How many trials should you expect to perform to estimate a free-energy difference?

Nicole Yunger Halpern[1] and Christopher Jarzynski[2, 3]

[1] *Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA*
[2] *Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*
[3] *Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

(Dated: January 13, 2016)

The difference ΔF between free energies has applications in biology, chemistry, and pharmacology. The value of ΔF can be estimated from experiments or simulations, via fluctuation theorems developed in statistical mechanics. Calculating the error in a (ΔF)-estimate is difficult. Worse, atypical trials dominate estimates. How many trials one should perform was estimated roughly in [1]. We enhance the approximation with information-theoretic strategies: We quantify "dominance" with a tolerance parameter chosen by the experimenter or simulator. We bound the number of trials one should expect to perform, using the order-∞ Rényi entropy. The bound can be estimated if one implements the "good practice" of bidirectionality, known to improve estimates of ΔF. Estimating ΔF from this number of trials leads to an error that we bound approximately. Numerical experiments on a weakly interacting dilute classical gas support our analytical calculations.

PACS numbers: 05.70.Ln, 05.40.-a, 05.70.Ce, 89.70.Cf

The numerical estimation of free-energy differences is an active area of research, having applications to chemistry, microbiology, pharmacology, and other fields. Fluctuation relations can be used to estimate equilibrium free-energy differences ΔF from nonequilibrium experimental and simulation data. One repeatedly measures the amount W of work extracted from, or performed on, a system during an experiment or simulation. Fluctuation relations express the value of ΔF in terms of averages over infinitely many trials. Finitely many trials are performed in practice, introducing errors into estimates of ΔF. Efforts to quantify these errors, and to promote "good practices" in estimating ΔF, have been initiated (e.g., [2–8]).

How many trials should one perform to estimate ΔF reliably? The work W extracted from a system is a random variable that assumes different values in different trials. Typical trials involve W-values that contribute little to the averages being estimated. *Dominant* W-values, which largely determine the averages, characterize few trials [1]. Until observing a dominant W-value, one cannot estimate ΔF with reasonable accuracy. The probability that some trial will involve a dominant W-value determines the number N of trials one should expect to perform.

An estimate of N was provided in [1]. The estimate followed from identifying which values of W are dominant, then calculating the rough probability of sampling a dominant work value during a given trial. Our aim, in the present paper, is to enhance that estimate's precision. First, we introduce fluctuation relations and *one-shot information theory*, a mathematical toolkit for quantifying efficiencies at small scales. Next, we quantify dominance in terms of a tolerance parameter w^δ, whose value the experimenter chooses. We bound the number N_δ of trials expected to be required to observe a dominant work value. This bound depends on the thermal order-∞ Rényi entropy H_∞^β, a quantity inspired by one-shot information theory [9]. The bound can be estimated during an implementation of the "bidirectionality good practice" recommended in [2]. Finally, we approximately bound the error in a ΔF-estimate inferred from N_δ trials. A weakly interacting dilute classical gas [10] illustrates our analytical results.

Technical introduction—Let us introduce nonequilibrium fluctuation relations and the thermal order-∞ Rényi entropy H_∞^β.

Nonequilibrium fluctuation relations—Nonequilibrium fluctuation relations govern statistical mechanical systems arbitrarily far from equilibrium. Consider a system in thermal equilibrium with a heat bath at inverse temperature $\beta \equiv \frac{1}{k_B T}$, wherein k_B denotes Boltzmann's constant. We focus on classical systems for simplicity, though fluctuation relations have been extended to quantum systems [11]. Suppose that a time-dependent external parameter λ_t determines the system's Hamiltonian: $H = H(\lambda_t, \mathbf{z})$, wherein $\mathbf{z}$ denotes a phase-space point. If the system consists of an ideal gas in a box, λ_t may denote the height of the piston that caps the gas. Suppose that, at time $t = -\tau$, the system begins with the equilibrium phase-space density $e^{-\beta H(\lambda_{-\tau}, \mathbf{z})}/Z_{-\tau}$, wherein the partition function $Z_{-\tau}$ normalizes the state. The external parameter is then varied according to a predetermined schedule λ_t, from $t = -\tau$ to $t = \tau$. The system evolves away from equilibrium if τ is finite. In the gas example, the piston is lowered, compressing the gas. We call this process the *forward protocol*.

The *reverse protocol* begins with the system at equilibrium relative to $H(\lambda_\tau, \mathbf{z})$. The external parameter is changed to $\lambda_{-\tau}$ along the time-reverse of the path followed during the forward protocol. In the gas example, the piston is raised, and the gas expands.

Changing the external parameter requires or outputs some amount of work. We use the following sign convention: The forward process tends to require an investment of a positive amount $W > 0$ of work, and the reverse process tends to output $W > 0$. The value of W varies from trial to trial. After performing many trials, one can estimate the probability $P_{\mathrm{fwd}}(W)$ that any particular forward trial will cost an amount W of work and the probability $P_{\mathrm{rev}}(-W)$ that any particular reverse trial will output an amount W.

These probabilities satisfy *Crooks' Theorem* [12],

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

Here, $\Delta F := F_\tau - F_{-\tau}$ denotes the difference between the free energy $F_\tau = -\beta^{-1}\log(Z_\tau)$ of the Gibbs distribution $e^{-\beta H(\lambda_\tau,\mathbf{z})}/Z_\tau$ corresponding to the final Hamiltonian and the free energy $F_{-\tau} = -\beta^{-1}\log(Z_{-\tau})$ of the Gibbs distribution corresponding to $H(\lambda_{-\tau},\mathbf{z})$. Multiplying each side of Crooks' Theorem by $P_{\mathrm{rev}}(-W)e^{\beta\Delta F}$, then integrating over W, yields a version of the *nonequilibrium work relation* [13]:

$$e^{\beta\Delta F} = \langle e^{\beta W}\rangle_{\mathrm{rev}} \tag{2}$$

$$:= \int_{-\infty}^{\infty} dW\, e^{\beta W} P_{\mathrm{rev}}(-W). \tag{3}$$

The angle brackets denote an average over infinitely many trials. To calculate ΔF, one performs many trials, estimates the average, and substitutes into Eq. (2).

Thermal order-∞ Rényi entropy (H_∞^β)—Entropies quantify uncertainties in statistical mechanics and in information theory. Let $P := \{p_i\}$ denote a probability distribution over a discrete random variable X. The *Shannon entropy* $H_S(P) := -\sum_i p_i \log(p_i)$ quantifies an average, over infinitely many trials, of the information one gains upon learning the value assumed by X in one trial [14].

H_S has been generalized to a family of *Rényi entropies* H_α. The parameter $\alpha \in [0,\infty)$ is called the *order*. The H_α's quantify uncertainties related to finitely many trials. In the limit as $\alpha \to \infty$, H_α approaches

$$H_\infty(P) = -\log(p_{\max}), \tag{4}$$

wherein $p_{\max}$ denotes the greatest p_i. This maximal entropy has applications to randomness extraction: The efficiency with which finitely many copies of P can be converted into a uniformly random distribution $(\underbrace{\frac{1}{d},\ldots,\frac{1}{d}}_{d})$ is quantified with $H_\infty(P)$ [15].

The distributions P_{fwd} and P_{rev} in Crooks' Theorem are continuous. Hence we need a continuous analog of H_∞. The definition

$$H_\infty^\beta(P) := -\log(p_{\max}/\beta) \tag{5}$$

has been shown to be useful in contexts that involve heat baths [9]. $p_{\max}$ denotes the greatest value of the probability density P. $p_{\max}$ can diverge, e.g., if P represents a Dirac delta function. But delta functions characterize the work distributions of quasistatic protocols, whose work $W = \Delta F$ in every trial. We focus on more-realistic, quick protocols. P_{fwd} and P_{rev} are short and broad, so $p_{\max}$ is finite.

The density $p_{\max}$ has dimensions of inverse energy, which are canceled by the β in Eq. (5). Hence the logarithm's argument is dimensionless. For further discussion about H_∞^β, see [9].

Quantification of dominance—Let us return to the nonequilibrium work relation (3). The exponential enlarges already-high W-values, which dominate the integral. To estimate the integral accurately, one must perform trials that output large amounts of work. Few trials do; dominant W-values are *atypical* [1]. How many trials should one expect to need to perform, to achieve reasonable convergence of the exponential average in Eq. (3)?

An approximate answer was provided in [1]:

$$N \sim e^{\beta(\langle W\rangle_{\mathrm{fwd}} - \Delta F)}, \tag{6}$$

wherein $\langle . \rangle_{\mathrm{fwd}}$ denotes an average with respect to $P_{\mathrm{fwd}}(W)$. The *average dissipated work* $\langle W\rangle_{\mathrm{fwd}} - \Delta F$ represents the mean amount of work wasted as heat. Switching λ_t quasistatically (infinitely slowly) would cost an amount ΔF of work. Switching at a finite speed costs more: Work is dissipated into the bath as heat when the system is driven away from equilibrium. The dissipated work $W - \Delta F$ signifies the extra work paid to switch λ_t in a finite amount of time.

How large must a W-value be to qualify as dominant? This question remained open in [1]. We propose a definition inspired by information-theoretic protocols in which an agent specifies an error tolerance. The experimenter who switches λ_t, or the programmer who simulates trials, chooses a threshold value of w^δ used to lower-bound the W-values considered large.

Definition 1. *A work value W extracted from a reverse-protocol trial is called* dominant *if $W \geq w^\delta$ for the fixed value w^δ chosen by the agent.*

This definition enables us to bound the number N_δ of trials expected to be performed before one trial outputs a dominant amount of work.

Bound on expected number N_δ of trials required—Imagine implementing reverse trials until extracting a dominant amount of work from one trial. You might get lucky and extract $W \geq w^\delta$ on your first try. But you would not expect to. You would expect the number of trials to equal the inverse $1/\int_{w^\delta}^{\infty} dW\, P_{\mathrm{rev}}(W)$ of the probability that any particular reverse trial will output $W \geq w^\delta$. In the notation of [9], $\int_{w^\delta}^{\infty} dW\, P_{\mathrm{rev}}(W) = 1 - \delta$ (see Fig. 1):

$$N_\delta = \frac{1}{1-\delta}. \tag{7}$$

Let us clarify what "expect to perform N_δ trials" means. Imagine performing M sets of reverse trials. In

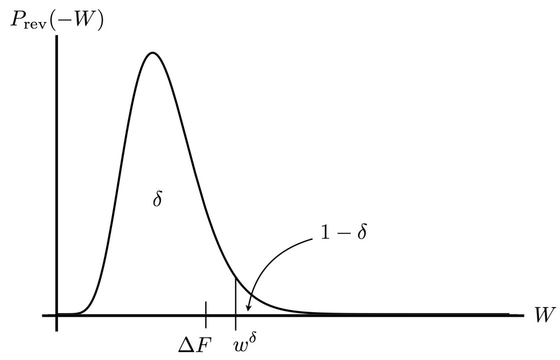

FIG. 1: Dominant values of work extractable from reverse-protocol trials: Large values W of work contribute the most to the integral in the nonequilibrium fluctuation relation (2). An amount W of extracted work is called "dominant" if it is at least as great as the threshold w^δ specified by the experimenter: $W \geq w^\delta$. The probability that any particular reverse trial will output a dominant amount of work is $\int_{w^\delta}^\infty dW \; P_{\rm rev}(-W) = 1-\delta$. This probability equals the area of the region under the distribution's right-hand tail.

each set, you perform trials until extracting $W \geq w^\delta$ from one trial. Let N_δ^i denote the number of trials performed during the $i^{\rm th}$ set. Consider averaging N_δ^i over the M sets of trials: $\frac{1}{M}\sum_{i=1}^M N_\delta^i$. As the number of sets grows large, the average of the number of required trials in a set approaches the "expected" value N_δ:

$$\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^M N_\delta^i = N_\delta. \tag{8}$$

This interpretation will facilitate our bounding of N_δ.

Theorem 1 (Bound on expected number of trials)**.** *The number N_δ of reverse trials expected to be performed before one trial outputs a dominant amount $W \geq w^\delta$ of work is bounded as*

$$N_\delta \geq e^{\beta(w^\delta - \Delta F) + H_\infty^\beta(P_{\rm fwd})}. \tag{9}$$

Proof. The inequality

$$w^\delta \leq \Delta F - \frac{1}{\beta}[H_\infty^\beta(P_{\rm fwd}) + \log(1-\delta)] \tag{10}$$

was derived in [9]. The derivation relies on the definitions of $1-\delta$ and H_∞^β, on Crooks' Theorem, and on the bound $P_{\rm fwd}(W) \leq p_{\rm max} \; \forall \, W$. Solving for $1-\delta$, then inverting the probability [Eq. (7)], yields Ineq. (9).

□

Inequality (9) implies that the bound on N_δ increases with w^δ, which makes sense. As we raise the threshold w^δ, fewer work values qualify as dominant. Hence more trials are expected to be required before a dominant work value is observed.

Inequality (9) resembles its inspiration, Relation (6), which states that the number N of trials required to achieve convergence of the average in Eq. (3) increases exponentially with the average dissipated work $\langle W \rangle_{\rm fwd} - \Delta F$. Similarly, the bound on N_δ increases exponentially with the "one-shot dissipated work" $w^\delta - \Delta F$. This $w^\delta - \Delta F$ represents the work sacrificed for time in a forward trial that costs an amount w^δ of work.

Moreover, N_δ is defined in terms of the reverse process. Yet the bound on N_δ given by Ineq. (9) depends on the forward work distribution, via $H_\infty^\beta(P_{\rm fwd})$. Similarly, in Relation (6), the number N of repetitions of the reverse process required for the convergence of Eq. (3) depends on the forward work distribution $P_{\rm fwd}(W)$, via $\langle W \rangle_{\rm fwd}$.

Despite its similarity to Relation (6), Ineq. (9) offers three advantages. First, Ineq. (9) quantifies dominance with δ, reflecting the agent's accuracy tolerance. Next, Relation (6) is a rough estimate. Inequality (9) is a strict bound on the number of trials expected to be performed before a dominant amount of work is extracted. Finally, Ineq. (9) contains an entropy that has no analog in Relation (6). The entropy tightens the bound when

$$p_{\rm max} < \beta. \tag{11}$$

This inequality is satisfied, for instance, in RNA-hairpin experiments used to test fluctuation theorems [16].

To appreciate these advantages over Relation (6), we can define dominant work values by choosing $w^\delta = \langle W \rangle_{\rm fwd}$, as in [1]. The bound becomes

$$N_\delta \geq e^{\beta(\langle W \rangle_{\rm fwd} - \Delta F) + H_\infty^\beta(P_{\rm fwd})}. \tag{12}$$

When $p_{\rm max} < \beta$ (such that $H_\infty^\beta > 1$), the number of trials required for Eq. (3) to converge exceeds the prediction in Relation (6).

We can gain further insight by rewriting Ineq. (9) as

$$N_\delta \geq \frac{\beta}{p_{\rm max}} e^{\beta(\langle W \rangle_{\rm fwd} - \Delta F)}, \tag{13}$$

using the definition of H_∞^β [Eq. (5)]. The fraction $\beta/p_{\rm max}$ represents approximately the number of forward trials performed before one trial's W-value falls within a width-($k_{\rm B}T$) window about the most probable work value $W_{\rm max}$: $W \in [W_{\rm max} - \frac{k_{\rm B}T}{2}, W_{\rm max} + \frac{k_{\rm B}T}{2}]$. The value of $\beta/p_{\rm max}$ generically increases with the width of the distribution $P_{\rm fwd}(W)$. Hence the bound on N_δ, as written in Ineq. (13), is a product of two factors. The first depends on the forward work distribution's width; and the second, on its mean. In contrast, Relation (6) depends only on the mean.

Evaluating the N_δ bound—Not only does Ineq. (9) have a theoretically satisfying form, but it can also be estimated in practice. We will discuss how to estimate the $H_\infty^\beta(P_{\rm fwd})$ and the ΔF in the bound. The bound can be estimated reasonably, we argue, from not too many trials.

The experimental set-up determines β, and the agent chooses w^δ. $H_\infty^\beta(P_{\rm fwd})$ and ΔF can be estimated if one

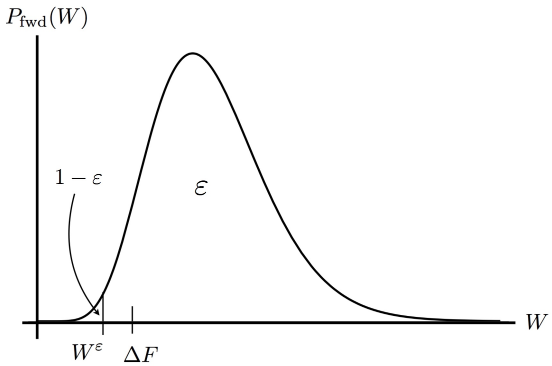

FIG. 2: Dominant values of work invested in forward-protocol trials: Small values W of work dominate the nonequilibrium work relation (14). An amount W of invested work is called "dominant" if it lies below or on the threshold W^ε chosen by the experimenter: $W \leq W^\varepsilon$. The probability that any particular forward trial will require a dominant amount of work is $1-\varepsilon = \int_{-\infty}^{W^\varepsilon} dW \, P_{\rm fwd}(W)$. This probability equals the area under the distribution's left-hand tail.

implements the "good practice" of bidirectionality. To mitigate errors in (ΔF)-estimates, one should perform forward trials, perform reverse trials, and combine all the data [2]. Upon performing several forward trials, one can estimate $H_\infty^\beta(P_{\rm fwd})$ and ΔF. One can estimate the N_δ bound, then perform (probably at least N_δ) reverse trials until observing a dominant work value, and improve the (ΔF)-estimate.[1]

H_∞^β depends on $p_{\rm max}$, the greatest probability (per unit energy) of any possible forward-trial outcome. This outcome will likely appear in many trials. Hence one expects to estimate H_∞^β well from finitely many forward trials.

Forward-protocol bound—Just as we bounded the number N_δ of reverse trials, we can bound the number of forward trials expected to be performed before a dominant amount of work is invested. The analysis is analogous to that of N_δ.

The nonequilibrium work relation for the forward process is

$$\langle e^{-\beta W} \rangle_{\rm fwd} = e^{-\beta \Delta F}. \tag{14}$$

The forward trials that dominate the average in Eq. (14) cost unusually small amounts of work. In the notation of [9], dominant work values satisfy $W \leq W^\varepsilon$, for a tolerance W^ε chosen by the agent. Each forward trial has a probability $1-\varepsilon$ of costing a dominant amount of work (see Fig. 2). Theorem 4 of [9] bounds W^ε in terms of $1-\varepsilon$. Solving for $1-\varepsilon$, then inverting, bounds the number $N_\varepsilon = 1/(1-\varepsilon)$ of forward trials expected to be performed before any trial costs a dominant amount of work:

$$N_\varepsilon \geq e^{-\beta(W^\varepsilon - \Delta F) + H_\infty^\beta(P_{\rm rev})}. \tag{15}$$

Error estimate: Calculating the error in a (ΔF)-estimate is crucial but difficult. Whenever one infers a value from data, the inference's reliability must be reported. Common error analyses do not suit estimates of (ΔF)-values, for two reasons. First, ΔF depends on the random variable W logarithmically [see Eq. (2)]. Second, W tends not to be Gaussian. Approaches such as an uncontrolled approximation, in the form of a truncation of a series expansion, have been proposed [2]. Our approach centers on the agent's choice of w^δ.

Consider choosing a w^δ-value and performing N_δ trials. With what accuracy can you estimate ΔF? We will bound the percent error

$$\epsilon := \left| \frac{\Delta F - (\Delta F)_{\rm est}}{\Delta F} \right| \tag{16}$$

roughly. To render the problem tractable, we assume that one knows the exact form of $P_{\rm rev}(-W)$ for all $W \leq w^\delta$.

Theorem 2 (Approximate error bound)**.** *Let the work tolerance be $w^\delta \in (-\infty, \infty)$. Let $(\Delta F)_{\rm est}$ denote the estimate of the free-energy difference ΔF inferred from data taken during N_δ trials. If $(\Delta F)_{\rm est}$ is calculated from the exact form of $P_{\rm rev}(-W) \quad \forall \, W \leq w^\delta$, the estimate has a percent error of*

$$\epsilon \leq \frac{1}{\beta(\Delta F)} \Big[\eta + O(\eta^2) \Big], \tag{17}$$

wherein

$$\eta := \frac{e^{\beta w^\delta}}{N_\delta \langle e^{\beta W} \rangle_{\rm rev}}. \tag{18}$$

Proof. Let us solve the nonequilibrium work relation (2) for ΔF:

$$\Delta F = \frac{1}{\beta} \log \Big(\langle e^{\beta W} \rangle_{\rm rev} \Big) \tag{19}$$

$$= \frac{1}{\beta} \log \left(\int_{-\infty}^{\infty} dW \, e^{\beta W} P_{\rm rev}(-W) \right). \tag{20}$$

The estimate has a similar form:

$$(\Delta F)_{\rm est} = \frac{1}{\beta} \log \left(\int_{-\infty}^{w^\delta} dW \, e^{\beta W} P_{\rm rev}(-W) \right) \tag{21}$$

$$= \frac{1}{\beta} \log \Bigg(\int_{-\infty}^{\infty} dW \, e^{\beta W} P_{\rm rev}(-W) - \int_{w^\delta}^{\infty} dW \, e^{\beta W} P_{\rm rev}(-W) \Bigg). \tag{22}$$

[1] N_δ can be estimated from reverse trials alone, less reliably. One could perform a few reverse trials, estimate $P_{\rm rev}(-W)$, and estimate ΔF. From these estimates and from Crooks' Theorem, one could estimate $P_{\rm fwd}(W)$. From $P_{\rm fwd}(W)$, one could estimate $H_\infty^\beta(P_{\rm fwd})$, then estimate the N_δ bound. One could repeat this process, improving one's estimate of the bound, until observing a dominant work value. But the estimate of ΔF is expected to jump repeatedly [1]. This sawtooth behavior, as well as the piling of estimate upon estimate, may taint the estimates of the bound.

We replace the first integral with $\langle e^{\beta W}\rangle_{\text{rev}}$, using Eq. (19). The second term, representing the error, is expected to be much smaller than the first term. This second term will serve as a small parameter in a Taylor expansion:

$$(\Delta F)_{\text{est}} = \frac{1}{\beta}\left[\log\left(\langle e^{\beta W}\rangle_{\text{rev}}\right) + \log\left(1 - \frac{\int_{w^\delta}^{\infty} dW\, e^{\beta W} P_{\text{rev}}(-W)}{\langle e^{\beta W}\rangle_{\text{rev}}}\right)\right] \quad (23)$$

$$= \Delta F - \frac{1}{\beta}\Big[\eta' + O(\,[\eta']^2\,)\Big\}, \quad (24)$$

wherein

$$\eta' := \frac{\int_{w^\delta}^{\infty} dW\, e^{\beta W} P_{\text{rev}}(-W)}{\langle e^{\beta W}\rangle_{\text{rev}}}. \quad (25)$$

We can bound the numerator, using Fig. 1:

$$\int_{w^\delta}^{\infty} dW\, e^{\beta W} P_{\text{rev}}(-W) \quad (26)$$

$$\leq e^{\beta w^\delta} \int_{w^\delta}^{\infty} dW\, P_{\text{rev}}(-W) \quad (27)$$

$$= e^{\beta w^\delta}(1-\delta) = \frac{e^{\beta w^\delta}}{N_\delta}. \quad (28)$$

Substituting into Eq. (25) yields $\eta' \leq \eta$. Hence Eq. (24) reduces to

$$(\Delta F)_{\text{est}} \geq \Delta F - \frac{1}{\beta}\Big[\eta + O(\,\eta^2\,)\Big]. \quad (29)$$

Substituting into the percent error's definition [Eq. (16)] yields Ineq. (17). $\square$

The approximate error bound can be estimated from agent-chosen parameters and from data: The experiment's set-up determines the value of β. The agent chooses the value of w^δ. For N_δ, one can substitute the number of trials performed [or can substitute from Ineq. (9)]. ΔF and $\langle e^{\beta W}\rangle_{\text{rev}}$ can be estimated from data.

Numerical experiments—To illustrate our analytical results, we considered the weakly interacting dilute classical gas. This system's forward and reverse work distributions can be calculated exactly [10]. The gas begins in equilibrium with a heat bath at inverse temperature $\beta \equiv \frac{1}{k_B T}$. During the forward protocol, the gas is isolated from the bath at $t = -\tau$. The gas is quasistatically compressed, its temperature rising from T. During the reverse protocol, the gas expands and cools. When discussing either direction, we denote the initial volume by V_0 and the final volume by V_1.

The probability densities over the possible work values were calculated in [10]:

$$P(W) = \frac{\beta}{|\alpha|\Gamma(k)}\left(\frac{\beta W}{\alpha}\right)^{k-1} e^{-\beta W/\alpha}\,\theta(\alpha W). \quad (30)$$

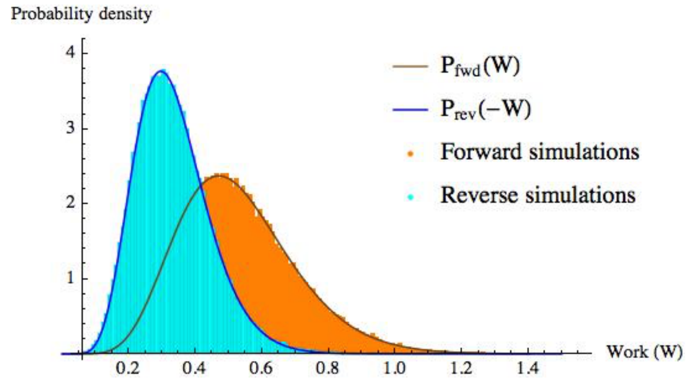

FIG. 3: Probability densities and numerical data for a weakly interacting dilute gas: We considered a gas undergoing compression (a forward protocol) and expansion (a reverse protocol). The probability per unit energy that any particular trial will involve an amount W of work [Eq. (30)] was calculated in [10]. The short, right-hand, brown curve represents $P_{\text{fwd}}(W)$; and the tall, left-hand, dark-blue curve represents $P_{\text{rev}}(-W)$. By sampling work values from these distributions, we effectively simulated each protocol 10^5 times. The cyan bars (under the left-hand curve) depict the data gathered from the forward-protocol samples; and the orange bars (under the right-hand curve) depict the data from the reverse-protocol samples.

During the forward protocol, $\alpha := (V_0/V_1)^{2/3} - 1 > 0$; during the reverse, $\alpha < 0$. The gamma function is denoted by $\Gamma(k)$; and its argument, by $k := \frac{3}{2}n$, wherein n denotes the number of particles. The theta function $\theta(\alpha W)$ ensures that $W \geq 0$ is invested in forward trials (for which $P = P_{\text{fwd}}$); and $W \leq 0$, in reverse trials (for which $P = P_{\text{rev}}$).

We sampled 10^5 values of W from the forward (compression) work distribution and 10^5 values from the reverse (expansion) work distribution. Figure 3 shows the probability densities and the sampled data. We chose $V_0/V_1 = 2$ and $n = 6$, following [10], and $\beta = 10$. Dividing a histogram of the forward-protocol data into 50 bins yielded $p_{\max} = 1.577$. Satisfying Ineq. (11), this $p_{\max}$ enables $H_\infty^\beta(P_{\text{fwd}})$ to tighten the N_δ bound.

Figure 4 illustrates our results. Possible values of w^δ appear along the abscissa. The blue curve shows the N_δ bound, calculated from forward-trial samples, in Theorem 1. The red curve, calculated from reverse-trial samples, shows after how many reverse trials (N_{true}) $W \geq w^\delta$ was extracted during one trial. N_{true} has a jagged, step-like shape, as one might expect.

The green curve depicts the estimate, in [1], of the number of reverse trials expected to be performed before one trial outputs a dominant work value, for an unspecified meaning of "dominant." We calculated $N_{\text{est}} = 3$ by simulating forward trials, calculating the average dissipated work, and substituting into Relation (6).

The curves' shapes and locations illustrate the N_δ bound's advantages. The bound (the blue curve) hugs the actual number N_{true} of trials required (red) more

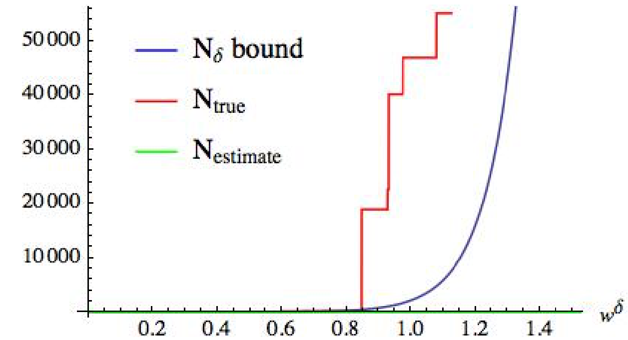

FIG. 4: Three number-of-required-trial measures: The abscissa shows possible choices of the threshold w^{δ} for dominant work values. The blue curve, calculated from 10^5 forward-trial samples, represents the bound on the number N_{δ} of reverse trials expected to be performed before any trial outputs a dominant amount $W \geq w^{\delta}$ of work (Theorem 1). The red curve, calculated from 10^5 reverse-trial samples, depicts the actual number N_{true} of trials performed before $W \geq w^{\delta}$ is extracted. The green curve, calculated from forward-trial samples, represents Relation (6): an estimate N_{est} of the number of trials required to extract a dominant amount of work, wherein the meaning of "dominant" is unspecified. The blue curve follows the red curve's shape more faithfully than the green does, illustrating the precision of Theorem 1. As expected, the blue curve lower-bounds the red at most w^{δ}-values.

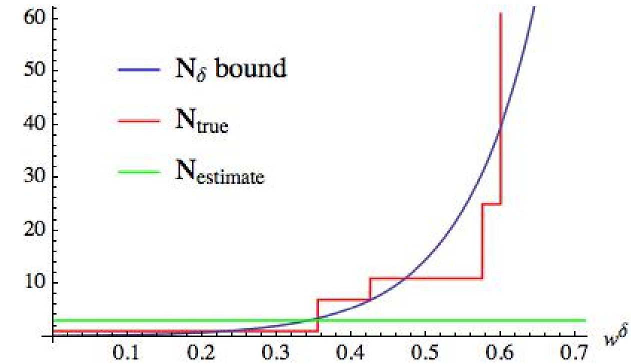

FIG. 5: Three number-of-required-trial measures at low threshold work values w^{δ}: At most threshold values w^{δ}, the N_{δ} bound (blue) lower-bounds the actual number N_{true} (red) of reverse trials performed before any trial outputs a dominant amount $W \geq w^{\delta}$ of work. At low w^{δ}-values, the red curve zigzags across the blue. This zigzagging stems from the technical definition of N_{δ}.

closely than N_{est} (green) does. N_{est} remains flat, whereas the N_{δ} bound rises as N_{true} rises. The N_{δ} bound often lower-bounds N_{true}, as expected. When w^{δ} is small, the N_{δ} bound weaves above and below N_{true}, as shown in Fig. 5. The reason was explained above Theorem 1: N_{δ} denotes the number of trials *expected*, in a sense defined by probability and frequency, to be required. One might get lucky and extract $W \geq w^{\delta}$ before performing N_{δ} trials. The dropping of the N_{true} curve below the N_{δ} bound represents such luck. But one expects to perform N_{δ} trials, and the N_{δ} bound lower-bounds N_{true} for most w^{δ}-values.

Conclusions—We have sharpened predictions about the number of experimental trials required to estimate ΔF from fluctuation relations. We improved the approximation in [1] to an inequality, tightened the bound (in scenarios of interest) with an entropy H_{∞}^{β}, freed the experimenter to choose a tolerance w^{δ} for dominance, and approximately bounded the error in an estimate of ΔF. How to choose w^{δ} merits further investigation. We wish to be able to specify the greatest error ϵ acceptable in an estimate of ΔF. From ϵ, we wish to infer the number N^{ϵ} of trials we should expect to perform. This entire investigation improves the rigor with which free-energy differences ΔF can be estimated from experimental and numerical-simulation data.

Acknowledgements—NYH thanks Yi-Kai Liu for conversations about error probability and thanks Alexey Gorshkov for hospitality at NIST. Part of this research was conducted while NYH was visiting the JQI, QuICS, and the UMD Department of Chemistry and Biochemistry. NYH was supported by an IQIM Fellowship and NSF grant PHY-0803371. The Institute for Quantum Information and Matter (IQIM) is an NSF Physics Frontiers Center supported by the Gordon and Betty Moore Foundation. CJ was supported by NSF grant DMR-1506969.

[1] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[2] A. Pohorille, C. Jarzynski, and C. Chipot, The Journal of Physical Chemistry B **114**, 10235 (2010), http://dx.doi.org/10.1021/jp102971x, PMID: 20701361.
[3] C. M. Rohwer, F. Angeletti, and H. Touchette, ArXiv e-prints (2014), 1409.8531.
[4] D. Wu and D. A. Kofke, The Journal of Chemical Physics **123**, (2005).
[5] D. Wu and D. A. Kofke, The Journal of Chemical Physics **123**, (2005).
[6] N. Lu and D. A. Kofke, The Journal of Chemical Physics **111**, 4414 (1999).
[7] J. Gore, F. Ritort, and C. Bustamante, **100**, 12564 (2003).
[8] A. M. Hahn and H. Then, Phys. Rev. E **80**, 031111 (2009).
[9] N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, New Journal of Physics **17**, 095003 (2015), 1409.3878.
[10] G. E. Crooks and C. Jarzynski, Phys. Rev. E **75**, 021116 (2007).
[11] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
[12] G. E. Crooks, Physical Review E **60**, 2721 (1999).
[13] C. Jarzynski, Physical Review Letters **78**, 2690 (1997).
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
[15] R. Renner and S. Wolf, Smooth Rényi entropy and applications, in *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, pp. 232–232, IEEE, 2004.
[16] D. Collin *et al.*, Nature **437**, 231 (2005).