

# Magnetic-Field Learning Using a Single Electronic Spin in Diamond with One-Photon Readout at Room Temperature

R. Santagati,<sup>1,\*</sup> A. A. Gentile,<sup>1,\*</sup> S. Knauer,<sup>1,\*||</sup> S. Schmitt,<sup>2</sup> S. Paesani,<sup>1</sup> C. Granade,<sup>3</sup> N. Wiebe,<sup>6</sup> C. Osterkamp,<sup>2</sup> L. P. McGuinness,<sup>2</sup> J. Wang,<sup>1,4</sup> M. G. Thompson,<sup>1</sup> J. G. Rarity,<sup>1</sup> F. Jelezko,<sup>2,5,†</sup> and A. Laing<sup>1,‡</sup>

<sup>1</sup>Quantum Engineering Technology Labs, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1FD, United Kingdom

<sup>2</sup>Institute of Quantum Optics, Ulm University, D-89081 Ulm, Germany

<sup>3</sup>Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA

<sup>4</sup>State Key Laboratory for Mesoscopic Physics and Collaborative Innovation Center of Quantum Matter, School of Physics, Peking University, Beijing 100871, China

<sup>5</sup>Center for Integrated Quantum Science and Technology (IQst), Ulm University, D-89081 Ulm, Germany

<sup>6</sup>Institute for Nuclear Theory (INT), University of Washington, 3720 15th Avenue NE, Seattle, Washington 98195, USA



(Received 3 October 2018; revised manuscript received 17 January 2019; published 29 April 2019)

Nitrogen-vacancy (NV) centers in diamond are appealing nanoscale quantum sensors for temperature, strain, electric fields, and, most notably, magnetic fields. However, the cryogenic temperatures required for low-noise single-shot readout that have enabled the most sensitive NV magnetometry reported to date are impractical for key applications, e.g., biological sensing. Overcoming the noisy readout at room temperature has until now demanded the repeated collection of fluorescent photons, which increases the time cost of the procedure, thus reducing its sensitivity. Here, we show how machine learning can process the noisy readout of a single NV center at room temperature, requiring on average only one photon per algorithm step, to sense magnetic-field strength with a precision comparable to those reported for cryogenic experiments. Analyzing large datasets from NV centers in bulk diamond, we report absolute sensitivities of  $60 \text{ nT s}^{1/2}$  including initialization, readout, and computational overheads. We show that dephasing times can be simultaneously estimated and that time-dependent fields can be dynamically tracked at room temperature. Our results dramatically increase the practicality of early-term single-spin sensors.

DOI: [10.1103/PhysRevX.9.021019](https://doi.org/10.1103/PhysRevX.9.021019)

Subject Areas: Quantum Physics

## I. INTRODUCTION

Quantum sensors are likely to be among the first quantum technologies to be translated from laboratory setups to commercial products [1]. The single electronic spin of a nitrogen-vacancy (NV) center in diamond operates with nanoscale spatial resolution as a sensor for electric and magnetic fields [2–6]. However, achieving high sensitivities for NV magnetometers has required a

low-noise mode of operation available only at cryogenic temperatures, which constitutes a major obstacle to real-world applications [7,8], or prior knowledge of correlations existing in the estimated signal [9]. Machine learning plays an enabling role for new generations of applications in conventional information-processing technologies, including pattern and speech recognition, diagnostics, and robot control [10,11]. Here, we show how machine-learning algorithms [12–15] can be applied to single-spin magnetometers at room temperature to give a sensitivity that scales with the Heisenberg limit and reduces overheads by requiring only one-photon readout, at each step of the algorithm. We go on to show that these methods allow a multiparameter estimation to simultaneously learn the decoherence time and implement a routine for the dynamical tracking of time-dependent fields.

dc-magnetic-field sensing with an NV center uses Ramsey interferometry [1,16,17]. With a microwave  $\pi/2$  pulse, the spin vector is rotated into an equal superposition of its  $\sigma_z$  spin eigenstates, such that its magnetic moment is perpendicular to the magnetic field ( $B$ ) to be sensed [18,19]. For some Larmor precession time  $\tau$  and frequency

\*These authors contributed equally to this work.

†fedor.jelezko@uni-ulm.de

‡anthony.laing@bristol.ac.uk

||Present address: Centre of Excellence for Quantum Computation and Communication Technology, School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia.

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

$f_B = \gamma B/2\pi$ , the relative phase between the eigenstates becomes  $\phi = 2\pi f_B \tau$ , where  $\gamma$  is the electron gyromagnetic ratio of the magnetic moment to the angular momentum. After a further  $\pi/2$  pulse to complete the Ramsey sequence, a measurement of the spin in its  $\sigma_z$  basis provides an estimate of  $\phi$ , the precision of which is usually improved by repeating the procedure. Collecting statistics for a series of different  $\tau$  produces a fringe of phase varying with time, from which  $B$  can be inferred.

Increasing the sensitivity of a magnetometer translates to increasing its rate of sensing precision with sensing time. The intrinsic resource cost in estimating  $B$  is the total phase accumulation time [20–22], which is the sum of every  $\tau$  performed during an experiment. A fundamental limitation on the sensitivity of an estimate of  $B$  is quantum projection noise—from the uncertain outcome of a  $\sigma_z$ -basis measurement—the effect of which is conventionally reduced through repeated measurements, at the cost of increasing the total sensing time. A further typical limitation on sensing precision is the timescale  $T_2^*$ , on which spin states decohere due to inhomogeneous broadening [23]. In an idealized setting, with an optimal sensing protocol, the Heisenberg limit (HL) [24] in sensitivity can be achieved, to arrive at a precision limited by  $T_2^*$  in the shortest time allowed by quantum mechanics. In practice, overheads such as the time required for initialization, computation, and readout must also be accounted for, while repeated measurements due to experimental inefficiencies and low-fidelity readout increase the time to reach the precision limited by  $T_2^*$ . The increase in the total sensing time due to overheads and repeated measurements thus decreases the sensitivity.

A particularly relevant overhead is the time taken to read out the state of the spin, which depends on the experimental conditions. At cryogenic temperatures, spin-selective optical transitions can be accessed such that, during optical pumping, fluorescence is almost completely absent for one of the spin states. This single-shot method allows the spin state to be efficiently determined with a high confidence for any given Ramsey sequence (up to collection and detection efficiencies), resulting in a relatively low readout overhead. At room temperature, in contrast, where spin-selective optical transitions are not resolved in a single shot, readout is typically performed by simultaneously exciting a spin triplet that includes both basis states and observing fluorescence from the subsequent decay, the probabilities for which differ by only approximately 35%. Overcoming this classical uncertainty (in addition to quantum projection noise) to allow a precise estimate of the relative spin-state probabilities after a given precession time  $\tau$  requires repeated Ramsey sequences to produce a large ensemble of fluorescent photons. Such a large readout overhead significantly reduces the sensitivity of NV magnetometry, and, so far, the high sensitivities reported at cryogenic temperatures are out of reach for room-temperature operation

by several orders of magnitude. Yet, NV sensing at cryogenic temperatures is impractical for biological applications such as *in vivo* measurements [2] and monitoring of metabolic processes [25].

A large body of work [8,19,21,24,26–31] has developed and improved quantum-sensing algorithms to surpass the classical standard measurement sensitivity (SMS). While the SMS bounds the sensitivity that can be achieved for NV magnetometry with a constant phase accumulation time, phase estimation algorithms using a set of different precession times  $\tau_i$  allow the SMS to be overcome [19,21,32]. Further improvements in sensitivity are possible by adapting measurement bases to require fewer Ramsey sequences [8,33,34]. However, sensing algorithms that use a standard Bayesian approach typically involve probability distributions that are computationally intensive to update or which contain outlying regions that significantly affect an estimate. An appealing alternative [13,14,35] uses techniques from machine learning to approximate a probability distribution with a relatively small collection of points, known as particles. These methods have been applied to the problem of learning a Hamiltonian [14,36] and to implement noise-tolerant quantum phase estimation [37].

Here, we experimentally demonstrate a magnetic-field learning (MFL) algorithm that operates with on average only one-photon readout from a single NV center at room temperature and achieves a level of sensitivity so far reported only for cryogenic operation [8]. MFL adapts efficient Bayesian phase estimation and Hamiltonian learning techniques for magnetometry to achieve a fast convergence to the correct value of the magnetic field and requires no adaptation of measurement bases. The parameters of our MFL algorithm, including the number of particles, can be optimized prior to operation without adding to the sensing time overhead. Each precession time  $\tau_i$  is chosen [38] as the inverse of the uncertainty  $\sigma_{i-1}$  in the algorithm’s previous estimate of  $B$ , allowing  $\tau$  to grow exponentially to achieve HL scaling in sensitivity. We tested MFL on a large dataset from 60 000 Ramsey interferometry experiments on a bulk diamond NV center. We benchmark the performance of MFL against standard FFT methods, as well as previous experimental results from other phase estimation algorithms. Simultaneous to the learning of  $B$ , MFL produces an estimate of  $T_2^*$ , which, in contrast to other phase estimation algorithms, allows MFL to lower bound its sensitivity to the SMS, however long its implementation run time. Remarkably, we show that MFL enables the dynamical tracking of time-varying magnetic fields at room temperature.

## II. MAGNETIC-FIELD LEARNING

In general, Hamiltonian learning algorithms estimate the parameters  $\vec{x}$  of a model Hamiltonian  $\hat{H}(\vec{x})$ , through iterations of informative measurements [14]. At each step, a prior probability distribution  $P(\vec{x})$  stores estimates of

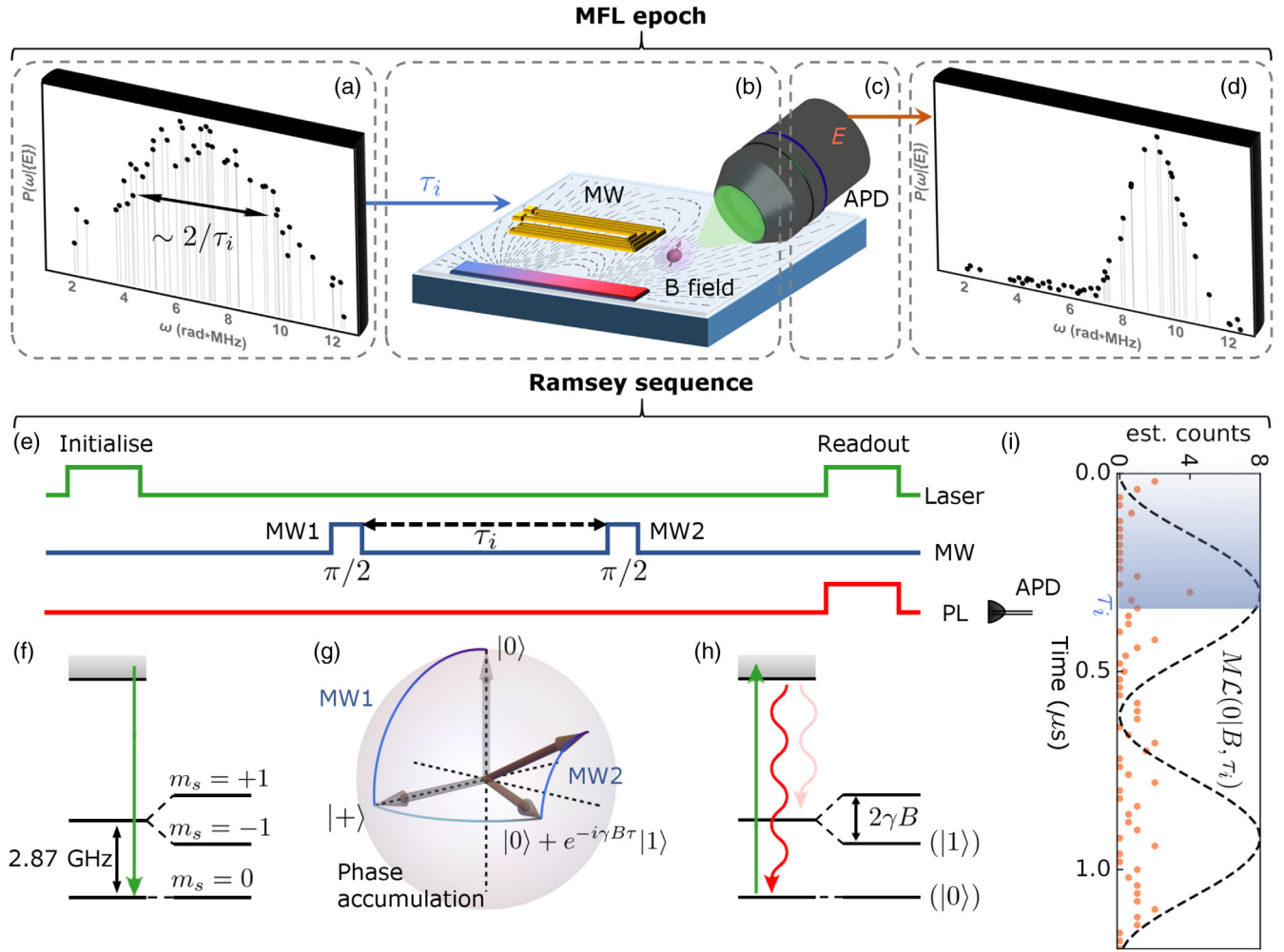


FIG. 1. An epoch of the MFL algorithm including a Ramsey sequence and readout. (a) The uncertainty encoded in the prior distribution  $P_{i-1}$  determines the phase accumulation time  $\tau_i$  for the next set of Ramsey sequences. (b) A number  $M$  of Ramsey sequences are implemented for  $\tau_i$ , with the precession driven by a  $B$  field from permanent magnets. Laser light is focused with a confocal microscope. A planar copper wire on the surface of the bulk diamond generates microwave pulses. (c) The outcomes  $E$  from the Ramsey sequences are measured. (d) The prior distribution is updated  $P_{i-1} \rightarrow P_i$  through Bayesian inference, from which the next phase accumulation time  $\tau_{i+1}$  is determined. (e) The NV spin vector is initialized with laser light, rotated with microwave pulses, and, using a second laser pulse, read out from photoluminescence (PL) with an avalanche photodiode (APD). (f) The electronic energy level triplet supports initialization and microwave manipulation between the  $m_s = 0$  and  $m_s = -1$  states, which encode the basis states  $|0\rangle$  and  $|1\rangle$ , respectively. (g) The Bloch sphere depicts the transit of the electronic state vector for the microwave rotations and Larmor precession. (h) Detection is performed by optically pumping the basis states to a higher-energy level triplet and measuring the decay via (non-spin-preserving) PL. (i) A representative PL fringe (theory plotted as a dashed line) with orange data points representing the number of detected photons for  $M = 8$ .

every parameter and its uncertainty [13]. Similarly, the four principal recursive steps of MFL, called an epoch and depicted in Figs. 1(a)–1(d), are as follows. (a) Choose  $\tau_i$  for the next Ramsey sequence from the heuristic  $\tau_i \approx 1/\sigma_{i-1}$ , where  $\sigma_{i-1}^2$  is the uncertainty embedded in the prior  $P(\vec{x}_{i-1})$ . (b) Allow the system to evolve under  $\hat{H}$  for a time  $\tau_i$ , using the Ramsey sequence shown in Figs. 1(e)–1(h). (c) Measure the outcome  $E$ , extracted from the photoluminescence count, e.g., Fig. 1(i). (d) Update the prior using Bayes' rule,  $P'(\vec{x}|E) \propto \mathcal{L}(E|\vec{x}; \tau)P(\vec{x})$ , where  $\mathcal{L}$  is the likelihood function [13]. The use of sequential Monte Carlo algorithms [13,14,35], where particles are reallocated

when required, makes the inference process practical and computationally efficient. Here, the Hamiltonian for the two relevant NV states is modeled as

$$\hat{H}(B) = \omega(B)\hat{\sigma}_z/2 = \gamma B\hat{\sigma}_z/2, \quad (1)$$

so that  $\omega$  is the only parameter to be estimated to learn the value of  $B$ .

In Ramsey interferometry, as performed here, we measure the magnetic-field component  $B$  parallel to the NV centers' symmetry axes. However, the MFL protocol could, in principle, be expanded to detect arbitrary oriented magnetic fields.

### III. EXPERIMENT

Experiments are performed using a confocal setup, at room temperature, with an external magnetic field of approximately 450 G, parallel to the NV center axis, giving a Zeeman shift of  $\omega = \gamma B$  [21], where  $\gamma \approx 2\pi \times 28$  MHz/mT [39].

The Ramsey interferometry experiments reported here use the  $m_s = 0$  and  $m_s = -1$  electronic sublevels of a few different  $^{12}\text{C}$  isotopically purified diamond samples (see Table S1 in Supplemental Material [40]). For each Ramsey sequence, the electronic spin is initialized and read out with 532 nm laser pulses, by detecting the photoluminescence (PL) signal with an avalanche photodiode (APD) for 350 ns. The PL signal is then normalized to extract an experimental estimate for  $\mathcal{L}$ . For every sequence, the experimental overhead is the sum of the times for the laser pulses length (3  $\mu\text{s}$ ), an idle time for relaxation (1  $\mu\text{s}$ ), a short transistor-transistor logic (TTL) pulse for synchronization (20 ns), and the duration of the two microwave pulses (together approximately 50 ns).

Data for several hundred Ramsey fringes are generated from experiments on three NV centers, labeled  $\alpha$ ,  $\beta$ , and  $\epsilon$  (Supplemental Material, Table S1 [40]). In particular, the dataset  $\epsilon_1$  comprises Ramsey sequences for precession times increasing from  $\tau_1$  to  $\tau_{500}$  in steps of 20 ns. For each  $\tau_i$ , 20 275 sequences are performed, and data are stored such that the results from each individual sequence could be retrieved. Therefore,  $\binom{20 \cdot 275}{M}^{500}$  subsets of data from  $\epsilon_1$  could be selected and combined to construct fringes comprised of  $M$  sequences. Running MFL on a sample of these subsets allows its performance to be compared over fringes with different (but fixed within a fringe) numbers of sequences including down to  $M = 8$ , where (due to low collection efficiencies) the average PL count ( $n_{\text{phot}}$ ) is approximately one photon. Additional experiments on the three NVs generate further datasets for several hundred fringes that each comprise tens of thousands of averaged sequences. All implementations of MFL are reported as representative behavior averaged over  $R = 1000$  independent protocol runs (unless otherwise stated), each using a single fringe from these datasets.

### IV. RESULTS

We begin by analyzing how the estimate of uncertainty in the magnetic field,  $\sigma(B_{\text{est}})$ , given by the variance of  $P(\vec{x})$ , scales with the number of MFL epochs. For this purpose, we use the dataset  $\alpha_1$ , with 120 fringes all obtained with  $M = 18\,500$  sequences. At every MFL epoch, given the adaptively chosen phase accumulation time  $\tau_i \simeq 1/\sigma_{i-1}$ , the experimental datum with  $\tau$  minimizing  $(|\tau - \tau_i|)$  is provided to the MFL updater. Figure 2(a) shows an exponential decrease in the scaling of  $\sigma(B_{\text{est}})$ , until approximately 50 epochs are reached. After this point, the precession times  $\tau$  selected by MFL saturate at  $\tau_{\text{max}} = 10 \mu\text{s}$ ,

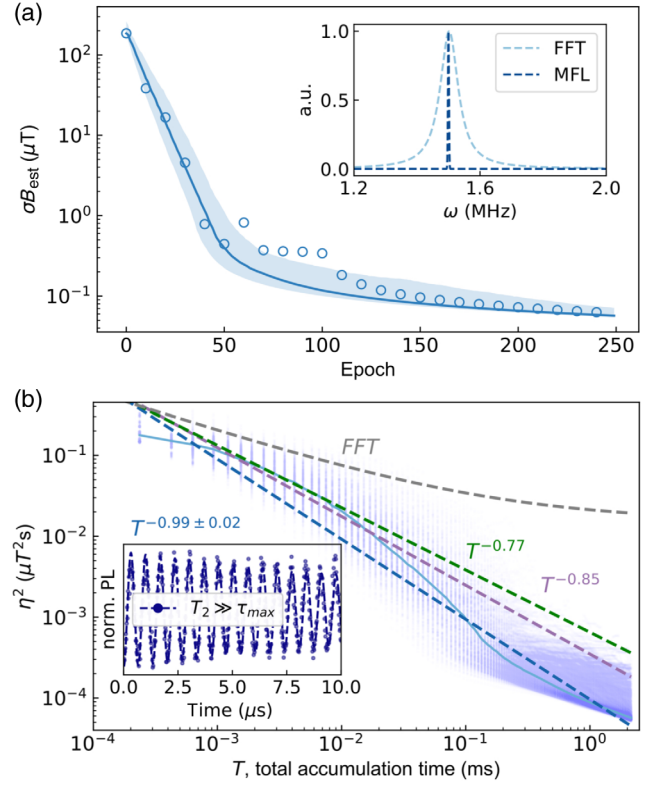


FIG. 2. Experimental results for scaling of precision. Lines represent median values, and performance within the 68.27% percentile range is shown as shaded areas. (a) Estimated uncertainty  $\sigma(B_{\text{est}})$  is plotted as a function of the epoch number; data from one sample run is shown as blue circles. In the inset, a plot of the final  $\sigma(\omega_{\text{est}})$  in the Ramsey frequency for a typical protocol run, from FFT (Lorentzian fit) and MFL (Gaussian fit). (b) The scaling of precision with total phase accumulation time  $T$ , excluding all overheads, is shown as density plots with a linear least-squares fit (blue dashed line). The FFT approach is plotted as a gray dashed curve. Scaling for phase estimation algorithms in Refs. [19,21] (respectively, green and violet lines) is also reproduced. The inset shows data from a Ramsey fringe in normalized PL, with a 20 ns sampling rate, up to  $\tau_{\text{max}} \sim 0.14T_2^*$ . A least-squares fit with a decaying sinusoid is shown as a blue dashed line.

and  $\sigma(B_{\text{est}})$  is reduced only polynomially fast, by accumulating statistics for  $\tau$  already retrieved. This slowdown is analogous to that occurring when the heuristic requires  $\tau$  exceeding the system dephasing time [13] (see Supplemental Material, Sec. S3, for details [40]).

The precision scaling is calculated using

$$\eta^2 = \delta B^2 = \sigma^2(B_{\text{est}})T, \quad (2)$$

where  $T := \sum_i^N \tau_i$  from  $N$  epochs represents the total phase accumulation time alone. Therefore,  $T$  neglects all overheads as in previous works [19,21]. Figure 2(b) plots  $\eta^2$  against  $T$ , for each epoch, and compares MFL with the standard FFT method, using the same  $\alpha_1$  set. The precision

of MFL scales as  $T^{-0.99 \pm 0.02}$ , which overlaps with HL scaling ( $\propto T^{-1}$ ). The FFT method rapidly approaches the SMS ( $\propto T^0$ ), whereas (neglecting overheads) the scaling reported for quantum phase estimation methods are qualitatively comparable to MFL, at the expense of more intensive postprocessing [30].

Finally, a comparison with FFT methods in terms of frequency resolution is performed and shown in the inset in Fig. 2(a), halting the estimation when  $\tau_{\max}$  is reached. At this moment in time, we find that  $\sigma(B_{\text{est}})$  is approximately 40 times smaller for MFL. In order to have a meaningful comparison, we allow for approximately similar  $T$  to be cumulated for both MFL and FFT protocols. Therefore, in this case, FFT adopts 500 equally spaced data points in the interval  $[0, \tau_{\max}]$  and, hence,  $T_{\text{FFT}} = 2.5$  ms, whereas the representative MFL run has  $T \simeq 2.2$  ms after analyzing 250 data points.

For a true measure of absolute sensitivity, experimental and computational overheads must be accounted for. Including all the relevant overheads, i.e., initialization, readout, and computation time, into the total running time  $\bar{T}$ , we redefine the precision Eq. (2) in terms of absolute precision:

$$\bar{\eta}^2 = \sigma^2(B_{\text{est}})\bar{T}, \quad (3)$$

where  $\bar{T}$  represents the total running time [see Eq. (B1) in Appendix B].

The average number of luminescent photons,  $n_{\text{phot}}$ , used for readout during each epoch, scales linearly with the number of sequences  $M$  ( $n_{\text{phot}} \propto M$ ); on average, one photon every  $M \simeq 8$  sequences is detected. As shown in Fig. 3, we use MFL to measure the scaling of  $\bar{\eta}$  with  $\bar{T}$  (up to 500 epochs) for decreasing numbers  $n_{\text{phot}}$  within each epoch. The plots have a shape characterized by an initial slow decrease, followed by a fast increase in precision. The relatively slow learning rate for the short phase accumulation times in the early stages of the algorithm leads to a slow increase in the phase accumulation time, since ( $\tau_i \propto 1/\sigma_{i-1}$ ). The algorithm is slowly learning, but the total measurement time is increasing faster than the decrease in uncertainty. However, when the particles start converging to a valid estimate of  $B$ , the uncertainty decreases exponentially, overcoming the corresponding increase in sensing times. Our analysis compares well with previous results performed under cryogenic conditions [8]; scaling parameters from linear least-squares fitting give a consistent overlap with HL scaling for protocol update rates up to 13 Hz. In Sec. S5 of Supplemental Material [40], we also provide, for comparison, functional simulations of another Bayesian method, quantum phase estimation [30], adopting characteristic parameters to reproduce our experimental setup.

Decreasing the number of sequences (thus  $n_{\text{phot}}$ ) per epoch increases the statistical noise, which extends the slow

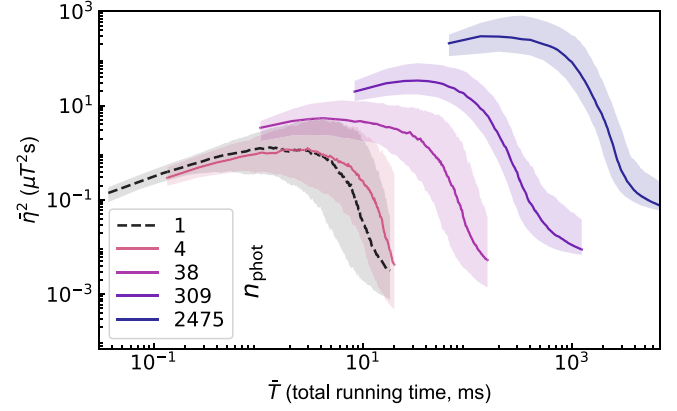


FIG. 3. The representative scaling of precision  $\bar{\eta}^2$ , inclusive of overheads, plotted against total running time  $\bar{T}$ . The colors identify different average numbers of photons detected per epoch ( $n_{\text{phot}}$ ), when correspondingly altering the  $M$  repetitions of Ramsey sequences. Each protocol run for  $n_{\text{phot}} > 1$  comprises  $N = 150$  epochs, and only Poissonian noise is modeled in the likelihood function. For  $n_{\text{phot}} = 1$ , each run comprises  $N = 500$  epochs, and an improved likelihood models also infidelities and losses.

learning period. However, the total time  $\bar{T}$  decreases with  $n_{\text{phot}}$  to produce an increased sensitivity in a shorter time. For  $n_{\text{phot}} < 4$ , readout infidelities and losses become the dominant noise mechanisms. In the case for  $n_{\text{phot}} = 1$ , therefore, these additional sources of noise are included as a precharacterized noise model (see Supplemental Material, Sec. S6 [40]). When using as little as  $n_{\text{phot}} = 1$ , we obtain a sensitivity of  $60 \text{ nT s}^{1/2}$  in approximately 10 ms, from a total of only 4000 single Ramsey sequences.

When an NV-sensing algorithm begins to request precession times  $\tau_i$  beyond  $T_2^*$ , where no information can be retrieved, the effectively wasted sensing time reduces the sensitivity. Knowledge of  $T_2^*$  can ensure that all  $\tau_i$  are less than  $T_2^*$ , to prevent this reduction in sensitivity and instead guaranteeing it to scale at the SMS for long sensing times. Learning  $T_2^*$  simultaneously with  $B$ , as part of a multi-parameter estimation strategy [41,42], can be more efficient than independently estimating  $T_2^*$  ahead of each sensing experiment. To describe experimental data from Ramsey fringes collected from an NV center with dephasing time  $T_2^*$ , immersed in a magnetic field of intensity  $B$ , we adopt the likelihood function:

$$\mathcal{L}(0|B, T_2^*; \tau) = \exp(-\tau/T_2^*) \cos^2(\gamma B \tau/2) + [1 - \exp(-\tau/T_2^*)]/2, \quad (4)$$

with unknown parameters  $\vec{x} = \{B, T_2^*\}$ . The exponential decay observed in the contrast of the Ramsey fringes, here phenomenologically observed, corresponds to approximately white noise affecting the sensing measurement here [43]. However, more general noise models can be adopted

by correspondingly adapting Eq. (4) to include information deriving, e.g., from noise spectroscopy experiments [44]. MFL naturally operates as a multiparameter estimation protocol when the prior probability distribution  $P(\vec{x})$  is multivariate [13], and the uncertainty in its joint probability distribution is captured by a normalized covariance matrix  $\Sigma$ .

Each precession time  $\tau_i$  is chosen proportional to the inverse of the (Frobenius) norm of the covariance matrix (see Appendices A and C). This choice can incur an initial slow learning period due to shorter  $\tau_i$  being initially most useful in estimating  $B$ , while longer precession times are better for an estimation of  $T_2^*$ . We therefore begin MFL in the single-parameter estimation mode for  $B$  and introduce the simultaneous learning of  $T_2^*$  at epoch  $N = 100$  (chosen empirically).

Figure 4 shows results from running the MFL algorithm on the  $\beta_1$  dataset, where  $\tau_{\max} > T_2^*$ . As is the case for single-parameter estimation results, we find an exponential scaling of the generalized uncertainty with the number of epochs, though the learning rate for  $B$  is faster than that for  $T_2^*$ . There is a discrepancy between the estimate of  $T_2^*$  from MFL shown in Fig. 4(a) and the fit (nonweighted least-squares) to the decaying sinusoid shown in Fig. 4(d). The discrepancy between these two estimates results from MFL preferentially requesting  $\tau_i < T_2^*$ , such that an estimate of  $T_2^*$  is more informed by data at these relatively shorter timescales (see Appendix C).

The strength of  $B$  may not be fixed in time for typical sensing experiments [45]. The Bayesian inference process is conceived to learn on-line when experimentally retrieved likelihoods  $P(E|\vec{x})$  conflict with its prior information. Thus, the ability to track time-varying magnetic fields follows naturally from the MFL's processing speed and adaptivity. With minor controls in the Bayesian inference procedure, including heuristics to trigger a reset of the prior, MFL accounts for not only fluctuations in the sensed  $B$ , but also high-amplitude changes invalidating the *a posteriori* credibility region, as estimated after a number of epochs. More details and a pseudocode are provided in Sec. S8 in Supplemental Material [40]. Here, we test an algorithm that tracks a  $B_{\text{set}}$  field using the  $e_3$  dataset, where  $B_{\text{set}}$  is experimentally modulated by changing the position of the permanent magnet [see Fig. 1(b)]. Data recording is paused during magnet adjustments, leading to stepwise transitions in this dataset, where the magnetic field instantly jumps to a new strength and then remains stable for a period of between hundreds and thousands of milliseconds.

Results are shown in Fig. 5(a), with a maximum approximately 30-fold instantaneous change in  $B$ . MFL detects when the posterior distribution has become nonrepresentative of the most recent measurements, by increasing the uncertainty  $\sigma(B_{\text{est}})$ . After approximately ten epochs, the estimate converges to the new value set for  $B$ . This analysis also emphasizes the high dynamic range  $B_{\max}/\sigma(B_{\text{est}})$  achievable with the protocol, that for this dataset is as high

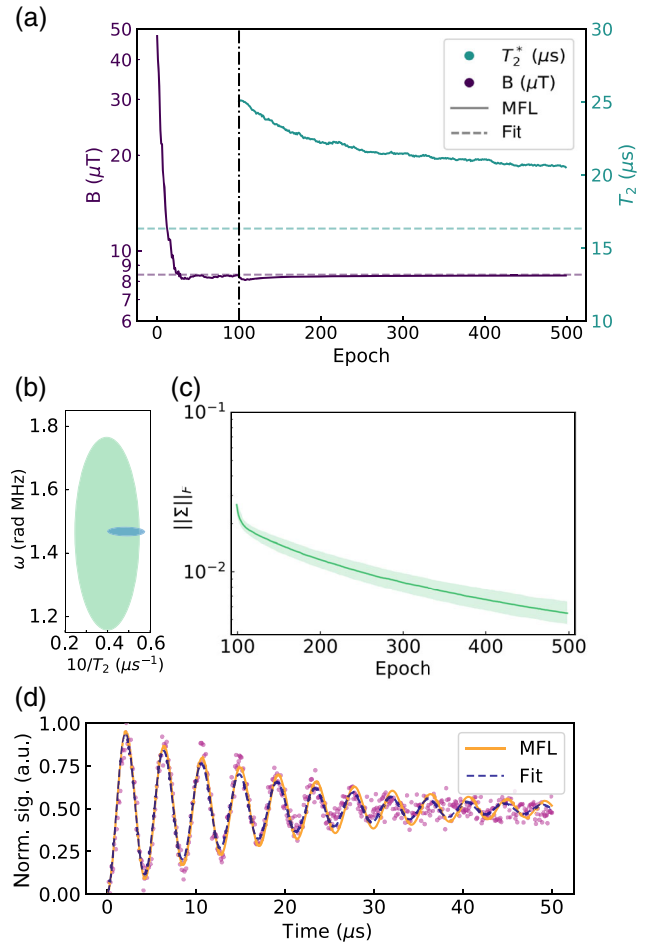


FIG. 4. Simultaneous learning of  $T_2^*$  and the magnetic field. (a) Simultaneous estimates of magnetic field  $B$  (purple) and decoherence time  $T_2^*$  (green) for epochs higher than 100. Solid lines are from MFL, and dashed lines are from a least-squares fit to the Ramsey fringe data in (d). (b) The 68.27% credible region at epoch 100 (green) and 500 (blue) for  $\omega$  and  $T_2^*$ , reported, respectively, on the  $y$  ( $x$ ) axes. The smaller area of the distribution at the final epoch indicates the decreased uncertainty on both parameters. (c) The norm of the covariance matrix  $\|\Sigma\|_F$ , representing the uncertainty in simultaneous estimates of  $B$  and  $T_2^*$ , is plotted against the epoch number. The median performance is shown as a solid line, with a shaded area representing the 68.27% percentile range. (d) Renormalized experimental data for a Ramsey fringe, along with a least-squares fit and an MFL-learned decay function showing decoherence.

as 12 300. Figure 5(b) summarizes the different computational and experimental contributions to the total running time per epoch (approximately 10 ms). The computational time cost of MFL is  $\tau^{\text{comp}} \simeq 0.2$  ms, with the remaining time costs coming from experimental routines. We note that the computational efficiency of MFL allows a computational overhead ( $\tau^{\text{comp}} = 0.21$  ms) that is smaller than the average phase accumulation time ( $\tau = 0.41$  ms) and 2 orders of magnitude smaller than the experimental overheads ( $\tau^{\text{exp}} = 16.28$  ms).

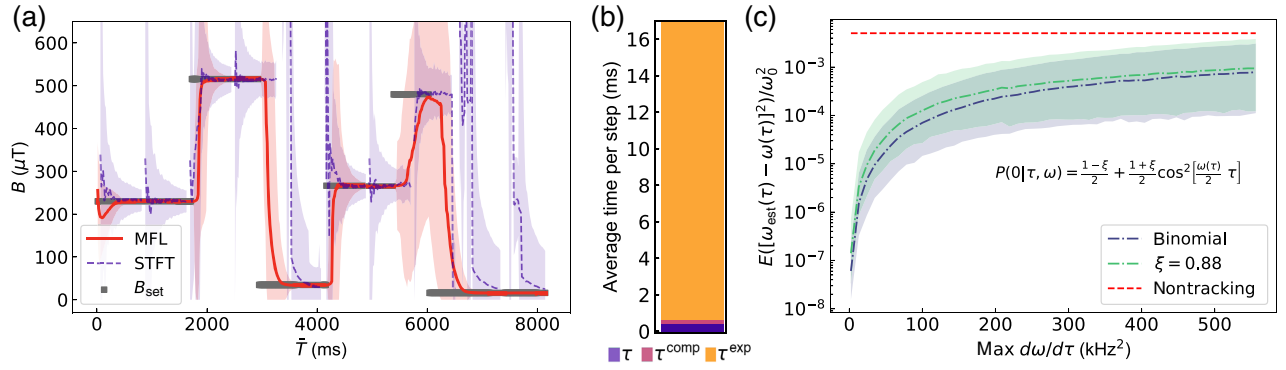


FIG. 5. Magnetic-field tracking. (a) Tracking with the MFL protocol is demonstrated on experimental data, where step changes in  $B$  are indicated by the gray bars (here, the number of sequences  $M = 4000$ ). The solid red line represents typical performances of MFL, with the shaded area indicating performance within an approximately 68.27% percentile range. For comparison, a dashed purple line indicates the outcome of a short-time Fourier transform (STFT) protocol applied progressively to all data points within time windows composed of 50 Ramsey sequences each, with the corresponding uncertainty from a Lorentzian fit as a shaded area of the same color. Results after fewer than three data points are omitted for STFT. (b) Itemization of the contributions to the average total time  $\bar{T}$  taken into account in (a): the precession time  $\tau$  and computational ( $\tau^{\text{comp}}$ ) and experimental ( $\tau^{\text{exp}}$ ) overheads. (c) Numerical study of MFL performance in tracking sinusoidally time-dependent magnetic fields  $B(\tau) = \omega(\tau)/\gamma$ , under ideal conditions ( $T_2^* = \infty$ ,  $\tau^{\text{exp}} = 0$ ). The y axis gives the median time-averaged squared error (nms $_{\omega}$ ) in the Ramsey frequency estimate, against the peak speed at which  $B$  changes along each simulated Ramsey sequence [ $\max d\omega(\tau)/d\tau$ ]. The blue dashed line refers to the case including only binomial noise in  $\mathcal{L}(B; \tau)$ , while the green line is the case with limited readout fidelity ( $\xi = 0.88$ ), as defined in Ref. [45]. The dashed red line indicates the error obtained via a nontracking strategy. Shaded areas indicate the approximately 68.27% percentile range.

Figure 5(c) shows numerical results demonstrating the resilience of MFL against a dynamic component of increasing frequency, when tracking an ac oscillating field  $B(\tau) = \omega(\tau)/\gamma$ , where we choose  $\omega(\tau) = \omega_0 + w \cos(\nu\tau)$ , with  $\nu$  a constant and  $w \ll \omega_0$ . The effectiveness of the tracking for each run is captured by a time-dependent normalized squared error  $\text{nms}_{\omega} := E[\omega_{\text{est}}(\tau) - \omega(\tau)]^2 / \omega_0^2 = \sum_i^N [\omega_{\text{est}}(\tau_i) - \omega(\tau_i)]^2 / (N\omega_0^2)$ , averaged for all  $N$  epochs performed, capturing the efficiency of the tracking as  $B$  is not constant along epochs. Typical estimation errors in  $B$  are lower than 3% for dynamic components up to 18  $\mu\text{T}/\text{ms}$ .

## V. CONCLUSION

The performance of magnetic-field learning found for our room temperature setup is comparable to other protocols in cryogenic environments [8]. These methods could be applied to other sensing platforms where noise has been a limiting factor. Alternatively, in pursuit of the fundamental limits in absolute sensing precision, they could be used together with single-shot readout [7], adaptive measurement bases [8], faster communication, and dynamical decoupling techniques [46,47]. Some features of the setup employed for this experiment, such as immersion lenses or bath polarization, might not be available in some applicative scenarios. However, promising developments that enhance photon collection efficiency [48–50] or that limit the environmental effects [3] might offer viable alternatives, thus removing such limitations. In conclusion, our methods would be particularly effective in applications

where single-spin sensing is desired for nanoscale resolution, but cryogenic conditions are prohibitive, such as biological sensing and new nano-MRI applications [6,51].

All data used to achieve the conclusions of this work are available for download in Refs. [52,53].

## ACKNOWLEDGMENTS

The authors thank Cristian Bonato and Marco Leonetti for useful discussion. M. G. T. acknowledges support from the ERC starting Grant No. ERC-2014-STG 640079 and an EPSRC Early Career Fellowship No. EP/K033085/1. J. G. R. is sponsored under EPSRC Grant No. EP/M024458/1. L. P. M. and F. J. acknowledge DFG (Deutsche Forschungsgemeinschaft Grants No. SFB/TR21 and No. FOR1493), BMBF, VW Stiftung, European Research Council DIADEMS (Simulators and Interfaces with Quantum Systems), European Union Initial Training Network Zero-to-ultra-low-field Nuclear Magnetic Resonance (ITN ZULF), and Diamond Quantum Devices for Biology ERC (European Research Council) project (BioQ). A. L. acknowledges support from an EPSRC Early Career Fellowship No. EP/N003470/1.

R. S., A. A. G, and S. K. contributed equally to this work.

## APPENDIX A: MFL EXECUTION

The data processing is performed by adapting the open PYTHON package QINFER [54] to the case of experimental metrology.

In dc-magnetometry experiments, such as those presented here, the reset of the sensing electron spin, and perturbation of its environment at the start of each Ramsey sequence, is expected to give mostly noise which is uncorrelated across epochs. The close agreement with an exponentially decaying envelope for the Ramsey signal suggests that the dephasing noise is induced by an approximately Markovian process [43]. In order to describe experimental data from Ramsey fringes collected from an NV center with dephasing time  $T_2^*$ , immersed in a magnetic field of intensity  $B$ , [13], we adopt the likelihood function in Eq. (4), where  $T_2^*$  is a known parameter or approximated by  $T_2^* = \infty$  in all cases when  $T_2^* \gg \tau_{\max}$ .

In cases when  $M > 1$ , the datum adopted is obtained from  $M$  combined sequences as stated in the main text. Results in Figs. 2(a), 2(b), and 5 are all obtained adopting a majority voting scheme to preprocess data from combined sequences [37], with no other assumptions about the noise model. Majority voting decides each single-shot datum according to the most frequent outcome, which is done by previously determining, during the characterization of the experimental setup, the average photoluminescence counts ( $\bar{n}$ ) detected throughout the execution of a Ramsey sequence. The datum of a single outcome is determined by comparing the number of photons detected during the measurement (extracted from  $M$  sweeps),  $n$ , and  $\bar{n}$ . If  $n > \bar{n}$ , then we set the value of the outcome to  $|1\rangle$  and otherwise to  $|0\rangle$ . Without this scheme in place, the outcome of a measurement is assigned sampling from the set  $\{|0\rangle, |1\rangle\}$ , with probabilities  $P \propto \{1 - n/n_{\max}, n/n_{\max}\}$ , respectively, with  $n_{\max}$  the maximum photoluminescence counts estimated during the characterization.

Other than the study of  $\bar{\eta}$  in Fig. 3, further examples of the performance of MFL with no majority voting scheme in place are reported in Sec. S3 in Supplemental Material [40].

Errors in the precision scaling are estimated from a bootstrapping procedure, involving a sampling with replacement from the available runs ( $R$ ). The cardinality of each resample matches  $R$ . The resampling is repeated  $[0.1R]$  times. Median precision scalings from each resample are estimated, and the standard deviation from this approximate population of scaling performances is provided as the precision scaling error.

## APPENDIX B: ABSOLUTE SCALING

In Fig. 3, we report the absolute scaling of  $\bar{\eta}^2 = \sigma^2(B_{\text{est}})\bar{T}$ , which requires one to take into account the main experimental and computational overheads contributing to the total running time  $\bar{T}$  of a phase estimation (PE) protocol (communication time  $\tau_{\text{comm}}$  is not considered here). In particular, these can be listed as the time required by the PE algorithm to compute the next experiment  $\tau^{\text{comp}}$  (here, approximately  $0.4 \mu\text{s}$  per step, per particle on a single-core

machine), the duration of the laser pulse  $\tau^{\text{las}}$  for initialization and readout ( $3 \mu\text{s}$  in total), the waiting time  $\tau^{\text{wait}}$  for relaxation ( $1 \mu\text{s}$ ), a short TTL pulse  $\tau^{\text{TTL}}$  for the photodetector ( $20 \text{ ns}$ ), and the duration  $\tau^{\text{MW}}$  of microwave pulses (approximately  $50 \text{ ns}$  in total). Including variable and constant overheads, we obtain

$$\bar{T} = \sum_i^N (M\tau_i + \tau_i^{\text{comp}}) + NM(\tau^{\text{las}} + \tau^{\text{wait}} + \tau^{\text{TTL}} + \tau^{\text{MW}}) \quad (\text{B1})$$

after  $N$  epochs of a PE algorithm.

In the  $n_{\text{phot}} = 1$  case, the final  $\sigma(B_{\text{est}}) \simeq 0.45 \mu\text{T}$  after 500 epochs, and  $\bar{T} \simeq 18 \text{ ms}$ , that is,  $\bar{\eta} \simeq 60 \text{ nT s}^{1/2}$ . In the  $n_{\text{phot}} = 20$  case, exhibiting a precision scaling that is essentially Heisenberg limited, the uncertainty saturates at protocol convergence (approximately 150 epochs) to  $\sigma(B_{\text{est}}) \simeq 0.3 \mu\text{T}$ , for a total running time  $\bar{T} \simeq 78 \text{ ms}$ . This result leads to a final sensitivity  $\bar{\eta} \simeq 84 \text{ nT s}^{1/2}$  and 12.8 Hz repetition rates.

## APPENDIX C: MULTIPARAMETER LEARNING

For the multiparameter case, we use again Eq. (4), but now considering the unknown parameters  $\vec{x} = \{B, T_2^*\}$ . Each precession time  $\tau_i$  is chosen proportionally to the inverse of the Frobenius norm of the covariance matrix,  $\|\Sigma\|_F = \|\text{cov}(B/b, T_2^*/t_2)\|_F$ . The parameters  $b = \max_{B:P(B) \neq 0} B$  and  $t_2 = \min_{T_2^*:P(T_2^*) \neq 0} T_2^*$  are introduced to render  $\|\Sigma\|_F$  dimensionless, with  $P$  the prior at epoch  $N = 100$ , when both parameters start to be learned simultaneously. In this analysis, this choice corresponds to  $b = 11 \mu\text{T}$  and  $t_2 = 20.2 \mu\text{s}$ ; however, we stress how different choices would be possible, with equivalent results for  $\|\Sigma\|_F$ , up to a normalization factor. We observe that MFL estimates of the dephasing time may differ consistently from a nonweighted least-squares fit. In the presence of dephasing, the heuristic of MFL preferentially adopts experiments with  $\tau < T_2^*$ . On the other hand, a least-squares fit attempts to equally mediate over data points where the contrast in the fringes is almost completely lost, underestimating  $T_2^*$ .

## APPENDIX D: FFT EXECUTION

For most analyzes, FFT estimates are run against the whole datasets available. For example, in the case of Fig. 2, the final estimate provided by a single run of FFT is performed using once all of the 500 phase accumulation times, recorded with  $20 \text{ ns}$  steps, for a representative subset among those available in  $\alpha_1$  (Supplemental Material, Table S1 [40]). We emphasize how these datasets are made of twice as many  $\tau$ 's as those actually used by the MFL



protocol (the single-run estimate being reported as converged after 250 epochs).

The only exception is the tracking in Fig. 5, where the data points are cumulatively added to the dataset. In such tracking applications, as long as  $B$  is kept constant, the estimate from FFT compares to MFL in a way similar to Fig. 2. However, FFT keeps estimating  $B$  from the prominent peak in the spectrum, corresponding to the  $\omega$  that is maintained for the longest time, not the most recent. Thus, it fails to track changes as they occur.

## APPENDIX E: PHOTON NUMBER ESTIMATION

After exciting a single NV center by a 532 nm laser pulse, the redshifted, individual photons are detected by an avalanche photodiode. To achieve a high collection efficiency, we use for initial measurements an oil immersion objective (NA 1.4), while for all measurements leading to absolute precision estimates, we exploit a solid-immersion lens on top of the NV centers and an air objective with NA 0.9. A time-tagged single-photon-counting card with nanosecond resolution is used for recoding. A TTL connection between the time tagger and the microwave pulse generator synchronizes the photon arrival time with respect to the pulse sequence and allows us to record the number of detected photons for every single laser pulse. Thereby, the photon detection efficiency is mainly limited by the collection volume, the total reflection within diamond (due to the high refractive index), and further losses due to the optics. These experimental conditions result a photon detected about every eighth laser pulse. Thus, to read out the NV state with high fidelity (and about 30% contrast), multiple measurements are usually required for meaningful statistics.

- 
- [1] J. M. Taylor, P. Cappellaro, L. Childress, L. Jiang, D. Budker, P. R. Hemmer, A. Yacoby, R. Walsworth, and M. D. Lukin, *High-Sensitivity Diamond Magnetometer with Nanoscale Resolution*, *Nat. Phys.* **4**, 810 (2008).
- [2] L. P. McGuinness, Y. Yan, A. Stacey, D. A. Simpson, L. T. Hall, D. Maclaurin, S. Prawer, P. Mulvaney, J. Wrachtrup, F. Caruso, R. E. Scholten, and L. C. L. Hollenberg, *Quantum Measurement and Orientation Tracking of Fluorescent Nanodiamonds inside Living Cells*, *Nat. Nanotechnol.* **6**, 358 (2011).
- [3] C. Müller, X. Kong, J. M. Cai, K. Melentijević, A. Stacey, M. Markham, D. Twitchen, J. Isoya, S. Pezzagna, J. Meijer, J. F. Du, M. B. Plenio, B. Naydenov, L. P. McGuinness, and F. Jelezko, *Nuclear Magnetic Resonance Spectroscopy with Single Spin Sensitivity*, *Nat. Commun.* **5**, 4703 (2014).
- [4] I. Lovchinsky, J. D. Sanchez-Yamagishi, E. K. Urbach, S. Choi, S. Fang, T. I. Andersen, K. Watanabe, T. Taniguchi, A. Bylinskii, E. Kaxiras, P. Kim, H. Park, and M. D. Lukin, *Magnetic Resonance Spectroscopy of an Atomically Thin Material Using a Single-Spin Qubit*, *Science* **355**, 503 (2017).
- [5] N. Zhao, J. L. Hu, S. W. Ho, J. T. K. Wan, and R. B. Liu, *Atomic-Scale Magnetometry of Distant Nuclear Spin Clusters via Nitrogen-Vacancy Spin in Diamond*, *Nat. Nanotechnol.* **6**, 242 (2011).
- [6] J. F. Barry, M. J. Turner, J. M. Schloss, D. R. Glenn, Y. Song, M. D. Lukin, H. Park, and R. L. Walsworth, *Optical Magnetic Detection of Single-Neuron Action Potentials Using Quantum Defects in Diamond*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14133 (2016).
- [7] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, *High-Fidelity Projective Read-Out of a Solid-State Spin Quantum Register*, *Nature (London)* **477**, 574 (2011).
- [8] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, *Optimized Quantum Sensing with a Single Electron Spin Using Real-Time Adaptive Measurements*, *Nat. Nanotechnol.* **11**, 247 (2016).
- [9] S. Schmitt, T. Gefen, F. M. Stürmer, T. Uden, G. Wolff, C. Müller, J. Scheuer, B. Naydenov, M. Markham, S. Pezzagna, J. Meijer, I. Schwarz, M. Plenio, A. Retzker, L. P. McGuinness, and F. Jelezko, *Submillihertz Magnetic Spectroscopy Performed with a Nanoscale Quantum Sensor*, *Science* **356**, 832 (2017).
- [10] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT, Cambridge, MA, 2012), ISBN 978-0-262-01802-9.
- [11] M. I. Jordan and T. M. Mitchell, *Machine Learning: Trends, Perspectives, and Prospects*, *Science* **349**, 255 (2015).
- [12] A. Hentschel and B. C. Sanders, *Machine Learning for Precise Quantum Measurement*, *Phys. Rev. Lett.* **104**, 063603 (2010).
- [13] C. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, *Robust Online Hamiltonian Learning*, *New J. Phys.* **14**, 103013 (2012).
- [14] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, *Hamiltonian Learning and Certification Using Quantum Resources*, *Phys. Rev. Lett.* **112**, 190501 (2014).
- [15] I. Hincks, C. Granade, and D. G. Cory, *Statistical Inference with Quantum Measurements: Methodologies for Nitrogen Vacancy Centers in Diamond*, *New J. Phys.* **20**, 013022 (2018).
- [16] L. Rondin, J.-P. Tetienne, T. Hingant, J.-F. Roch, P. Maletinsky, and V. Jacques, *Magnetometry with Nitrogen-Vacancy Defects in Diamond*, *Rep. Prog. Phys.* **77**, 056503 (2014).
- [17] F. Jelezko and J. Wrachtrup, *Single Defect Centres in Diamond: A Review*, *Phys. Status Solidi (a)* **203**, 3207 (2006).
- [18] C. L. Degen, *Scanning Magnetic Field Microscope with a Diamond Single-Spin Sensor*, *Appl. Phys. Lett.* **92**, 243111 (2008).
- [19] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, *High-Dynamic-Range Magnetometry with a Single Electronic Spin in Diamond*, *Nat. Nanotechnol.* **7**, 109 (2012).
- [20] K. Arai, C. Belthangady, H. Zhang, N. Bar-Gill, S. J. DeVience, P. Cappellaro, A. Yacoby, and R. L. Walsworth, *Fourier Magnetic Imaging with Nanoscale Resolution and Compressed Sensing Speed-Up Using Electronic Spins in Diamond*, *Nat. Nanotechnol.* **10**, 859 (2015).

- [21] G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, *High-Dynamic-Range Magnetometry with a Single Nuclear Spin in Diamond*, *Nat. Nanotechnol.* **7**, 105 (2012).
- [22] G. Puentes, G. Waldherr, P. Neumann, G. Balasubramanian, and J. Wrachtrup, *Efficient Route to High-Bandwidth Nanoscale Magnetometry Using Single Spins in Diamond*, *Sci. Rep.* **4**, 4677 (2014).
- [23] G. Balasubramanian, I. Y. Chan, R. Kolesov, M. Al-Hmoud, J. Tisler, C. Shin, C. Kim, A. Wojcik, P. R. Hemmer, A. Krueger, T. Hanke, A. Leitenstorfer, R. Bratschitsch, F. Jelezko, and J. Wrachtrup, *Nanoscale Imaging Magnetometry with Diamond Spins under Ambient Conditions*, *Nature (London)* **455**, 648 (2008).
- [24] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, *How to Perform the Most Accurate Possible Phase Measurements*, *Phys. Rev. A* **80**, 052114 (2009).
- [25] C. L. Degen, F. Reinhard, and P. Cappellaro, *Quantum Sensing*, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [26] V. Giovannetti, S. Lloyd, and L. Maccone, *Quantum-Enhanced Measurements: Beating the Standard Quantum Limit*, *Science* **306**, 1330 (2004).
- [27] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, *Entanglement-Free Heisenberg-Limited Phase Estimation*, *Nature (London)* **450**, 393 (2007).
- [28] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, *Demonstrating Heisenberg-Limited Unambiguous Phase Estimation without Adaptive Measurements*, *New J. Phys.* **11**, 073023 (2009).
- [29] V. Giovannetti, S. Lloyd, and L. Maccone, *Advances in Quantum Metrology*, *Nat. Photonics* **5**, 222 (2011).
- [30] R. S. Said, D. W. Berry, and J. Twamley, *Nanoscale Magnetometry Using a Single-Spin System in Diamond*, *Phys. Rev. B* **83**, 125410 (2011).
- [31] N. Aharon, A. Rotem, L. P. McGuinness, F. Jelezko, A. Retzker, and Z. Ringel, *NV Center Based Nano-NMR Enhanced by Deep Learning*, [arXiv:1809.02583](https://arxiv.org/abs/1809.02583).
- [32] T. E. O'Brien, B. Tarasinski, and B. M. Terhal, *Quantum Phase Estimation of Multiple Eigenvalues for Small-Scale (Noisy) Experiments*, *New J. Phys.* **21**, 023022 (2019).
- [33] I. Schwartz, J. Roskopf, S. Schmitt, B. Tratzmiller, Q. Chen, L. P. McGuinness, F. Jelezko, and M. B. Plenio, *Blueprint for Nanoscale NMR*, [arXiv:1706.07134v1](https://arxiv.org/abs/1706.07134v1).
- [34] H. T. Dinani, D. W. Berry, R. Gonzalez, J. R. Maze, and C. Bonato, *Bayesian Estimation for Quantum Sensing in the Absence of Single-Shot Detection*, *Phys. Rev. B* **99**, 125413 (2019).
- [35] N. Wiebe and C. Granade, *Efficient Bayesian Phase Estimation*, *Phys. Rev. Lett.* **117**, 010503 (2016).
- [36] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, *Experimental Quantum Hamiltonian Learning*, *Nat. Phys.* **13**, 551 (2017).
- [37] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, *Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip*, *Phys. Rev. Lett.* **118**, 100503 (2017).
- [38] T. Ruster, H. Kaufmann, M. A. Luda, V. Kaushal, C. T. Schmiegelow, F. Schmidt-Kaler, and U. G. Poschinger, *Entanglement-Based dc Magnetometry with Separated Ions*, *Phys. Rev. X* **7**, 031050 (2017).
- [39] I. Jakobi, P. Neumann, Y. Wang, D. B. R. Dasari, F. El Hallak, M. A. Bashir, M. Markham, A. Edmonds, D. Twitchen, and J. Wrachtrup, *Measuring Broadband Magnetic Fields on the Nanoscale Using a Hybrid Quantum Register*, *Nat. Nanotechnol.* **12**, 67 (2017).
- [40] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.9.021019> for further details on the methods reported and additional analyses supporting this work.
- [41] J. F. Ralph, S. Maskell, and K. Jacobs, *Multiparameter Estimation along Quantum Trajectories with Sequential Monte Carlo Methods*, *Phys. Rev. A* **96**, 052306 (2017).
- [42] M. A. Ciampini, N. Spagnolo, C. Vitelli, L. Pezzè, A. Smerzi, and F. Sciarrino, *Quantum-Enhanced Multiparameter Estimation in Multiarm Interferometers*, *Sci. Rep.* **6**, 28881 (2016).
- [43] L. Chirolli and G. Burkard, *Decoherence in Solid-State Qubits*, *Adv. Phys.* **57**, 225 (2008).
- [44] C. Ferrie, C. Granade, G. Paz-Silva, and H. M. Wiseman, *Bayesian Quantum Noise Spectroscopy*, *New J. Phys.* **20**, 123005 (2018).
- [45] C. Bonato and D. W. Berry, *Adaptive Tracking of a Time-Varying Field with a Quantum Sensor*, *Phys. Rev. A* **95**, 052348 (2017).
- [46] E. R. MacQuarrie, T. A. Gosavi, S. A. Bhave, and G. D. Fuchs, *Continuous Dynamical Decoupling of a Single Diamond Nitrogen-Vacancy Center Spin with a Mechanical Resonator*, *Phys. Rev. B* **92**, 224419 (2015).
- [47] D. Farfurnik, A. Jarmola, D. Budker, and N. Bar-Gill, *Spin Ensemble-Based ac Magnetometry Using Concatenated Dynamical Decoupling at Low Temperatures*, *J. Opt.* **20**, 024008 (2018).
- [48] T. Schröder, F. Gädeke, M. J. Banholzer, and O. Benson, *Ultrabright and Efficient Single-Photon Generation Based on Nitrogen-Vacancy Centres in Nanodiamonds on a Solid Immersion Lens*, *New J. Phys.* **13**, 055017 (2011).
- [49] L. Li, E. H. Chen, J. Zheng, S. L. Mouradian, F. Dolde, T. Schröder, S. Karaveli, M. L. Markham, D. J. Twitchen, and D. Englund, *Efficient Photon Collection from a Nitrogen Vacancy Center in a Circular Bullseye Grating*, *Nano Lett.* **15**, 1493 (2015).
- [50] F. Lenzini, N. Gruhler, N. Walter, and W. H. P. Pernice, *Diamond as a Platform for Integrated Quantum Photonics*, *Adv. Quantum Technol.* **1**, 1800061 (2018).
- [51] A. Boretti and S. Castelletto, *Nanometric Resolution Magnetic Resonance Imaging Methods for Mapping Functional Activity in Neuronal Networks*, *MethodsX* **3**, 297 (2016).
- [52] <https://doi.org/10.6084/m9.figshare.7855118>.
- [53] <https://doi.org/10.6084/m9.figshare.7855124>.
- [54] C. Granade, C. Ferrie, I. Hincks, S. Casagrande, T. Alexander, J. Gross, M. Kononenko, and Y. Sanders, *Qinfer: Statistical Inference Software for Quantum Applications*, *Quantum* **1**, 5 (2017).