$ of outputting less work than w^δ .

The failure probability has two interpretations. Suppose, in the work-investment case, that an agent invests only the amount W^ε of work in a forward trial. The external parameter λ_t has a probability ε of failing to reach λ_τ . Alternatively, suppose the agent invests all the work required to evolve λ_t to λ_τ . The agent has a probability ε of overshooting the 'work budget' W^ε . The failure probability δ associated with work extraction can be interpreted similarly.

3.2. One-shot Jarzynski equalities

Even if only forward trials have been performed, the reverse process's w^δ can be calculated from Crooks' theorem.

Lemma 1. Each reverse trial has a probability $1 - \delta$ of outputting at least the amount w^δ of work that satisfies

$$\chi_{\text{fwd}}^\delta(\beta) = (1 - \delta)e^{-\beta\Delta F}, \quad (9)$$

wherein

$$\chi_{\text{fwd}}^\delta(\beta) \equiv \int_{w^\delta}^{\infty} dW P_{\text{fwd}}(W) e^{-\beta W} \quad (10)$$

generalizes the characteristic function $\chi_{\text{fwd}}(\beta)$.

Proof. Upon multiplying each side of Crooks' theorem (equation (1)) by $P_{\text{rev}}(-W)e^{-\beta W}$, we integrate from w^δ to infinity. The lhs equals $\chi_{\text{fwd}}^\delta(\beta)$ by definition (equation (10)). The right-hand integral evaluates to $1 - \delta$ by definition 2. \square

We can calculate W^ε from $P_{\text{rev}}(-W)$ via Crooks' theorem in the same way:

Lemma 2. Each forward trial has a probability $1 - \varepsilon$ of requiring no more work than the W^ε that satisfies

$$\chi_{\text{rev}}^\varepsilon(\beta) = (1 - \varepsilon)e^{\beta\Delta F}, \quad (11)$$

wherein

$$\chi_{\text{rev}}^\varepsilon(\beta) \equiv \int_{-W^\varepsilon}^{\infty} dW P_{\text{rev}}(W) e^{-\beta W} \quad (12)$$

generalizes the characteristic function $\chi_{\text{rev}}(\beta)$.

These one-shot generalizations extend Jarzynski's equality (equation (3)), rendering it more robust against unlikely (probability less than ε) but highly expensive (work cost more than W^ε) fluctuations in work. The

generalizations characterize every quantum or classical process that produces a work distribution governed by Crooks' theorem. An alternative proof of these general lemmata—specialized for quantum systems undergoing unitary evolution and hence producing discrete work distributions—is presented in appendix B.

3.3. Bounding one-shot work quantities

We can use Crooks' theorem, via the above lemmata, to derive bounds on the one-shot required and extractable work. These bounds depend on characteristics of the work distributions:

Theorem 3. *The work δ -extractable from each reverse trial satisfies*

$$w^\delta \leq \Delta F - \frac{1}{\beta} \left[H_\infty^\beta(P_{\text{fwd}}) + \log(1 - \delta) \right] \quad (13)$$

for $\delta \in [0, 1)$, wherein we have defined

$$H_\infty^\beta(P) \equiv -\log(P^{\text{max}}/\beta) \quad (14)$$

for continuous work distributions.

Proof. Let $P_{\text{fwd}}^{\text{max}}$ denote the greatest value of $P_{\text{fwd}}(W) : P_{\text{fwd}}^{\text{max}} \geq P_{\text{fwd}}(W) \quad \forall W$. We can upper-bound the integral implicit in the $\chi_{\text{fwd}}^\delta(\beta)$ of equation (9) in lemma 1:

$$\begin{aligned} (1 - \delta)e^{-\beta\Delta F} &= \chi_{\text{fwd}}^\delta(\beta) \equiv \int_{w^\delta}^{\infty} dW P_{\text{fwd}}(W) e^{-\beta W} \\ &\leq P_{\text{fwd}}^{\text{max}} \int_{w^\delta}^{\infty} dW e^{-\beta W} \\ &= e^{\log(P_{\text{fwd}}^{\text{max}}/\beta)} e^{-\beta w^\delta} \\ &= e^{-H_\infty^\beta(P_{\text{fwd}}) - \beta w^\delta}. \end{aligned} \quad (15)$$

Solving for w^δ yields inequality (13). □

Analogous statements, which we present without further proof, describe W^ε :

Theorem 4. *The work ε -required during each forward trial is bounded by*

$$W^\varepsilon \geq \Delta F + \frac{1}{\beta} \left[H_\infty^\beta(P_{\text{rev}}) + \log(1 - \varepsilon) \right] \quad (16)$$

for failure probability $\varepsilon \in [0, 1)$.

These theorems demonstrate our central claim: that one-shot statistical mechanics can be applied to settings governed by fluctuation theorems. We have related the one-shot work quantities W^ε and w^δ to the one-shot entropy H_∞^β .

Operationally when one handles data arising from a simulation or experiment, one does not directly observe a work distribution. Rather, one obtains a list of values that could be divided into bins of finite range (i.e., presented as a histogram) to approximate the true probability distribution. As want to consider a theoretical bound that reflects the behavior of the underlying thermal process independently of the choice of binning, we consider the entropy expressed in terms of the probability density $P(W)$. For theorems 3 and 4 to be applicable, it must be the case that the experimental data can be fitted to a theoretical model—a weak assumption for any scientific experiment.

There are further considerations that must be taken into account when considering the entropy of a continuous distribution in contrast with the entropy of a histogram taken from that distribution. (Throughout the following paragraph, we refer with P_{fwd} and P_{rev} jointly as P .) For a histogram taken of $P(W)$, in the limit of small enough bin width dW , the probability of each bin is well approximated by the product $P(W)dW$ as evaluated at a point W inside that bin. However, in the limit of decreasing bin size, the probability of being in any particular bin will tend to zero, and the order- ∞ entropy, as previously defined, diverges:

$H_\infty(P) = \lim_{dW \rightarrow 0} [-\log(P^{\text{max}} dW)] \rightarrow \infty$. An unmodified H_∞ is not a useful quantity in this limit. By this measure, there is an infinite amount of information in *any* continuous distribution, and hence it can not be used to quantify the amount by which some continuous distributions are more entropic than others. To circumvent this problem, we employ a type of renormalization. H_∞ can be split into a finite part that varies with the distribution under consideration, and an infinite part that does not: $H_\infty = -\log(P^{\text{max}} dW) = -\log(P^{\text{max}} k^{-1}) - \log(kdW)$, where k is an arbitrary factor with units of inverse energy chosen such that the argument of each logarithm is dimensionless. When one considers the *difference* in entropy between distributions, the latter term always cancels

out. As such the quantity $H_\infty^k = -\log(P^{\max} k^{-1})$, by omitting the latter term, defines the amount by which the continuous order- ∞ entropy differs from a reference distribution of uniform probability density k over width k^{-1} . Our choice here of

$$H_\infty^\beta(P) \equiv -\log(P^{\max}/\beta) \quad (17)$$

amounts to comparing the entropy of $P(W)$ with that of a uniform distribution over range β^{-1} with probability density β . We remark this technique is not unique to the order- ∞ Rényi entropy, but is also necessary if one wishes to arrive at the differential entropy as a limiting case of the discrete Shannon entropy⁵.

Although any quantity with units of inverse energy could be used as k , β is a natural choice: $-\log \beta$ appears everywhere $\log P^{\max}$ appears in our calculations; any alternative choice of normalization k would result in the need for an additional correction term of $\log(k/\beta)$ in inequalities (13) and (16). (This extra term can be interpreted as how the entropy of the new reference distribution compares to that of the uniform distribution of range β^{-1} and probability β).

The bounds in theorems 3 and 4 shed light on the physical contributions to W^ε . To a first approximation, the ε -required work equals ΔF , the work needed to complete the process quasistatically. The negative contribution from $\log(1 - \varepsilon)$ accounts for the tradeoff between work and failure probability: the agent can lower the bound on the required work by accepting a higher failure probability ε .

Outside the quasistatic limit, the system leaves equilibrium, and W fluctuates from trial to trial. This fluctuation necessitates a protocol-specific correction $H_\infty^\beta(P_{\text{rev}})$. In cryptography applications, $H_\infty(P)$ quantifies the uniform randomness (and hence resources usable to ensure privacy) extractable from a distribution P [35]. A distribution might have more randomness, but H_∞ quantifies the minimum value (being the lowest-valued Rényi entropy). In our result, $H_\infty^\beta(P_{\text{rev}})$ can be thought of as the uniform randomness intrinsic to the work distribution. $H_\infty^\beta(P_{\text{rev}})$ thus quantifies fluctuations in work, caused by irreversibility, that raise the lower bound on W^ε .

Whereas some one-shot results (e.g. [18, 22]) involve Rényi entropies of states, $H_\infty^\beta(P_{\text{fwd}})$ is an entropy of a *work distribution*. $H_\infty^\beta(P_{\text{fwd}})$ captures fluctuation information from all sources that might affect the work distribution. These sources include the initial state and the manner in which the protocol is executed (e.g., quickly or quasistatically). In contrast, an entropy evaluated on states encodes only some of this information. By containing an entropy of a work distribution, rather than an entropy of a state, the above results can be applied in a more general setting: they can be related to the output of any procedure, as opposed to the worth of a particular input state under a fixed procedure (usually taken to be the optimal one [18]). Consequently, our results remain independent of the work-extraction model used. Theorems 3 and 4 govern (semi)classical and quantum systems, so long the protocol produces a work distribution consistent with Crooks' theorem.

As theoretical limits, these bounds will always hold true. However, to be operationally useful for bounding w^δ (or W^ε) they must be applied to models where P^{\max} can be upper-bounded following enough trials of the experiment. This is possible if the distribution $P(W)$ is smooth such that beyond a certain narrowness of bin width, any further division of each bin results in approximately equal probability densities.

Applying information about the protocol executed in one direction, we have used Crooks' theorem to infer about the opposite direction. This information tightens the bound on W^ε when⁶ $H_\infty^\beta(P_{\text{rev}}) \geq 0$, i.e., when

$$P^{\max} < \beta. \quad (18)$$

Systems described poorly by conventional statistical mechanics tend to satisfy inequality (18). Such systems' work distributions have significant spreads relative to the characteristic energy scale β^{-1} , such that the distribution lacks tall peaks. We will present DNA-hairpin experiments as an example.

3.4. Crooks' theorem in specific one-shot work-extraction models

3.4.1. Tightening a bound in the work-extraction game

Egloff *et al* calculate the optimal amount w_{best}^δ of work δ -extractable from a state via the most efficient strategy [18]. Their calculation implies an upper bound on the work extractable via arbitrary strategies. By applying theorems 3 and 4 to the Egloff *et al* framework, we can tighten the bound for protocols that satisfy the assumptions used to derive Crooks' theorem and for which $P^{\max} < \beta$.

In the Egloff *et al* setting, we can consider a forward protocol that consists of two stages: first, the thermal state $\gamma_{-T} \equiv e^{-\beta H_{-T}}/Z_{-T}$ transforms into some nonequilibrium state σ as the externally driven Hamiltonian

⁵ For the standard formulation $H(X) = -\int dX P(X) \log P(X)$, dimensions have been ignored, and the implicit reference is a uniform distribution with width 1 and probability density 1. The unnormalized expression is $\lim_{dX \rightarrow 0} -\sum_X dX P(X) \log [P(X) dX]$, which diverges due to the dX in the logarithm.

⁶ One-shot entropies of continuous probability density functions are known to assume negative values [40] when they are less entropic than the implicit reference distribution.

changes. The system either can remain thermally isolated or can thermalize, provided that the thermalization satisfies detailed balance (see appendix A). The agent can choose one of many possible strategies, e.g., by alternating Hamiltonian changes and thermalizations or by thermally isolating the system throughout the first stage. Second, σ thermalizes to $\gamma_\tau \equiv e^{-\beta H_\tau}/Z_\tau$. This thermalization neither costs nor produces work. The entire protocol is encapsulated in $(\gamma_{-\tau}, H_{-\tau}) \mapsto (\sigma, H_\tau) \mapsto (\gamma_\tau, H_\tau)$. At the start of the reverse protocol, the system begins in the thermal state of H_τ . Under the time-reversed process (in which the drive is reversed, such that the Hamiltonian retraces its path through configuration space), the system is transformed into some nonequilibrium state σ' . Then the state thermalizes to the thermal state of H_τ .

For protocols which fall into the above category, knowing about one protocol, we can bound the work extractable from, or the work cost of, the opposite protocol:

Corollary 5. *The work δ -extractable from each implementation of the reverse protocol, in terms of the forward protocol's $H_\infty^\beta(P_{\text{fwd}})$, satisfies*

$$w^\delta \leq \frac{1}{\beta} \left[\log M \left(\frac{G^T(\gamma_\tau)}{1-\delta} \parallel G^T(\gamma_{-\tau}) \right) - H_\infty^\beta(P_{\text{fwd}}) \right] \quad (19)$$

for $\delta \in [0, 1)$.

Corollary 6. *The work ε -required during each implementation of the forward protocol, in terms of the reverse protocol's $H_\infty^\beta(P_{\text{rev}})$, satisfies*

$$W^\varepsilon \geq \frac{1}{\beta} \left[\log M \left(\frac{G^T(\gamma_{-\tau})}{1-\varepsilon} \parallel G^T(\gamma_\tau) \right) + H_\infty^\beta(P_{\text{rev}}) \right] \quad (20)$$

for $\varepsilon \in [0, 1)$.

The proofs appear in appendix C.2. Each corollary consists of a bound derived from [18] and an H_∞^β correction attributable to Crooks' theorem (introduced via theorems 3 and 4). The H_∞^β quantifies the protocol's suboptimality, caused by dissipation due to the protocol's speed [29]. Positive values of H_∞^β tighten the bounds. Hence incorporating information about the forward (reverse) process into the reverse (forward) bound improves the bound when the process deviates sufficiently from the quasistatic ideal.

3.4.2. Modeling fluctuation-relation problems with resource theories

One can formulate scenarios governed by Crooks' theorem in thermodynamic resource theories. Such scenarios involve a sequence of *thermal operations* that obey detailed balance. Such operations form a strict subset of the set of all thermal operations (see appendix A). Hence Crooks' theorem does not govern all thermal operations. The application of Crooks' theorem requires the introduction of work and time into the resource theories. Work can be defined in terms of a battery [21, 23]; and time, in terms of a clock [20, 26]. Resource-theory results can be used to derive the work cost of a sequence of transformations that a system governed by Crooks' theorem can follow (see appendix D.2).

We leave for future work the derivation, from resource-theory results, of testable predictions about Crooks' problem. Considerable mathematical tools, such as monotones [20, 22, 41] and catalysts [22, 41], have been developed within the resource-theory framework. Having demonstrated the applicability of Crooks' theorem to resource theories, we look to use Crooks' theorem to bridge these mathematical tools to experiments.

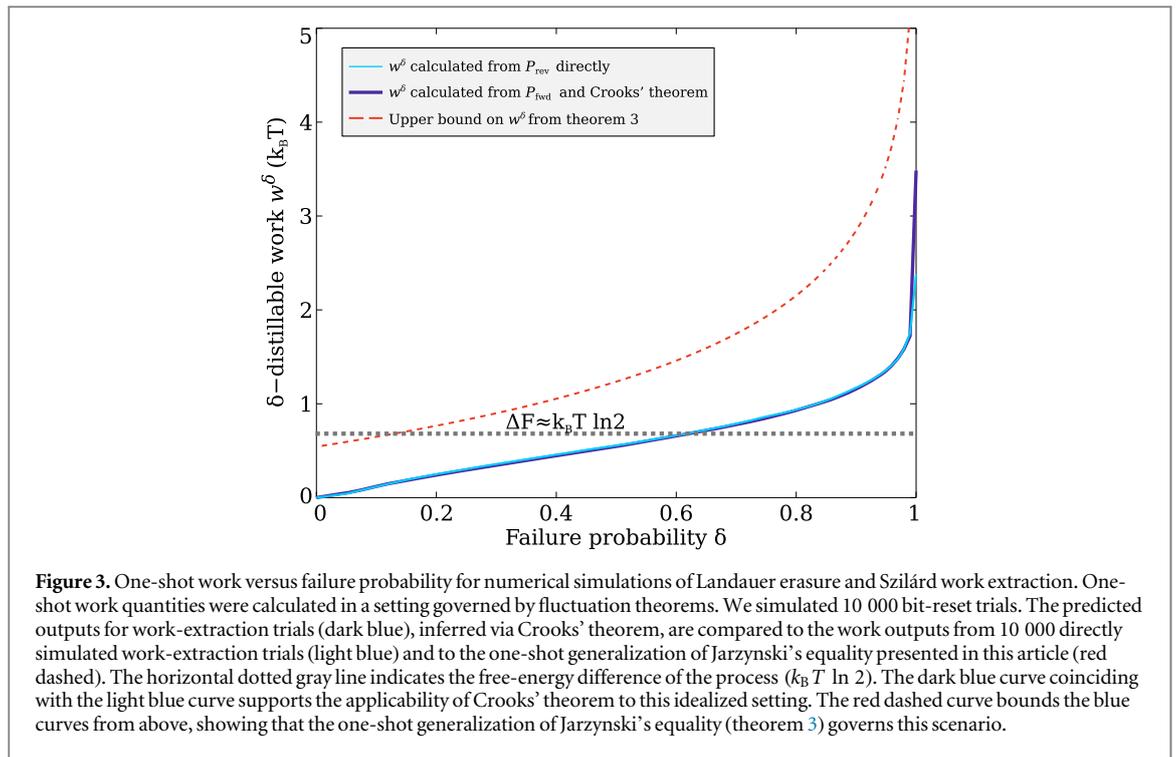
4. Examples of one-shot work quantities in fluctuation contexts

4.1. Landauer bit reset and Szilárd work extraction

A simple example involves the heat-exchanging portion of *Landauer bit reset* and its reverse, *Szilárd work extraction*. The set-up consists of a two-level system \mathcal{S} governed by the Hamiltonian $H(\lambda_t) = E(t)|E\rangle\langle E|$. Suppose \mathcal{S} exchanges energy with a heat bath whose temperature is $T = 1/\beta$. At time $t = -\tau$, $E(t) = 0$, and \mathcal{S} is in thermal equilibrium, i.e., in the maximally mixed state $\rho(-\tau) = \frac{1}{2}(|0\rangle\langle 0| + |E\rangle\langle E|)$. If ρ represents the location of a particle in a two-compartment box, the agent has no idea which compartment the particle occupies.

Transforming $\rho(-\tau)$ into a pure state—forcing the particle into one half of the box—is called *bit reset*, or *Landauer erasure*. Resetting the bit quasistatically costs, on average,

$$\langle W \rangle = \int_0^\infty \frac{e^{-\beta E}}{Z} dE = k_B T \log 2. \quad (21)$$



If the bit is reset in a finite time, $\langle W \rangle$ might exceed $k_B T \log 2$ [29]. Such a protocol has appeared in fluctuation contexts before [18, 19, 29] and has been realized experimentally (e.g., in a test of Jarzynski's equality by Brownian motion [42, 43]).

We define failure under the assumption that every started trial is completed: a forward trial *fails* if it consumes more work than the budgeted work W^ϵ . (Alternatively, one could define 'failure' under the assumption that too-costly trials would not be completed. A trial would be said to fail if the budgeted work were consumed but the bit had not been reset.)

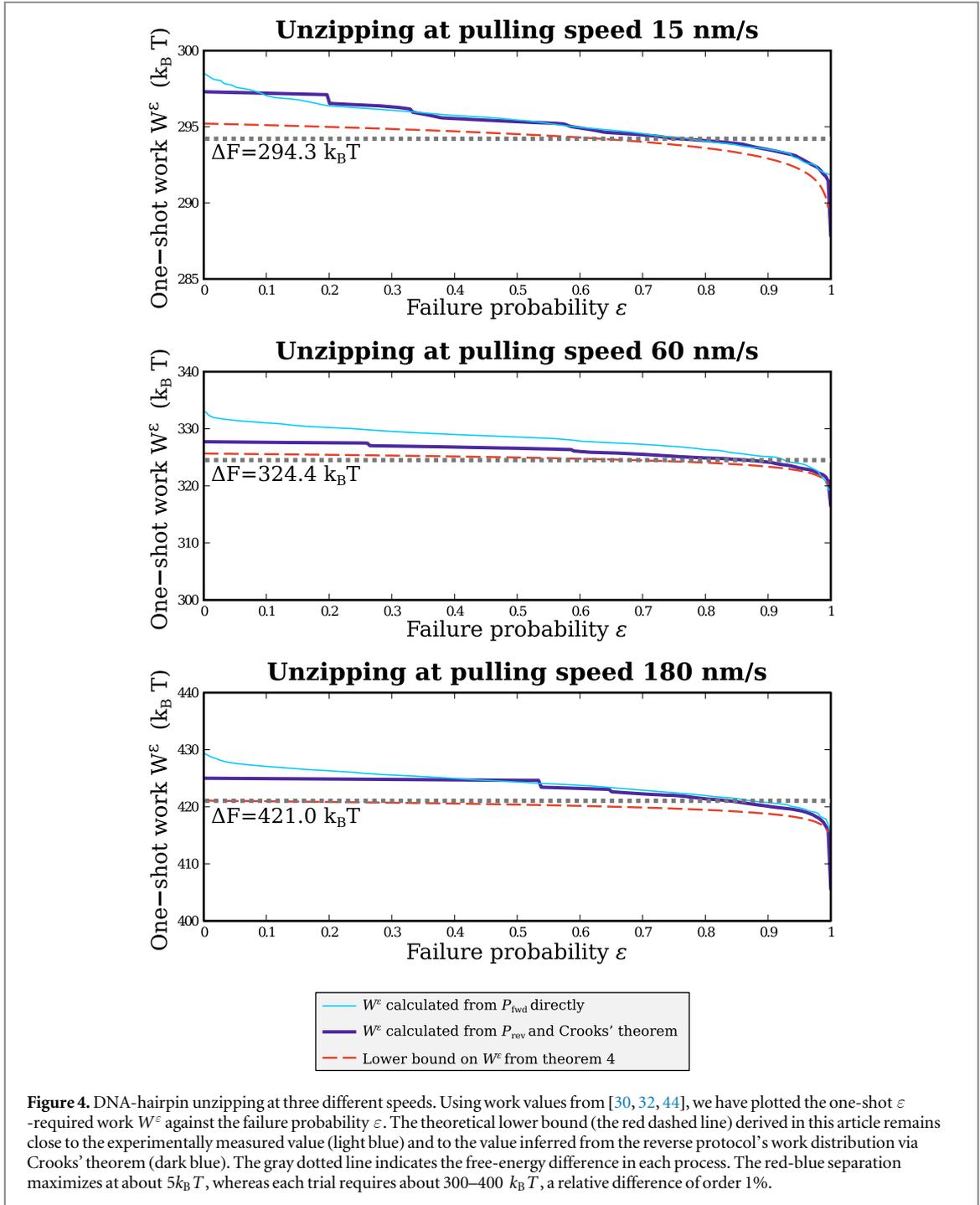
Reversal of the bit reset amounts to *Szilárd work extraction*. Leó Szilárd envisioned the conversion of information into work in 1929 [27]. \mathcal{S} begins thermally isolated, in the pure state $|0\rangle$, and governed by $H(\lambda_t) = 0$. During the first leg of Szilárd work extraction, the agent raises $E(t)$ to infinity. The raising costs no work because \mathcal{S} occupies the lower level. In the second leg, \mathcal{S} is coupled to the bath, then performs positive work as $E(t)$ decreases to zero. To be strictly the reverse of the bit reset presented above, we consider only the second leg as an implementation of the reverse protocol. The initial state, although it is a pure state, is thermal since only the lower energy level has non-zero occupation probability according to the Gibbs distribution. As such this work-extraction step can be linked to the bit reset via Crooks' theorem; their Hamiltonians are the time-reverse of each other, and each protocol starts in the appropriate thermal state.

These two processes, forming a forward-and-reverse pair governed Crooks' theorem, provide a natural example with which to test our one-shot results. We performed a Monte Carlo simulation of Landauer erasure and Szilárd work extraction. Details appear in appendix E. The simulation produced results consistent with theorem 3, as shown in figure 3.

4.2. Experiment: DNA-hairpin unzipping

When single molecules are manipulated experimentally, 'fluctuations are relevant and deviations from the average behavior are observable' [32]. Some such experiments are known to obey fluctuation relations. We show that data from DNA-hairpin experiments used previously to test Crooks' theorem [30–32] agree with the one-shot results in section 3. The agreement suggests that one-shot statistical mechanics might shed light on similar single-molecule experiments and applications. Alternatively, such experiments might be used to test one-shot statistical mechanics.

A *DNA hairpin* is a short double helix of about 21 base pairs [30–32]. The helix's two strands are called *legs*. One end of one leg is attached to one end of the other leg, forming a shape like a hairpin's. The other end of each leg ends in a *handle* formed from DNA. To each handle is attached a polystyrene or silica bead. One bead remains anchored on a micropipette. The other is caught in an optical trap that exerts a force. During the forward protocol, these optical tweezers pull the legs apart, unzipping the DNA into one strand. The more quickly the



hairpin is split (the greater the *pulling speed*), the more work is dissipated. During the reverse protocol, the helix is rezipped.

We combined work distributions provided by Ritort and Alemany [30, 32, 44] with theorem 4 (illustrated in figure 4). Each graph shows the work W^ϵ that a given unzipping trial is ϵ -guaranteed to require, plotted against the failure probability ϵ . We converted the data (a list of work values) into a probability distribution by forming a histogram with 50 equally sized bins that span the range of work costs. Energy is given in units of $k_B T$, such that $\beta = 1$. For pulling speeds of 15, 60, and 180 nm s^{-1} , the binning resulted in distributions whose $P^{\max} = 0.465 \beta$, 0.277β , and 0.162β respectively. In all three cases $P^{\max} < \beta$, such that the H_{\max} term in theorem 4 tightens the bound.

Whereas a work investment of about 300–400 $k_B T$ is required to complete the procedure, the jittering between the light blue curve (the directly measured value of W^ϵ) and the dark blue curve (the value of W^ϵ predicted from the reverse work distribution P_{rev} via lemma 2) is of a scale less than 5 $k_B T$. Hence Crooks' theorem interrelates the work probability distributions up to a discrepancy of around 1%, as argued in [30, 32].

The red curve remains below the dark blue and light blue curves, confirming that the one-shot lower bound in theorem 4 governs this experimental setting. The red curve remains close to—always within about $5k_B T$ of (around 1% of ΔF)—the light blue curve that represents directly simulated work investment. This agreement between theory and experiment suggests that the application of one-shot results may shed light on similar single-molecule experiments and on applications such as molecular motors, thermal ratchets, and nanoscale engines (e.g., [45–47]).

5. Conclusions and outlook

Crooks' theorem relates probability distributions between a process and its reverse. We can manipulate these distributions using tools from one-shot statistical mechanics. As demonstrated in this article, combining the toolkits leads to bounds on the work likely to be required (or produced) in classical and quantum process. Information about fluctuations tighten the bounds. Fluctuation relations and one-shot statistical mechanics are not competitors, but are mutually compatible. Combining the approaches yields statements about quite general thermal systems. The combination illustrates a possible bridge from one-shot theory to experimental settings through fluctuation theorems.

One experimental application is the cost of bit reset in modern microprocessors. As miniaturization reduces the size of transistors further into the nanoscale (e.g., [48]), limiting only the *average* heat dissipation does not ensure that devices work. Of increasing importance is a guarantee that no single bit reset dissipates any amount of heat (costs any amount of work) above some threshold that could damage the nanoscale device. The relevant fluctuations can be studied with Crooks' theorem and related, via the results in this article, to the one-shot maximum work cost.

The results in this article might be useful also when the work available to be spent on each trial is limited, or if the work extracted from each trial must exceed a certain threshold, except with bounded failure probability. Such quantities might have uses also in a paranoia setting. An agent might have a known amount of work to invest, and one might need to ensure that the agent can not erase some information, except with some small probability. Similarly, one-shot work might be applicable in a verification scenario. Suppose an agent claims to be able to provide some amount of work. To test the claim, one can request a transformation that costs more than this amount of work, except in a bounded number of cases.

Future research might reveal further links between one-shot statistical mechanics and fluctuation theorems. Here through the analysis of fluctuation theorems, we identified a relationship between one-shot work quantities and the order- ∞ Rényi entropy. By considering the Rényi divergence between the work distributions of a process and its reverse, one might find a relationship with one-shot *dissipated work*, following from the observation that the average dissipated work is proportional to the Kullback–Leibler divergence (average relative entropy) between the forwards and reverse work distributions [49]. Considering divergences between distributions avoids the issue of infinitely high entropy when the work distribution contain sharp peaks that have a finite ratio of height to each corresponding peak in the reverse distribution. As such, this approach could provide general and robust tools for bridging one-shot statistical mechanics to fluctuation settings that hold for discrete work distributions in addition to the continuous distributions focused on in this article.

Note added: between the first presentation of these results and the current version of this article, related results have appeared in [50, 51].

Acknowledgments

The authors are grateful for conversations with Anna Alemany, Janet Anders, Cormac Browne, Tanapat Deesuwan, Alex Lucas, Jonathan Oppenheim, Felix Pollock, and Ibon Santiago. This work was supported by a Virginia Gilloon Fellowship, an IQIM Fellowship, support from NSF grant PHY-0803371, the FQXi Large Grant for 'Time and the Structure of Quantum Theory,' the EPSRC, the John Templeton Foundation, the Leverhulme Trust, the Oxford Martin School, the National Research Foundation (Singapore), and the Ministry of Education (Singapore). The Institute for Quantum Information and Matter (IQIM) is an NSF Physics Frontiers Center with support from the Gordon and Betty Moore Foundation. VV and OD acknowledge funding from the EU Collaborative Project TherMiQ (Grant Agreement 618074). NYH was visiting the University of Oxford under the auspices of Jonathan Barrett and the Department of Atomic and Laser Physics while much of this paper was developed.

Appendix A. Relationships among thermalization models

Consider a system \mathcal{S} governed by a discrete N -level Hamiltonian H . Suppose that \mathcal{S} interacts with a heat bath whose inverse temperature is β . An N -dimensional probability vector \vec{s} represents the system's state. Whole or partial thermalization of \mathcal{S} can be modeled as a sequence of discrete steps, each represented by a stochastic matrix. Different possible properties of such matrices characterize different models of heat exchanges. We address the properties of Gibbs-preservation, detailed balance, and thermalization. By \vec{g} , we denote the probability vector that represents the Gibbs state associated with H and β :

$$\vec{g} = \left(\frac{e^{-\beta E_1}}{Z}, \frac{e^{-\beta E_2}}{Z}, \dots, \frac{e^{-\beta E_N}}{Z} \right).$$

A matrix M is *Gibbs-preserving* relative to H and β if M maps the corresponding Gibbs state to itself:

$$M\vec{g} = \vec{g}. \quad (\text{A1})$$

Gibbs preservation constrains the unit-eigenvalue eigenspace of M . The set \mathcal{G} of Gibbs-preserving matrices on quasiclassical states is equivalent to the set of resource-theory thermal operations on quasiclassical states [20] and to the set of thermal interactions in the game [18].

A strict subset of \mathcal{G} is the set \mathcal{D} of detailed-balanced matrices: $\mathcal{D} \subset \mathcal{G}$. Let A and B denote microstates associated with the energies E_A and E_B . M encodes the probabilities that \mathcal{S} transitions from A to B , and vice versa, during one heat-exchange step. If these probabilities satisfy

$$P(A \mapsto B) = P(B \mapsto A)e^{-\beta(E_B - E_A)}, \quad (\text{A2})$$

M obeys *detailed balance* [2].

If the steps in an extended heat exchange obey detailed balance, the extended heat exchange obeys *microscopic reversibility* [33]. From the assumption that heat exchanges are microscopically reversible, Crooks derives his theorem [2]. Hence if the heat exchanges in a process obey detailed balance (and the other assumptions used to derive Crooks' theorem, such as initialization to a thermal state), the process obeys Crooks' theorem.

Crooks defines microscopic reversibility as follows while deriving his theorem [2]. Let $P(x(t)|\lambda_t)$ denote the probability that, if the external parameter varies as λ_t during some forward trial, the state of the classical system \mathcal{S} follows the phase-space trajectory $x(t)$. The 'corresponding time reversed path' is denoted by $(\bar{\lambda}(-t), \bar{x}(-t))$. Let the functional $Q[x(t), \lambda_t]$ denote the heat that \mathcal{S} ejects if λ_t and $x(t)$ characterize the forward trial. The heat exchange obeys *microscopic reversibility* if

$$\frac{P(x(t)|\lambda_t)}{P(\bar{x}(-t)|\bar{\lambda}(-t))} = e^{-\beta Q[x(t), \lambda_t]}. \quad (\text{A3})$$

Another strict subset of Gibbs-preserving matrices is the set \mathcal{T} of thermalizing matrices: $\mathcal{T} \subset \mathcal{G}$. We call a matrix M *thermalizing* if it evolves every state \vec{s} of \mathcal{S} toward the Gibbs state associated with H and β :

$$\lim_{n \rightarrow \infty} M^n \vec{s} = \vec{g}. \quad (\text{A4})$$

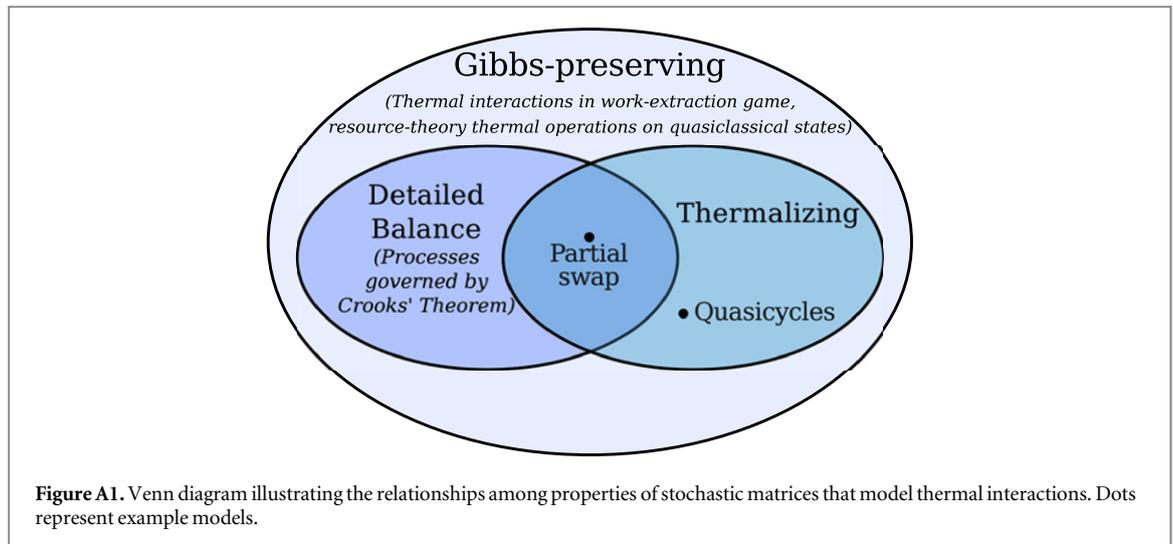
Equation (A4) encapsulates intuitions about what 'thermalization' means. Some matrices that model thermal interactions in the game and in the thermodynamic resource theories violate equation (A4), as do some thermal interactions governed by Crooks' theorem. \mathcal{T} overlaps with \mathcal{D} .

The properties we have introduced—Gibbs preservation, detailed balance, and thermalization—imply relationships among Crooks' theorem, theorems about the game, and resource-theory theorems. The game, as well as the thermodynamic resource theories, model the heat exchanges in some processes governed by Crooks' theorem. Crooks' theorem does not necessarily govern all heat exchanges possible in the game or in the resource theories.

A.1. Proof of Venn diagram

Let us justify our modeling of processes governed by Crooks' theorem with the work-extraction game in [18] and with resource theories. Different *frameworks* (Crooks' theorem, the game, and the resource theories) model interactions with heat baths differently. One step in an interaction can be represented by a stochastic matrix that has at least one of three properties: Gibbs-preservation (\mathcal{G}), detailed balance (\mathcal{D}), and thermalization (\mathcal{T}). The relationships among these matrices are summarized in figure A1 and in the following statements:

- (i) $\mathcal{T} \subset \mathcal{G}$: all thermalizing matrices are Gibbs-preserving (lemma 7), but not vice versa (lemma 8).
- (ii) $\mathcal{D} \subset \mathcal{G}$: all detailed-balanced matrices are Gibbs-preserving (lemma 9), but not vice versa (lemma 10).



(iii) $\mathcal{D} \neq \mathcal{T}$, $\mathcal{D} \not\subset \mathcal{T}$, and $\mathcal{T} \not\subset \mathcal{D}$: obeying detailed balance is not equivalent to being thermalizing, and neither category is a subset of the other (lemma 11).

(iv) $\mathcal{D} \cap \mathcal{T} \neq \emptyset$: some matrices are detailed-balanced and thermalizing (lemma 12).

While proving these claims, we justify the inclusion of two example matrices, the partial swap and quasicycles, in figure A1.

The proofs contain the following notation: \mathcal{S} denotes a quasiclassical system that evolves under a Hamiltonian H and that exchanges heat with a bath whose inverse temperature is β . By $\vec{s} = (s_1, s_2, \dots, s_d)$, we denote the state of \mathcal{S} . The vector's elements are the diagonal elements of a density matrix relative to the eigenbasis of H . The Gibbs state relative to H and to β is denoted by \vec{g} .

To prove some of the foregoing claims, we characterize thermalizing matrices with the Perron–Frobenius theorem [52]. The theorem governs irreducible aperiodic non-negative matrices M^7 . Consider the eigenvalue λ of M that has the greatest absolute value. According to the Perron–Frobenius Theorem, λ is the only positive real eigenvalue of M , and λ is associated with the only non-negative eigenvector \vec{v}_λ of M . Suppose that M is stochastic, such that $\lambda = 1$. By the spectral decomposition theorem, $\lim_{n \rightarrow \infty} M^n \vec{s} = \vec{v}_\lambda$. If $\vec{v}_\lambda = \vec{g}$, the matrix is thermalizing.

Lemma 7. *All thermalizing matrices are Gibbs-preserving.*

Proof. Let M denote a thermalizing matrix associated with the same Hamiltonian and β as \vec{g} . For all states \vec{s} of \mathcal{S} ,

$$\lim_{n \rightarrow \infty} M^n \vec{s} = \vec{g}. \quad (\text{A5})$$

To prove the lemma by contradiction, we suppose that M does not map \vec{g} to itself: $\vec{g} \not\propto M\vec{g}$.

Premultiplying equation (A5) by M generates

$$\lim_{n \rightarrow \infty} MM^n \vec{s} = M\vec{g} \neq \vec{g}. \quad (\text{A6})$$

This equation contradicts

$$\lim_{n \rightarrow \infty} MM^n \vec{s} = \lim_{n \rightarrow \infty} M^{n+1} \vec{s} = \lim_{n \rightarrow \infty} M^n \vec{s} = \vec{g}. \quad (\text{A7})$$

By the contrapositive, all thermalizing matrices are Gibbs-preserving. \square

Lemma 8. *Not all Gibbs-preserving matrices are thermalizing.*

Proof. In general, this will be the case for matrices which have the Gibbs state as an eigenvector but do not otherwise satisfy the conditions of irreducibility or aperiodicity required for the Perron–Frobenius theorem to govern their behavior. To prove the lemma by example, we construct one Gibbs-preserving matrix that is not thermalizing. Consider a block-diagonal stochastic $N \times N$ matrix M . (Being block-diagonal, M is reducible.)

⁷ By *non-negative*, we mean that every element of M is no less than zero.

Let M decompose into two submatrices: $M = M_1 \oplus M_2$. Let M_1 be defined on the first n_1 energy levels, and let M_2 be defined on the remaining n_2 energy levels.

Denote by \vec{g}_1 the Gibbs state associated with the first n_1 energies (and the partition function Z_1), and by g_2 the Gibbs state associated with the final n_2 energies (and the partition function Z_2). Suppose that

$$\vec{g}_1 \equiv g_1 \oplus \underbrace{(0, 0, \dots, 0)}_{n_2} \quad \text{and} \quad \vec{g}_2 \equiv \underbrace{(0, 0, \dots, 0)}_{n_1} \oplus g_2 \quad (\text{A8})$$

are normalized probability eigenvectors of M , each associated with the unit eigenvalue. Every vector of the form

$$\vec{v}_\alpha = \alpha \vec{g}_1 + (1 - \alpha) \vec{g}_2 \quad (\text{A9})$$

is also a normalized probability eigenvector of M associated with the unit eigenvalue.

The possible forms of \vec{v}_α form a family. One member of the family is the Gibbs state \vec{g} , which corresponds to $\alpha = Z_1/Z$ (wherein Z denotes the total partition function). Hence $\vec{g} \in \{\vec{v}_\alpha\}_{\alpha \in [0,1]}$ is an eigenvector of M , and M is Gibbs-preserving. However, \vec{g} is not the only eigenvector associated with the unit eigenvalue. M does not evolve every initial state toward \vec{g} . In general, $\lim_{n \rightarrow \infty} M^n \vec{s} = \nu_\alpha$, wherein α is the total occupation probability of the first n_1 energy levels of \vec{s} . As some states \vec{s} correspond to $\alpha \neq Z_1/Z$ and to $\nu_\alpha \neq g$, M does not map every initial state to the Gibbs state. Hence M is not thermalizing. Our claim has been proved by example. \square

Together, lemmas 7 and 8 imply the strict relation $\mathcal{T} \subset \mathcal{G}$.

Lemma 9. *All matrices that obey detailed balance relative to the Hamiltonian H and the inverse temperature β preserve the Gibbs state \vec{g} associated with H and β .*

Proof. We will write the forms of the elements in an arbitrary detailed-balanced stochastic $N \times N$ matrix M . By performing matrix multiplication explicitly, we show that $M\vec{g} = \vec{g}$.

Let M_{ij} denote the element in the i th row and j th column of M . This element equals the probability that, upon beginning in the j th energy level, a system \mathcal{S} transitions to the i th level. Let g_i denote the thermal population of level i (the i th element of \vec{g}).

Detailed balance and stochasticity constrain the relationships among the M_{ij} . By the definition of detailed balance (equation (A2)), the elements in the lower left-hand triangle of M are related to the elements in the upper right-hand triangle by

$$M_{ji} = M_{ij} \frac{g_j}{g_i} \quad \forall j > i. \quad (\text{A10})$$

The matrix has the form

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} & \dots \\ M_{12} \frac{g_2}{g_1} & M_{22} & M_{23} & \dots \\ M_{13} \frac{g_3}{g_1} & M_{23} \frac{g_3}{g_2} & M_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{A11})$$

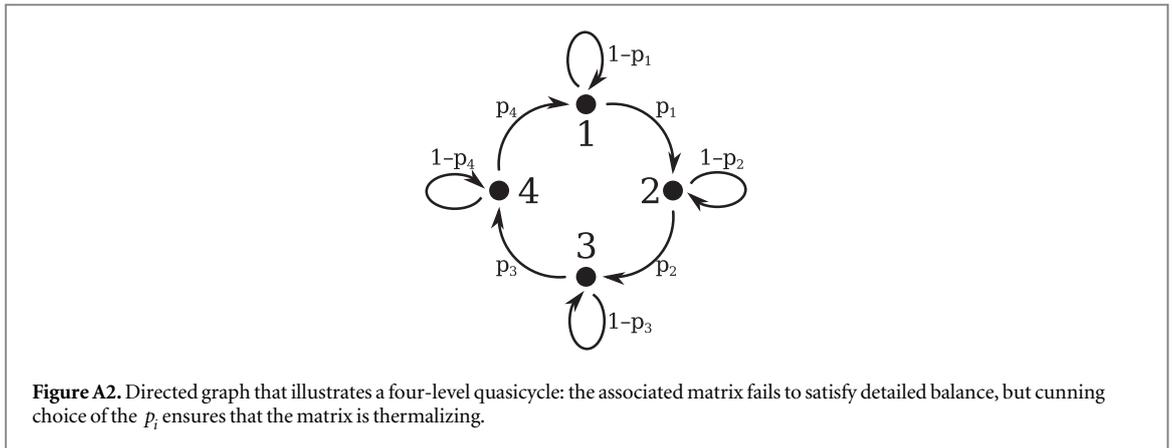
Because M is stochastic, the elements in each column sum to one. This normalization condition fixes each diagonal element M_{ii} as a function of the other M_{ij} that occupy the same column:

$$M_{ii} = 1 - \sum_{j < i} M_{ji} - \sum_{j > i} M_{ij} \frac{g_j}{g_i}. \quad (\text{A12})$$

Using index notation, we ascertain how M transforms the Gibbs state \vec{g} :

$$\begin{aligned} \sum_j M_{ij} g_j &= M_{ii} g_i + \sum_{j < i} M_{ij} g_j + \sum_{j > i} M_{ij} g_j \\ &= g_i - \sum_{j < i} M_{ji} g_i - \sum_{j > i} M_{ij} \frac{g_j}{g_i} g_i + \sum_{j < i} M_{ij} g_j + \sum_{j > i} M_{ij} g_j \\ &= g_i - \sum_{j < i} M_{ij} g_j - \sum_{j > i} M_{ij} g_j + \sum_{j < i} M_{ij} g_j + \sum_{j > i} M_{ij} g_j \\ &= g_i \end{aligned} \quad (\text{A13})$$

The second line follows from the substitution of equation (A12) for M_{ii} . The third line follows from the substitution of equation (A10) into the elements of first sum.



We have shown that \vec{g} is an eigenvector of M that corresponds to the unit eigenvalue. An $N \times N$ matrix M that obeys detailed balance relative to H and β preserves the Gibbs state associated with H and β . \square

Lemma 10. *Not every Gibbs-preserving matrix for some Hamiltonian H and inverse temperature β satisfies detailed balance for H and β .*

Proof. Gibbs-preservation only places a restriction on one eigenvalue of a matrix, and so there remains enough freedom to choose a matrix exhibiting this property, but not detailed balance. An example of such is a quasicycle, described in [20]. A *quasicycle* is a process that has a probability $P(i \mapsto (i + 1) \bmod N) \equiv p_i$ of evolving a system \mathcal{S} that occupies energy eigenstate i to eigenstate $i + 1$ and has a probability $1 - p_i$ of keeping \mathcal{S} in state i . All $p_i > 0$, and for at least one value of i , $p_i < 1$. The directed graph of a quasicycle forms a ring in which at least one node also has a loop to itself, corresponding to a value of $(1 - p_i) > 0$. An example appears in figure A2. The probability that i evolves to j forms element M_{ij} of matrix M . The matrix fails to satisfy detailed balance if \mathcal{S} has more than three energy eigenstates, because $P(i \mapsto (i + 1) \bmod N)$ is finite, though $P((i + 1) \bmod N \mapsto i) = 0$.

We will show that, if the p_i assume certain values, the Gibbs state \vec{g} is an eigenvector of M . Let level $i = 1$ correspond to the lowest energy eigenvalue. Solutions of the form

$$p_i = p \frac{g_1}{g_i}, \tag{A14}$$

wherein $p \in (0, 1)$ denotes a free parameter in the range $(0, 1)$, are Gibbs-preserving. Roughly, the greater the value of p , the more quickly the quasicycle is traversed.

To verify that equation (A14) describes a Gibbs-preserving matrix, we express the matrix multiplication $M\vec{g}$ in index form:

$$(M\vec{g})_1 = (1 - p_1)g_1 + p_N g_N \tag{A15}$$

$$(M\vec{g})_i = (1 - p_i)g_i + p_{i-1}g_{i-1} \quad i = 2 \dots N. \tag{A16}$$

Upon substituting in from equation (A14), we can simplify the equations to

$$(M\vec{g})_1 = g_1 \tag{A17}$$

$$(M\vec{g})_i = g_i \quad i = 2 \dots N. \tag{A18}$$

Hence $M\vec{g} = \vec{g}$, so M preserves Gibbs states. \square

Together, lemmas 9 and 10 imply the strict relation $\mathcal{D} \subset \mathcal{G}$.

Lemma 11. *Obeying detailed balance is not equivalent to thermalizing: $\mathcal{D} \neq \mathcal{T}$, nor is one category a subset of the other.*

Proof. We will show that the quasicycle matrix M described in the proof of lemma 10—a matrix that does not obey detailed balance—is thermalizing. Because M is stochastic by construction, its greatest eigenvalue equals one.

In addition to being stochastic, M is irreducible, aperiodic⁸ and non-negative. By the Perron–Frobenius Theorem, the greatest eigenvalue λ of M corresponds to the only non-negative eigenvector \vec{v}_λ of M . This $\lambda = 1$, because M is stochastic. As shown in the proof of lemma 10, $\vec{v}_\lambda = \vec{g}$. As explained below the proof of lemma 7, $\lim_{n \rightarrow \infty} M^n$ maps every vector \vec{s} to \vec{g} : the matrices that represent quasicycles thermalize. M does not obey detailed balance, as discussed in the proof of lemma 9.

A simple example of a matrix that obeys detailed balance, but is not thermalizing, is the identity matrix. Less trivially, it is possible in general to engineer a block-diagonal matrix of the form given in lemma 8, and if each block obeys detailed balance, the matrix as a whole will also obey detailed balance (noting that $P(A \mapsto B) = P(B \mapsto A) = 0$ trivially satisfies detailed balance). Such a matrix will not be thermalizing. Hence we see that thermalizing is not equivalent to obeying detailed balance: $\mathcal{D} \neq \mathcal{T}$, and neither category is a subset of the other. \square

Lemma 12. *Some matrices are detailed-balanced and thermalizing: $\mathcal{D} \cap \mathcal{T} \neq \emptyset$.*

Proof. We can prove this lemma by example. After reviewing the form of the partial-swap matrix M , we show that M thermalizes, then show that M obeys detailed balance.

A partial-swap operation has some probability p of replacing the operated-on state \vec{s} with a thermal state and a probability $1 - p$ of preserving \vec{s} . If \vec{s} denotes the state of an N -level system,

$$M = (1 - p)\mathbb{1}_N + pG, \quad (\text{A19})$$

wherein $\mathbb{1}_N$ denotes the $N \times N$ identity and every column of the matrix G is the thermal state \vec{g} .

Let us prove that M thermalizes. M is stochastic, as it is the probabilistic combination of $\mathbb{1}_N$ and G , which are stochastic. If N is finite, G is positive; so when $p > 0$, M is positive. Positivity implies irreducibility and aperiodicity. Hence the Perron–Frobenius Theorem⁹ implies that M has just one non-negative eigenvector \vec{v}_λ and that this eigenvector corresponds to $\lambda = 1$. Direct multiplication shows $M\vec{g} = \vec{g}$. Thus, $\vec{g} = \vec{v}_\lambda$ is the only non-negative eigenvector of M and corresponds to the largest eigenvalue. By the argument above lemma 7, M thermalizes.

To show that M obeys detailed balance, we compare the matrix elements that represent the probabilities of transitions between states i and j :

$$\begin{aligned} \frac{P(i \mapsto j)}{P(j \mapsto i)} &= \frac{M_{ji}}{M_{ij}} \\ &= \frac{(1 - p)\delta_{ij} + pg_j}{(1 - p)\delta_{ij} + pg_i} \\ &= e^{-\beta(E_j - E_i)}, \end{aligned} \quad (\text{A20})$$

wherein δ_{ij} denotes the Kronecker delta. This equation recapitulates the definition of detailed balance (equation (A2)). Hence matrices—such as the partial swap—can obey detailed balance while thermalizing. \square

Appendix B. Quantum derivation of generalized Jarzynski equalities

The results in section 3.2 apply to classical and quantum systems. To shed extra light on quantum applications, we present an alternative derivation of lemma 2 for a quantum system whose energy spectrum is discrete and that lacks contact with the heat bath while its Hamiltonian changes.

Work is defined as the difference between the outcomes of energy measurements near the protocol’s start and end. This definition of work, which appears in [5, 13, 15], differs from the definition in [6]. The discrete version of $\chi_{\text{rev}}^\epsilon(\beta)$ will be defined via analogy with equation (12):

$$\chi_{\text{rev}}^\epsilon(\beta) \equiv \sum_{W \geq -W^\epsilon} P_{\text{rev}}(W) e^{-\beta W}. \quad (\text{B1})$$

Let \mathcal{S} denote a quantum system characterized by an external parameter λ_t and governed by a Hamiltonian $H(\lambda_t)$ whose energy spectrum is discrete. Let β denote the inverse temperature of the heat bath with which \mathcal{S} interacts at times $t \in (-\infty, -\tau)$. At $t = -\tau$, \mathcal{S} is projectively measured in the energy eigenbasis, then isolated

⁸ Though quasicycles look cyclic, they are aperiodic. For the purposes of the Perron–Frobenius Theorem, a matrix’s period is the maximum value k_{max} of k that satisfies the statement ‘a system prepared in level A has a nonzero probability of evolving to level A only (but not necessarily) after multiples of k steps’. For irreducible matrices, the possible values of k do not depend on the choice of A [52]. When this index $k_{\text{max}} = 1$, the matrix is *aperiodic*. Every quasicycle contains at least one node that transitions to itself (not all $p_i = 1$, so at least one loop satisfies $P(i \mapsto i) = (1 - p_i) > 0$). Thus any value of k satisfies the above statement. Hence $k_{\text{max}} = 1$, and quasicycles are aperiodic.

⁹ For strictly positive matrices, the earlier Perron theorem implies the same result.

from the bath. Until $t = \tau$, a unitary $U(2\tau)$ evolves $H(\lambda_t)$ to H_τ , and \mathcal{S} is perturbed out of equilibrium. At $t = \tau$, the energy of \mathcal{S} is measured projectively. Define the work W performed on \mathcal{S} as the difference between the measurements' outcomes.

Lemma. *The ε -required work satisfies*

$$\chi_{\text{rev}}^\varepsilon(\beta) = (1 - \varepsilon)e^{\beta\Delta F} \quad \forall \varepsilon \in [0, 1]. \quad (\text{B2})$$

Proof. Let $\{|\phi_m(-\tau)\rangle\}$ and $\{E_m(-\tau)\}$ denote the eigenstates and eigenvalues of $H(\lambda_{-\tau})$, and let $\{|\phi_n(\tau)\rangle\}$ and $\{E_n(\tau)\}$ denote those of $H(\lambda_\tau)$. If the measurements yield outcomes m and n , the forward trial consumes $W \equiv E_n(\tau) - E_m(-\tau)$.

The time-reversed protocol proceeds from $t = \infty$ to $t = -\infty$ and is defined as in section 2.1. Let $p_n(\tau)$ denote the probability that the first measurement during a reverse trial yields $E_n(\tau)$; and let $p_{\text{rev}}(m|n)$ denote the probability that, if the first measurement yields $E_n(\tau)$, the second yields $E_m(-\tau)$. By definition,

$$\chi_{\text{rev}}^\varepsilon(\beta) = \sum_{m,n} p_{\text{rev}}(m|n)p_n(\tau)e^{-\beta[E_m(-\tau)-E_n(\tau)]} \Theta(W^\varepsilon - E_n(\tau) + E_m(-\tau)), \quad (\text{B3})$$

wherein

$$\Theta(W^\varepsilon - W_0) \equiv \begin{cases} 1 & \text{if } W^\varepsilon \geq W_0 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B4})$$

Invoking $p_n(\tau) = e^{-\beta E_n(\tau)}/Z_\tau$, we cancel the $E_n(\tau)$ -dependent exponentials. $p_{\text{rev}}(m|n)$ equals the probability $p_{\text{fwd}}(n|m)$ that, if an energy measurement at $t = -\tau$ yields $E_m(-\tau)$ during a forward trial, an energy measurement at τ yields $E_n(\tau)$:

$$\begin{aligned} p_{\text{rev}}(m|n) &= \text{Tr}\left(|\phi_m(-\tau)\rangle\langle\phi_m(-\tau)| U^\dagger(2\tau) |\phi_n(\tau)\rangle\langle\phi_n(\tau)| U(2\tau)\right) \\ &= \text{Tr}\left(|\phi_n(\tau)\rangle\langle\phi_n(\tau)| U(2\tau) |\phi_m(-\tau)\rangle\langle\phi_m(-\tau)| U^\dagger(2\tau)\right) \\ &= p_{\text{fwd}}(n|m). \end{aligned} \quad (\text{B5})$$

Substitution into equation (B3) yields

$$\chi_{\text{rev}}^\varepsilon(\beta) = \frac{1}{Z_\tau} \sum_{m,n} p_{\text{fwd}}(n|m)e^{-\beta E_m(-\tau)} \Theta(W^\varepsilon - E_n(\tau) + E_m(-\tau)). \quad (\text{B6})$$

Upon multiplying by $Z_{-\tau}/Z_{-\tau}$, we replace $e^{-\beta E_m(-\tau)}/Z_{-\tau}$ with $p_m(-\tau)$:

$$\begin{aligned} \chi_{\text{rev}}^\varepsilon(\beta) &= \frac{Z_{-\tau}}{Z_\tau} \sum_{m,n} p_m(-\tau)p_{\text{fwd}}(n|m) \Theta(W^\varepsilon - E_n(\tau) + E_m(-\tau)) \\ &= (1 - \varepsilon)e^{\beta\Delta F}. \end{aligned} \quad (\text{B7})$$

The final equality follows from $F(\gamma) = -T \log Z$ and from the definition of ε . \square

An analogous argument yields equation (9) in lemma 2.

Appendix C. Details of the work-extraction game

C.1. Description of the game

Let us briefly review the bound presented by Egloff *et al* [18]. Consider the most efficient transformation $(\rho, H_\rho) \mapsto (\sigma, H_\sigma)$ that has a probability $1 - \delta$ of failing. That is, one sacrifices the certainty that the transformation will succeed, in hopes of extracting more work than can be gained from a certain-to-succeed transformation. All work that the system can output is collected; none is wasted. The transformation has a probability $1 - \delta$ of outputting at least the work

$$w_{\text{best}}^\delta(\rho, H_\rho \mapsto \sigma, H_\sigma) = T \log \left(M \left(\frac{G^T(\rho)}{1 - \delta} \parallel G^T(\sigma) \right) \right). \quad (\text{C1})$$

We will briefly overview the geometric definitions of *Gibbs-rescaling* (G^T) and the *relative mixedness* (M).

Let ρ have the spectral decomposition $\sum_{i=1}^{d_\rho} r_i |E_i\rangle\langle E_i|$ such that

$$r_1 e^{\beta E_1} \geq r_2 e^{\beta E_2} \geq \dots \geq r_{d_\rho} e^{\beta E_{d_\rho}}. \quad (\text{C2})$$

Consider the histogram that represents the r_i . Gibbs-rescaling ρ resizes each box in the histogram. The width of box i changes from unity to $e^{-\beta E_i}$, and the box's height increases by a factor of $e^{\beta E_i}$. Denote by $h_\rho^T(u)$ the height

of the point, on the rescaled histogram, whose x -coordinate is $u \in [0, Z(H_\rho)]$. Integrating $h_\rho^T(u)$, we define the *Gibbs-rescaled Lorenz curve* as the set of points

$$\left\{ \left(u, L_\rho^T(u) \right) \mid u \in [0, Z(H_\rho)] \right\}, \quad \text{wherein} \quad L_\rho^T(u) \equiv \int_0^u h_\rho^T(u) du. \quad (\text{C3})$$

The (unscaled) Lorenz curve L_ρ is equivalent to L_ρ^0 . Upon Gibbs-rescaling ρ and σ , we can compare the states' resourcefulness even though different Hamiltonians govern the states.

To incorporate the failure probability into the curve, we stretch L_ρ^T upward by a factor of $1/(1 - \delta)$. The resulting curve, $L_\rho^{T,\delta}$, encodes more reliable resourcefulness than (ρ, H_ρ) possesses, because extractable work trades off with the failure probability δ . Consider plotting $L_\rho^{T,\delta}$ on the same graph as L_σ^T . The curves are concave, bowing outward from the x -axis or stretching straight from $(0, 0)$ to $y = 1$. Consider compressing L_σ^T leftward. M denotes the inverse of the greatest factor by which L_σ^T can compress without popping above $L_\rho^{T,\delta}$:

$$L_\rho^T(u) \geq \left[M \left(\frac{G^T(\rho)}{1 - \delta} \parallel G^T(\sigma) \right) \right]^{-1} L_\sigma^T(u) \quad \forall u \in [0, \max(Z(H_\rho), Z(H_\sigma))]. \quad (\text{C4})$$

Illustrations appear in [18]. While transforming (ρ, H_ρ) into (σ, H_σ) , the player can extract no more work than $T \log M$: according to equation (C1),

$$w^\delta(\rho, H_\rho \mapsto \sigma, H_\sigma) \leq T \log \left(M \left(\frac{G^T(\rho)}{1 - \delta} \parallel G^T(\sigma) \right) \right). \quad (\text{C5})$$

C.2. Tightening and generalizing a one-shot bound with Crooks' theorem

Theorem 5 strengthens the above inequality (C5) (in the appropriate parameter regime), and theorem 6 generalizes this to work investment. These theorems are proved below:

Theorem. *The work δ -extractable from each Crooks-type reverse trial satisfies*

$$w^\delta \leq \frac{1}{\beta} \left[\log M \left(\frac{G^T(\gamma_\tau)}{1 - \delta} \parallel G^T(\gamma_{-\tau}) \right) - H_\infty(P_{\text{fwd}}) \right] \quad \forall \delta \in [0, 1). \quad (\text{C6})$$

Proof. During the reverse protocol, the state of a system \mathcal{S} transforms as

$$(\gamma_\tau, H_\tau) \mapsto (\sigma, H_{-\tau}) \mapsto (\gamma_{-\tau}, H_{-\tau}), \quad (\text{C7})$$

wherein σ denotes some density operator that likely is not an equilibrium state.

The set of strategies for transforming from (γ_τ, H_τ) to $(\gamma_{-\tau}, H_{-\tau})$ contains all strategies that achieve this transformation via $(\sigma, H_{-\tau})$. As such, when optimizing to maximize the work production allowing an error rate of δ ,

$$w_{\text{best}}^\delta(\gamma_\tau, H_\tau \mapsto \gamma_{-\tau}, H_{-\tau}) \geq w_{\text{best}}^\delta(\gamma_\tau, H_\tau \mapsto \sigma, H_{-\tau}) + w_{\text{best}}^0(\sigma, H_{-\tau} \mapsto \gamma_{-\tau}, H_{-\tau}). \quad (\text{C8})$$

The final term (which only involves thermalization) always succeeds, does not contribute either to the failure probability or the work cost, and hence can be eliminated. An arbitrary, possibly suboptimal, strategy from (γ_τ, H_τ) to $(\sigma, H_{-\tau})$ that generates work w^δ will be bounded by this value:

$$w^\delta \leq w_{\text{best}}^\delta(\gamma_\tau, H_\tau \mapsto \gamma_{-\tau}, H_{-\tau}). \quad (\text{C9})$$

Hence, by equation (C1),

$$w^\delta \leq \frac{1}{\beta} \log M \left(\frac{G^T(\gamma_\tau)}{1 - \delta} \parallel G^T(\gamma_{-\tau}) \right). \quad (\text{C10})$$

Let us calculate M . $L_{\gamma_{-\tau}}^T$ stretches straight from $(0, 0)$ to $(Z_{-\tau}, 1)$, whereas $L_{\gamma_\tau}^T/(1 - \delta)$ stretches straight to $(Z_\tau, \frac{1}{1 - \delta})$. Compressing $L_{\gamma_\tau}^T(u)/(1 - \delta)$ leftward by a factor of $M^{-1} = \frac{Z_\tau(1 - \delta)}{Z_{-\tau}}$ keeps the latter curve from dipping below $L_{\gamma_{-\tau}}^T(u)$. Hence

$$\begin{aligned} \frac{1}{\beta} \log M \left(\frac{G^T(\gamma_\tau)}{1-\delta} \parallel G^T(\gamma_{-\tau}) \right) &= \frac{1}{\beta} \left[\log \left(\frac{Z_{-\tau}}{Z_\tau} \right) - \log(1-\delta) \right] \\ &= \Delta F - \frac{1}{\beta} \log(1-\delta), \end{aligned} \quad (\text{C11})$$

wherein $\Delta F \equiv F(\gamma_\tau) - F(\gamma_{-\tau})$. We substitute from equation (C11) into the inequality (13) derived from Crooks' theorem in theorem 3. \square

Crooks' theorem introduces an H_∞ into the bound, derived from [18], on extractable work. If P_{fwd} satisfies inequality (18), this H_∞ strengthens the bound. A work cost can similarly be derived from [18], then enhanced with Crooks' theorem.

Appendix D. Details of thermodynamic resource theories

D.1. Description of thermodynamic resource theories

Each thermodynamic resource theory models energy-preserving transformations performed with a heat bath characterized by an inverse temperature β . To specify a state, one specifies a density operator and a Hamiltonian: (ρ, H) . Sums of Hamiltonians will be denoted by $H_1 + H_2 \equiv H_1 \otimes \mathbb{1} + \mathbb{1} \otimes H_2$.

Thermal operations can be performed for free. Each consists of three steps: (1) a Gibbs state relative to β and to any Hamiltonian H_γ can be drawn from the bath:

$$(\gamma, H_\gamma), \quad \text{wherein} \quad \gamma \equiv \frac{e^{-\beta H_\gamma}}{Z}. \quad (\text{D1})$$

(Below, the Gibbs state relative to β and to H_γ will be denoted also by $\gamma(H_\gamma)$.) (2) Any unitary U that conserves the total energy can be implemented, and (3) any subsystem A associated with its own Hamiltonian H_A can be discarded. Each thermal operation on (ρ, H) has the form

$$(\rho, H) \mapsto \left(\text{Tr}_A(U[\rho \otimes \gamma]U^\dagger), H + H_\gamma - H_A \right), \quad (\text{D2})$$

wherein $[U, H + H_\gamma] = 0$.

D.2. Applicability of Crooks' theorem

Some resource-theory operations obey detailed balance and Markovianity. We can use resource theories to model processes governed by Crooks' theorem if we define a battery and a clock. Our model for the battery appears in [23] and resembles the model in [21]. The quasiclassical battery B has closely spaced energy levels and occupies an energy eigenstate:

$$B_i \equiv \left(|B_i\rangle\langle B_i|, H_B \right), \quad \text{wherein} \quad H_B \equiv \sum_i E_{B_i} |B_i\rangle\langle B_i|. \quad (\text{D3})$$

If E_{B_i} is large, a work-costing (forward) process can transfer work from the battery to \mathcal{S} . If E_{B_i} is small, a work-extraction (reverse) process can transfer work from \mathcal{S} to the battery.

We model the evolution of H with a clock C that occupies a pure state $|C_j\rangle$ [20, 26]. The changing of $|C_j\rangle$, like the movement of a clock hand, models the passing of instants. In processes governed by Crooks' theorem, $H = H(\lambda_t)$. We discretize t such that the system's Hamiltonian is $H(\lambda_{t_i})$ when the clock occupies the state $|C_j\rangle$. In the notation introduced earlier, $t_1 = -\tau$, and $t_n = \tau$. The composite-system Hamiltonian

$$H_{\text{tot}} \equiv \sum_{i=1}^n H(\lambda_{t_i}) \otimes |C_i\rangle\langle C_i| \otimes \mathbb{1} + \mathbb{1} \otimes \mathbb{1} \otimes \sum_j E_{B_j} |B_j\rangle\langle B_j| \quad (\text{D4})$$

remains constant.

Having defined the battery and clock, we define the work extractable from, and the work cost of, a protocol. Let $E_{B_0} = 0$. The most work extractable from the reverse protocol equals the greatest E_{B_m} for which some sequence of thermal operations evolves the state of SCB as

$$\begin{aligned} \gamma(H_{-\tau}) \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| &\mapsto \rho(t_2) \otimes |2\rangle\langle 2| \otimes |E_{B_2}\rangle\langle E_{B_2}| \mapsto \dots \\ &\mapsto \gamma(H_\tau) \otimes |m\rangle\langle m| \otimes |E_{B_m}\rangle\langle E_{B_m}|, \end{aligned} \quad (\text{D5})$$

wherein $\rho(t_i)$ represents the state occupied by \mathcal{S} at time t_i . The forward protocol's minimum work cost equals the least E_{B_n} for which a sequence of thermal operations implements

$$\gamma(H_\tau) \otimes |n\rangle\langle n| \otimes |E_{B_n}\rangle\langle E_{B_n}| \mapsto \dots \mapsto \gamma(H_{-\tau}) \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0|. \quad (\text{D6})$$

Results in [20] can be applied if all the states commute with their Hamiltonians. Horodecki and Oppenheim have calculated the maximum work yield, or minimum work cost, of any quasiclassical transformation $(\rho, H_\rho) \mapsto (\sigma, H_\sigma)$ by thermal operations. They have calculated also faulty transformations' work yields and work costs. A faulty transformation generates a state (σ', H_σ) that differs from the desired state. The discrepancy is quantified by the trace distance between the density operators:

$$\frac{1}{2} \|\sigma - \sigma'\|_1 \leq \epsilon \in [0, 1]. \quad (\text{D7})$$

According to [20], this ϵ can be interpreted as the probability that the process fails to accomplish its mission, similarly to ε and δ . Generalizations to nonclassical states appear in [38, 39].

Appendix E. Details of numerical simulation

Our simulation of Landauer bit reset and Szilard work extraction resembles the scenario presented in [29], that models a two-level quasiclassical system \mathcal{S} . At each time t , the energy $\mathcal{E}(t)$ of \mathcal{S} equals E_0 or $E_1(t)$. The state of \mathcal{S} is represented by a vector $\vec{s}(t) = (p(t), 1 - p(t))$, wherein $p(t)$ equals the probability that $\mathcal{E}(t) = E_0$.

If observers have different amounts of information about $\mathcal{E}(t)$, they ascribe different values to $p(t)$. Suppose an agent draws \mathcal{S} from a temperature- $(1/\beta)$ heat bath. According to this *ignorant agent*,

$$\vec{s}(t) = \left(\frac{e^{-\beta E_0}}{Z(t)}, \frac{e^{-\beta E_1(t)}}{Z(t)} \right). \quad (\text{E1})$$

According to an *omniscient observer*, $\vec{s}(t) = (1, 0)$ or $(0, 1)$. The code is written from the perspective of an omniscient observer. On average, the code's predictions coincide with the predictions that code written by an ignorant agent would make.

While $t \in (-\infty, -\tau)$ during the forward (erasure) protocol, $E_1(t) = E_0 = 0$, and \mathcal{S} is thermally equilibrated. According to the ignorant agent, $\vec{s}(t) = \left(\frac{1}{2}, \frac{1}{2}\right)$. Beginning at $t = -\tau$, the agent raises E_1 by the infinitesimal amount dE while preserving $\vec{s}(t)$. Then, the agent couples \mathcal{S} to the bath for some time interval. The raising and coupling are repeated until $t = \tau$ and $E_1(\tau) = E_{\max}$.

The agent's actions are simulated as follows: our code has a probability $\frac{1}{2}$ of representing the initial state $\vec{s}(-\tau)$ with $(1, 0)$ and a probability $\frac{1}{2}$ of representing $\vec{s}(-\tau)$ with $(0, 1)$. Consider one thermal interaction that occurs at some $t \in (-\tau, \tau)$. If $\vec{s}(t) = (1, 0)$ before the thermal interaction, the agent invests no work to raise E_1 . If $\vec{s}(t) = (0, 1)$, the agent invests work dE .

A probabilistic swap models each interaction with the heat bath [29]. $\vec{s}(t)$ has a probability P_{swap} of being exchanged with a pure state sampled from a Gibbs distribution. That is, $\vec{s}(t)$ has a probability $P_{\text{swap}} e^{-\beta E_0}/Z(t)$ of being interchanged with $(1, 0)$, a probability $P_{\text{swap}} e^{-\beta E_1(t)}/Z(t)$ of being interchanged with $(0, 1)$, and a probability $1 - P_{\text{swap}}$ of remaining unchanged: $\vec{s}(t + dt) = \vec{s}(t)$. The longer \mathcal{S} couples to the reservoir, the greater the P_{swap} . The ignorant agent represents this thermalization with $\vec{s}(t + dt) = M(t; P_{\text{swap}})\vec{s}(t)$, wherein $M(t; P_{\text{swap}})$ is a thermalizing matrix that obeys detailed balance (see the proof of lemma 12). Because $\vec{s}(t + dt)$ depends on no earlier state except $\vec{s}(t)$, the evolution is Markovian.

Ideally, the agent increases $E_1(t)$ and thermalizes \mathcal{S} repeatedly until $t = \tau$, $E_1(\tau) = \infty$, and $\vec{s}(\tau) = (1, 0)$ according to both observers. The simulated $E_1(t)$ peaks at some large E_{\max} , and the final state has a high probability of being $(1, 0)$ [29]. During stage two of erasure, \mathcal{S} is thermally isolated, and E_1 decreases to zero. Because \vec{s} has no weight on E_1 , this stage costs no work.

Reversing erasure amounts to extracting work. Initially, $E_1 = E_0 = 0$, and $\vec{s} = (1, 0)$. As \mathcal{S} remains thermally isolated, E_1 rises to infinity (approximated by E_{\max}) without consuming work. During stage two of work extraction, the agent repeatedly lowers $E_1(t)$ by dE and thermalizes \mathcal{S} . Whenever the agent lowers $E_1(t)$ while $\vec{s}(t) = (0, 1)$ to the omniscient observer, \mathcal{S} outputs work dE . Once $t = -\tau$ such that $E_1(-\tau) = 0$, \mathcal{S} thermalizes until the probability that $\vec{s}(t) = (1, 0)$ equals the probability that $\vec{s}(t) = (0, 1)$.

To produce figure 3, we simulated a bit-reset process where the energy gap between the two levels increases linearly from 0 to $40 k_B T$ across 100 000 equal time-steps. After each step, the partial swap probability of thermalizing is 0.002, and $\beta = 10 k_B^{-1} K^{-1}$. This process was repeated 10 000 times, and the resulting work values binned into a histogram with 50 divisions. The maximum probability for the bit-reset work distribution was $P^{\max} = 8.60$, satisfying $P^{\max} < \beta$.

References

- [1] Jarzynski C 1997 Nonequilibrium equality for free energy differences *Phys. Rev. Lett.* **78** 2690–3
- [2] Crooks G E 1999 Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences *Phys. Rev. E* **60** 2721–6
- [3] Kurchan J 2000 A quantum fluctuation theorem (arXiv:cond-mat/0007360)
- [4] Tasaki H 2000 Jarzynski relations for quantum systems and some applications (arXiv:cond-mat/0009244)
- [5] Engel A and Nolte R 2007 Jarzynski equation for a simple quantum system: comparing two definitions of work *Europhys. Lett.* **79** 10003
- [6] Talkner P, Lutz E and Hänggi P 2007 Fluctuation theorems: work is not an observable *Phys. Rev. E* **75** 050102
- [7] Talkner P and Hänggi P 2007 The Tasaki–Crooks quantum fluctuation theorem *J. Phys. A: Math. Theor.* **40** 569–71
- [8] Quan H T and Dong H 2008 Quantum Crooks fluctuation theorem and quantum Jarzynski equality in the presence of a reservoir (arXiv:0812.4955)
- [9] Campisi M, Talkner P and Hänggi P 2009 Fluctuation theorem for arbitrary open quantum systems *Phys. Rev. Lett.* **102** 210401
- [10] Talkner P, Campisi M and Hänggi P 2009 Fluctuation theorems in driven open quantum systems *J. Stat. Mech.: Theory Exp.* P02025
- [11] Campisi M, Hänggi P and Talkner P 2011 Colloquium: quantum fluctuation relations: foundations and applications *Rev. Mod. Phys.* **83** 771–91
- [12] Esposito M, Harbola U and Mukamel S 2009 Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems *Rev. Mod. Phys.* **81** 1665–702
- [13] Hide J and Vedral V 2010 Detecting entanglement with Jarzynski equality *Phys. Rev. A* **81** 062303
- [14] Cohen D and Imry Y 2012 Straightforward quantum-mechanical derivation of the Crooks fluctuation theorem and the Jarzynski equality *Phys. Rev. E* **86** 011111
- [15] Dörner R, Clark S R, Heaney L, Fazio R, Goold J and Vedral V 2013 Extracting quantum work statistics and fluctuation theorems by single-qubit interferometry *Phys. Rev. Lett.* **110** 230601
- [16] Renner R 2005 Security of quantum key distribution *PhD Thesis* ETH Zürich (arXiv:quant-ph/0512258)
- [17] Dahlsten O C O, Renner R, Rieper E and Vedral V 2011 Inadequacy of von Neumann entropy for characterizing extractable work *New J. Phys.* **13** 053015
- [18] Egloff D, Dahlsten O C O, Renner R and Vedral V 2015 A measure of majorization emerging from single-shot statistical mechanics *New J. Phys.* **17** 073001
- [19] Åberg J 2013 Truly work-like work extraction via a single-shot analysis *Nat. Commun.* **4** 1925
- [20] Horodecki M and Oppenheim J 2013 Fundamental limitations for quantum and nanoscale thermodynamics *Nat. Commun.* **4** 2059
- [21] Skrzypczyk P, Short A J and Popescu S 2014 Work extraction and thermodynamics for individual quantum systems *Nat. Commun.* **5** 4185
- [22] Brandão F, Horodecki M, Ng N, Oppenheim J and Wehner S 2015 The second laws of quantum thermodynamics *Proc. Natl Acad. Sci. USA* **112** 201411728
- [23] Yunger Halpern N and Renes J M 2014 Beyond heat baths: generalized resource theories for small-scale thermodynamics (arXiv:1409.3998)
- [24] Janzing D, Wocjan P, Zeier R, Geiss R and Beth T 2000 Thermodynamic cost of reliability and low temperatures: tightening Landauer’s principle and the second law *Int. J. Theor. Phys.* **39** 2717–53
- [25] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 Quantum entanglement *Rev. Mod. Phys.* **81** 865
- [26] Brandão F G S L, Horodecki M, Oppenheim J, Renes J and Spekkens R W 2013 Resource theory of quantum states out of thermal equilibrium *Phys. Rev. Lett.* **111** 250404
- [27] Szilard L 1929 Über die entropieerminderung in einem thermodynamischen system bei eingriffen intelligenter wesen *Z. Phys.* **53** 840–56
- [28] Landauer R 1961 Irreversibility and heat generation in the computer process *IBM J. Res. Dev.* **5** 183–91
- [29] Browne C, Garner A J P, Dahlsten O C O and Vedral V 2014 Guaranteed energy-efficient bit reset in finite time *Phys. Rev. Lett.* **113** 100603
- [30] Mossa A, Manosas M, Forns N, Hugué J M and Ritort F 2009 Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes *J. Stat. Mech.: Theory Exp.* P02060
- [31] Manosas M, Mossa A, Forns N, Hugué J M and Ritort F 2009 Dynamic force spectroscopy of DNA hairpins: II. Irreversibility and dissipation *J. Stat. Mech.: Theory Exp.* P02061
- [32] Alemany A and Ritort F 2010 Fluctuation theorems in small systems: extending thermodynamics to the nanoscale *Europhys. News* **41** 27–30
- [33] Crooks G E 1998 Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems *J. Stat. Phys.* **90** 1481–7
- [34] Jarzynski C 2008 Nonequilibrium work relations: foundations and applications *Eur. Phys. J. B* **64** 331–40
- [35] Renner R and Wolf S 2004 Smooth Rényi entropy and applications *Int. Symp. on Information Theory, 2004. ISIT 2004. Proc.* (Piscataway, NJ: IEEE) pp 232–232 (<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1365269>)
- [36] Tomamichel M 2012 A framework for non-asymptotic quantum information theory *PhD Thesis* ETH Zürich (arXiv:1203.2142)
- [37] Dupuis F, Kraemer L, Faist P, Renes J M and Renner R 2012 Generalized Entropies *17th Int. Congress on Mathematical Physics* pp 134–53
- [38] Lostaglio M, Jennings D and Rudolph T 2015 Description of quantum coherence in thermodynamic processes requires constraints beyond free energy *Nat. Commun.* **6** 6383
- [39] Lostaglio M, Korzekwa K, Jennings D and Rudolph T 2015b Quantum coherence, time-translation symmetry, and thermodynamics *Phys. Rev. X* **5** 021001
- [40] Hanel R, Thurner S and Tsallis C 2009 On the robustness of q -expectation values and Rényi entropy *Europhys. Lett.* **85** 20005
- [41] Gour G, Müller M P, Narasimha V, Spekkens R W and Yunger Halpern N 2013 The resource theory of informational nonequilibrium in thermodynamics *Phys. Rep.* **583** 1–58
- [42] Bérut A, Petrosyan A and Ciliberto S 2013 Detailed Jarzynski equality applied to a logically irreversible procedure *Europhys. Lett.* **103** 60002
- [43] Jun Y, Gavrilov M and Bechhoefer J 2014 High-precision test of Landauer’s principle in a feedback trap *Phys. Rev. Lett.* **113** 190601
- [44] Alemany A and Ritort F 2013 private communication
- [45] Lacoste D, Lau A and Mallick K 2008 Fluctuation theorem and large deviation function for a solvable model of a molecular motor *Phys. Rev. E* **78** 011915

- [46] Cheng J, Sreelatha S, Hou R, Efremov A, Liu R, van der Maarel J R C and Wang Z 2012 Bipedal nanowalker by pure physical mechanisms *Phys. Rev. Lett.* **109** 238104
- [47] Serreli V, Lee C-F, Kay E R and Leigh D A 2007 A molecular information ratchet *Nature* **445** 523–7
- [48] Baldo M 2010 6.701 *Introduction to Nanoelectronics* MIT OpenCourseWare, Massachusetts Institute of Technology (<http://ocw.mit.edu>)
- [49] Gomez-Marin A, Parrondo J M R and van den Broeck C 2008 The footprints of irreversibility *Europhys. Lett.* **82** 50002
- [50] Dahlsten O, Choi M-S, Braun D, Garner A J P, Yunger Halpern N and Vedral V 2015 Equality for worst-case work at any protocol speed (arXiv:1504.05152)
- [51] Salek S and Wiesner K 2015 Fluctuations in single-shot ϵ -deterministic work extraction (arXiv:1504.05111)
- [52] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press) ISBN 0-521-38632-2