

# Optimal Cold Atom Thermometry Using Adaptive Bayesian Strategies

Glatthard, Jonas; Rubio, Jesús; Sawant, Rahul; Hewitt, Thomas; Barontini, Giovanni; Correa, Luis A.

DOI:

[10.1103/PRXQuantum.3.040330](https://doi.org/10.1103/PRXQuantum.3.040330)

License:

Creative Commons: Attribution (CC BY)

## Document Version

Publisher's PDF, also known as Version of record

## Citation for published version (Harvard):

Glatthard, J, Rubio, J, Sawant, R, Hewitt, T, Barontini, G & Correa, LA 2022, 'Optimal Cold Atom Thermometry Using Adaptive Bayesian Strategies', *PRX Quantum*, vol. 3, no. 4, 040330.  
<https://doi.org/10.1103/PRXQuantum.3.040330>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.


If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Optimal Cold Atom Thermometry Using Adaptive Bayesian Strategies

Jonas Glatthard<sup>1,\*</sup>, Jesús Rubio<sup>1,†</sup>, Rahul Sawant,<sup>2</sup> Thomas Hewitt,<sup>2</sup> Giovanni Barontini,<sup>2,‡</sup> and Luis A. Correa<sup>1</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Exeter, Stocker Road, Exeter EX4 4QL, United Kingdom*

<sup>2</sup>*School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*

 (Received 16 May 2022; revised 21 October 2022; accepted 16 November 2022; published 19 December 2022)

Precise temperature measurements on systems of few ultracold atoms is of paramount importance in quantum technologies, but can be very resource intensive. Here, we put forward an adaptive Bayesian framework that substantially boosts the performance of cold atom temperature estimation. Specifically, we process data from real and simulated release-recapture thermometry experiments on few potassium atoms cooled down to the microkelvin range in an optical tweezer. From simulations, we demonstrate that adaptively choosing the release-recapture times to maximize information gain does substantially reduce the number of measurements needed for the estimate to converge to a final reading. Unlike conventional methods, our proposal systematically avoids capturing and processing uninformative data. We also find that a simpler nonadaptive method exploiting all the *a priori* information can yield competitive results, and we put it to the test on real experimental data. Furthermore, we are able to produce much more reliable estimates, especially when the measured data are scarce and noisy, and they converge faster to the real temperature in the asymptotic limit. Importantly, the underlying Bayesian framework is not platform specific and can be adapted to enhance precision in other setups, thus opening new avenues in quantum thermometry.

DOI: [10.1103/PRXQuantum.3.040330](https://doi.org/10.1103/PRXQuantum.3.040330)

## I. INTRODUCTION

Precise low-temperature thermometry is a key enabler for cold-atom-based quantum technologies. In particular, analog quantum simulation in optical lattices [1,2], computation on large-scale programmable simulators [3,4], the study of thermalization in closed quantum systems [5–7], or the realization of thermodynamic cycles [8–10]. The most common thermometric technique for cold atoms is time-of-flight imaging, which infers temperature from the velocity distribution of a free-expanding atomic cloud. However, this is not sufficiently accurate in optical lattices. It also becomes unsuitable in more advanced setups like optical tweezers, which allow us to independently trap and manipulate only one or few atoms [3,4,11–13]. In these cases, alternatives such as Raman sideband spectroscopy [12], release-recapture thermometry [14–16], or selective

atom spilling [11] must be employed. Often, however, such techniques require a large number of experimental repetitions to be accurate, rendering thermometry expensive in terms of machine time.

On the theoretical side, the application of estimation-theoretic methods to low-temperature thermometry has consolidated into the novel field of “quantum thermometry” [17,18]. Specifically, progress has been made on establishing fundamental scaling laws for the signal-to-noise ratio of low-temperature estimates [19–21], or on the identification of design prescriptions that can make a probe more responsive to temperature fluctuations [22–28]. As a result, precision tuning in sensing applications with atomic impurities is starting to be informed by estimation theory [7,29,30].

Most of the existing literature on quantum thermometry relies on *local* estimation theory [17,18], and adopts the (quantum) Fisher information  $\mathcal{F}$  as the figure of merit to be maximized [31]. Indeed, in the asymptotic limit of a large number of measurements (i.e.,  $\mu \rightarrow \infty$ ), and for unbiased estimators,  $\mathcal{F}$  controls the scaling of the best-case signal-to-noise ratio of temperature estimates, as per  $(T/\delta T)^2 \sim \mu \mathcal{F} T^2$ . However, in the experimentally relevant scenario of *finite* measurement records, with just tens or hundreds of shots,  $\mathcal{F}$  does not always capture—even qualitatively—the behavior of optimal

\*J.Glatthard@exeter.ac.uk

†J.Rubio-Jimenez@exeter.ac.uk

‡G.Barontini@bham.ac.uk

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

temperature estimates. In these cases, one must adopt the more general Bayesian framework [32–34]. This has recently led to a wealth of new results in quantum thermometry [35–41], although there had been earlier applications [42,43].

While experimental proposals have been put forward within quantum thermometry [7,23,25,30,44,45], the field has remained eminently theoretical and mostly concerned with setting ultimate precision bounds [18]. With this work we want to take a more practical turn by exploiting the theory to process actual experimental data. This has been facilitated by the adoption of the Bayesian framework that, not only provides precision bounds as in local estimation theory, but also allows for the construction of optimal estimators from measurement outcomes [32–34,40].

By exploiting Bayesian methods, here we develop an adaptive strategy to optimally assess the temperature of single- or few-atom systems. Namely, given a set of measurement outcomes and a model for the experiment, we find the estimate that minimizes, on average, a suitably defined error. Crucially, we can compute the expected information gain shot by shot, which allows us to adjust the experiment in real time so that each measurement is maximally informative.

We illustrate the benefits of our adaptive Bayesian protocol by optimizing a release-recapture thermometry experiment on a few  $^{41}\text{K}$  atoms tightly confined in an optical tweezer at microkelvin temperatures. This is a particularly challenging system in which precise thermometry is usually exceedingly expensive in terms of the number of measurements needed. Indeed, as already advanced, single-shot procedures like time-of-flight absorption imaging cannot be used with a single or few atoms, due to the vanishing optical density. On the other hand, alternative methods like Raman sideband spectroscopy operate at far lower temperatures, which leaves release and recapture as the only option. Our aim here is to show how resource optimization can help to overcome the two main weaknesses of release-recapture thermometry, namely, that it inherently needs many experimental repetitions, and that it is usually less accurate than other methods. To do so, we use the light-off time of free expansion of the atoms as the control parameter in our optimization.

Working with both real and simulated data, we show that our adaptive Bayesian protocols largely outperform conventional release-recapture thermometry in practice. Specifically, our methods showcase a much reduced estimate variability when applied to finite measurement records with an equal number of data points. Our approach is thus more dependable. Furthermore, we find that, in the asymptotic limit of many data, the estimates from our Bayesian protocols converge faster to the true temperature than conventional estimates. Finally, we show how guiding data acquisition by information gain can substantially reduce the number of measurements needed

for an estimate to converge to a stable value, thus helping to optimize resources. Ultimately, in this work we demonstrate in practice how quantum thermometry can provide concrete design prescriptions capable of making a difference in experiments.

This paper is structured as follows. In Sec. II A we introduce our Bayesian method. In Sec. II B we describe a simple analytical model for the recapture probability. Next, in Sec. II C, we describe three different thermometric protocols and compare their performance. These consist in the mere Bayesian processing of data captured at unoptimized times, the repeated measurement at the *a priori* optimal recapture time, and a fully adaptive method. Details on our experimental setup and on how real data are processed are given in Sec. III. In Sec. IV, we rank the protocols in terms of estimate variability and asymptotic convergence speed and show how information-maximization reduces the required number of measurements for estimates to converge to a target precision. Finally, in Sec. V we draw our conclusions.

## II. ADAPTIVE BAYESIAN THERMOMETRY

### A. Bayesian data analysis and global thermometry

We start by outlining the Bayesian-global paradigm introduced by some of us in Refs. [35,40]. Even though we later focus on the specific release-recapture setup, this framework is fully applicable to any thermometry experiment in which temperature behaves as a *scale parameter* [40,46]. For a more detailed account of this paradigm and why it applies to release-recapture thermometry, see Appendix A.

Our aim is to estimate a temperature  $T$  from a set of measurement outcomes  $\mathbf{n} = (n_1, \dots, n_\mu)$  recorded in  $\mu$  runs of a given experiment. In addition, we consider a controllable parameter  $t$  that we set to  $\mathbf{t} = (t_1, \dots, t_\mu)$  in each subsequent run. In order to extract maximum information from these data, we need a likelihood function [33,34]. Assuming all shots to be independent, this amounts to finding a probability  $p(n_i|T, t_i)$  for obtaining the outcome  $n_i$  given  $t_i$  and  $T$ . Furthermore, since the true temperature is unknown, we work with a hypothesis  $\theta \in [\theta_{\min}, \theta_{\max}]$  about its value.

The idea is then to assign a weight to the different  $\theta$  in light of the measurement record  $(\mathbf{n}, \mathbf{t})$ . This is achieved by constructing the probability  $p(\theta|\mathbf{n}, \mathbf{t})$  [33,34]. Using Bayes theorem, such a distribution can be cast as (cf. Appendix B)

$$p(\theta|\mathbf{n}, \mathbf{t}) \propto p(\theta) \prod_{i=1}^{\mu} p(n_i|\theta, t_i), \quad (1)$$

where  $p(n_i|\theta, t_i)$  is our likelihood model evaluated at  $\theta$ , and  $p(\theta)$  is a probability representing information about  $T$  *prior* to collecting the data.

Since one does not, in general, have any prior knowledge about  $T$ , we choose the distribution  $p(\theta)$

corresponding to maximum ignorance within  $[\theta_{\min}, \theta_{\max}]$  [33,47]. As we will see next, the energy scale in which our model  $p(n_i|T, t_i)$  for release-recapture thermometry is applicable is determined by the unknown temperature itself. We thus say that temperature is a scale parameter [35,40,42,46], and adopt Jeffreys's prior

$$p(\theta) = \left[ \theta \log \left( \frac{\theta_{\max}}{\theta_{\min}} \right) \right]^{-1} \quad (2)$$

to represent maximum ignorance about it [33,34,48].

Having fixed  $p(\theta)$  and  $p(n_i, t_i|\theta)$ , Eq. (1) is used as follows. Upon recording  $(n_1, t_1)$ , the likelihood  $p(n_1|\theta, t_1)$  is calculated as a function of  $\theta$ . The functions  $p(\theta)$  and  $p(n_1|\theta, t_1)$  are multiplied and the result normalized—this operation is interpreted as having *updated* the initial information with the first pair of data. Next,  $(n_2, t_2)$  is recorded and the resulting probability, multiplied by  $p(n_2|\theta, t_2)$ , and normalized. After  $\mu$  iterations, one finally arrives at  $p(\theta|\mathbf{n}, \mathbf{t})$ , which does encode all the available temperature information. That is why  $p(\theta|\mathbf{n}, \mathbf{t})$  is referred to as the posterior probability.

At this point, one can already extract an estimate and its associated error. As shown in Ref. [35], whenever  $T$  behaves as a scale parameter, the *optimal* estimator is given by

$$\tilde{\vartheta}(\mathbf{n}, \mathbf{t}) = \theta_u \exp \left[ \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) \log \left( \frac{\theta}{\theta_u} \right) \right], \quad (3)$$

where  $\theta_u$  simply neutralizes the units within the logarithm [49] without affecting the value of  $\tilde{\vartheta}(\mathbf{n}, \mathbf{t})$ . In this work, we set  $\theta_u = 1 \mu\text{K}$ . In turn, the “error bar” on estimator (3) is

$$\Delta \tilde{\vartheta}(\mathbf{n}, \mathbf{t}) = \tilde{\vartheta}(\mathbf{n}, \mathbf{t}) \sqrt{\bar{\epsilon}_{\text{MLE}}(\mathbf{n}, \mathbf{t})}, \quad (4a)$$

where

$$\bar{\epsilon}_{\text{MLE}}(\mathbf{n}, \mathbf{t}) = \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) \log^2 \left[ \frac{\tilde{\vartheta}(\mathbf{n}, \mathbf{t})}{\theta} \right] \quad (4b)$$

is the mean logarithmic error—a relative error—evaluated for the specific dataset at hand. Ultimately, we report Bayesian temperature estimates as  $\tilde{\vartheta}(\mathbf{n}, \mathbf{t}) \pm \Delta \tilde{\vartheta}(\mathbf{n}, \mathbf{t})$ .

We also exploit the notion of mean information gain for a single shot. As per Ref. [35], this can be quantified as

$$\mathcal{K}(t) = \sum_n p(n|t) \log^2 \left[ \frac{\tilde{\vartheta}(n, t)}{\tilde{\vartheta}_p} \right], \quad (5)$$

where

$$\tilde{\vartheta}_p = \theta_u \exp \left[ \int d\theta p(\theta) \log \left( \frac{\theta}{\theta_u} \right) \right], \quad (6a)$$

$$p(n|t) = \int d\theta p(\theta) p(n|\theta, t). \quad (6b)$$

This allows us to adaptively adjust the control parameter  $t$  so as to maximize  $\mathcal{K}(t)$  at every step, thus making each measurement maximally informative.

The Bayesian-global paradigm summarized here will suffice to optimize release-recapture experiments. Nevertheless, we note that alternative formulations of Bayesian thermometry [36,38,39,41,50] exist. For a perspective on when it is appropriate to employ these, see Refs. [35,40].

## B. The release-recapture method

In a release-recapture thermometry experiment, the trap confining an ultracold gas at temperature  $T$  is switched off during a time  $t$ , so that the atoms expand ballistically. Upon reactivating the trap, the number of recaptured atoms  $n$  is measured by fluorescence imaging. The temperature is obtained by fitting measurement outcomes at different times to a model  $f(T, t)$  for the fraction of recaptured atoms [16].

Let us start with the simple case of a trap deterministically loaded with one atom, deferring the multiatom scenario to Sec. III C. In that case, the probabilities for successfully recapturing or losing the atom after time  $t_i$  are respectively

$$p(n_i = 1|T, t_i) = f(T, t_i), \quad (7a)$$

$$p(n_i = 0|T, t_i) = 1 - f(T, t_i); \quad (7b)$$

this gives us the required likelihood model.

Next, we need to calculate the recaptured fraction  $f(T, t)$ . The trapping potential is created by focusing a laser beam to a narrow waist, which produces tight confinement in the transverse  $x$ - $y$  plane. This is described by

$$U(r) = -U_0 \exp \left( -\frac{2r^2}{w_0^2} \right), \quad (8)$$

where  $U_0$  is the trap depth,  $w_0$  is the beam waist, and  $r^2 = x^2 + y^2$ . At low  $T$ , i.e.,  $U_0/k_B T \gg 1$ , Eq. (8) leads to [16, 51] (cf. Appendix C)

$$f(T, t) = \frac{1}{g(\eta)} g \left[ \frac{\eta W(\tilde{r}^2)}{\tilde{r}^2} \right], \quad (9)$$

where  $g(s) = 1 - e^{-s}$ ,  $\eta = U_0/(k_B T)$ ,  $\tilde{r}^2 = 4U_0 t^2/(mw_0^2)$ ,  $m$  is the mass of the atom, and  $W(\cdot)$  is the Lambert function, defined implicitly as  $W(s)e^{W(s)} := s$ . By resorting to numerical simulations, one could further account for the axial motion of the atoms, i.e., their initial position distribution and the effect of gravity. However, Eq. (9) is in very

good agreement with such simulations in the low- $T$  regime (see Appendix C).

To estimate  $T$  in practice, one conventionally chooses a range of free-expansion times  $\{t_1, \dots, t_v\}$  so that the decay of  $f(T, t)$  is well sampled [14,15]. At each time  $t_i$ ,  $\alpha_i$  independent photon-count measurements are performed, so as to infer the number of recaptured atoms  $\{n_{ij}\}_{j=1}^{\alpha_i}$ . Eventually, one records the average  $\bar{n}_i$  for every  $t_i$ , and fits these to  $\lambda f(T, t)$  via least squares. The light-off time  $t$  then acts as the independent variable, while  $T$  and  $\lambda$  are fitting parameters [16].

In spite of the simplicity of the method, two issues must be noted. First, assuming that the empirical average  $\bar{n}_i$  is close to its true value presupposes an indeterminate “large” number of measurements  $\alpha_i$ . Hence, should  $\alpha_i$  not be “large enough,” the final estimate could become unreliable [52,53]. Second, the method discards most of the information available—the full measurement record  $\{n_{ij}\}$  is known and yet one chooses to work only with the averages  $\{\bar{n}_i\}$ , which likely leads to precision loss [33]. Here,  $i \in \{1, \dots, v\}$  and  $j \in \{1, \dots, \alpha_i\}$ . Both issues are naturally bypassed by the Bayesian estimation framework of global thermometry [35], as discussed below.

### C. Adaptive maximization of the information gain

We now move on to discuss how Bayesian methods can be employed to process the data from conventional release-recapture experiments and, more importantly, how Bayesian data analysis can inform the measurement strategy. Before we proceed, some remarks are in order.

First, we have just seen that the validity of Eq. (9) rests on the condition  $U_0/(k_B T) \gg 1$ , so that it is the unknown  $T$  itself that determines the energy scale in which our model for the fraction of recaptured atoms is applicable. Hence, we are dealing with a scale-estimation problem [40,46] for which Eqs. (2–5) are appropriate.

As for the range  $[\theta_{\min}, \theta_{\max}]$ , a conservative choice for the lower bound is 5% of the trap depth in units of the Boltzmann constant (i.e.,  $U_0/k_B$ ), as the final cooling stage that the atoms undergo in the trap is not expected to reach such low temperatures (see Sec. III A). On the other hand, we choose  $\theta_{\max}$  as the temperature at which Eq. (9) starts to break down (see Appendix C).

#### 1. Unoptimized Bayesian protocol

Let us consider a release-recapture experiment such as those described in Sec. II B. In this case, the vector  $\mathbf{t}$  of expansion times would be

$$\mathbf{t} = (\underbrace{t_1, \dots, t_1}_{\alpha_1 \text{ times}}, \dots, \underbrace{t_v, \dots, t_v}_{\alpha_v \text{ times}}),$$

where the values  $t_i$  would be chosen “by eye” such that the decay of the survival probability (7) is well sampled. Processing the resulting measurement record  $(\mathbf{n}, \mathbf{t})$  with Eqs. (2)–(5) may return slightly better estimates than the conventional least-squares fit of the recaptured atom averages  $\{\bar{n}_i\}$ , since we make use of *all* the available data (see Secs. II D and IV B).

Not all expansion times  $t_i$  are equally informative, however. Intuitively, if the temperature is low, the recapture probability at short enough times is roughly constant, as the atoms have no time to move far—it is those measurements at longer times that carry most of the temperature information. The opposite is true if  $T$  is large. Hence, given that the temperature is unknown, by evenly sampling a time interval we are likely to generate and process many uninformative data points. Instead, as the temperature information is updated by feeding in measurement outcomes, efforts should be concentrated in sampling only those times  $t_i$  that maximize information gain. As we will see next, this speeds up the convergence of the temperature estimate and, more importantly, it largely reduces the spread of estimates drawn from small measurement records.

#### 2. A priori optimization of the expansion time

Even in the absence of measured data and starting from maximum ignorance about  $T$ , one can make the most of the available information—namely, the parameters of the trap—in order to determine an optimal expansion time  $t_s$  for the first shot of the protocol. This amounts to searching for the time at which  $\mathcal{K}(t)$  in Eq. (5) is maximal, which can be calculated *before* any measurement is carried out. The simplest strategy would then consist in performing every subsequent measurement at the same recapture time  $t_s$ , i.e.,

$$\mathbf{t} = (\underbrace{t_s, \dots, t_s}_{\mu \text{ times}}).$$

The intuition behind this approach is that, by keeping the uncertainty of the most uninformed shot—the first one—to a minimum, one can boost the convergence of the estimate to the true temperature as knowledge is updated, as per Eq. (1). This kind of *a priori* optimization was first explored in Mach-Zehnder interferometry [54], and it has also been proven informative in distributed sensing [55]. For a discussion about the construction and application of single-shot optimized protocols, see Ref. [56].

#### 3. Adaptive optimization of the expansion time

Instead of continuing to measure at the *a priori* optimal time  $t_s$ , we can change the recapture intervals at each step so as to maximize the information content of the posterior probability *adaptively* [37]. Specifically,

1. given the prior  $p(\theta)$  and the likelihood  $p(n|\theta, t)$  for the first shot, maximize  $\mathcal{K}(t)$  over  $t$  to find  $t_1 = t_s$ ;
2. perform a measurement at  $t_1 = t_s$  and record  $n_1$ ;
3. normalize  $p(\theta)p(n_1|\theta, t_1)$  and use it as the new “prior” for a second run [33,34]; then apply step 1 to find the optimal expansion time  $t_2$ , and measure  $n_2$ ;
4. iterate  $\mu$  times. The resulting data can then be processed using Eqs. (3) and (4).

#### D. Performance comparison on single-atom simulations

Let us now make a preliminary analysis of how these Bayesian strategies compare to each other, and with the nonlinear fit described in Sec. II B. To that end, we simulate a trap loaded deterministically with one atom, choosing a trap depth of  $U_0/k_B = 29 \mu\text{K}$  and a beam waist  $w_0 = 1.971 \mu\text{m}$ . In these simulations, the temperature is set to  $T = 40 \mu\text{K}$ . The parameters are taken to be the same as in our actual experiment.

For the nonlinear fit and the unoptimized Bayesian protocol, we generate 30 measurement outcomes at each of the seven expansion times (5, 10, 20, 30, 50, 70, 100)  $\mu\text{s}$ . The *a priori* optimized and fully adaptive protocols are applied on simulated records with the same number of entries as the other two methods (i.e., 210). All three Bayesian strategies use Jeffreys’s prior in Eq. (2), with support between

$\theta_{\min} = 14.5 \mu\text{K}$  and  $\theta_{\max} = 125 \mu\text{K}$ . For these parameters, the *a priori* optimal expansion time evaluates to  $t_s = 14 \mu\text{s}$ .

Figure 1 illustrates the estimate variability of each method. The advantage of time optimization seems clear, as both the *a priori* and adaptively optimized methods (rhombs and squares, respectively) yield much more accurate estimates than the fits to data collected at unoptimized recapture times (circles and triangles). Furthermore, in insets (a)–(c) we show how the time-optimized protocols converge to the true temperature with significantly less data than their unoptimized counterpart. In the next two sections we put these preliminary observations on a more solid basis.

### III. EXPERIMENTAL REALIZATION

#### A. The experimental setup

We now apply our Bayesian framework on actual experimental data. Details on our experimental setup can be found in Ref. [57]. In this work,  $^{41}\text{K}$  atoms are loaded into an optical tweezer from a dark magneto-optical trap. The number of atoms transferred into the tweezer can be controlled by varying the magneto-optical trap loading time. The tweezer is realized by focusing a beam at 790nm, produced by a Ti:sapphire laser, to a waist of  $\simeq 1.9 \mu\text{m}$  using a  $20\times$  apoplanar infinity corrected objective. This realizes an elongated tweezer trap, with a radial-to-axial dimension ratio of 1 : 249. As discussed in Sec. II B, such a setup can be described by the model in Eq. (9) to a very good

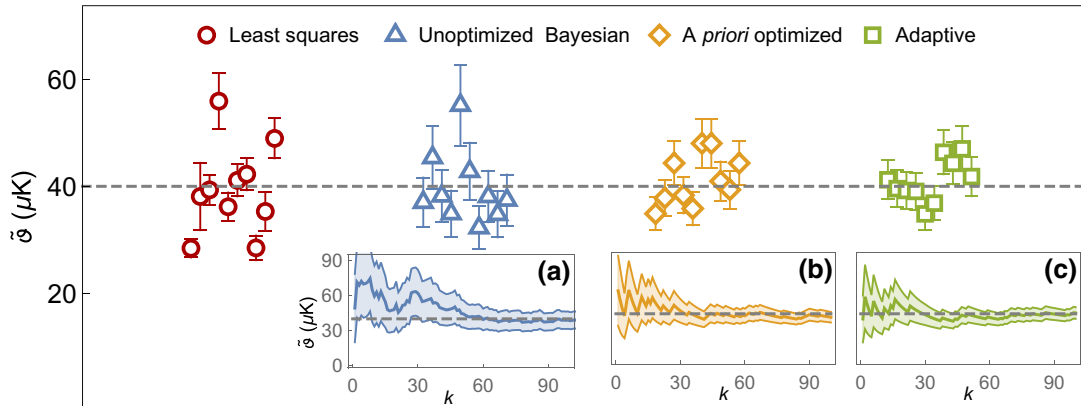


FIG. 1. Reliability of estimation strategies for a trap loaded with a single atom. Temperature estimates and their error bars, obtained from processing simulated release-recapture measurements on a single  $^{41}\text{K}$  atom with different protocols (see Sec. II D for details). The results obtained from ten different realizations are shown. The underlying true temperature is set to  $40 \mu\text{K}$  (dashed gray). First, a record with 30 measurements at seven unoptimized expansion times is fitted to the recapture probability, as discussed in Sec. II B (red circles), as well as processed with our Bayesian-global technique (blue triangles). The time-optimized methods are also applied to the same overall number of data points (i.e., 210), respectively simulated at the optimal *a priori* recapture time  $t_s$  (orange rhombs) and at adaptively chosen variable optimal times (green squares). Optimization of the recapture times noticeably reduces the variability of the estimates. Note that there is no abscissa axis in the main plot; the estimates have been squeezed together to help visualize their variability. Inset (a) shows the typical convergence of global estimates with the number of measurements  $k$ . As shown, time-optimized methods typically require much less data to converge, when compared with their unoptimized counterparts. Also, the final error bars are smaller in (b) and (c).

approximation. The temperature of the few-atom system is determined by the depth of the optical tweezer. Evaporation cooling brings down the temperature from its initial higher value of the atom source.

The release-recapture protocol is realized by switching off the tweezer’s light for a variable amount of time and then switching it on again. To measure the number of atoms left in the tweezer, we utilize fluorescence imaging switching on a series of dedicated resonant beams. The light emitted by the recaptured atoms is collected with the same microscope objective that is used to create the tweezer, and detected with a scientific CMOS camera. In order to avoid antitrapping while imaging [58], we chop the tweezer and imaging lights out of phase at 1 MHz for the duration of the measurement. Once the detection is completed, the  $^{41}\text{K}$  atoms are lost. A new batch of atoms needs to be loaded in the tweezer, and the process above is repeated.

### B. Mapping photon counts into atom numbers

The experimental measurement record is a list of photon counts  $n_{p,i}$  at each shot at time  $t_i$ . We must, however, convert each photon count into a number of recaptured atoms  $n_i$  to be fed into the vector  $\mathbf{n}$ . To do so, we take a set of calibration fluorescence images. The histogram of the resulting photon counts produces a multimodal distribution, where each peak is associated with a different number of atoms in the trap. Applying a multimodal Gaussian fit to these data gives us the position  $m$  of the peak corresponding to an empty trap, as well as the spacing  $\Delta$  between the (equispaced) peaks. Hence, the photon-count-to-atom conversion is simply

$$n_i = \text{round}\left(\frac{n_{p,i} - m}{\Delta}\right);$$

that is, we map each photon-count measurement to the atom number associated with the nearest peak.

### C. Multiatom model

Although the initial atom number can be controlled to some degree, our assumption of deterministic single-atom loading certainly needs to be relaxed before we can process experimental data. This can be achieved by upgrading the measurement model  $p(n_i|T, t_i)$  in Eq. (7).

Assuming distinguishable atoms with negligible mutual interactions, the probability for recapturing  $n_i$  of them during the  $i$ th run, given that the trap is loaded with  $N_0$  atoms, can be cast as

$$p(n_i|N_0, T, t_i) = B[n_i|N_0, f(T, t_i)], \quad (10)$$

where  $B(n_i|N_0, q)$  is the binomial distribution

$$B(n_i|N_0, q) = \binom{N_0}{n_i} q^{n_i} (1 - q)^{N_0 - n_i} \quad (11)$$

with success probability  $q = f(T, t_i)$ , as defined in Eq. (9).

Here  $N_0$  appears as a nuisance parameter and thus needs to be marginalized over. To do so, we first note that, in this type of problem, the probability of starting with  $N_0$  atoms can be described by a Poisson distribution [59,60], denoted

$$P(N_0|\lambda) = \frac{\lambda^{N_0}}{N_0!} e^{-\lambda}, \quad (12)$$

where  $\lambda$  is the mean initial atom number. In practice, this can be inferred from a sufficiently large set of calibration measurements at zero recapture time (i.e., on *unreleased* atoms). In such a case (cf. Appendix D),

$$\begin{aligned} p(n_i|T, t_i) &= \sum_{N_0=0}^{\infty} P(N_0|\lambda) B[n_i|N_0, f(T, t_i)] \\ &= P[n_i|\lambda f(T, t_i)]. \end{aligned} \quad (13)$$

The model for the multiatom scenario is thus a Poisson distribution with mean equal to  $\lambda f(T, t_i)$ . With the sole exception of this change, the Bayesian protocols from Sec. II C can be directly applied to experimental data.

## IV. RESULTS AND DISCUSSION

### A. Processing real data at unoptimized recapture times

Let us start by analyzing experimental measurement records taken at unoptimized light-off times. We apply the conventional method of least-squares fitting, as well as our Bayesian technique. Specifically, we work on two sets of measurements on  $^{41}\text{K}$  atoms, which are performed in a shallow and a deep trapping potential. These have depths  $U_0/k_B = 110 \mu\text{K}$  and  $290 \mu\text{K}$ , respectively. For the shallow configuration, we perform 50 measurements of the number of recaptured atoms at each of the expansion times  $\mathbf{t}_a = (5, 10, 15, 20, 30, 40, 50) \mu\text{s}$ , plus 100 calibration measurements without release. In turn, for the deep trap, we take 30 measurements at expansion times  $\mathbf{t}_b = (5, 10, 20, 30, 50, 70, 100) \mu\text{s}$ , and 60 calibration measurements at zero expansion time.

Let us now process these data via the nonlinear fit to the mean number of recaptured atoms (cf. Sec. II B). We note that the average atom number at  $t = 0$  is also a point for the fit. This returns temperature estimates of  $15.8 \pm 32 \mu\text{K}$  and  $57 \pm 11 \mu\text{K}$  for the shallow and deep configurations, respectively. The fits are shown in Fig. 2 (solid red lines).

We next apply our Bayesian methods to the *same* data. From the calibration measurements, we find the initial

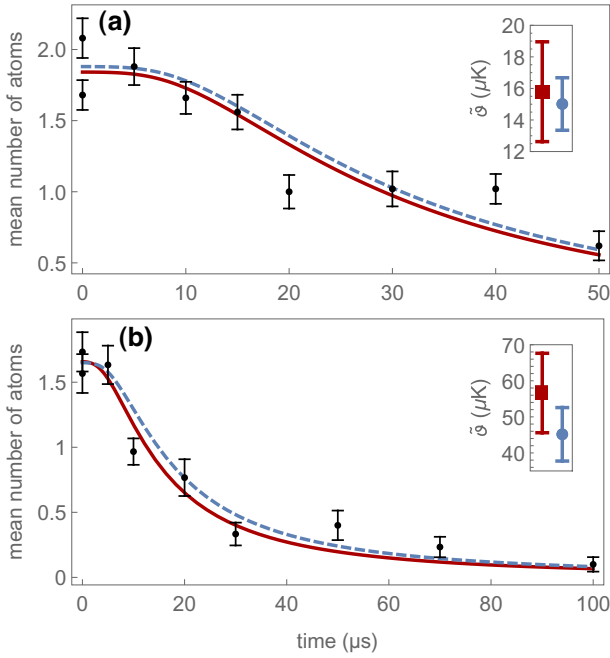


FIG. 2. Temperature estimation from measurements at unoptimized recapture times. Empirical mean and standard error of the number of recaptured  $^{41}\text{K}$  atoms (black dots) at each free-expansion time in the two experiments described in Sec. IV A for a (a) shallow and (b) deep trapping potential. The solid red curves are two-parameter nonlinear fits to  $\lambda f(T, t)$  in  $\lambda$  and  $T$ , where  $f(T, t)$  is the survival probability from Eq. (9). These yield (a)  $T_a = 15.8 \pm 32 \mu\text{K}$  and  $\lambda_a = 1.84 \pm 11$  and (b)  $T_b = 57 \pm 11 \mu\text{K}$  and  $\lambda_b = 1.66 \pm 9$ . In turn, the Bayesian strategy (cf. Sec. II C 1) applied to the full measurement record gives (a)  $T'_a = 15.0 \pm 17 \mu\text{K}$  and (b)  $T'_b = 45 \pm 7 \mu\text{K}$ . The insets show the estimates. The dashed blue curves correspond to  $\lambda f(t, T)$  evaluated at temperatures  $T'_a$  and  $T'_b$ , with initial average atom numbers  $\lambda'_a = 1.88$  and  $\lambda'_b = 1.65$ , as calculated from the calibration measurements.

mean atom numbers  $\lambda = 1.88$  and  $\lambda = 1.65$  for the shallow and the deep traps, respectively. As discussed in Sec. II A, the support of our prior  $p(\theta)$  is set to  $5.5 \leq \theta \leq 30 \mu\text{K}$  (shallow) and  $14.5 \leq \theta \leq 125 \mu\text{K}$  (deep). Following the Bayesian procedure we arrive at the estimates  $15.0 \pm 17 \mu\text{K}$  and  $45 \pm 7 \mu\text{K}$  (dashed blue lines in Fig. 2).

Although the Bayesian and non-Bayesian estimates seem roughly comparable, we note that the Bayesian temperature estimate for the deep-trap experiment does fall below the error bars of the conventional method. Indeed, some discrepancy could have been expected in the second experiment, due to the smaller number of measurements performed. It is precisely in these cases when the Bayesian analysis pays off.

### B. Finding the most reliable strategy

Since the analysis above has shown that conventional and Bayesian processing of the same experimental data can

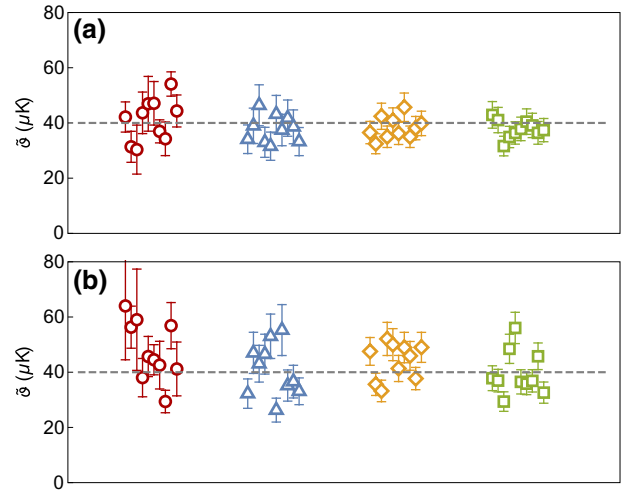


FIG. 3. Estimate variability of different strategies. Temperature estimates drawn from simulated release-recapture measurements on  $^{41}\text{K}$ . The parameters for the simulations are those of the deep-trap experiment from Sec. IV A and Fig. 2(b), and the true temperature is set to  $40 \mu\text{K}$  (dashed gray). In (a), the mean number of atoms loaded into the trap is assumed to be known, and equal to  $\lambda = 1.65$ . In each case, ten independent runs of the experiment are simulated, consisting of 30 measurements generated at each of the seven unoptimized recapture times  $t_b$ . The estimates are calculated via the conventional least-squares method (red circles) and the Bayesian approach (blue triangles). Since  $\lambda$  is known, the least-squares method becomes a one-parameter fit. Orange rhombs and green squares correspond to estimates drawn from the *a priori* optimized and fully adaptive protocols, respectively. These also run on datasets of 210 measurements. The *a priori* optimal recapture time here is  $t_s = 22 \mu\text{s}$ . In (b)  $\lambda$  is not known (its true value continues to be  $\lambda = 1.65$ ). A calibration set of 60 measurements at  $t = 0$  is thus needed, in addition to the 210 data points. In this case, the least-squares method becomes again a two-parameter fit in  $T$  and  $\lambda$ . For all three Bayesian variants,  $\lambda$  is recalculated at every run from a fresh calibration set. Hence, the  $\lambda$  used for estimation differs slightly between independent simulations. The overall variability of the estimates thus increases. This is particularly apparent for both the *a priori* optimized and adaptive protocols, as the optimizations are done with the estimated  $\lambda$ , resulting in slightly off optimal times. Markers in (b) are the same as in (a).

yield different temperature estimates, we need to establish which one is more dependable and should thus be trusted. To that end, we now simulate multiple runs of the deep-trap experiment (see Fig. 3 for details), setting the temperature to  $40 \mu\text{K}$ .

Let us start by focusing on the red circles and blue triangles in Fig. 3(b), which correspond to ten simulated deep-trap experiments processed with the conventional and Bayesian methods, respectively. Each run thus consists of 30 data points simulated at each of the seven recapture times in  $t_b$ , plus 60 calibration measurements, at  $t = 0$ . As we can see, the estimate variability is substantial, although



the Bayesian estimates have a narrower spread and more consistent error bars.

To quantify such variability, we use the empirical error [61]  $\text{Var } \tilde{\vartheta} / \langle \tilde{\vartheta} \rangle^2$  of a set of estimates, where

$$\text{Var } \tilde{\vartheta} = \langle \tilde{\vartheta}^2 \rangle - \langle \tilde{\vartheta} \rangle^2, \quad (14a)$$

$$\langle \tilde{\vartheta} \rangle = \frac{1}{N} \sum_i^N \tilde{\vartheta}_i. \quad (14b)$$

For these parameters, and considering 100 simulated runs, the variability of the conventional least-squares method evaluated to 0.064, while the Bayesian analysis proved slightly more reliable, with  $\text{Var } \tilde{\vartheta} / \langle \tilde{\vartheta} \rangle^2 = 0.057$ .

However, the full potential of the Bayesian method is unleashed when the light-off times are picked to maximize information gain. Back to Fig. 3(b), we now turn our attention to the results for the *a priori* optimized and fully adaptive protocols (orange rhombs and green squares, respectively). These use datasets of the same length as the unoptimized schemes (i.e., 210 points and 60 calibration measurements). In order to evaluate the information gain in Eq. (5), we must truncate the Poisson distribution (12) for the number of atoms initially loaded into the trap. Specifically, we set  $N_0 = 7$  as the maximum, since the probability of loading more atoms is negligible for  $\lambda = 1.65$ , which is the true value of the initial number of atoms in our simulations. It is important to emphasize that  $\lambda$  is assumed unknown when processing data and, thus, needs to be recalculated from the calibration data at every simulated run (see the caption of Fig. 3). Furthermore, in order to account for the finite resolution when adjusting the recapture times in the lab, in our calculations we replace the exact optimal times by integer multiples of a minimum resolution of 2  $\mu\text{s}$ .

As it could be expected in light of Fig. 1, the fully adaptive protocol is the most dependable, visibly outperforming its unoptimized counterparts. In terms of variability,  $\text{Var } \tilde{\vartheta} / \langle \tilde{\vartheta} \rangle^2 = 0.034$  for 100 runs. That is a 40% improvement with respect to the Bayesian processing of unoptimized data, and over 50%, with respect to the conventional method. Interestingly, the *a priori* optimized protocol is comparable in variability to the fully adaptive one and yet it is much simpler to implement. This is our first main result, and it supports using the *a priori* optimized protocol as the new default when it comes to practical release-recapture thermometry.

The fact that the initial average number of atoms in the tweezer is unknown certainly interferes with estimate variability. Namely, the empirical average of the simulated calibration data differs slightly from run to run, even if  $\lambda$  is fixed in the simulation. In order to single out the temperature-estimation aspect of the calculations, we also run variability checks, treating  $\lambda$  as a *known* parameter. As we can see in Fig. 3(a), the overall variability is much

smaller in this case, but the hierarchy among the protocols is maintained.

As an alternative, we might want to adapt our Bayesian-global processing to a multiparameter scenario, so that both  $\lambda$  and  $T$  can be estimated. The procedure is summarized in Appendix E. Although this certainly deserves further investigation, Bayesian multiparameter adaptive estimation seems, at this stage, too time-consuming to be practical.

### C. Assessing convergence speed

So far, we have shown that taking release-recapture data at the fixed *a priori* optimal expansion time  $t_s$ , and processing them with our global thermometry technique yields a clear advantage—in terms of reliability—over the conventional unoptimized protocol. However, in order to optimize the experimental resources it is also important to consider the speed of convergence that is, the amount of data needed for the variability to settle into an asymptotic regime, and the rate at which the variability is reduced once in that limit.

In Fig. 4 we compare the conventional non-linear-fit method and our *a priori* optimized Bayesian-global protocol in terms of convergence. We do this using the same parameters as in Sec. IV B. Specifically, we compute the estimate variability from Eqs. (14) for 5000 simulated runs of the experiment, and gradually increase the number of measurements per run  $k$ . As can be seen from the figure, as  $k$  grows, the variability of both protocols eventually scales as

$$\frac{\text{Var } \tilde{\vartheta}}{\langle \tilde{\vartheta} \rangle^2} \sim \frac{1}{kF},$$

where  $F$  is a free parameter [62]. To get a rough idea of the onset of this scaling, we take the logarithm of the calculated variabilities and fit it to  $-a \log k - \log F$ . We start from the largest  $k$  and work backwards, adding points at lower  $k$  to the fit. We then search for the number of shots at which the exponent  $a$  first deviates from 1 by more than 2.5%. This gives  $k_c \simeq 350$  for the conventional method and  $k_{\text{ap}} \simeq 200$  for our *a priori* optimized protocol.

Hence, for  $k > k_c$ , we can be certain that both protocols have entered the asymptotic regime in  $k$ . From their ratio we find that the *a priori* optimized method affords a 43% variability reduction with respect to the common practice in the release-recapture method, even in the asymptotic limit. We have thus shown that our method is not only more reliable when processing short data sets, but that it also requires less measurements to attain the same precision as the conventional protocol. In our experimental setup, this roughly translates into *halving* the number of measurements needed to hit any target precision. This is our second main result.

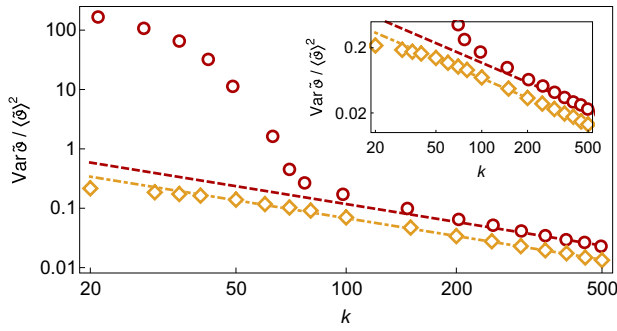


FIG. 4. Convergence of conventional and *a priori* optimized protocols. Log-log plot of variability versus the number of measurements  $k$ , calculated from 5000 simulated runs of the deep-trap experiment from Sec. IV B. The measurement count  $k$  excludes any calibration measurements at  $t = 0$ . Specifically,  $(2/7)k$  calibration outcomes are generated for each run. All other parameters are the same as in Fig. 3. The conventional non-linear-fit release-recapture (red circles) is compared to our *a priori* optimized protocol (orange rhombs). Note the vast superiority of the optimized method for low  $k$ , i.e., scarce data. Asymptotically, the variability takes the approximate form  $1/(kF)$ . This occurs for  $k > 350$  for the conventional protocol and  $k > 200$  for the optimized one. The dashed and dash-dot lines highlight the asymptotic variability scaling for both methods. The offset between these lines indicates that the asymptotic convergence speed to the true temperature is roughly twice as fast for the *a priori* optimized protocol. The insets enlarges the region in which our Bayesian technique enters the asymptotic regime.

#### D. Time optimization on an individual experiment

We now illustrate what an *a priori* optimized release-recapture experiment might look like (cf. Fig. 5). To that end, we retrospectively analyze the data from the shallow-trap experiment processed in Fig. 2(a). Just like in Sec. IV B, we cap the distribution of initial atom numbers in the trap at  $N_0 = 7$ , which allows us to compute  $t_s = 42 \mu\text{s}$  (see the inset of Fig. 5). We then process, with our Bayesian-global technique, those measurement outcomes captured at the times closest to  $t_s$ . In this case, those would be the 50 measurements at  $40 \mu\text{s}$ . The *a priori* optimized method from Sec. II C 2 would demand to continue measuring at the same time. However, having run out of data, we move on to process the measurements at the second closest time in the dataset (i.e.,  $50 \mu\text{s}$ ), and so on.

The estimate and error bars resulting from the Bayesian processing of the first  $k$  entries of the reordered record are plotted in orange in Fig. 5. We observe that over 100 points seemingly add no information about temperature, and could thus have been spared. That is, by aiming to uniformly sample the decay of the survival probability, nearly 30% of the experimental resources are wasted. To further emphasize the important difference that time optimization can make to the convergence of the estimate, we overlay (in blue) a similar plot, in which the measurement record is ordered from lowest to largest information gain per entry.

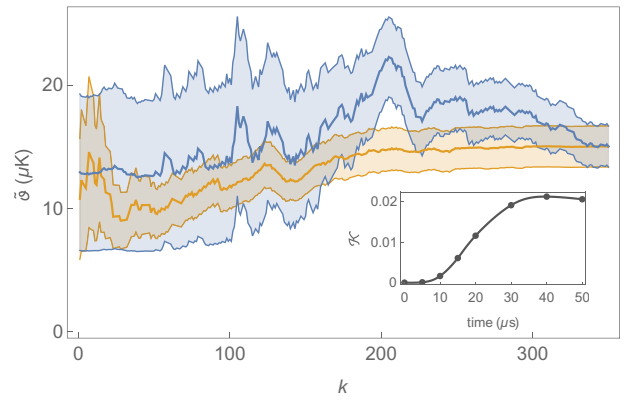


FIG. 5. Recapture-time optimization. Estimate and error bar from the first  $k$  entries of the measurement record of the shallow-trap experiment from Fig. 2(a), reordered to mimic the *a priori* optimized protocol (orange). The optimal recapture time  $t_s = 42 \mu\text{s}$  is calculated as per Sec. II C 2. The *a priori* information gain  $\mathcal{K}$  as a function of the recapture time is shown in the inset for this model (gray dots). The gray line is a mere guide to the eye. The 50 measurements at the closest recorded time ( $40 \mu\text{s}$ ) are moved to the top of the list. The set of measurements at the second closest recorded time ( $50 \mu\text{s}$ ) are picked next, and so forth, until exhausting the 350 available data. The estimate and error bars of that exact same list, but processed in the reverse order, are shown in blue for comparison. As we can see, the orange curve already converges to the final estimate by processing the 200 most informative data. At that point, the estimate is  $14.7 \pm 1.7 \mu\text{K}$ . The remaining 150 data carry practically no information, since the estimate drawn from the full measurement record is  $15 \pm 1.7 \mu\text{K}$ .

Hence, the effect of optimizing over the recapture time to increase information gain is threefold. First, it produces more reliable estimates when data are scarce. Second, it ensures a faster convergence towards the true temperature in the asymptotic limit of many measurements. Finally, it allows us to significantly reduce the number of data points needed for the point estimate to converge (i.e., to stabilize at a certain value) in a *single* experimental run. The latter observation is the final main result of this work.

#### E. A priori optimized protocol versus fully adaptive protocol

As illustrated in Sec. IV B, the *a priori* optimized and the adaptive protocols can have almost identical estimate variability. This advised us to focus on the former, due to its simplicity. In some situations, however, the effort of adaptively recalculating optimal expansions in real time might pay off. To see this, we go back to the deep-trap configuration of Fig. 2(b), to make another showdown between these two protocols. Proceeding exactly as in Sec. IV B, we simulate ten independent experiments for each method but, in this case, we work with a wider support

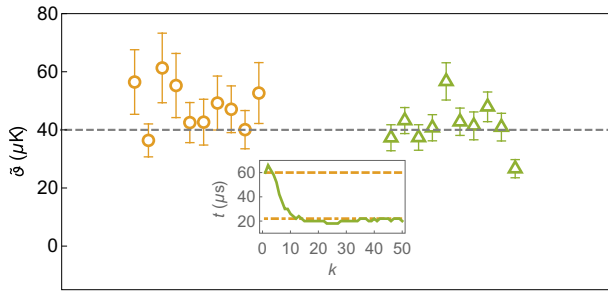


FIG. 6. *A priori* optimized protocol versus adaptive protocol. Comparison between the *a priori* optimized (orange rhombs) and fully adaptive (green squares) protocols. Ten independent runs of the deep-trap experiment are simulated for each method. All parameters are the same as in Fig. 3(b), except for  $\theta_{\min}$ , which is set here to 0.5% of the trap depth  $U_0/k_B$  (see Sec. IV E). The inset shows the optimal expansion time as a function of  $k$  for one of the simulated measurement records (solid green). The *a priori* optimal recapture time for this wider hypothesis support is indicated by the orange dashed line. In contrast, the *a priori* optimal time for the narrower support used in Figs. 1–5 is the orange dash-dot line. Our observation that the *a priori* optimized method seems to perform equally well than the arguably superior fully adaptive scheme is explained by the fact that, for the narrower prior,  $t_s$  almost coincides with the true optimum for most of the experiment.

for our prior, stretching it by one order of magnitude into colder temperatures. Namely,  $1.45 \leq \theta \leq 125 \mu\text{K}$ .

As we can see in Fig. 6, the adaptive protocol exhibits here a reduced variability with respect to the *a priori* optimized method. This is due to the fact that the *a priori* optimal time is nearly  $60 \mu\text{s}$ , while the true optimal release-recapture delay stabilizes around  $20 \mu\text{s}$  after few measurements (see the inset of Fig. 6). This quickly renders the *a priori* optimized method inefficient when it comes to information acquisition. The fully adaptive protocol, however, does not share this problem.

## V. CONCLUSIONS

In this paper we have demonstrated how temperature measurements can be substantially improved in practice by actively exploiting Bayesian-global thermometry techniques. We do so on a concrete experimental setup, namely, release-recapture measurements on few tightly confined  $^{41}\text{K}$  atoms at  $\mu\text{K}$  temperatures. The conventional protocol samples the fraction of recaptured atoms evenly and fits the data to a model for the recapture probability. Instead, we propose to choose the delay between release and recapture so as to maximize information gain, and to process the measured data with our Bayesian technique [35].

Working with real experimental data as well numerical simulations, we have shown that our method has three important advantages. First, the variability of estimates extracted from small data sets is largely reduced when compared to the conventional protocol. This results in

much more dependable thermometry. Second, time optimization also leads to a faster convergence towards the true temperature in the asymptotic limit of many measurements. Finally, the optimization of the recapture time substantially reduces the number of measurements needed for estimates to stabilize to a final temperature reading. Put differently, the conventional approach wastes resources by performing and processing a sizeable proportion of uninformative measurements.

We have also demonstrated that the best results are obtained with a fully adaptive approach, in which each new release-recapture interval takes in all previous measurements. Interestingly, a simpler protocol can yield surprisingly competitive estimates, namely, measuring repeatedly at a single expansion time—the *a priori* optimal one. That is, the delay that maximizes the information gain in the first measurement. This follows from the parameters of the experiment and the range of temperatures being probed. Other than an arbitrarily broad temperature range, we have assumed no prior knowledge about the temperature.

We have shown how quantum thermometry can deliver practical solutions leading to quantifiable precision enhancements, once the quantum Fisher information-based local paradigm [18] is abandoned in favor of the global framework. While the former has proven useful when studying fundamental precision limits and maximizing the *responsiveness* of temperature sensors, the more general Bayesian framework can, in addition to that [40,47], process finite experimental records optimally. Importantly, the exact same fundamental principles can be applied to other techniques in different temperature ranges and experimental platforms, which opens a new exciting avenue in quantum thermometry.

## ACKNOWLEDGMENTS

We thank Robert Smith for useful comments. J.G. is funded by the College of Engineering, Mathematics and Physical Sciences of the University of Exeter. J.R. acknowledges support from EPSRC (Grants No. EP/T002875/1 and No. EP/R045577/1). The work of R.S., T.H., and G.B. is supported by the Leverhulme Trust Research Project Grant UltraQuTe (Grant No. RGP-2018-266).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, J.G., upon reasonable request.

## APPENDIX A: SCALE-INVARIANT THERMOMETRY

This appendix applies the framework of (quantum) scale estimation in Refs. [35,40] to release-recapture thermometry.

## 1. Temperature as a scale parameter

Let us first examine the multiatom measurement model for our experiment. That is, the probability function  $p(n_i|T, t_i)$  for recapturing  $n_i$  atoms after time  $t_i$  during the  $i$ th run, given that their temperature is  $T$ . As shown in Secs. II B and III C and in Appendix C,  $p(n_i|T, t_i)$  takes the form

$$p(n_i|T, t_i) = \frac{e^{-\lambda f(T, t_i)}}{n_i!} [\lambda f(T, t_i)]^{n_i}. \quad (\text{A1})$$

Since both temperature and the other dimensioned quantities appear only within the fraction of recaptured atoms, we can base our analysis on the function  $f(T, t)$  alone. This implies that the discussion in this appendix also applies to the single-atom model in the main text.

Let us now define the kinetic energy

$$E_k := \frac{1}{2} m \left( \frac{l}{t} \right)^2 \quad (\text{A2})$$

of an atom having moved by the characteristic length scale  $l = w_0/\sqrt{2}$  at time  $t$ . Using this, the fraction of recaptured atoms can be cast as

$$f(T, t) = \frac{g\{[E_k/(k_B T)]W(U_0/E_k)\}}{g[U_0/(k_B T)]}. \quad (\text{A3})$$

That is, our statistics is invariant under transformations  $U_0 \mapsto \gamma U_0$ ,  $E_k \mapsto \gamma E_k$ , and  $T \mapsto \gamma T$ , where  $\gamma$  is an arbitrary positive constant. This implies that temperature sets the energy scale of the problem, and so it is a *scale parameter* from the point of view of estimation theory [33,35,40,42,46].

## 2. Protocol optimization

This work optimizes release-recapture experiments in two ways. First, by constructing an estimator function mapping the measurement outcomes  $\mathbf{n} = (n_1, \dots, n_\mu)$  and the release times  $\mathbf{t} = (t_1, \dots, t_\mu)$  to the *optimal* temperature estimate  $\tilde{\vartheta}(\mathbf{n}, \mathbf{t})$ , where  $\mu$  is the number of experimental runs. Second, by identifying the most informative values for the release times  $\mathbf{t} = (t_1, \dots, t_\mu)$ .

Given that  $T$  is unknown, both optimizations need to be carried out with respect to a temperature-independent measure of uncertainty. In addition, they need to be performed *before* the experiment is run (see Sec. II C), meaning that the uncertainty must also be outcome independent. We thus use the uncertainty quantifier

$$\bar{\epsilon}(\mathbf{t}) = \int d\mathbf{n} p(\mathbf{n}|\mathbf{t}) \bar{\epsilon}(\mathbf{n}, \mathbf{t}), \quad (\text{A4})$$

where

$$\bar{\epsilon}(\mathbf{n}, \mathbf{t}) = \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) \mathcal{D}[\tilde{\theta}(\mathbf{n}, \mathbf{t}), \theta]. \quad (\text{A5})$$

Here,  $\theta$  is a hypothesis about the true value of  $T$ ,  $\tilde{\theta}(\mathbf{n}, \mathbf{t})$  is a generic temperature estimator, and  $\mathcal{D}[\tilde{\theta}(\mathbf{n}, \mathbf{t}), \theta]$  is a deviation function gauging the deviation of  $\tilde{\theta}(\mathbf{n}, \mathbf{t})$  from  $\theta$ . Furthermore,  $p(\mathbf{n}|\mathbf{t}) = \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) p(\mathbf{n}|\theta, \mathbf{t})$ , while the probabilities  $p(\theta)$ ,  $p(\mathbf{n}|\theta, \mathbf{t})$ , and  $p(\theta|\mathbf{n}, \mathbf{t})$  are defined in the main text and in Appendix B.

To use Eqs. (A4) and (A5), we must choose a deviation function  $\mathcal{D}$ . As shown in Ref. [35], the scale-invariant nature of the protocols in this work (see Sec. A 1) leads to the logarithmic deviation function  $\mathcal{D}[\tilde{\theta}(\mathbf{n}, \mathbf{t}), \theta] = \log^2[\tilde{\theta}(\mathbf{n}, \mathbf{t})/\theta]$ . Using this, Eqs. (A4) and (A5) become the mean logarithmic errors  $\bar{\epsilon}(\mathbf{t}) \mapsto \bar{\epsilon}_{\text{MLE}}(\mathbf{t})$  and

$$\bar{\epsilon}(\mathbf{n}, \mathbf{t}) \mapsto \bar{\epsilon}_{\text{MLE}}(\mathbf{n}, \mathbf{t}) = \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) \log^2 \left[ \frac{\tilde{\theta}(\mathbf{n}, \mathbf{t})}{\theta} \right], \quad (\text{A6})$$

respectively. Unlike the more familiar square errors, which are based on the deviation function  $[\tilde{\theta}(\mathbf{n}, \mathbf{t}) - \theta]^2$  [33,53], logarithmic errors respect the scale invariance of the problem, i.e.,  $\mathcal{D}(\gamma\tilde{\theta}, \gamma\theta) = \mathcal{D}(\tilde{\theta}, \theta)$ , and they do so while relying on minimal assumptions [35,40]; hence, their use in this work. We can then minimize  $\bar{\epsilon}_{\text{MLE}}(\mathbf{t})$  with respect to  $\tilde{\theta}(\mathbf{n}, \mathbf{t})$ , finding the optimal temperature estimator [35]

$$\tilde{\vartheta}(\mathbf{n}, \mathbf{t}) = \theta_u \exp \left[ \int d\theta p(\theta|\mathbf{n}, \mathbf{t}) \log \left( \frac{\theta}{\theta_u} \right) \right] \quad (\text{A7})$$

used in the main text.

To find the optimal release times  $\mathbf{t}$ , we first insert the estimator above into the expression for  $\bar{\epsilon}_{\text{MLE}}(\mathbf{t})$ , arriving at the fundamental lower bound  $\bar{\epsilon}_{\text{MLE}}(\mathbf{t}) \geq \bar{\epsilon}_p - \mathcal{K}(\mathbf{t})$  [35,40]. Such a bound is the result of taking the uncertainty prior to performing the experiment, denoted by  $\bar{\epsilon}_p$ , and subtracting the mean information gain  $\mathcal{K}(\mathbf{t})$  associated with the measurement protocol under consideration. The expression for  $\mathcal{K}(\mathbf{t})$  is given in the main text for a single shot, while that for  $\bar{\epsilon}_p$  can be found in Refs. [35,40]. For our purposes, the key property of  $\bar{\epsilon}_p$  is its lack of dependence on the release times  $\mathbf{t}$ . Given this, if we wish to optimize our protocol by minimizing the bound above for a single shot, as done in this work, we simply need to replace  $\mathbf{t} \mapsto t$  and maximize  $\mathcal{K}(t)$  with respect to  $t$ .

We finally note that, once the optimal strategy has been found, the theoretical error in Eq. (A4) ceases to be relevant from an experimental point of view. Indeed, since our final goal is processing real data, the associated error bars must rely on a temperature-independent but outcome-dependant uncertainty, i.e., that in Eq. (A6).

## APPENDIX B: POSTERIOR PROBABILITY

In this appendix we derive Eq. (1). First, the posterior probability  $p(\theta|\mathbf{n}, \mathbf{t})$  is given by Bayes theorem as [33,34]

$$p(\theta|\mathbf{n}, \mathbf{t}) \propto p(\theta|\mathbf{t}) p(\mathbf{n}|\theta, \mathbf{t}).$$

We now note that the expansion times  $\mathbf{t} = (t_1, \dots, t_\mu)$  do not (by themselves) inform how likely different hypotheses  $\theta$  about  $T$  are. Hence, we take  $p(\theta|\mathbf{t}) \mapsto p(\theta)$ . Assuming statistical independence between measurements further leads to

$$p(\mathbf{n}|\theta, \mathbf{t}) = \prod_{i=1}^{\mu} p(n_i|\theta, t_i).$$

But we also see that expansion times other than  $t_i$  do not inform the likelihood for the number of recaptured atoms  $n_i$  recorded in the  $i$ th measurement. One thus finally arrives at

$$p(\theta|\mathbf{n}, \mathbf{t}) = \frac{1}{A} p(\theta) \prod_{i=1}^{\mu} p(n_i|\theta, t_i),$$

as stated in the main text. The normalization factor  $A$  is found by integrating over the hypothesis range  $[\theta_{\min}, \theta_{\max}]$ , i.e.,

$$A = \int_{\theta_{\min}}^{\theta_{\max}} d\theta p(\theta) \prod_{i=1}^{\mu} p(n_i|\theta, t_i).$$

### APPENDIX C: CALCULATION OF THE RECAPTURED FRACTION

For the sake of completeness, we summarize the main steps of the derivation of Eq. (9) as given in Ref. [51]. Let us load  $N_0$  atoms of mass  $m$  at low temperature, i.e.,  $k_B T \ll U_0$ , into the tweezer trap. Specifically, the temperature is low enough so that the atoms are well localized at the center of the trap. Furthermore, we neglect interatomic interactions, the influence of gravity, and the movement along the  $z$  axis. After a free-expansion time  $t$ , the trap is reactivated. An atom may only be recaptured then if

$$\frac{1}{2} m \left( \frac{r}{t} \right)^2 \leq U_0 \exp\left(-\frac{2r^2}{w_0^2}\right),$$

where  $r$  is the distance from the center of the trap (see Sec. II B). One then defines  $r_{\max}$  as the distance that saturates the above inequality. It is convenient to introduce the dimensionless variables  $\tilde{r}^2 = 2r^2/w_0^2$  and  $\tilde{t}^2 = 4U_0 t^2/(mw_0^2)$ , so that  $\tilde{t}^2 = \tilde{r}_{\max}^2 \exp(\tilde{r}_{\max}^2)$  or, equivalently,  $\tilde{r}_{\max}^2 = W(\tilde{t}^2)$ , where  $W(\cdot)$  is the product-logarithm or Lambert- $W$  function defined in the main text. Hence,  $r_{\max}^2 = w_0^2 W(\tilde{t}^2)/2$ , so the maximal speed of a recaptured atom is

$$v_{\max}^2(t) = \left( \frac{r_{\max}}{t} \right)^2 = w_0^2 \frac{W(\tilde{t}^2)}{2t^2}.$$

Only those particles with velocity below  $v_{\max}(0) = \sqrt{2U_0/m}$  would be initially captured by the trap. On the other hand, only those with velocity under  $v_{\max}(t)$  would

be recaptured after free expansion during time  $t$ . Therefore, the recapture fraction simply reads

$$f(T, t) = \frac{\int_0^{v_{\max}(t)} dv F(v)}{\int_0^{v_{\max}(0)} dv F(v)} = \frac{1}{g(\eta)} g\left[\frac{\eta W(\tilde{t}^2)}{\tilde{t}^2}\right], \quad (\text{C1})$$

where  $F(v) = (mv/k_B T) \exp[-mv^2/(2k_B T)]$  is the Maxwell-Boltzmann distribution in two dimensions,  $g(s) = 1 - e^{-s}$ , and  $\eta = U_0/(k_B T)$ . Equation (C1) follows from

$$\int_0^u dv F(v) = g\left(\frac{mu^2}{2k_B T}\right).$$

Equation (C1) is useful due to its simplicity. However, we may loosen most of the underlying assumptions by simulating many individual trajectories of atoms. At low-enough temperatures, these can be sampled from a Gaussian distribution in position as well as in velocity in spite of the anharmonicity of the trap. The effects of gravity and the subsequent axial motion can also be included.

As we can see in Fig. 7, our simple model from Eq. (C1) is in very good agreement with the numerically calculated recaptured fraction over a broad range of temperatures and recapture times. As expected, some deviations start to be appreciated as the temperature nears  $U_0/k_B$ .

It must be noted that limiting ourselves to the regime in which Eq. (9) is applicable is not as restrictive as it might seem. As we can see, the numerically calculated recaptured fraction becomes almost independent of  $T$  when the model and simulation start to diverge. Hence, when Eq. (9) ceases to apply, release-recapture thermometry itself becomes inefficient.

### APPENDIX D: PROOF OF EQ. (13) IN THE MAIN TEXT

Starting from the left-hand side of Eq. (13), we have

$$\begin{aligned} p(n_i|T, t_i) &= \sum_{N_0=0}^{\infty} P(N_0|\lambda) B[n_i|N_0, q = f(T, t_i)] \\ &= \sum_{N_0=0}^{\infty} \frac{\lambda^{N_0} e^{-\lambda}}{N_0!} \binom{N_0}{n_i} q^{n_i} (1-q)^{N_0-n_i}, \end{aligned}$$

where we recall that  $q = f(T, t_i)$ . Given that

$$\binom{N_0}{n_i} = 0$$

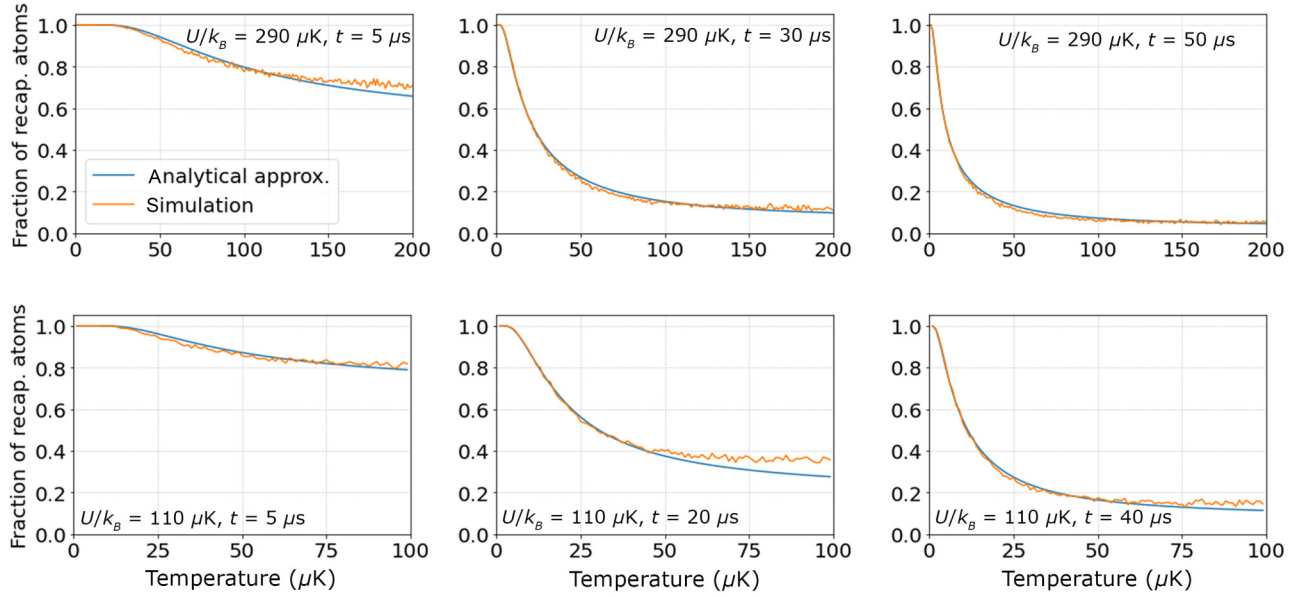


FIG. 7. Regime of validity of Eq. (C1) for the recaptured fraction. The analytical approximation (C1) (blue) is compared to the recaptured fraction computed from 5000 individual simulated trajectories of atoms sampled from a thermal distribution (orange), as a function of temperature, for different recapture times. Axial motion due to gravity is also factored in the calculation. As can be seen, for a deep trap with  $U_0/k_B = 290 \mu\text{K}$  (top row), the approximation holds well up to about  $125 \mu\text{K}$ , while it breaks down at about  $30 \mu\text{K}$  in a shallow-trap configuration (bottom row) with  $U_0/k_B = 110 \mu\text{K}$ . Simulations are for  $^{41}\text{K}$  and the trap geometry described in Sec. III A.

for  $N_0 < n_i$ , we only need to sum from  $N_0 = n_i$ , i.e.,

$$\begin{aligned} p(n_i|T, t_i) &= \sum_{N_0=n_i}^{\infty} \frac{\lambda^{N_0} e^{-\lambda}}{N_0!} \binom{N_0}{n_i} q^{n_i} (1-q)^{N_0-n_i}, \\ &= \sum_{N_0=n_i}^{\infty} \frac{\lambda^{N_0} e^{-\lambda}}{N_0!} \frac{N_0!}{n_i!(N_0-n_i)!} q^{n_i} (1-q)^{N_0-n_i} \\ &= \frac{q^{n_i} e^{-\lambda}}{n_i!} \sum_{N_0=n_i}^{\infty} \frac{\lambda^{N_0}}{(N_0-n_i)!} (1-q)^{N_0-n_i}. \end{aligned}$$

Next, we introduce the new variable  $j = N_0 - n_i$ , so that

$$\begin{aligned} p(n_i|T, t_i) &= \frac{q^{n_i} e^{-\lambda}}{n_i!} \sum_{j=0}^{\infty} \frac{\lambda^{j+n_i}}{j!} (1-q)^j \\ &= \frac{(\lambda q)^{n_i} e^{-\lambda}}{n_i!} \sum_{j=0}^{\infty} \frac{[\lambda(1-q)]^j}{j!}. \end{aligned}$$

But the remaining sum is just the Taylor expansion of the exponential function  $e^x$  around  $x = 0$ . Therefore,

$$\begin{aligned} p(n_i|T, t_i) &= \frac{(\lambda q)^{n_i} e^{-\lambda}}{n_i!} e^{\lambda(1-q)} \\ &= \frac{(\lambda q)^{n_i} e^{-\lambda q}}{n_i!} \\ &= P(n_i|\lambda q), \end{aligned} \quad (\text{D1})$$

which is a Poisson distribution with mean  $\lambda q = \lambda f(T, t_i)$ .

## APPENDIX E: BAYESIAN MULTIPARAMETER ESTIMATION

Our calculations assume that  $\lambda$  in the likelihood model  $p(n_i|T, t_i) = P[n_i|\lambda f(T, t_i)]$  is perfectly known. Yet, our value for  $\lambda$  comes from calibration measurements, and its uncertainty has not been taken into account explicitly. While, as shown in Sec. IV B, our results would remain unchanged even if  $\lambda$  were exactly known, here we summarize, for the sake of completeness, how our global-Bayesian technique can be generalized for multiparameter estimation.

Firstly, we must think about the likelihood as a two-parameter model, i.e.,  $p(n_i|\lambda, T, t_i) = P[n_i|\lambda f(T, t_i)]$ . The corresponding posterior would then be constructed as

$$p(\theta, \lambda | \mathbf{n}, \mathbf{t}) \propto p(\theta, \lambda) \prod_{i=1}^{\mu} p(n_i|\lambda, \theta, t_i), \quad (\text{E1})$$

in which we additionally allow for uncertainty in  $\lambda$  through  $p(\theta, \lambda)$ . We may then treat  $\lambda$  as a nuisance parameter, which leads to the temperature estimator

$$\tilde{\vartheta}(\mathbf{n}, \mathbf{t}) = \theta_u \exp \left[ \int d\lambda d\theta p(\lambda, \theta | \mathbf{n}, \mathbf{t}) \log \left( \frac{\theta}{\theta_u} \right) \right]$$

with mean logarithmic error

$$\bar{\epsilon}_{\text{MLE}}(\mathbf{n}, \mathbf{t}) = \int d\lambda d\theta p(\lambda, \theta | \mathbf{n}, \mathbf{t}) \log^2 \left[ \frac{\tilde{\vartheta}(\mathbf{n}, \mathbf{t})}{\theta} \right].$$

If we instead wish to estimate  $\lambda$ , we can do so by marginalizing over  $\theta$  in Eq. (E1).

- 
- [1] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, Single-atom-resolved fluorescence imaging of an atomic Mott insulator, *Nature* **467**, 68 (2010).
- [2] I. Bloch, J. Dalibard, and S. Nascimbene, Quantum simulations with ultracold quantum gases, *Nat. Phys.* **8**, 267 (2012).
- [3] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, and W. W. Ho, *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, *Nature* **595**, 227 (2021).
- [4] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, and T. Lahaye, *et al.*, Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms, *Nature* **595**, 233 (2021).
- [5] T. Langen, R. Geiger, M. Kuhnert, B. Rauer, and J. Schmiedmayer, Local emergence of thermal correlations in an isolated quantum many-body system, *Nat. Phys.* **9**, 640 (2013).
- [6] T. Langen, R. Geiger, and J. Schmiedmayer, Ultracold atoms out of equilibrium, *Annu. Rev. Condens. Matter Phys.* **6**, 201 (2015).
- [7] Q. Bouton, J. Nettersheim, D. Adam, F. Schmidt, D. Mayer, T. Lausch, E. Tiemann, and A. Widera, Single-Atom Quantum Probes for Ultracold Gases Boosted by Nonequilibrium Spin Dynamics, *Phys. Rev. X* **10**, 011018 (2020).
- [8] W. Niedenzu, I. Mazets, G. Kurizki, and F. Jendrzejewski, Quantized refrigerator for an atomic cloud, *Quantum* **3**, 155 (2019).
- [9] G. Barontini and M. Paternostro, Ultra-cold single-atom quantum heat engines, *New J. Phys.* **21**, 063019 (2019).
- [10] Q. Bouton, J. Nettersheim, S. Burgardt, D. Adam, E. Lutz, and A. Widera, A quantum heat engine driven by atomic collisions, *Nat. Commun.* **12**, 1 (2021).
- [11] F. Serwane, G. Zürn, T. Lompe, T. Ottenstein, A. Wenz, and S. Jochim, Deterministic preparation of a tunable few-fermion system, *Science* **332**, 336 (2011).
- [12] A. M. Kaufman, B. J. Lester, and C. A. Regal, Cooling a Single Atom in an Optical Tweezer to Its Quantum Ground State, *Phys. Rev. X* **2**, 041014 (2012).
- [13] G. Li, S. Zhang, L. Isenhowe, K. Maller, and M. Saffman, Crossed vortex bottle beam trap for single-atom qubits, *Opt. Lett.* **37**, 851 (2012).
- [14] S. Chu, L. Hollberg, J. E. Bjorkholm, A. Cable, and A. Ashkin, Three-Dimensional Viscous Confinement and Cooling of Atoms by Resonance Radiation Pressure, *Phys. Rev. Lett.* **55**, 48 (1985).
- [15] P. D. Lett, R. N. Watts, C. I. Westbrook, W. D. Phillips, P. L. Gould, and H. J. Metcalf, Observation of Atoms Laser Cooled below the Doppler Limit, *Phys. Rev. Lett.* **61**, 169 (1988).
- [16] M. Mudrich, S. Kraft, K. Singer, R. Grimm, A. Mosk, and M. Weidemüller, Sympathetic Cooling with Two Atomic Species in an Optical Trap, *Phys. Rev. Lett.* **88**, 253001 (2002).
- [17] A. De Pasquale and T. M. Stace, in *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions*, edited by F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso (Springer International Publishing, Cham, 2018), p. 503.
- [18] M. Mehboudi, A. Sanpera, and L. A. Correa, Thermometry in the quantum regime: Recent theoretical progress, *J. Phys. A: Math. Theor.* **52**, 011611 (2019).
- [19] K. V. Hovhannisyanyan and L. A. Correa, Measuring the temperature of cold many-body quantum systems, *Phys. Rev. B* **98**, 045101 (2018).
- [20] P. P. Potts, J. B. Brask, and N. Brunner, Fundamental limits on low-temperature quantum thermometry with finite resolution, *Quantum* **3**, 161 (2019).
- [21] I. Henao, K. V. Hovhannisyanyan, and R. Uzdin, Thermometric machine for ultraprecise thermometry of low temperatures, (2021), [ArXiv:2108.10469](https://arxiv.org/abs/2108.10469).
- [22] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, Individual Quantum Probes for Optimal Thermometry, *Phys. Rev. Lett.* **114**, 220405 (2015).
- [23] M. Płodzień, R. Demkowicz-Dobrzański, and T. Sowiński, Few-fermion thermometry, *Phys. Rev. A* **97**, 063619 (2018).
- [24] V. Mukherjee, A. Zwick, A. Ghosh, X. Chen, and G. Kurizki, Enhanced precision bound of low-temperature quantum thermometry via dynamical control, *Commun. Phys.* **2**, 162 (2019).
- [25] M. T. Mitchison, T. Fogarty, G. Guarnieri, S. Campbell, T. Busch, and J. Goold, In Situ Thermometry of a Cold Fermi Gas via Dephasing Impurities, *Phys. Rev. Lett.* **125**, 080402 (2020).
- [26] J. Glatthard and L. A. Correa, Bending the rules of low-temperature thermometry with periodic driving, *Quantum* **6**, 705 (2022).
- [27] L. A. Correa, M. Perarnau-Llobet, K. V. Hovhannisyanyan, S. Hernández-Santana, M. Mehboudi, and A. Sanpera, Enhancement of low-temperature thermometry by strong coupling, *Phys. Rev. A* **96**, 062103 (2017).
- [28] S. Seah, S. Nimmrichter, D. Grimmer, J. P. Santos, V. Scarani, and G. T. Landi, Collisional Quantum Thermometry, *Phys. Rev. Lett.* **123**, 180602 (2019).
- [29] J. Nettersheim, Q. Bouton, D. Adam, and A. Widera, Sensitivity of a collisional single-atom spin probe, (2022), [ArXiv:2203.13656](https://arxiv.org/abs/2203.13656).
- [30] M. Mehboudi, A. Lampo, C. Charalambous, L. A. Correa, M. A. García-March, and M. Lewenstein, Using Polarons for sub-nK Quantum Nondemolition Thermometry in a Bose-Einstein Condensate, *Phys. Rev. Lett.* **122**, 030403 (2019).
- [31] S. L. Braunstein and C. M. Caves, Statistical Distance and the Geometry of Quantum States, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [32] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

- [33] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).
- [34] U. von Toussaint, Bayesian inference in physics, *Rev. Mod. Phys.* **83**, 943 (2011).
- [35] J. Rubio, J. Anders, and L. A. Correa, Global Quantum Thermometry, *Phys. Rev. Lett.* **127**, 190402 (2021).
- [36] W.-K. Mok, K. Bharti, L.-C. Kwek, and A. Bayat, Optimal probes for global quantum thermometry, *Commun. Phys.* **4**, 1 (2021).
- [37] M. Mehboudi, M. R. Jørgensen, S. Seah, J. B. Brask, J. Kołodyński, and M. Perarnau-Llobet, Fundamental Limits in Bayesian Thermometry and Attainability via Adaptive Strategies, *Phys. Rev. Lett.* **128**, 130502 (2022).
- [38] M. R. Jørgensen, J. Kołodyński, M. Mehboudi, M. Perarnau-Llobet, and J. B. Brask, Bayesian quantum thermometry based on thermodynamic length, *Phys. Rev. A* **105**, 042601 (2022).
- [39] J. Boeyens, S. Seah, and S. Nimmrichter, Uninformed Bayesian quantum thermometry, *Phys. Rev. A* **104**, 052214 (2021).
- [40] J. Rubio, Quantum scale estimation, *Quantum Sci. Technol.* **8**, 015009 (2023).
- [41] G. O. Alves and G. T. Landi, Bayesian estimation for collisional thermometry, *Phys. Rev. A* **105**, 012212 (2022).
- [42] H. B. Prosper, Temperature fluctuations in a heat bath, *Am. J. Phys.* **61**, 54 (1993).
- [43] T. Johnson, F. Cosco, M. T. Mitchison, D. Jaksch, and S. R. Clark, Thermometry of ultracold atoms via nonequilibrium work distributions, *Phys. Rev. A* **93**, 053619 (2016).
- [44] K. D. B. Higgins, S. C. Benjamin, T. M. Stace, G. J. Milburn, B. W. Lovett, and E. M. Gauger, Superabsorption of light via quantum engineering, *Nat. Commun.* **5**, 4705 (2014).
- [45] M. Mehboudi, M. Moreno-Cardoner, G. De Chiara, and A. Sanpera, Thermometry precision in strongly correlated ultracold lattice gases, *New J. Phys.* **17**, 055020 (2015).
- [46] E. T. Jaynes, Prior probabilities, *IEEE Trans. Syst. Sci. Cybern.* **4**, 227 (1968).
- [47] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multi-parameter estimation beyond quantum Fisher information, *J. Phys. A Math. Theor.* **53**, 363001 (2020).
- [48] H. Jeffreys, *Theory of Probability* (Clarendon P., Oxford, 1961), 3rd ed.
- [49] C. F. Matta, L. Massa, A. V. Gubskaya, and E. Knoll, Can one take the logarithm or the sine of a dimensioned quantity or a unit? Dimensional analysis involving transcendental functions, *J. Chem. Educ.* **88**, 67 (2011).
- [50] M. E. Pearce, E. T. Campbell, and P. Kok, Optimal quantum metrology of distant black bodies, *Quantum* **1**, 21 (2017).
- [51] M. Mudrich, *Interactions in an optically trapped mixture of ultracold lithium and cesium atoms: Thermalization, spin-exchange collisions and photoassociation*, Ph.D. thesis, Ruprecht-Karls Universität, Heidelberg 2003.
- [52] S. L. Braunstein, A. S. Lane, and C. M. Caves, Maximum-Likelihood Analysis of Multiple Quantum Phase Measurements, *Phys. Rev. Lett.* **69**, 2153 (1992).
- [53] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, Quantum limits in optical interferometry, *Prog. Opt.* **60**, 345 (2015).
- [54] J. Rubio and J. A. Dunningham, Quantum metrology in the presence of limited data, *New J. Phys.* **21**, 043037 (2019).
- [55] J. Rubio and J. Dunningham, Bayesian multiparameter quantum metrology with limited data, *Phys. Rev. A* **101**, 032114 (2020).
- [56] J. Rubio Jiménez, *Non-asymptotic quantum metrology: Extracting maximum information from limited data*, Ph.D. thesis, University of Sussex 2020.
- [57] J. M. Muñoz, X. Wang, T. Hewitt, A. U. Kowalczyk, R. Sawant, and G. Barontini, Dissipative Distillation of Supercritical Quantum Gases, *Phys. Rev. Lett.* **125**, 020403 (2020).
- [58] L. W. Cheuk, M. A. Nichols, M. Okan, T. Gersdorf, V. V. Ramasesh, W. S. Bakr, T. Lompe, and M. W. Zwierlein, Quantum-Gas Microscope for Fermionic Atoms, *Phys. Rev. Lett.* **114**, 193001 (2015).
- [59] F. Ruschewitz, D. Bettermann, J. L. Peng, and W. Ertmer, Statistical investigations on single trapped neutral atoms, *Europhys. Lett.* **34**, 651 (1996).
- [60] D. Haubrich, H. Schadwinkel, F. Strauch, B. Ueberholz, R. Wynands, and D. Meschede, Observation of individual neutral atoms in magnetic and magneto-optical traps, *Europhys. Lett.* **34**, 663 (1996).
- [61] Here we ignore the bias  $(\langle \hat{\vartheta} \rangle - T)$  with respect to the true temperature [53,63] because we are primarily concerned with the *variability* of the estimate rather than its *accuracy*. Nevertheless, we have verified that our conclusions remain valid even if a simulated bias term is included, i.e.,  $[\text{Var } \hat{\vartheta} + (\langle \hat{\vartheta} \rangle - T)^2]/\langle \hat{\vartheta} \rangle^2$ , with  $T = 40 \mu\text{K}$ . As expected, this extra term vanishes asymptotically regardless of the estimation technique employed. Furthermore, in the limit of few data, the bias term has been found to be considerably smaller for Bayesian estimates.
- [62] The parameter  $F$  does not necessarily coincide with the Fisher information. It would indeed stand for the Fisher information associated with  $T$  if  $\lambda$  were known exactly. However, here we are dealing with a multiparameter problem, as both  $T$  and  $\lambda$  need to be determined. In principle, to guarantee optimality, one would perform a two-parameter Bayesian estimation (see Appendix E), which would further allow for the Cramér-Rao bound to be saturated in the limit of many measurements [53]. In that case,  $F$  would correspond to the diagonal element of the inverse of the Fisher information matrix that is associated with temperature, thus involving a combination of the off-diagonal elements of such a matrix [63]. Yet, in this work,  $\lambda$  is separately estimated by taking the empirical mean on unreleased atom numbers. Therefore, our overall estimator needs not be efficient and the Cramér-Rao bound may not be saturated.
- [63] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, Englewood Cliffs, NJ, 1993).