Supplemental material

This Supplemental Material contains some additional details relating to the calculations presented in the main text and is organized in the following way. We briefly introduce the relationship between the characteristic function and moments of the work distribution in Sec. A. Then, in Sec. B, we give details regarding how estimates of the work distributions are obtained from estimates of the characteristic function, together with properties of these estimates. Section C concerns how these estimates are used to infer the temperature, in both a frequentist and Bayesian approach. Then finally in Sec. D we explain how the characteristic function is calculated for the specific case of the Bose-Hubbard model, using Bogoliubov theory and tensor network theory.

Appendix A: Properties of the work distribution

In what follows we will refer to a few generic properties of the work distribution $P_Q(W)$, namely its cumulants κ_m and principally its first $\kappa_{Q1} = \mu_Q$ and second $\kappa_{Q2} = \sigma_Q^2$ cumulants. The cumulants κ_{Qm} of the work distribution $P_Q(W)$ are related to the derivatives of the logarithm of the corresponding characteristic function $\chi_Q(u)$, according to

$$\kappa_{Qm} = \frac{1}{\mathrm{i}^m} \frac{\mathrm{d}^m}{\mathrm{d}u^m} \log \chi_Q(u) \Big|_{u=0}. \tag{A1}$$

We note two properties of the work distribution $P_Q(W)$, one relevant to each of μ_Q and σ_Q^2. First, the second law of thermodynamics expressed in terms of mean work and free energy ensures that $\mu_F \geq \Delta F \leq -\mu_B$, where $\Delta F = F(\lambda_f) - F(\lambda_i)$ is the free energy difference. Second, though this is just a re-expression of Eq. (A1) for $m = 2$, the second cumulant is related to the so-called dephasing time $\tau_{Q\mathrm{deph}}$ by $\sigma_Q = 1/\tau_{Q\mathrm{deph}}$, where $\tau_{Q\mathrm{deph}}$ is here defined by the short-time behavior of the dephasing function $|\chi_Q(u)| = \exp(-\Gamma_Q(u))$, with $\Gamma_Q(u) = (u/\tau_{Q\mathrm{deph}})^2/2 + \mathcal{O}(u^3)$. This gives us a way of assessing the size of the second cumulant σ_Q^2 from observing a small part of $\chi_Q(u)$ directly. We assume that this is done, at a small overhead to our thermometry protocol discussed in the next section, and thus σ_Q and $\tau_{Q\mathrm{deph}}$ are quantities that are, at least approximately, known when implementing the protocol.

Appendix B: Estimating the work distribution

1. Deterministic errors

Here we discuss how to estimate the work distribution

$$P_Q(W) = \mathcal{F}\{\chi_Q(u)\}(W) = \frac{1}{2\pi} \int \mathrm{d}u \mathrm{e}^{-\mathrm{i}Wu} \chi_Q(u),$$

from estimates of the characteristic function $\chi_Q(u)$ obtained in a (numerical or actual) experiment, where $\mathcal{F}$ denotes a Fourier transform.

It is infeasible to study the characteristic function $\chi_Q(u)$ continuously over all times. More realistically the characteristic function $\chi_Q(u_j)$ is studied at a finite number N_{steps} of discrete times $u_j = j\Delta u$ for integer $j = 1, \ldots, N_{\mathrm{steps}}$ in some domain $[-T, T]$ with $T = N_{\mathrm{steps}}\Delta u$. Noting that $\chi_Q(0) = 1$ and $\chi_Q(u_j) = \chi_Q^*(-u_j)$, we then, instead of $P_Q(W)$, construct a discrete and finite-time Fourier transform

$$\begin{aligned} p_Q(W) =& \frac{\Delta u}{2\pi} \sum_{j=-N_{\mathrm{steps}}}^{N_{\mathrm{steps}}} \mathrm{e}^{-\mathrm{i}Wu_j} \chi_Q(u_j) w(u_j) \\ =& \frac{\Delta u}{2\pi} \left(1 + 2\Re \left\{ \sum_{j=1}^{N_{\mathrm{steps}}} \mathrm{e}^{-\mathrm{i}Wu_j} \chi_Q(u_j) w(u_j) \right\} \right), \end{aligned}$$

where we have introduced a windowing function $w(u)$.

As we will now discuss, $p_Q(W)$ differs significantly from $P_Q(W)$. However, the forward and backward distributions only enter into our thermometry protocol through their ratio. We show below that the ratios $p_F(W)/p_B(-W)$ and $R(W) = P_F(W)/P_B(-W)$ can be made identical, meaning we are able to work with the alternative distributions $p_Q(W)$ rather than $P_Q(W)$. For the rest of this subsection, we consider how to choose T and N_{steps} for our protocol such that this is the case.

The two Fourier transforms $p_Q(W)$ and $P_Q(W)$ are related, using the Poisson summation formula and the convolution theorem, by

$$p_Q(W) = \left\{ \mathcal{F}\{w(u)\}(W) \star \sum_{k=-\infty}^{\infty} P_Q(W + k\Delta W) \right\}(W),$$

where $\Delta W = 2\pi/\Delta u$ and $\star$ represents a convolution. This demonstrates the two sources of error in going from $P_Q(W)$ to $p_Q(W)$. First, aliasing, where frequencies differing by ΔW cannot be distinguished by looking at a function at a discretized set of points. Second, spectral leakage, where contributions from one frequency leak to those nearby on a scale π/T due to the finite resolution offered by the window of finite size T. The former leads to the sum and the latter to the convolution.

For the effects of aliasing to be small, we must have that $\sigma_Q/\Delta W \ll 1$ or $\sigma_Q \Delta u \ll 2\pi$ so that the widths σ_Q of the work distributions is much smaller than the periodicity of its approximation $p_Q(W)$. Another way to write this, in terms of the dephasing time $\tau_{Q\text{deph}} = 1/\sigma_Q$ is $\Delta u/\tau_{Q\text{deph}} \ll 2\pi$ or $N_{\text{steps}} \gg T/\tau_{Q\text{deph}}$. In all examples used in our work, the number of time-steps N_{steps} is easily large enough to ensure the effects of aliasing are negligible.

The effects of spectral leakage on $p_Q(W)$ will always be significant for a finite system with a discrete set of energy levels. Importantly for our protocol, the same is not true for their ratios, as we now show. Consider aliasing to have a negligible effect, then

$$\begin{aligned}\frac{p_F(W)}{p_B(-W)} =& \frac{\{\mathcal{F}\{w(u)\}(W) \star P_F(W)\}(W)}{\{\mathcal{F}\{w(u)\}(W) \star P_B(-W)\}(W)} \\ =& \frac{\{\mathcal{F}\{w(u)\}(W) \star P_B(-W)R(W)\}(W)}{\{\mathcal{F}\{w(u)\}(W) \star P_B(-W)\}(W)} \\ \approx& \frac{\{\mathcal{F}\{w(u)\}(W) \star P_B(-W)\}(W)R(W)}{\{\mathcal{F}\{w(u)\}(W) \star P_B(-W)\}(W)} \\ =& R(W).\end{aligned}$$

In going from the second to the third line we have only used that $R(W) = \mathrm{e}^{\beta(W-\Delta F)}$ does not vary too much on the scale of the characteristic width $\Delta W_{\mathcal{F}w} \approx \pi/T$ of the smoothing kernel $\mathcal{F}\{w(u)\}(W)$, i.e., $\beta \Delta W_{\mathcal{F}w} \ll 1$ or $T \gg \pi\beta$. However, even this criterion is sometimes overly strict and may be loosened depending on the relationship between σ_Q, β and $\Delta W_{\mathcal{F}w}$. For example, consider the case that both $p_Q(W)$ are effectively flat on scale $\Delta W_{\mathcal{F}w}$ i.e. $\sigma_Q/\Delta W_{\mathcal{F}w} \gg 1$ assuming $p_Q(W)$ to be unimodal. Then even having $\beta \Delta W_{\mathcal{F}w} \approx 1$ would lead only to small errors, as errors due to variations in $R(W)$ within the convolution would largely cancel. Note that if an incorrect choice is made and spectral leakage does introduce errors, then this will be clear from the non-linear behavior of $L(W) = \ln(R(W))$ and thus a larger T can be chosen.

In this work, we find that a good rule of thumb is to choose T to be sufficiently long that the qubit has fully dephased. Specifically, for the superfluid calculations, we use a phenomenological and unoptimized expression based on the above discussion

$$T = \pi\beta[1 + (5-1)\mathrm{e}^{-\sigma\beta}], \tag{B1}$$

with $\sigma = \sigma_F = \sigma_B$. The effect is that spectral leakage, like aliasing, has a negligible effect on errors. In a real experiment, β is not known in advance and so T must be chosen using the considerations above and some prior expectations about β. We find that the thermometry protocol is typically robust to changes in T by factors of the order of unity.

2. Random errors

From now on, we assume that T and N_{steps} have been chosen, such that $R(W) = p_F(W)/p_B(-W)$. We focus on the fact that $\chi_Q(u_j)$ will not be known exactly and that instead we only have access to an estimator $\bar{\chi}_Q(u)$ based on expectation values estimated from a finite number N_{meas} of measurements each. We will always use a bar to indicate an estimate that is a random variable obtained stochastically from measurements made during the protocol. Propagating this forward according to

$$\bar{p}_Q(W) = \frac{\Delta u}{2\pi}\left(1 + 2\Re\left\{\sum_{j=1}^{N_{\text{steps}}} \mathrm{e}^{-\mathrm{i}Wu_j}\bar{\chi}_Q(u_j)w(u_j)\right\}\right), \tag{B2}$$

we obtain an estimate $\bar{p}_Q(W)$ of $p_Q(W)$. The remainder of this subsection addresses how to choose a set of work values W_k such that $\bar{p}_Q(W_k)$ are independent, unbiased $\mathrm{E}[\bar{p}_Q(W_k)] = p_Q(W_k)$, and of known variance $\mathrm{Var}[\bar{p}_Q(W_k)]$. Here expectation values are always taken with respect to the distributions generating the measurement outcomes.

Let us begin by considering how $\chi_Q(u_j)$ are estimated. In an experiment we estimate

$$\chi_Q(u) = \frac{\mathrm{e}^{\mathrm{i}\phi(u)}}{2s_\uparrow^* s_\downarrow}\left(\langle\hat{\sigma}_x\rangle + \mathrm{i}\langle\hat{\sigma}_y\rangle\right),$$

with $\phi(u) = \Delta(\tau + u)$ by first estimating expectation values $\langle\hat{\sigma}_\mu\rangle$, for $\mu = x, y$, with respect to the state $\hat{\rho}_q$ of the qubit at the end of the interferometric protocol. Specifically, each $\langle\hat{\sigma}_\mu\rangle$ is estimated from N_{meas} independent measurements of σ_μ, which returns 1 or -1 with probability $p = (1 + \langle\hat{\sigma}_\mu\rangle)/2$ and $1 - p$, respectively. Then the average $\bar{\sigma}_\mu$ of the measurements is an estimator of $\langle\hat{\sigma}_\mu\rangle$ that is unbiased, i.e., its mean is $\mathrm{E}[\bar{\sigma}_\mu] = \langle\hat{\sigma}_\mu\rangle$, and has variance $\mathrm{Var}[\bar{\sigma}_\mu] = (1 - \langle\hat{\sigma}_\mu\rangle^2)/N_{\text{meas}}$. In accordance with the central limit theorem, for large enough N_{meas} the estimator $\bar{\sigma}_\mu$ is normal and so its properties are fully characterized by its mean and variance.

The linear combination

$$\bar{\chi}_Q(u) = \frac{\mathrm{e}^{\mathrm{i}\phi(u)}}{2s_\uparrow^* s_\downarrow}\left(\bar{\sigma}_x + \mathrm{i}\bar{\sigma}_y\right),$$

is thus also an unbiased ($\mathrm{E}[\bar{\chi}_Q(u)] = \chi_Q(u)$) Gaussian estimator of $\chi_Q(u)$ and has variance (using the generalized definition $\mathrm{Var}[\bar{z}] = \mathrm{E}\left[|\bar{z} - \mathrm{E}[\bar{z}]|^2\right]$ of variance for a complex random variable $\bar{z}$)

$$\begin{aligned}\mathrm{Var}[\bar{\chi}_Q(u)] =& \frac{\mathrm{Var}[\bar{\sigma}_x] + \mathrm{Var}[\bar{\sigma}_y]}{4|s_\uparrow^* s_\downarrow|^2} \\ =& \frac{2 - 4|s_\uparrow^* s_\downarrow|^2|\chi_Q(u)|^2}{4|s_\uparrow^* s_\downarrow|^2 N_{\text{meas}}},\end{aligned} \tag{B3}$$

where the first line holds due to the independence of $\bar{\sigma}_x$ and $\bar{\sigma}_y$.

Note that this variance $\mathrm{Var}[\bar{\chi}_Q(u)] = \mathrm{Var}[\Re\{\bar{\chi}_Q(u)\}] + \mathrm{Var}[\Im\{\bar{\chi}_Q(u)\}]$ is divided between those of the real and complex parts, but how exactly this division occurs depends on the phase $\phi(u) = \Delta(\tau + u)$, and thus Δ, and the phase of $s_\uparrow^* s_\downarrow$. The effect of Δ is quantitative but not qualitative, and the Δ for a particular implementation can be known. For the purposes of our plots, we assume $\phi(u) = \Delta = 0$. Further, Eq. (B3) makes it obvious that $s_\downarrow^*$ and $s_\uparrow^*$ should be chosen such that $|s_\uparrow^* s_\downarrow|^2$ takes its maximum possible value, $1/4$, and in all plots we assume the choice $s_\uparrow^* s_\downarrow = 1/2$ is made.

This variance propagates forward into our estimator $\bar{p}_Q(W)$ of $p_Q(W)$ according to Eq. (B2). Since this again a linear sum, the estimators are unbiased and Gaussian, characterized by the expectations $\mathrm{E}[\bar{p}_Q(W)] = p_Q(W)$ and covariances

$$\begin{aligned}
&\mathrm{E}\big[\,(\bar{p}_Q(W) - p_Q(W))\,(\bar{p}_Q(W') - p_Q(W'))\,\big] \\
&= \frac{2}{4|s_\uparrow^* s_\downarrow|^2 N_{\mathrm{meas}}} \left(\frac{\Delta u}{2\pi}\right)^2 \sum_j |w(u_j)|^2 \cos\left((W - W')u_j\right)\left(2 - 4|s_\uparrow^* s_\downarrow|^2 |\chi_Q(u_j)|^2\right) \\
&\quad - \frac{2}{4(s_\uparrow^* s_\downarrow)^2 N_{\mathrm{meas}}} \left(\frac{\Delta u}{2\pi}\right)^2 \sum_j w^2(u_j) \Re\left\{\mathrm{e}^{-\mathrm{i}(W+W')u_j} \mathrm{e}^{2\mathrm{i}\phi(u_j)} \Re\left\{\mathrm{e}^{-2\mathrm{i}\phi(u_j)} 4(s_\uparrow^* s_\downarrow)^2 \chi_Q^2(u_j)\right\}\right\}.
\end{aligned} \tag{B4}$$

The first term arises from the interference of random deviations in $\chi_Q(u_j)$ at different u_j. The second term, which is much smaller, arises due to the fact that, to ensure that $\bar{p}_Q(W)$ is real, we have used $\chi_Q(-u_j) = \chi_Q^*(u_j)$ rather than generate independent estimators for $\chi_Q(u_j)$ and $\chi_Q(-u_j)$.

When W and W' are close, approximating $W = W'$, we obtain the variance. The dominant term is given by

$$\begin{aligned}
\mathrm{Var}\left[\bar{p}_Q(W)\right] \approx &\frac{2}{4|s_\uparrow^* s_\downarrow|^2 N_{\mathrm{meas}}} \left(\frac{\Delta u}{2\pi}\right)^2 \sum_j |w(u_j)|^2 \\
&\times \left(2 - 4|s_\uparrow^* s_\downarrow|^2 |\chi_Q(u_j)|^2\right).
\end{aligned} \tag{B5}$$

Notably, this scales as $\mathrm{Var}\left[\bar{p}_Q(W)\right] \sim T^2/N_{\mathrm{steps}} N_{\mathrm{meas}}$. Compare this to the conditions $T/\tau_{Q\mathrm{deph}} \ll 2\pi N_{\mathrm{steps}}$ and $T \gtrsim \pi\beta$ for avoiding systematic errors due to aliasing and spectral leakage. The form of $\mathrm{Var}\left[\bar{p}_Q(W)\right]$ suggests increasing N_{steps}, which would also reduce the aliasing error. However, $\mathrm{Var}\left[\bar{p}_Q(W)\right]$ suggests taking a smaller window size T, which would come at the cost of higher spectral leakage. We leave a discussion of the trade-off of these two errors until later.

When W and W' are separated by more than roughly π/T, the cosine term in Eq. (B4) will oscillate rapidly enough that the covariance is significantly reduced. Thus π/T represents the range in W over which our estimates $\bar{p}_Q(W)$ are correlated. Thus there is little information added by generating estimates $\bar{p}_Q(W)$ for W separated by less than this correlation range π/T. As a result, in our inference protocol we consider only the values $\bar{p}_Q(W_k)$ taken at an array of points $W_k = k\pi/T$ separated by this amount. These values should only be weakly-correlated. Conveniently, but not crucially, these are exactly the same values W_k for which $\bar{p}_Q(W_k)$ can be found using the fast Fourier transform. This discussion suggests increasing T in order to increase the density of points W_k. Again, a discussion of the trade-off of this with the other factors affecting the choice of T is left until later.

Appendix C: Inferring the temperature

Here we detail the core of the thermometry protocol, giving the precise procedure to go from the estimates of the work distributions $p_Q(W_k)$ discussed in the previous section to an estimate of the inverse temperature β. Note that the protocol uses no knowledge specific about the distributions $P_Q(W)$ and characteristic functions $\chi_Q(u)$ other than the width σ_Q and dephasing time $\tau_{Q\mathrm{deph}}$ described above.

1. Outline

The essential fact on which we base our inference is that

$$L(W) = \ln(R(W)) = \ln\left\{\frac{p_F(W)}{p_B(-W)}\right\} = \beta(W - \Delta F), \tag{C1}$$

where we refer to the work distributions $p_Q(W)$, ratio $R(W)$ and log-ratio $L(W)$ discussed in the previous section.

The information we collect during our protocol is a set of near-independent estimates $\bar{p}_Q(W_k)$ of $p_Q(W_k)$ and their variances at a discrete set of energies W_k. The essential idea is to use these estimated values of the work distribution, together with our knowledge of how they

relate to β and ΔF via Eq. (C1), to infer the values, or distribution of values, of β (and ΔF if desired).

We consider two approaches to this inference problem, frequentist and Bayesian. The Bayesian approach is used for all results we present in the main text, but methods along the lines of the frequentist approach are more commonly found in the literature. Our presentation of the frequentist approach thus serves to highlight the benefits our Bayesian approach, and also provides a simpler setting in which to analyze the dependence of errors on parameters e.g. T used in the protocol.

a. *Frequentist*

The frequentist approach is most similar to how ΔF has previously been estimated from work distributions. The approach is to use $\bar{p}_Q(W_k)$ to obtain estimates $\bar{L}(W_k)$ of $L(W_k)$ for many values of work W_k. For now, let's assume, these estimates $\bar{L}(W_k)$ are independent, unbiased $\mathrm{E}[\bar{L}(W_k)] = L(W_k)$, and Gaussian with known variance $\mathrm{Var}[\bar{L}(W_k)]$. Assuming this, the knowledge of Eq. (C1) means that we can use weighted linear regression to construct the maximally likely $\bar{\beta}$ and $\overline{\Delta F}$ as the pair of values that minimize the least-squared error

$$\sum_k \frac{(\bar{L}(W_k) - \beta(W_k - \Delta F))^2}{\mathrm{Var}[\bar{L}(W_k)]}, \tag{C2}$$

and take them as our estimates of β and ΔF. This is shown in Fig. 1. The procedure then boils down to obtaining estimates $\bar{L}(W_k)$ whilst ensuring no bias and estimating $\mathrm{Var}[\bar{L}(W_k)]$.

As a starting point, consider the estimator $\bar{L}'(W) = \ln\{\bar{p}_F(W)/\bar{p}_B(-W)\}$. Due to the non-linear nature of the inverse and logarithm functions, $\bar{L}'(W)$ is neither Gaussian nor an unbiased estimator of $L(W)$. The errors resulting from the bias can be small, but ideally we would like to construct an unbiased estimator $\bar{L}(W) = \bar{L}'(W) - \Delta L(W)$, which removes the bias $\Delta L(W) = \mathrm{E}[\bar{L}'(W)] - L(W)$. We would also like to characterize the remaining mean-zero random errors in $\bar{L}(W)$ and find when they are approximately Gaussian.

To do this, we approximate the bias and variance by simulating the sampling of $\bar{p}_Q(W)$, by adding Gaussian noise of zero mean and variance $\mathrm{Var}[\bar{p}_Q(W)]$ to $\bar{p}_Q(W)$ before taking the logarithm. From these values, we then estimate the bias $\Delta L(W)$ and variance $\mathrm{Var}[\bar{L}(W)]$, and assess the non-Gaussian character of the distribution of $\bar{L}(W)$. As is expected, we find $\bar{L}(W)$ is reasonably Gaussian when $\mathrm{Var}[\bar{p}_Q(W)]/p_Q^2(W)$ is small. For larger values the distribution is very skewed and least-squares minimization may not correspond to the maximally likely parameters.

A further flaw in this approach is the possibility of obtaining negative values $\bar{p}_Q(W)$, whose logarithm is undefined. We simply ignore values of W_k for which $\bar{p}_F(W_k)$ or $\bar{p}_B(-W_k)$ is negative, and we ignore negative values arising in our estimate of the bias and variance. This comes at a cost of potentially inducing a bias. Balancing the need to avoid such a bias with the desire to have as many points as possible for the fit, in our frequentist calculations presented here we only consider values W_k where $\mathrm{Var}[\bar{p}_F(W_k)]/p_F^2(W_k) > 1$ and $\mathrm{Var}[\bar{p}_B(-W_k)]/p_B^2(-W_k) > 1$. The results in Fig. 2 show that this bias is perhaps acceptable but not small. A better approach for dealing with negative values is the Bayesian approach of the next subsection, which makes full use of the information provided by obtaining a negative value for $\bar{p}_F(W_k)$ or $\bar{p}_B(-W_k)$. For now, we continue our analysis of the frequentist approach ignoring any bias introduced by these negative values.

We have shown how to choose a set of W_k for which we have approximately independent, unbiased, and Gaussian random estimators $\bar{L}(W_k)$ of $L(W_k)$ whose variances $\mathrm{Var}[\bar{L}(W_k)]$ we know approximately. From these, we then obtain a maximally likely $\bar{\beta}$ and $\overline{\Delta F}$ as the pair of values that minimize the least-squared error of Eq. (C2) and take them as our estimates of β and ΔF. Let us now turn to discussing how the variance in $\bar{\beta}$ should behave, including how it depends on some of the parameter choices we make in our protocol, particularly T, since the dependence of the error on N_{meas} and N_{steps} is clear.

It is well known that, for least-squares estimation, the fitted parameter $\bar{\beta}$ provides an unbiased estimate of β with variance

$$\begin{aligned}
\mathrm{Var}[\bar{\beta}] =& \frac{1/\varsigma_1}{\varsigma_{W^2}/\varsigma_1 - (\varsigma_W/\varsigma_1)^2}, \\
\varsigma_1 =& \sum_k \frac{1}{\mathrm{Var}[\bar{L}(W_k)]}, \\
\varsigma_W =& \sum_k \frac{W_k}{\mathrm{Var}[\bar{L}(W_k)]}, \\
\varsigma_{W^2} =& \sum_k \frac{W_k^2}{\mathrm{Var}[\bar{L}(W_k)]}.
\end{aligned}$$

We obtain a simpler expression for the fractional error

$$\frac{\sqrt{\mathrm{Var}[\bar{\beta}]}}{\beta} = \frac{\sqrt{\mathrm{Var}[\bar{L}]/N_W}}{\beta \Delta W_k},$$

that qualitatively captures the basic behavior if we assume uniform variance $\mathrm{Var}[\bar{L}(W_k)] \sim \mathrm{Var}[\bar{L}]$ with $\mathrm{Var}[\bar{L}]$ the typical value of the variance over the values of W_k used. Appearing in the denominator of this equation is the spread $\Delta W_k = \sqrt{\sum_k W_k^2/N_W - (\sum_k W_k/N_W)^2}$ of the values W_k, where we have written $N_W = \sum_k$ for the number of points used in the estimation.

We can now insert some of the findings from the error analysis of the previous sections into this expression. First, demanding independence of points required us to choose that W_k were spaced by π/T and so $N_W \sim$

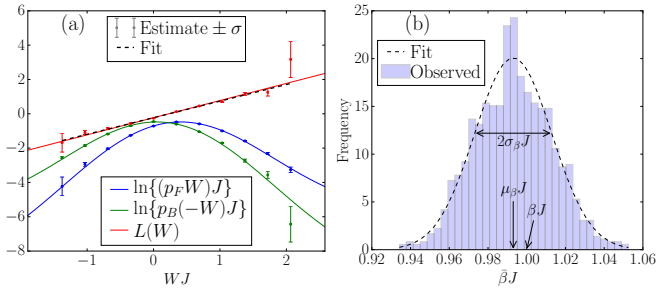

FIG. 1: *Superfluid phase; frequentist.* (a) The logarithms $\ln\{p_F(W)\}$ and $\ln\{p_B(-W)\}$ of the work distributions and their difference $L(W) = \ln\{p_F(W)\} - \ln\{p_B(-W)\}$. Solid lines are known values, points are estimates obtained in a single simulated experiment, and error bars calculated from those estimates mark expected errors of one standard deviation. The black dashed line is obtained from a weighted least squares fit. (b) A histogram of $\bar{\beta}$ estimates obtained in 1000 simulated experimental runs, together with their mean μ_β and standard deviation σ_β, and Gaussian fit. All parameters are identical to those for Fig. 1 in the main text.

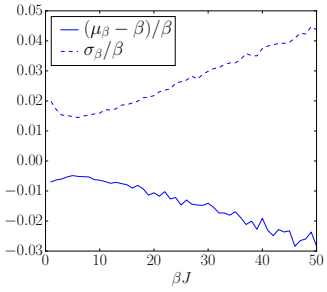

FIG. 2: *Accuracy at low temperatures; frequentist.* Over 1000 simulated experiments for each β, the average fractional standard deviation σ_β/β and bias $(\mu_\beta - \beta)/\beta$. All other parameters are identical to those for Fig. 1(d) in the main text.

$\Delta W_k T/\pi$, giving us

$$\frac{\sqrt{\mathrm{Var}[\bar{\beta}]}}{\beta} \sim \frac{\sqrt{\mathrm{Var}[\bar{L}]/T}}{\beta(\Delta W_k)^{3/2}}.$$

Second, the range of work values ΔW_k we include is the range satisfying $\mathrm{Var}[\bar{p}_F(W_k)]/p_F^2(W_k) > 1$ and $\mathrm{Var}[\bar{p}_B(-W_k)]/p_B^2(-W_k) > 1$, which corresponds to ensuring that $T^2/N_{\mathrm{meas}}N_{\mathrm{steps}}p_F^2(W_k)$ and $T^2/N_{\mathrm{meas}}N_{\mathrm{steps}}p_B^2(-W_k)$ are smaller than a threshold amount. This tells us that this range ΔW_k can decay quickly with increasing T if $p_Q(W)$ depends sharply on W at the edge of the region W_k. In turn this means that T should be chosen to be as small as possible without introducing spectral leakage, which is the solution to one of the main questions remaining from the above discussion. We found these optimum values of T to behave roughly as Eq. (B1), essentially linearly in β, leaving us with

$$\frac{\sqrt{\mathrm{Var}[\bar{\beta}]}}{\beta} \sim \frac{\sqrt{\mathrm{Var}[\bar{L}]}}{(\beta\Delta W_k)^{3/2}}.$$

It is difficult to ascertain from this simplified analysis how this fractional error depends on β. In our examples we find that the relevant width ΔW_k of work values decreases roughly linearly with β for large β. The variances captured by $\mathrm{Var}[\bar{L}]$ are the deciding factor. We find that in our examples they increase with β^2 as might be predicted from $\mathrm{Var}[\bar{p}_Q(W)] \sim T^2$ and our setting of $T \sim \beta$. Thus the fractional variance of our β estimate increases roughly linearly with β [Fig. 2].

b. *Bayesian*

The Bayesian approach is to capture probabilistically the knowledge we have about β and ΔF conditioned upon obtaining estimates $\bar{p}_Q(W)$. It also allows us to include our prior expectations about the system into the statistical analysis. However, we do not make use of that feature here, assuming nothing about the system. Further, the Bayesian approach outputs a probability distribution of the values of β and ΔF, which need not be Gaussian, rather than only returning maximally likely parameters and an estimate of their variance. The Bayesian approach therefore can use and provide more information than the frequentist approach.

In the following we use the shorthand notation $p_F = p_F(W_k)$, $p_B = p_B(-W_k)$, $R = R(W_k)$ and $W = W_k$. We write O_k to denote the observations of estimates $\bar{p}_F$ and $\bar{p}_B$ for some value of k and O for the combined set of observations.

The Bayesian approach is based on the expression

$$\mathcal{P}(\beta, \Delta F|O) = \frac{\mathcal{P}(O|\beta, \Delta F)\mathcal{P}(\beta, \Delta F)}{\int \mathrm{d}\beta \mathrm{d}\Delta F \mathcal{P}(O|\beta, \Delta F)\mathcal{P}(\beta, \Delta F)}, \quad \text{(C3)}$$

for our assessment of the probability of the system having temperature β and free energy difference ΔF given observations O. It uses our assumptions about $\mathcal{P}(O|\beta, \Delta F)$, the probability we would have obtained those observations for all possible values of β and ΔF, and the prior $\mathcal{P}(\beta, \Delta F)$, capturing our knowledge of the system in advance of the experiment.

In this paper we make what is called a null prior, setting $\mathcal{P}(\beta, \Delta F)$ to be constant, essentially assuming we have no information about β and ΔF. An experimentalist who does have more prior information could easily adapt our approach to include that information in the inference scheme.

We are left then to come up with an expression for the conditional probability $\mathcal{P}(O|\beta, \Delta F)$, which, assuming independence, is just a product of the conditional probability of obtaining pairs of observations at each W. The next few paragraphs deal with the evaluation of this

conditional probability. We perform this evaluation by breaking it up into several parts, in three stages. First we use that

$$\begin{aligned}\mathcal{P}(\bar{p}_F,\bar{p}_B|\beta,\Delta F) &= \int \mathrm{d}R\mathcal{P}(\bar{p}_F,\bar{p}_B|R)\mathcal{P}(R|\beta,\Delta F)\\ &= \mathcal{P}(\bar{p}_F,\bar{p}_B|\mathrm{e}^{\beta(W-\Delta F)}).\end{aligned}$$

Here we have conditioned the obtaining of the estimates $\bar{p}_F$ and $\bar{p}_B$ on the ratio R of their true values, and used fact that this ratio is always equal to $\mathrm{e}^{\beta(W-\Delta F)}$. It is this piece of information that is the key to the whole protocol.

The second step is to use

$$\mathcal{P}(\bar{p}_F,\bar{p}_B|R) = \int \mathrm{d}p_F\mathrm{d}p_B\mathcal{P}(\bar{p}_F,\bar{p}_B|p_F,p_B)\mathcal{P}(p_F,p_B|R).$$

Here we have conditioned on the exact values of the work distributions p_Q. We do this because we know that $\bar{p}_Q$ are unbiased and Gaussian with known variance σ_Q i.e. $\mathcal{P}(\bar{p}_F,\bar{p}_B|p_F,p_B) = \mathcal{P}(\bar{p}_F|p_F)\mathcal{P}(\bar{p}_B|p_B)$ with

$$\mathcal{P}(\bar{p}_Q|p_Q) = \frac{1}{\sqrt{2\pi\sigma_Q^2}}\mathrm{e}^{-(\bar{p}_Q-p_Q)^2/2\sigma_Q^2}.$$

The distribution $\mathcal{P}(p_F,p_B|R)$ of the work distributions given their ratio can be built in our third and final step. We use Bayes' law again

$$\mathcal{P}(p_F,p_B|R) = \frac{\mathcal{P}(R|p_F,p_B)\mathcal{P}(p_F,p_B)}{\int \mathrm{d}p_F\mathrm{d}p_B\mathcal{P}(R|p_F,p_B)\mathcal{P}(p_F,p_B)},$$

since we know that

$$\mathcal{P}(R|p_F,p_B) = \delta(R-p_F/p_B) = p_B\delta(Rp_B-p_F).$$

This leaves us having to define some prior $\mathcal{P}(p_F,p_B)$ representing our prior knowledge of the work distribution values. We again essentially assume no prior knowledge, assuming independence $\mathcal{P}(p_F,p_B) = \mathcal{P}(p_F)\mathcal{P}(p_B)$ and a uniform distribution over positive values

$$\mathcal{P}(p_Q) = \begin{cases}\frac{1}{C} & \text{if } 0\le p_Q < C\\ 0 & \text{otherwise}\end{cases},$$

for some C taken to be much bigger than all other values.

We now have all we need. Collecting all of the above together, after performing two integrals, we obtain

$$\begin{aligned}&\mathcal{P}(p_F,p_B|R) =\\ &\quad 2\max\{1,1+R^2\}p_B\delta(Rp_B-p_F)\mathcal{P}(p_F)\mathcal{P}(p_B),\end{aligned}$$

and

$$\mathcal{P}(\bar{p}_F,\bar{p}_B|\beta,\Delta F) = \frac{\max\{1,1+R^2\}}{2\pi C^2 R}\left(\exp\left[-\frac{p_F'^2}{2\sigma_F'^2}-\frac{p_B^2}{2\sigma_B^2}\right]\frac{\sigma_F'\sigma_B}{\pi\sigma_{FB}'^2}+\exp\left[-\frac{(p_F'-p_B)^2}{2\sigma_{FB}'^2}\right]\frac{p_F'\sigma_B^2+p_B\sigma_F'^2}{\sqrt{2\pi}\sigma_{FB}'^3}\left[1+\mathrm{erf}\left(\frac{p_F'\sigma_B^2+p_B\sigma_F'^2}{\sqrt{2}\sigma_F'\sigma_B\sigma_{FB}'}\right)\right]\right),$$

where we have used the shorthand $p_F' = p_F/R$, $\sigma_F' = \sigma_F/R$ and $\sigma_{FB}'^2 = \sigma_F'^2+\sigma_B^2$.

All of this can be fed back into our expression for $\mathcal{P}(\beta,\Delta F|O)$ [Eq. (C3)] and allows us to plot this distribution and calculate its properties. Upon testing, the Bayesian prediction is found to be remarkably well calibrated. We have performed 1000 simulated experiments and the fraction of times in which the true β value lies in each decile of the Bayesian prediction. Ideally this would be exactly 0.1 for each decile and our predictions conform to this, staying between 0.8 and 1.2 for β between 1 and 50, as shown in Fig. 3.

Appendix D: Theoretical values for $\chi_Q(u)$

In this part of the supplemental material, we provide details of the two methods used to calculate the characteristic function (as well as other relevant quantities) of our example system. Specifically, we calculate

$$\chi_Q(u) = \mathrm{tr}\left\{\hat{U}_Q^\dagger\hat{U}^\dagger(\lambda_Q(\tau),u)\hat{U}_Q\hat{U}(\lambda_Q(0),u)\hat{\rho}_\beta(\lambda_Q(0))\right\}, \tag{D1}$$

where

$$\begin{aligned}\hat{U}_Q &= \mathcal{T}\exp\left[-\mathrm{i}\int_0^\tau \mathrm{d}t\hat{H}(\lambda_Q(t))\right], &\text{(D2)}\\ \hat{U}(\lambda,u) &= \mathrm{e}^{-\mathrm{i}\hat{H}(\lambda)u},\\ \hat{\rho}_\beta(\lambda) &= \mathrm{e}^{-\beta\hat{H}(\lambda)}/\mathcal{Z}_\beta(\lambda),\\ \mathcal{Z}_\beta(\lambda) &= \mathrm{tr}\{\exp[-\beta\hat{H}(\lambda)]\},\\ \hat{H}(\lambda) &= \hat{H}_S+\lambda\hat{V}.\end{aligned}$$

We begin by discussing the Bogoliubov treatment, relevant to the superfluid phase, and then move on to the tensor network approach that is necessary when interactions are stronger.

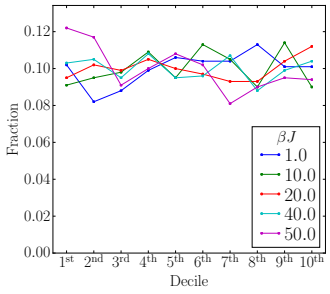

FIG. 3: *Accuracy and precision; Bayesian.* The fraction, over 1000 simulated experiments, of experiments in which the true value of β lies in the deciles of $\mathcal{P}(\beta)$ predicted using Bayesian analysis. The flatness of the plots, especially for the lower temperatures, reveals that our null priors are well calibrated. All parameters are identical to those for Fig. 2.

1. Bogoliubov treatment

a. The free energy

In the Bogoliubov treatment [1] we have the simplified Hamtilonian $\hat{H}(\lambda) = \hat{H}_S + \lambda\hat{V}$, where

$$\begin{aligned}\hat{H}_S =& \sum_k \omega_k \hat{b}_k^\dagger \hat{b}_k,\\ \hat{V} =& \eta n + \sum_k (\eta_k^* \hat{b}_k^\dagger + \eta_k \hat{b}_k).\end{aligned}$$

Here $\hat{b}_k^\dagger$ and $\hat{b}_k$, introduced in the main text along with the other terms, satisfy bosonic commutation relations.

This Hamiltonian can be diagonalized by defining a new displaced creation operator $\hat{d}_k^\dagger = \hat{b}_k^\dagger + \lambda\eta_k/\omega_k$, and its annihilating conjugate $\hat{d}_k$, leading to

$$\hat{H}(\lambda) = \sum_k \omega_k \hat{d}_k^\dagger \hat{d}_k + \lambda\eta\left(n - \lambda\sum_k \frac{|\eta_k|^2}{\eta\omega_k}\right).$$

We can thus immediately write the free energy as

$$\begin{aligned}F(\lambda) &= -\frac{1}{\beta}\ln \mathcal{Z}_\beta(\lambda)\\ &= \lambda\eta\left(n - \lambda\sum_k \frac{|\eta_k|^2}{\eta\omega_k}\right) + F(0),\\ F(0) &= -\frac{1}{\beta}\ln \mathrm{tr}\left\{\exp\left[-\beta\sum_k \omega_k \hat{d}_k^\dagger \hat{d}_k\right]\right\}\\ &= \frac{1}{\beta}\sum_k \ln(1 - \mathrm{e}^{-\beta\omega_k}).\end{aligned}$$

b. The characteristic function

Our first step in evaluating $\chi_Q(u)$ [Eq. (D1)] is to simplify by absorbing any displacement of the Bogoliubov phonons present in the initial state $\hat{\rho}_\beta(\lambda_Q(0))$ into the Hamiltonian $\hat{H}(\lambda)$. We do this using a method similar to that used for the free energy. We define a new displaced creation operator $\hat{d}_k^\dagger = \hat{b}_k^\dagger + \lambda_Q(0)\eta_k/\omega_k$ and its annihilating conjugate $\hat{d}_k$. In terms of these operators, we re-write the Hamiltonian as

$$\begin{aligned}\hat{H}(\lambda) =& \hat{H}'(\lambda') = \hat{H}'_S + \lambda'\hat{V}',\\ \lambda' =& \lambda - \lambda_Q(0),\\ \hat{H}'_S =& \sum_k \omega_k \hat{d}_k^\dagger \hat{d}_k,\\ \hat{V}' =& \eta n' + \sum_k (\eta_k^* \hat{d}_k^\dagger + \mathrm{h.c.}),\\ n' =& n - 2\lambda_Q(0)\sum_k \frac{|\eta_k|^2}{\eta\omega_k}.\end{aligned}$$

Here n' represents the reduced background density due to the initial perturbation, and we have omitted a constant term $\lambda_Q(0)\eta(n - \lambda_Q(0)\sum_k |\eta_k|^2/\eta\omega_k)$ representing its energy.

We now have

$$\chi_Q(u) = \mathrm{tr}\left\{\hat{U}_Q'^\dagger \hat{U}'^\dagger(\lambda'_Q(\tau), u)\hat{U}'_Q\hat{\rho}'_\beta(0)\right\},$$

where

$$\begin{aligned}\hat{U}'_Q =& \mathcal{T}\exp[-\mathrm{i}\int_0^\tau \mathrm{d}t\hat{H}'(\lambda'_Q(t))],\\ \hat{U}'(\lambda', u) =& \mathrm{e}^{-\mathrm{i}\hat{H}'(\lambda')u},\\ \hat{\rho}'_\beta(\lambda') =& \mathrm{e}^{-\beta\hat{H}'(\lambda')}/\mathcal{Z}'_\beta(\lambda'),\\ \mathcal{Z}'_\beta(\lambda') =& \mathrm{tr}\{\exp[-\beta\hat{H}'(\lambda')]\}.\end{aligned}$$

Note that, by design, $\lambda'_Q(0) = 0$ and thus we have used $\hat{U}'(\lambda'_Q(0), u) = \hat{U}'(0, u) = 1$. In what follows, for clarity, we drop the primes.

Our second step is to move to the interaction picture, whereupon

$$\chi_Q(u) = \mathrm{tr}\left\{\tilde{U}_Q^\dagger(0)\tilde{U}^\dagger(\lambda_Q(\tau), u, \tau)\tilde{U}_Q(u)\hat{\rho}_\beta(0)\right\}.$$

Here the tilde indicates an operator in the interaction

picture, specifically

$$\begin{aligned}\tilde{U}_Q(t) =&\mathcal{T}\exp\left[-\mathrm{i}\int_0^\tau \mathrm{d}t'\lambda_Q(t')\tilde{V}(t'+t)\right],\\ \tilde{U}(\lambda,u,t) =&\mathcal{T}\exp\left[-\mathrm{i}\lambda\int_0^u \mathrm{d}t'\tilde{V}(t'+t)\right],\\ \tilde{V}(t) =&\mathrm{e}^{\mathrm{i}\hat{H}_S t}\hat{V}\mathrm{e}^{-\mathrm{i}\hat{H}_S t}\\ =&\eta n+\sum_k(\eta_k^*\hat{d}_k^\dagger\mathrm{e}^{\mathrm{i}\omega_k t}+\mathrm{h.c.}).\end{aligned}$$

The third step is to simplify the time-ordered exponentials. For this we appeal to the Magnus expansion

$$\tilde{U} =\mathcal{T}\exp\left[-\mathrm{i}\int_0^t \mathrm{d}t'\lambda(t')\tilde{V}(t')\right]=\mathrm{e}^{-\mathrm{i}\tilde{A}},$$

in terms of Hermitian operators

$$\begin{aligned}\tilde{A} =&\tilde{A}^{[1]}+A^{[2]}+\cdots,\\ \tilde{A}^{[1]} =&\int_0^t \mathrm{d}t'\lambda(t')\tilde{V}(t'),\\ A^{[2]} =&\frac{-\mathrm{i}}{2}\int_0^t \mathrm{d}t'\int_0^{t'}\mathrm{d}t''\lambda(t')\lambda(t'')[\tilde{V}(t'),\tilde{V}(t'')],\\ &\vdots\end{aligned}$$

The simplifying feature of the Bogoliubov Hamiltonian, which we will use repeatedly, is that the commutator

$$[\tilde{V}(t),\tilde{V}(t')]=-2\mathrm{i}\sum_k|\eta_k|^2\sin(\omega_k(t-t')),$$

is a pure imaginary c-number. Hence we have omitted the tilde from $A^{[2]}$ to highlight that it is a real c-number only. This also ensures that all terms $\tilde{A}^{[m]}$ for $m>2$ vanish from the Magnus expansion. The result is that

$$\begin{aligned}\chi_Q(u) =&\mathrm{e}^{\mathrm{i}A^{[2]}(\lambda_Q(\tau),u)}\\ &\times\mathrm{tr}\left\{\mathrm{e}^{\mathrm{i}\tilde{A}_Q^{[1]}(0)}\mathrm{e}^{\mathrm{i}\tilde{A}^{[1]}(\lambda_Q(\tau),u,\tau)}\mathrm{e}^{-\mathrm{i}\tilde{A}_Q^{[1]}(u)}\hat{\rho}_\beta(0)\right\}.\end{aligned}$$

Here, we have used a simplification of the form $\mathrm{e}^{\mathrm{i}A_Q^{[2]}(0)}\mathrm{e}^{-\mathrm{i}A_Q^{[2]}(u)}=1$, and the other integrals appearing in the expression are as follows

$$\begin{aligned}\tilde{A}_Q^{[1]}(t) =&\int_0^\tau \mathrm{d}t'\lambda_Q(t')\tilde{V}(t'+t)\\ =&\eta n\int_0^\tau \mathrm{d}t'\lambda_Q(t')\\ &+\sum_k\left(\frac{\eta_k^*}{\omega_k}\Lambda_{Qk}^*\hat{d}_k^\dagger\mathrm{e}^{\mathrm{i}\omega_k t}+\mathrm{h.c.}\right),\\ \Lambda_{Qk} =&\omega_k\int_0^\tau \mathrm{d}t\lambda_Q(t)\mathrm{e}^{-\mathrm{i}\omega_k t},\\ \tilde{A}^{[1]}(\lambda,u,t) =&\lambda\int_0^u \mathrm{d}t'\tilde{V}(t'+t)\\ =&\eta n\lambda u\\ &+\sum_k\left(\frac{\eta_k^*}{\omega_k}\Lambda_k^*(\lambda,u)\hat{d}_k^\dagger\mathrm{e}^{\mathrm{i}\omega_k t}+\mathrm{h.c.}\right),\\ \Lambda_k(\lambda,u) =&\omega_k\lambda\int_0^u \mathrm{d}t\mathrm{e}^{-\mathrm{i}\omega_k t}\\ =&-\mathrm{i}\lambda(1-\mathrm{e}^{-\mathrm{i}\omega_k u}),\\ A^{[2]}(\lambda,u) =&-\lambda^2\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2(\omega_k u-\sin(\omega_k u)).\end{aligned}$$

Having simplified each time-ordered exponential, our third step is to combine them using the Baker-Campbell-Hausdorff formula

$$\mathrm{e}^{\tilde{A}}\mathrm{e}^{\tilde{B}}=\mathrm{e}^{\tilde{A}+\tilde{B}+[\tilde{A},\tilde{B}]/2},$$

for the case that $[\tilde{A},\tilde{B}]$ is a c-number. Explicitly, we use

$$\begin{aligned}&\mathrm{e}^{\mathrm{i}\tilde{A}_Q^{[1]}(0)}\mathrm{e}^{\mathrm{i}\tilde{A}^{[1]}(\lambda_Q(\tau),u,\tau)}\mathrm{e}^{-\mathrm{i}\tilde{A}_Q^{[1]}(u)}\\ &=\mathrm{e}^{\mathrm{i}(\tilde{A}_Q^{[1]}(0)-\tilde{A}_Q^{[1]}(u))+\mathrm{i}\tilde{A}^{[1]}(\lambda_Q(\tau),u,\tau)}\\ &\quad\times\exp\Big\{-\frac{1}{2}\Big([\tilde{A}_Q^{[1]}(u),\tilde{A}_Q^{[1]}(0)]\\ &\qquad+[\tilde{A}_Q^{[1]}(u),\tilde{A}^{[1]}(\lambda_Q(\tau),u,\tau)]\\ &\qquad+[\tilde{A}_Q^{[1]}(0),\tilde{A}^{[1]}(\lambda_Q(\tau),u,\tau)]\Big)\Big\},\end{aligned}$$

together with

$$\begin{aligned}&[\tilde{A}_Q^{[1]}(t),\tilde{A}_Q^{[1]}(t')]\\ &=-2\mathrm{i}\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2|\Lambda_{Qk}|^2\sin(\omega_k(t-t')),\\ &[\tilde{A}_Q^{[1]}(t),\tilde{A}^{[1]}(\lambda,u,t')]\\ &=-2\mathrm{i}\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2\Im\left\{\Lambda_{Qk}^*\Lambda_k(\lambda,u)\mathrm{e}^{\mathrm{i}\omega_k(t-t')}\right\}.\end{aligned}$$

This leaves us with

$$\begin{aligned}\chi_Q(u) =& \exp\left(\mathrm{i}\eta n\lambda_Q(\tau)u\right)\\ &\times\exp\left(-\mathrm{i}\lambda_Q^2(\tau)\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2(\omega_k u-\sin(\omega_k u))\right)\\ &\times\exp\left(-\mathrm{i}\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2 h_{Qk}(u)\right)\\ &\times\mathrm{tr}\left\{\exp\left(-\mathrm{i}\sum_k\left(\frac{\eta_k^*}{\omega_k}g_{Qk}^*(u)\hat{d}_k^\dagger+\mathrm{h.c.}\right)\right)\hat{\rho}_\beta(0)\right\},\end{aligned}\tag{D3}$$

where

$$\begin{aligned}h_{Qk}(u) =& -|\Lambda_{Qk}|^2\sin(\omega_k u)\\ &-\Im\left\{\Lambda_{Qk}^*\left(\mathrm{e}^{\mathrm{i}\omega_k u}+1\right)\Lambda_k(\lambda_Q(\tau),u)\mathrm{e}^{-\mathrm{i}\omega_k\tau}\right\}\\ =&H_{Qk}\sin(\omega_k u),\\ H_{Qk} =& -|\Lambda_{Qk}|^2-2\lambda_Q(\tau)\Im\left\{\Lambda_{Qk}^*\mathrm{e}^{-\mathrm{i}\omega_k\tau}\right\},\\ g_{Qk}(u) =& -\Lambda_{Qk}\left(1-\mathrm{e}^{-\mathrm{i}\omega_k u}\right)-\Lambda_k(\lambda_Q(\tau),u)\mathrm{e}^{-\mathrm{i}\omega_k\tau}\\ =&G_{Qk}\left(1-\mathrm{e}^{-\mathrm{i}\omega_k u}\right),\\ G_{Qk} =& -\Lambda_{Qk}+\mathrm{i}\lambda_Q(\tau)\mathrm{e}^{-\mathrm{i}\omega_k\tau}.\end{aligned}$$

The final step is to evaluate the trace $\mathrm{tr}\{\cdot\hat{\rho}_\beta(0)\}$, or expected value $\langle\cdot\rangle$ with respect to state $\hat{\rho}_\beta(0)$. We use that the exponent of the first term in the trace contains only terms that are linear in $\hat{d}_k^\dagger$ and $\hat{d}_k$. The state $\hat{\rho}_\beta(0)$ with respect to which the expected value is taken, the other term in the trace, consists of a mixture of different occupations of these phonon modes, with all non-number conserving combinations of $\hat{d}_k^\dagger$ and $\hat{d}_k$ thus having zero expected value and $\langle(\hat{d}_k^\dagger\hat{d}_k)^m\rangle=(n_k)^m$ with $n_k=(\exp(\beta\omega_k)-1)^{-1}$ the mean occupation. This means that for an operator formed for linear combinations of $\hat{d}_k^\dagger$ and $\hat{d}_k$, we have

$$\begin{aligned}&\left\langle\exp\left(-\mathrm{i}\sum_k\left(\frac{\eta_k^*}{\omega_k}g_{Qk}^*(u)\hat{d}_k^\dagger+\mathrm{h.c.}\right)\right)\right\rangle\\ &=\exp\left(-\frac{1}{2}\left\langle\left(\sum_k\left(\frac{\eta_k^*}{\omega_k}g_{Qk}^*(u)\hat{d}_k^\dagger+\mathrm{h.c.}\right)\right)^2\right\rangle\right),\end{aligned}$$

and the particular expectation value is

$$\begin{aligned}&\left\langle\left(\sum_k\left(\frac{\eta_k^*}{\omega_k}g_{Qk}^*(u)\hat{d}_k^\dagger+\mathrm{h.c.}\right)\right)^2\right\rangle\\ &=\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2|g_{Qk}(u)|^2\coth\left(\frac{\beta\omega_k}{2}\right),\end{aligned}$$

where we note that

$$|g_{Qk}(u)|^2=4|G_{Qk}|^2\sin^2\left(\frac{\omega_k u}{2}\right).$$

Inserting this back into Eq. (D3) we arrive at our final expression

$$i\ln\chi_Q(u)=-\eta n\lambda_Q(\tau)u+\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2\left(\lambda_Q^2(\tau)\omega_k u+(H_{Qk}-\lambda_Q^2(\tau))\sin(\omega_k u)\right)-2\mathrm{i}\sum_k\left|\frac{\eta_k}{\omega_k}\right|^2|G_{Qk}|^2\frac{\sin^2\left(\frac{\omega_k u}{2}\right)}{\tanh\left(\frac{\beta\omega_k}{2}\right)}.$$

c. The work distribution

As discussed above, the cumulants κ_m of the work distribution $P_Q(W)$ are related to the derivatives of the logarithm of its characteristic function, by Eq. (A1). We may use this to calculate all cumulants of the work distribution, but here we report only the first three

$$\begin{aligned}\mu_Q=\kappa_{Q1} =&\lambda_Q(\tau)\eta n-\sum_k\frac{|\eta_k|^2}{\omega_k}H_{Qk},\\ \sigma_Q^2=\kappa_{Q2} =&\sum_k|\eta_k|^2|G_{Qk}|^2\coth\left(\frac{\beta\omega_k}{2}\right),\\ \kappa_{Q3} =&-\sum_k|\eta_k|^2\omega_k H_{Qk}.\end{aligned}$$

d. Symmetry

A close inspection of the results reveals that reversing the quench has no effect on either H_{Qk} or $|G_{Qk}|^2$ and thus also on the cumulants κ_{Qm} for $m>1$. This represents the fact that in the Bogoliubov description the deviations of the condensate, assumed small, do not to interact resulting in a symmetry between repulsive and attractive interactions. This means that the only difference between the two characteristic functions lies in the phase $\eta n\lambda_Q(\tau)u$. The effect of this is merely to shift the associated probability distributions by the amount $\eta n\lambda_Q(\tau)$, as we can see in our expression for μ_Q. Evaluating these phase shifts, we find that, for the Bogoliubov case, the forward and backward distributions are identical up to a relative shift $\mu_F-\mu_B=2\Delta F$, where $\Delta F=F(\lambda_f)-F(\lambda_i)$ is the free energy difference.

This symmetry could be exploited to provide a better

estimate for β. However, we do not do this here as we wish to keep our estimation protocol general, so that it is applicable to situations where this symmetry does not exist or is not known to exist.

2. Tensor network theory

In regimes where the condition necessary for the validity of Bogoliubov theory $Un \ll J$ is not satisfied, we numerically calculate the characteristic function using a tensor network representation. We proceed by writing the quantum state as a matrix product operator (MPO) [2, 3] and then performing imaginary- and real-time evolution using the time-evolving block decimation (TEBD) algorithm [4, 5]. In the following, we concisely outline our method. See Ref. [6] for a detailed pedagogical introduction to tensor-network algorithms.

The state of a quantum lattice system with M sites and local Hilbert space dimension d can be represented in MPO form

$$\hat{\rho} = \sum_{i_1,\ldots,i_M=1}^{d} \sum_{j_1,\ldots,j_M=1}^{d} \mathrm{tr}\left[A_1^{(i_1 j_1)} \cdots A_M^{(i_L j_M)}\right] \times |i_1 \ldots i_M\rangle\langle j_1 \ldots j_M|. \tag{D4}$$

Here, the matrix elements of the density operator are given by the trace over a product of matrices $A_l^{(i_l j_l)}$ of maximum dimension $D \times D$, where the bond dimension D quantifies the correlations between lattice sites. The states $\{|i_l\rangle\}$ constitute a complete orthonormal basis for the Hilbert space on lattice site l. The set of matrices $A_l^{(i_l j_l)}$ may also be considered as elements of a single combined tensor A_l of dimension $D \times d \times D \times d$. Together the tensors A_l can represent an arbitrary quantum state, for sufficiently large d and D. In a bosonic system, however, both the physical dimension d and the bond dimension D must be truncated, in general, in order to obtain a tractable numerical representation. In our calculations we use $d = 4$ and $D = 200$.

The time-evolution operator over a small time step δt is approximated by a product of two-site gates using a second-order Suzuki-Trotter "staircase" decomposition, see Ref. [7] for details. The tensor network representing the time evolution of the quantum state is depicted schematically in Fig. 4(a). The TEBD algorithm proceeds by sweeping across this tensor network and applying two-site gates sequentially to pairs of nearest-neighbor tensors $A_l A_{l+1}$ appearing in the MPO representation of the quantum state in Eq. (D4). The result is a new tensor $\Theta_{l,l+1}$ formed by contracting the A tensors with two-site gates above and below. A singular value decomposition (SVD) is then performed on $\Theta_{l,l+1}$, resulting in a new pair of tensors $A'_l A'_{l+1}$ [Fig. 4(b)]. Only the largest D singular values at most are retained after the SVD, so that an efficient MPO representation of the quantum state is maintained at each time step. Repeating this procedure across the entire system over N small time steps δt leads to the desired numerical evolution over a time duration $t = N\delta t$.

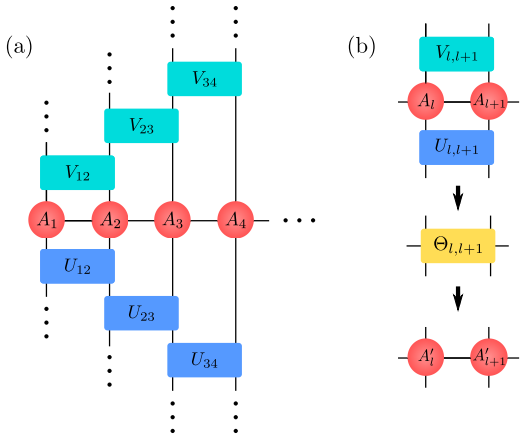

FIG. 4: (a) Schematic depicting one part of a tensor network representing the time evolution of an MPO (red circles) generated by a product of two-site gates $U_{l,l+1}$ and $V_{l,l+1}$ acting on the MPO from the left and right, respectively. (b) Time evolution proceeds by contraction of nearest-neighbor pairs of MPO tensors $A_l A_{l+1}$ with two-site gates above and below, resulting in a new tensor $\Theta_{l,l+1}$. This tensor is then decomposed into a product of tensors $A'_l A'_{l+1}$ via an SVD. Retaining only the largest D singular values in the SVD maintains an efficient MPO representation at each time step.

In order to find the initial thermal state, we use the identity

$$\hat{\rho}_\beta(\lambda) = \frac{1}{\mathcal{Z}_\beta(\lambda)} \mathrm{e}^{-\beta\hat{H}/2} \hat{\mathbb{1}} \mathrm{e}^{-\beta\hat{H}/2}.$$

The right-hand side of this equation can be calculated using the TEBD algorithm as outlined above, after a Wick rotation to imaginary time $t \to -\mathrm{i}\beta/2$ and taking the initial state to be the system-wide identity operator $\hat{\mathbb{1}}$. The MPO representation of the identity operator is given simply by $A_l^{(i_l j_l)} = \delta_{i_l j_l}$.

The characteristic function [Eq. (D1)] may then be calculated by performing real-time evolution to find the operator

$$\hat{X}_Q(u) = \hat{U}_Q \mathrm{e}^{-\mathrm{i}u\hat{H}(\lambda_Q(0))} \hat{\rho}_\beta(\lambda_Q(0)) \hat{U}_Q^\dagger \mathrm{e}^{\mathrm{i}u\hat{H}(\lambda_Q(\tau))},$$

such that $\chi_Q(u) = \mathrm{tr}\left\{\hat{X}_Q(u)\right\}$. The unitary $\hat{U}_Q = \mathcal{T}\exp[-\mathrm{i}\int_0^\tau \mathrm{d}t \hat{H}(\lambda_Q(t))$ [Eq. (D2)] is performed by discretizing the quench path into steps $\lambda_{Qm} = \lambda_Q(m\delta t)$, with $m = 1, \ldots, M$ and $\tau = M\delta t$. Discrete time evolution under the TEBD algorithm reproduces the quench unitary, since

$$\hat{U}_Q \approx \prod_{m=M}^{1} \mathrm{e}^{-\mathrm{i}\delta t H(\lambda_{Qm})},$$

for sufficiently small δt.

In order to ensure numerical stability, each matrix $A_l^{(i_l j_l)}$ is divided by its matrix norm (the square root of the sum of its elements) after each time step. The accumulated product of these norms then multiplies expectation values to give the correct final result.

[1] D. Van Oosten, P. Van der Straten, and H. T. C. Stoof, Phys. Rev. A **63**, 053601 (2001).

[2] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).

[3] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004).

[4] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).

[5] S. R. White and A. E. Feiguin, Phys. Rev. Lett. **93**, 076401 (2004).

[6] U. Schöllwock, Ann. Phys. **326**, 96 (2011).

[7] T. H. Johnson, S. R. Clark, and D. Jaksch, Phys. Rev. E **82**, 036702 (2010)