Magnetic-field-learning using a single electronic spin in diamond with one-photon-readout at room temperature

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

Quantum Engineering Technology Labs, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1FD, UK.

†Centre of Excellence for Quantum Computation and Communication Technology, School of Physics, University of New South Wales, Sydney, NSW 2052, Australia.

2Institute of Quantum Optics, Ulm University, 89081 Ulm, Germany.

3Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA.

†These authors contributed equally to this work. and

*email to: fedor.jelezko@uni-ulm.de, anthony.laing@bristol.ac.uk

Nitrogen-vacancy (NV) centres in diamond are appealing nano-scale quantum sensors for temperature, strain, electric fields and, most notably, for 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 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 centre 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. Analysing large data sets from NV centres in bulk diamond, we report absolute sensitivities of 60 nT/s^{1/2} including initialisation, 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.

Quantum sensors are likely to be among the first quantum technologies to be translated from laboratory set-ups to commercial products [1]. The single electronic spin of a nitrogen-vacancy (NV) centre in diamond operates with nano-scale 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]. Machine learning has played an enabling role for new generations of applications in conventional information processing technologies, including pattern and speech recognition, diagnostics, and robot control [9, 10]. Here we show how machine learning algorithms [11–14] 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. We go on to show that these methods allow multiparameter estimation to simultaneously learn the decoherence time, and implement a routine for the dynamical tracking of time-dependent fields.

Magnetic field sensing with an NV centre uses Ramsey interferometry [1, 15, 16]. With a microwave π/2-pulse the spin vector is rotated into an equal superposition of its σ_z spin eigenstates, such that its magnetic moment is perpendicular to the magnetic field (B) to be sensed [17, 18]. For some Larmor precession time, τ, and frequency, f_B = γB/2π, the relative phase between the eigenstates becomes φ = 2πf_BT, where γ is the electron gyromagnetic ratio of magnetic moment to angular momentum. After a further π/2-pulse to complete the Ramsey sequence, a measurement of the spin in its σ_z basis provides an estimate of φ, the precision of which is usually improved by repeating the procedure. Collecting statistics for a series of different τ, 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 [19–21], which is the sum of every τ 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 σ_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 (though spin-echo methods could extend this [22]). In an idealised setting, with an optimal sensing protocol, the Heisenberg limit (HL) [23] 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 initialisation, 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 total sensing time due to overheads and repeated mea-
measurements thus decreases the sensitivity.

A particularly relevant overhead is the time taken to readout 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 subsequent decay, the probabilities for which differ by only $\approx 35\%$. 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 $\tau$, required 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 have been 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 [24].

A large body of work [8, 18, 20, 23, 25–29] 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 constant phase accumulation time, phase estimation algorithms using a set of different precession times $\tau_i$ allow the SMS to be overcome [18, 20]. Further improvements in sensitivity are possible by adapting measurement bases, to require fewer Ramsey sequences [8]. 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 [12, 13, 30] 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 [13, 31], and to implement noise-tolerant quantum phase estimation [32].

Here, we experimentally demonstrate a magnetic field learning (MFL) algorithm that operates with on average only one photon readout from a single NV centre at room temperature, and achieves a level of sensitivity so far only reported 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 optimised prior to operation without adding to the sensing time overhead. Each precession time $\tau_i$ is chosen $\in [33]$ as the inverse of the uncertainty $\sigma_{\tau_i}$ 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 data set from 60,000 Ramsey interferometry experiments on a bulk diamond NV centre. 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 runtime. Remarkably, we show that MFL enables the dynamical tracking of time-varying magnetic fields at room temperature.

In general, Hamiltonian learning algorithms estimate the parameters $\vec{x}$ of a model Hamiltonian $\hat{H}(\vec{x})$, through iterations of informative measurements [13]. At each step, a prior probability distribution $P(\vec{x})$ stores estimates of every parameter and its uncertainty [12]. Similarly, the four principal recursive steps of MFL, called an epoch and depicted in Fig. 1(a-d), are: (a) Choose $\tau_i$ for the next Ramsey sequence from the heuristic $\tau_i \approx 1/\sigma_{\tau_i}$, where $\sigma_{\tau_i}$ 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 Fig. 1(e-h). (c) Measure the outcome $E$, extracted from the photoluminescence count, e.g. Fig. 1(i). (d) Update the prior using Bayes’ rule, $P(\vec{x}\mid E) \propto L(E|\vec{x}; \tau)P(\vec{x})$, where $L$ is the likelihood function [12]. The use of sequential Monte Carlo algorithms [12, 13, 30] where particles are reallocated when required, makes the inference process practically and computationally efficient. Here, the Hamiltonian for the two relevant NV states is modelled as

$$\hat{H}(B) = \omega(B) \vec{\sigma} \times /2 = \gamma B \vec{\sigma} \times /2,$$

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

Experiments were performed using a confocal set-up, at room temperature, with an external magnetic field of $\approx 450$ G, parallel to the NV centre axis, giving a Zeeman shift of $\omega = \gamma B$ [20], where $\gamma \approx 2\pi \times 28$ MHz/mT [35]. For each Ramsey sequence, the electronic spin is initialised and readout with 532 nm laser pulses, by detecting the photoluminescence (PL) signal with an avalanche photodiode (APD) for 350 ns. The PL signal is then normalised to extract an experimental estimate for $L$. For every sequence, the experimental overhead is the sum of the times for the laser pulses length (3 µs), an idle time for relaxation (1 µs), a short TTL pulse for synchronization (20 ns) and the duration of the two MW-pulses (together $\approx 50$ ns).

Data for several hundred Ramsey fringes were generated from experiments on three NV centres, labelled $\alpha$, $\beta$ and $\epsilon$ (see Supplementary Table S1). In particular,
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 MW pulses. (c) The outcome $E$ 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 initialised with laser light, rotated with MW pulses, and, using a second laser pulse, readout from photoluminescence (PL) with an avalanche photodiode (APD). (f) The electronic energy level triplet supports initialisation and MW 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 MW 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 dashed line) with orange data-points representing the number of detected photons for $M = 8$.

the dataset $\epsilon_1$ comprises Ramsey sequences for precession times increasing from $\tau_1$ to $\tau_{500}$ in steps of 20 ns. For each $\tau_i$, 20275 sequences were performed, and data were stored such that the results from each individual sequence could be retrieved. Therefore, $\binom{20275}{M}$ 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 allowed 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 generated further data sets for several hundred fringes that each comprised tens of thousands of averaged sequences. All implementations of MFL are reported as representative behaviour averaged over $R = 1000$ independent protocol runs (unless otherwise stated) each using a single fringe from these data sets.

We begin by analysing 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
Neglecting overheads, the sensitivity \( \eta \) of a magnetometer, is calculated from

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

where \( T := \sum_i^N \tau_i \) from \( N \) epochs. 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 \( T^{-0.99\pm0.02} \), 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 post-processing [29].

For a true measure of absolute sensitivity, experimental and computational overheads must be accounted for. Including initialisation, read-out and computation time, into the total running time \( \bar{T} \), we redefine Eq. 2 for absolute scaling of \( \eta \) (see Methods for details). 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 \( M \approx 8 \) sequences is detected. As shown in Fig. 3, we use MFL to measure the scaling of \( \bar{\eta} \) with \( \bar{T} \) (up to 250 epochs) for decreasing numbers \( n_{\text{phot}} \) within each epoch. The plots have a shape characterised 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 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.

To that purpose, we use the dataset \( \alpha_1 \), with 120 fringes all obtained with \( M = 18500 \) sequences. At every MFL epoch, given the adaptively chosen phase accumulation time \( \tau_1 \approx 1/\sigma_{i-1} \), the experimental datum with \( \tau \) minimising \( (|\tau - \tau_i|) \) is provided to the MFL updater. Figure 2(a) shows an exponential decrease in the scaling of \( \sigma(B_{\text{est}}) \), until \( \sim 50 \) epochs are reached. After this point, the precision times \( \tau \) selected by MFL saturate at \( \tau_{\text{max}} = 10 \) \( \mu \)s, 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 [12] (see Supplementary Information for details). A comparison with FFT methods, inset in Fig. 2(a), finds that \( \sigma(B_{\text{est}}) \) is \( \sim 40 \) times smaller for MFL.
FIG. 4. Simultaneous learning of $T_2^*$ and 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) 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 epoch number. The median performance is shown as a solid line, with a shaded area representing the 68.27% percentile range. (d) Renormalised experimental data for a Ramsey fringe, along with a least-square-fit and an MFL-learned decay function showing decoherence.

Our analysis compares well with previous results performed under cryogenic conditions, and scaling parameters for linear least squares fitting obtain a consistent overlap with HL scaling for protocol update rates up to 13 Hz.

Decreasing the number of sequences (thus $n_{\text{phot}}$) per epoch increases the statistical noise, which extends the slow 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 were included in the model. For $n_{\text{phot}} = 1$ we obtain a sensitivity of $60\,\text{nT s}^{1/2}$ in $\sim 10\,\text{ms}$ (see also the Supplementary Information).

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 [36, 37], can be more efficient than independently estimating $T_2^*$ alone, ahead of each sensing experiment. MFL naturally operates as a multi-parameter estimation protocol when the prior probability distribution $P(\bar{x})$ is multivariate [12], with the uncertainty in its joint probability distribution captured by a normalised covariance matrix $\Sigma$.

Each precession time $\tau_i$ is chosen proportionality to the inverse of the (Frobenius) norm of the covariance matrix (see Methods). This 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 $\beta_1$ data set, where $\tau_{\text{max}} > T_2^*$. As is the case for single parameter estimation results, we find an exponential scaling of the generalised 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 4(a) and the fit (non-weighted least-square) to the decaying sinusoid shown in 4(d). The discrepancy between these two estimates results from the PGH preferentially requesting $\tau_i < T_2^*$, such that an estimate of $T_2^*$ is more informed by data at these relatively shorter time scales (see Methods).

The strength of $B$ may not be fixed in time for typical sensing experiments [34]. The Bayesian inference process is conceived to learn on-line when experimentally retrieved likelihoods $P(E|\bar{x})$ 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, MFL can account for such fluctuations and high-amplitude changes in the sensed $B$ (See Methods for details). Here, we test an algorithm that tracks a $B_{\text{sea}}$-field using the $c_3$ dataset, where $B_{\text{sea}}$ was experimentally modulated by changing the position of the permanent magnet (see Fig 1b). Data recording was paused during magnet adjustments, leading to stepwise transitions in this data set, where the magnetic field instantly jumps to a new strength then remains stable for a period of between hundreds and thousands of milliseconds.
Results are shown in Fig. 5(a), with a maximum ≈ 30-fold instantaneous change in $B$. MFL detects when the posterior distribution has become non-representative of the most recent measurements, by increasing the uncertainty, $\sigma(B_{\text{est}})$. After approximately 10 epochs, the estimate converges to the new value set for $B$. Figure 5(b) summarises the different computational and experimental contributions to the total running time per epoch ($\approx 10$ ms). The computational time cost of MFL is $\tau_{\text{comp}} \approx 0.2$ ms, with the remaining time costs coming from experimental routines. We note 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 two orders of magnitude smaller than the experimental overheads ($\tau_{\text{exp}} = 16.28$ ms).

Figure 5(c) shows numerical results demonstrating the resilience of MFL against a dynamic component of increasing frequency, when tracking an A.C. 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 normalised squared error $nms_{\omega} := E[(\omega_{\text{est}}(\tau) - \omega(\tau))^2]/\omega_0^2 = \sum_i (\omega_{\text{est}}(\tau_i) - \omega(\tau_i))^2/(N\omega_0^2)$, 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 $\mu$T/ms.

The performance of magnetic field learning found for our room temperature set-up 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 [38, 39]. Our methods would be particularly effective in applications where single-spin sensing is desired for nanoscale resolution, but where cryogenic conditions are prohibitive, such as biological sensing and in new nano-MRI applications [6, 40].

[1] J M Taylor, P Cappellaro, L Childress, L Jiang, D Budker, P R Hammer, A Yacoby, R Walsworth, and M D Lukin. High-sensitivity diamond magnetometer with nanoscale resolution. Nature Physics, 4(10):810–816, October 2008. doi:10.1038/nphys1075. URL https://www.nature.com/articles/nphys1075.

[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. Nature Nanotechnology, 6(6):358–363, June 2011. doi:10.1038/nnano.2011.64.
[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. *Nature Communications*, 5:4703, August 2014. doi:10.1038/ncomms5703 URL https://www.nature.com/articles/ncomms5703.

[4] I. Lovchinskiy, J. D. Sanchez-Yamagishi, E. K. Urbach, S. Choi, S. Fang, T. I. Andersen, K. Watanabe, T. Taniguchi, A. Bylinski, 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(6324):503–507, February 2017. ISSN 0036-8075. doi:10.1126/science.aal5238 URL http://science.sciencemag.org/content/355/6324/503.

[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. *Nature Nanotechnology*, 6(4):242–246, April 2011. doi:10.1038/nnano.2011.22 URL https://www.nature.com/articles/nnano.2011.22.

[6] J F Barry, M J Turner, J M Schloss, David 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. *Proceedings of the national academy of sciences*, 113(49):14133–14138, November 2016. ISSN 0027-8424. doi:10.1073/pnas.1601513113 URL http://www pnas org/content/113/49/14133.

[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*, 477(7366):574–578, September 2011. doi:10.1038/nature10401 URL https://www.nature.com/articles/nature10401.

[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. *Nature Nanotechnology*, 11(3):247–252, November 2015. doi:10.1038/nnano.2015.261 URL http://dx.doi.org/10.1038/nnano.2015.261.

[9] K P Murphy. *Machine Learning: a Probabilistic Perspective*. MIT Press, 2012. ISBN 978-0-262-01802-9. URL https://mitpress.mit.edu/books/machine-learning-1.

[10] M I Jordan and T M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, July 2015. ISSN 0036-8075. doi:10.1126/science.aaa8415 URL http://science.sciencemag.org/content/349/6245/255.

[11] A Hentschel and B C Sanders. Machine learning for precise quantum measurement. *Physical Review Letters*, 104(6):063603, February 2010. doi:10.1103/PhysRevLett.104.063603 URL https://link.aps.org/doi/10.1103/PhysRevLett.104.063603.

[12] C E Granade, C Ferrie, N Wiebe, and D G Cory. Robust online hamiltonian learning. *New Journal of Physics*, 14(10):103013, October 2012. doi:10.1088/1367-2630/14/10/103013 URL http://iopscience.iop.org/article/10.1088/1367-2630/14/10/103013.

[13] N Wiebe, C Granade, C Ferrie, and D G Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, 112(19):190501–5, May 2014. doi:10.1103/PhysRevLett.112.190501 URL http://link.aps.org/doi/10.1103/PhysRevLett.112.190501.

[14] I Hincks, C Granade, and D G Cory. Statistical inference with quantum measurements: methodologies for nitrogen vacancy centers in diamond. *New Journal of Physics*, 20(1):013022, January 2018. doi:10.1088/1367-2630/aa9c9f URL http://iopscience.iop.org/article/10.1088/1367-2630/aa9c9f.

[15] L Rondin, J-P Tetienne, T Hingant, J-F Roch, P Maletinsky, and V Jacques. Magnetometry with nitrogen-vacancy defects in diamond. *Reports on progress in physics*, 77(5), May 2014. doi:10.1088/1367-2630/14/10/103033 URL http://iopscience.iop.org/article/10.1088/0004-4885/77/5/056503/meta.

[16] F Jelezko and J Wrachtrup. Single defect centres in diamond: A review. *Physica status solidi (a)*, 203(13):3207–3225, October 2006. doi:10.1002/pssa.200671403 URL http://onlinelibrary.wiley.com/doi/full/10.1002/pssa.200671403.

[17] C L Degen. Scanning magnetic field microscope with a diamond single-spin sensor. *Applied Physics Letters*, 92(24):243111, June 2008. doi:10.1063/1.2943282 URL http://aip.scitation.org/doi/10.1063/1.2943282.

[18] N M Nusran, M Umoul Momeen, and M V G Dutt. High-dynamic-range magnetometry with a single electronic spin in diamond. *Nature Nanotechnology*, 7(2):109–113, February 2012. doi:10.1038/nnano.2011.225 URL https://www.nature.com/articles/nnano.2011.225.

[19] K Arai, C Belthangady, H Zhang, N Bar-Gill, S J Devi, P Cappellaro, A Yacoby, and R L Walsworth. Fourier magnetic imaging with nanoscale resolution and compressed sensing speed-up using electronic spins in diamond. *Nature Nanotechnology*, 10(10):859–864, October 2015. doi:10.1038/nnano.2015.171 URL https://www.nature.com/articles/nnano.2015.171.

[20] 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. *Nature Nanotechnology*, 7(2):105–108, February 2012. doi:10.1038/nnano.2011.224 URL https://www.nature.com/articles/nnano.2011.224.

[21] G Puentes, G Waldherr, P Neumann, G Balasubramanian, and J Wrachtrup. Efficient route to high-bandwidth nanoscale magnetometry using single spins in diamond. *Scientific Reports*, 4(1):4677, April 2014. doi:10.1038/srep04677 URL https://www.nature.com/articles/srep04677.

[22] 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*, 455(7213):648, October 2008. doi:10.1038/nature07278 URL https://www.nature.com/articles/nature07278.

[23] D W Berry, B L Higgins, S D Bartlett, M W Mitchell, GJ Pryde, and H M Wiseman. How to perform the most accurate possible phase measurements. *Physical Review A*, 80(5):5–22, November 2009. doi:10.1103/PhysRevA.80.052114 URL https://link.aps.org/doi/10.1103/PhysRevA.80.052114.
[24] C L Degen, F Reinhard, and P Cappellaro. Quantum sensing. Reviews of Modern Physics, 89(3):035002, July 2017. doi:10.1103/RevModPhys.89.035002. URL http://link.aps.org/doi/10.1103/RevModPhys.89.035002.

[25] V Giovannetti, S Lloyd, and L Maccone. Quantum-enhanced measurements: The standard quantum limit. Science, 306(5700):1330–1336, November 2004. doi:10.1126/science.1104149 URL http://www.sciencemag.org/cgi/doi/10.1126/science.1104149.

[26] B L Higgins, D W Berry, S D Bartlett, M W Mitchell, H M Wiseman, and GJ Pryde. Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements. New Journal of Physics, 11(7):073023, July 2009. doi:10.1088/1367-2630/11/7/073023 URL http://iopscience.iop.org/article/10.1088/1367-2630/11/7/073023.

[27] V Giovannetti, S Lloyd, and L Maccone. Advances in quantum metrology. Nature Photonics, 5(4):222–229, April 2011. doi:10.1038/nphoton.2011.35 URL https://www.nature.com/articles/nphoton.2011.35.

[28] R. S. Said, D. W. Berry, and J. Twamley. Nanoscale magnetometry using a single-spin system in diamond. Phys. Rev. B, 83:125410, Mar 2011. doi:10.1103/PhysRevB.83.125410 URL https://link.aps.org/doi/10.1103/PhysRevB.83.125410.

[29] N Wiebe and C Granade. Efficient bayesian phase estimation. Physical Review Letters, 117(1):010503, June 2016. doi:10.1103/PhysRevLett.117.010503 URL https://link.aps.org/doi/10.1103/PhysRevLett.117.010503.

[30] 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. Nature Physics, 13(6):551–555, June 2017. doi:10.1038/nphys4074 URL http://www.nature.com/doifinder/10.1038/nphys4074.

[31] 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. Physical Review Letters, 118(10):100503, March 2017. doi:10.1103/PhysRevLett.118.100503 URL https://link.aps.org/doi/10.1103/PhysRevLett.118.100503.

[32] 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. Physical Review X, 7(3):031050, September 2017. doi:10.1103/PhysRevX.7.031050 URL https://link.aps.org/doi/10.1103/PhysRevX.7.031050.

[33] C Bonato and D W Berry. Adaptive tracking of a time-varying field with a quantum sensor. Physical Review A, 95(5):052348, May 2017. doi:10.1103/PhysRevA.95.052348 URL http://link.aps.org/doi/10.1103/PhysRevA.95.052348.

[34] J F Ralph, S Maskell, and K Jacobs. Multiparameter estimation along quantum trajectories with sequential monte carlo methods. Phys. Rev. A, 96:052306, Nov 2017. doi:10.1103/PhysRevA.96.052306 URL http://link.aps.org/doi/10.1103/PhysRevA.96.052306.

[35] M A Ciampini, N Spagnolo, C Vitelli, L Pezzè, A Smerzi, and F Sciarrino. Quantum-enhanced multiparameter estimation in multiarm interferometers. Scientific Reports, 6(1):222, July 2016. doi:10.1038/srep28881 URL https://www.nature.com/articles/srep28881.

[36] D Farfurnik, A Jarmola, D Budker, and N Bar-Gill. Spin ensemble-based ac magnetometry using concatenated dynamical decoupling at low temperatures. Journal of Optics, 20(2):024008, January 2018. doi:10.1088/2040-8986/aaa1bf URL http://iopscience.iop.org/article/10.1088/2040-8986/aaa1bf.

[37] Albert Boretti and Stefania Castelletto. Nanometric resolution magnetic resonance imaging methods for mapping functional activity in neuronal networks. MethodsX, 3:297 – 306, 2016. ISSN 2215-0161. doi:10.1016/j.mex.2016.04.003 URL http://www.sciencedirect.com/science/article/pii/S2215016116300085.

[38] C Granade, C Ferrie, Ian Hincks, S Casagrande, T Alexander, J Gross, M Kononenko, and Y Sanders. Qinfer: Statistical inference software for quantum applications. Quantum, 1, 2017. doi:10.22331/q-2017-04-25-5 URL http://quantum-journal.org/papers/q-2017-04-25-5/.

[39] T M Cover and J A Thomas. Elements of information theory. John Wiley & Sons, Inc., Hoboken, NJ, USA, May 1999. ISBN 0-471-20065-6. doi:10.1002/047174882X. URL http://www.wiley.com/go/covers/cover0047174882X.

[40] C Ferrie. Data-processing inequalities for quantum metrology. Physical Review A, 90(1):014101, July 2014. doi:10.1103/PhysRevA.90.014101 URL https://link.aps.org/doi/10.1103/PhysRevA.90.014101.

[41] R D Gill and B Y Levit. Applications of the van trees inequality: a bayesian cramér-rao bound. Bernoulli, 1(2):59-79, March 1995. doi:10.2307/3318681 URL http://www.jstor.org/stable/3318681.

[42] M Opper. A Bayesian approach to on-line learning. Cambridge University Press, New York, May 1999. ISBN 0-521-65263-4. URL http://dl.acm.org/citation.cfm?doid=304710.304756.

[43] T Alexander, J Gross, M Kononenko, and Y Sanders. Qinfer: Statistical inference software for quantum applications. Quantum, 1, 2017. doi:10.22331/q-2017-04-25-5 URL http://quantum-journal.org/papers/q-2017-04-25-5/.
Methods

MFL execution The data processing was performed by adapting the open Python package QInfer [41] to the case of experimental metrology.

In order to describe experimental data from Ramsey fringes collected from an NV centre with dephasing time $T_2^*$, immersed in a magnetic field of intensity $B$, we adopt the likelihood function as in [12]:

$$L(\theta; B, T_2^*; \tau) = \exp(-\tau/T_2^*) \cos^2(\gamma B r/2) + [1 - \exp(-\tau/T_2^*)]/2,$$

where $T_2^*$ is a known parameter, or approximated by $T_2^* = \infty$ in all cases when $T_2^* \gg \tau_{\text{max}}$.

In cases when $M > 1$, the datum adopted was obtained from $M$ combined sequences as stated in the main text. Results in Fig. 2a&b and Fig. 5 were all obtained adopting a majority voting scheme to pre-process data from combined sequences [32]. Major- ity voting decides each single-shot datum according to the most frequent outcome. This is done by previously determining, during the characterisation of the experimental set-up, 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 $|1\rangle$, 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 $P \propto (1 - n/\eta_{\text{max}}, n/\eta_{\text{max}})$, respectively, with $\eta_{\text{max}}$ the maximum photoluminescence counts estimated during the characterisation.

Other than the study of $\eta$ in Fig. 3, further examples of the performance of MFL with no majority voting scheme in place are reported in the Supplementary Information.

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 $R$. 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.

Absolute scaling In Fig. 3 we reported the absolute scaling of $\bar{\eta}^2 = \eta^2 (\text{Base})T$, which requires to take into account the main experimental and computational overheads contributing to the total running time $T$ 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 $\tau_{\text{comp}}^\text{next}$ (here $\sim 0.4$ ms per step, per particle on a single-core machine), the duration of the laser pulse $\tau_{\text{las}}$ for initialisation and readout (3 ms in total), the waiting time $\tau_{\text{wait}}$ for relaxation (1 ms), a short TTL pulse $\tau_{\text{TL}}$ for the photodetector (20 ns) and the duration $\tau_{\text{MW}}$ of MW-pulses (approximately 50 ns in total). Including variable and constant overheads, we obtain:

$$T = \sum_i^N \left( \tau_i + \tau_{\text{comp}}^\text{next} + \tau_{\text{MW}}^\text{phot} + \tau_{\text{TL}} + \tau_{\text{wait}} + \tau_{\text{las}} \right)$$

after $N$ epochs of a PE algorithm.

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

Multi-parameter Learning For the multi-parameter case, we use again Eq. 3, 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, $|\Sigma|_F = |\text{cov}(B, \bar{T}_2^*)|_F$. The parameters $b = \max_B \rho(B, \bar{T}_2^*)$ and $t_2 = \min_B \rho(B, \bar{T}_2^*)$ are introduced to render $|\Sigma|_F$ dimensionless, with $\bar{P}$ the prior at epoch $N = 100$, when both parameters start to be learnt simultaneously. In this analysis this corresponded to $b = 11$ µT and $t_2 = 20.2$ µs, however we stress how different choices would be possible, with equivalent results for $|\Sigma|_F$, up to a normalisation factor. We observed that MFL estimates of the dephasing time may differ consistently from a non-weighted least-square fit. In the presence of dephasing, the heuristic of MFL will preferentially adopt experiments with $\tau < T_2^*$. This relation is similar to a weighting mechanism of the data (see also Supplementary Information), preferring more consistent observations. On the other hand, a least-square fit will attempt to equally mediate over data-points where the contrast in the fringes is almost completely lost, underestimating $T_2^*$.

MFL tracking We mentioned that Bayesian inference processes are ideally suited for tracking purposes. However, we observe that in cases where the characteristics of the changes in the parameter $\vec{x}$ completely invalidates the a-posteriori credibility region, the recovery time of a standard Hamiltonian learning protocol might be unsuitable for practical applications. To tackle also this situation, we modified here the standard update procedure to reset its prior when the effective sample size of the particles’ ensemble is not restored above the resampling threshold by a sequence of resampling events. Details and a pseudocode are provided in Supplementary Information.

FFT execution For most analyses, FFT estimates were run against the whole datasets available. For example in the case of Fig. 2, the final estimate provided by a single run of FFT was performed using once all of the 500 phase accumulation times, recorded with 20 ns steps, for a representative subset among those available in $\alpha_1$ (Supplementary Table 1). We emphasise how this amounts to twice as many $\tau$’s as those actually used by the MFL algorithm (being the single-run estimate reported as converged after 250 epochs).

The only exception is the tracking in Fig. 5, where the data-points were 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 was maintained for the longest time, not the most recent. Thus, it fails to track changes as they occur.

Experimental details In Ramsey interferometry, as performed here, we measure the magnetic field component parallel to the NV centres’ symmetry axes. However, the MFL protocol can be expanded to differently orientated NV centres, to detect arbitrary orientated magnetic fields.

The experiments are performed here using two different 12C isotopically purified diamond samples. For the Ramsey interferometry we use the $m_a = 0$ and $m_a = -1$ electronic sublevels.

Photon number estimation After exciting a single NV centre by a 532nm laser pulse, the red-shifted, individual photons were detected by an avalanche photodiode. A time-tagged single photon counting card with nanosecond resolution was used for recording. A TTL-connection between the time-tagging and the MW pulse generator synchronises the photon arrival time with respect to the pulse sequence and allows 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. This results in a photon detected about every eighth laser pulse. Thus, to readout the NV state with high-fidelity (and about 30% contrast) multiple measurements are usually required for meaningful statistics.

Acknowledgements The authors thank Cristian Bonato and Marco Leonetti for useful discussion. M.G.T. acknowledges support from the ERC starting grant ERC-2014-STG 640079 and an EPSRC Early Career Fellowship EP/K033085/1. J.G.R. is sponsored under EPSRC grant EP/M024458/1. L.M. and F.J. would like to acknowledge DFG (Deutsche Forschungsgemeinschaft grants SFB/TR21 and FOR1493), BMBF, VW Stiftung, European Research Council DIADEMS (Simulators and Interfaces with Quantum Systems), ITN ZULF and BioQ. A.Laing acknowledges support from an EPSRC Early Career Fellowship EP/N003470/1.
Supplementary Information

CLASSICAL AND QUANTUM LIKELIHOOD ESTIMATION

In the main text we have introduced the Bayesian inference process underlying the MFL protocol (known as CLE, Classical Likelihood Estimation[12]) as composed by four main steps. Here we expand the discussion to provide additional details and comments about the adoption of CLE.

1. At each epoch the prior distribution \( P(\vec{x}) \) is used to choose what experimental setting to use for the next iteration. In MFL, the only experimental setting is the phase accumulation time \( \tau \), that can be updated effectively using the so called particle guess heuristic (see the section below).

2. The quantum system undergoes an appropriate evolution, according to the Hamiltonian \( \hat{H} \).

   (a) The system is prepared in an appropriate initial state \( |\psi\rangle \), chosen such to have informative evolution under \( \hat{H} \). E.g. a state orthogonal to the Hilbert subspace spanned by the Hamiltonian eigenstates. We remark how \( |\psi\rangle \) is not adaptive in CLE. In this work, the NV centre is always prepared in \( |\psi\rangle = |+\rangle \).

   (b) Let the system evolve according to its Hamiltonian \( \hat{H} \) for the chosen time \( \tau \).

3. A measurement is performed on the system (here the quantum sensor). In MFL we perform a projective measurement on the \( \{|0\rangle, |1\rangle\} \) computational basis, obtaining a bipartite outcome \( E \in \{0,1\} \).

4. The computed likelihoods are used to update the probability distribution of the Hamiltonian \( \vec{x} \) parameters

   (a) The same experiment is also performed on a simulator, implementing a generic parametrised Hamiltonian \( \hat{H}(\vec{x}) \), thus providing an estimate for the likelihood \( P(E|\vec{x},\tau) \), i.e. the probability of measuring outcome \( E \) when \( \vec{x} \) is chosen as parameter.

   (b) It is thus possible to apply Bayes’ rule:

\[
P'(\vec{x}|E) = \frac{P(E|\vec{x},\tau)P(\vec{x})}{P(E)},
\]

where \( P(\vec{x}) \) can be immediately inferred from the prior at the corresponding epoch, while \( P(E) \) is a normalization factor.

Steps 1 – 4 are repeated until the variance of the probability distribution \( \sigma(\vec{x}_{\text{est}}) \) converges, or falls below a predefined target threshold. In cases with limited readout fidelity like for the NV centre set-up presented here, in step 3 a meaningful statistics might be cumulated for \( E \) repeating the same measurement a number of times \( M \), as suggested in the main text. Evidences from the text suggest that in most cases, this is a suboptimal choice for the absolute scaling performance of the MFL protocol. Note how only steps 2 & 3 involve the quantum sensor. All other steps require instead a simulator. In particular, step 4(a) can be performed on a classical or quantum simulator, the choice of the second being justified whenever the size of the sensor, and the eventual lack of an analytical model to simplify the evolution, make the system simulation classically expensive. In this case, the inference process is known as Quantum Likelihood Estimation (QLE, [30]).

| label | NV centre | sets (n) | sequences (M) | \( \Delta \tau \) (ns) | \( \bar{B}_{\text{est}} \) (\( \mu\)T) | \( T_2^* \) (from fit, \( \mu\)s) | \( \eta \) (\( \bar{\eta} \)) |
|-------|-----------|----------|---------------|-----------------|-----------------|-----------------|-----------------|
| \( \alpha_1 \) | \( \alpha \) | 120 | 18500 | 20 | 52 | - | \( T^{-0.99 \pm 0.02} \) (-) |
| \( \alpha_2 \) | 115 | 18500 | 20 | 710 | - | \( T^{-0.97 \pm 0.03} \) (-) |
| \( \beta_1 \) | \( \beta \) | 67 | 30000 | 100 | 8.3 | 16 | - (-) |
| \( \epsilon_1 \) | \( \epsilon \) | 1 | 20275 | 20 | 58 | - | \( T^{-0.73 \pm 0.03} \rightarrow T^{-0.6 \pm 0.02} \) |
| \( \epsilon_2 \) | 1 | 8876 | 200 | 6.0 | 64 | \( T^{-0.91 \pm 0.03} \) (-) |
| \( \epsilon_3 \) | 1 | 44000 | 20 | 10 \rightarrow 550 | - | - (-) |

TABLE S1. **Synopsis of available data.** Table summarising the different data sets and systems used in this analysis, along with representative MFL performances in precision scaling. Datasets \( \alpha_2, \epsilon_2 \) are discussed only in the Supplementary Information.
FIG. S1. **Analysis of the behaviour of the PGH** for datasets where $T^*_2 \gg \tau_{max}$ or $T^*_2 \simeq \tau_{max}$, reported in the plots in darker and brighter colours, respectively. The first dataset is collected with $B \simeq 58 \mu T$, whereas the second has $B \simeq 6 \mu T$. a. Renormalized photon counts along two different Ramsey experiments with the same NV centre (scatter plots). Superimposed a least-square fit (dashed lines), adopting the oscillatory function with depolarizing noise as in Eq. 3 of Methods. b, Estimated uncertainty $\sigma(B_{est})$ and ratio between PGH-generated time and $\tau_{max}$ as available from the first dataset, plotted against each epoch of the MFL algorithm. A majority voting method is adopted, under the hypothesis that $T^*_2 \gg \tau_{max}$. Solid lines are median values calculated over 1000 independent runs, whereas shaded areas are 68.27% percentile ranges centred around the median. Superimposed as a scatter plot, a sample of times generated by the PGH during a representative run. c, Same as in b, for the case where $T^*_2 \simeq \tau_{max}$. No majority voting is in place, and data from the experiment are extracted probabilistically from the experimentally estimated likelihoods.

**SEQUENTIAL MONTE CARLO APPROXIMATION AND PARTICLE GUESS HEURISTIC**

The protocol performances in terms of computational overhead are made possible by adopting advanced approximate Bayesian inference methods. In particular, MFL inherits from CLE the Sequential Monte Carlo (SMC) approximation [12–14]. Within this approximation only a finite number of values $\vec{x}_i$ (called particles) are sampled from the prior distribution, and thus used to approximate the prior in each update step. This approximation makes the Bayesian update as in Eq. S1 numerically feasible.

If the particle positions $\vec{x}_i$ were held constant throughout the inference process, and starting the protocol from a uniform prior, the cardinality of their set should scale approximately as $n_{part} := |\{\vec{x}_i\}| \propto \Delta B/\tilde{\sigma}(B_{est})$, where $\Delta B$ is the expected magnetic field range to be sensed, and $\tilde{\sigma}(B_{est})$ is the targeted uncertainty upon convergence. This is inefficient, as with the inference progressing through epochs, many particles will provide very limited support to the updated prior approximation. Indeed, for a successful learning process $\lim_{N \to \infty} P(\vec{x}_i) \to 0$ for the weights $w_i \propto P(\vec{x}_i)$ of most particles, as they have been effectively ruled out by the observations.

This inefficiency can be addressed with resampling methods, that allow the particles to be sampled again from the updated posterior, whenever their weights signify that the effective size of the sampled particles $\sum_i w_i^2$ has fallen below a user-defined threshold. Following [41], here we adopt a Liu-West resampler (LWR) with optimised resampling threshold $t_{resample} = 0.5$ and smoothing parameter $a = 0.9$. These parameters allow to tune when and to what extent the positions of the particles can be altered by the LWR [12]. Hence, it was possible to accurately represent $P(\vec{x}_i)$ throughout the whole protocol execution, whilst employing not more than 1500 particles for the discretisation in most cases. The only exceptions were limited fidelity cases, as for the absolute scaling we chose the number of particles $n_{part}$ according to the empirical rule

$$n_{part} = \frac{25000}{(\log(M) + 1)}$$
$$t_{resample} = \max(0.1, 0.5 - 0.4/\log(M))$$
$$a = 0.9 + 0.08 M/M^{max},$$

with $M$ the number of averaged single sequences selected. This increase in the number of particles can be justified
by a corresponding reduction in the risk of “aggressive” resampling leading to inference failures, heralded especially by multi-modality in the parameter distribution [14].

The particle guess heuristic (PGH) plays a fundamental role in the effectiveness of the MFL protocol. PGH was introduced in [13] to provide optimal choice of the experimental \( \tau \) (here the phase accumulation time) in analytically tractable cases of Hamiltonian Learning protocols. Such cases happen to include the sensing Hamiltonian presented in the main paper as Eq. 1. The PGH samples two particles \( \{\vec{x}_0, \vec{x}_1\} \) from the particle distribution \( P(\vec{x}_i) \), and then chooses:

\[
\tau = \frac{1}{\|\hat{H}(\vec{x}_0) - \hat{H}(\vec{x}_1)\|}.
\]  

(S3)

In the single parameter where only \( B \) is sensed, \( \tau \approx 1/\sigma(B_{\text{est}}) \), where \( \sigma(B_{\text{est}}) \) represents the standard deviation of the Gaussian-approximated posterior distribution \( P(B) \). Intuitively, this corresponds to selecting longer, more informative accumulation times, as the estimated uncertainty about the parameter to be learned shrinks.

**THE ROLE OF T\(_2^*\) IN TIME ADAPTIVITY**

In the main text, we observed the emergence of a slowdown in the learning rate, when MFL chooses accumulation times \( \tau_i \geq \tau_{\text{max}} \). This slowdown appears when plotting either the scaling in \( \sigma(B_{\text{est}}) \) as well as \( \eta \) (respectively Fig. 2akb, referring for example to the dataset \( \alpha_1 \)). This dataset represents a situation, where \( \tau_{\text{max}} \ll T^*_2 \) is chosen as a maximum time budget per-epoch. In this case, once the PGH encounters the \( \tau_{\text{max}} \) limit, learning by statistical accumulation of data-points with \( \tau \approx \tau_{\text{max}} \) will occur, and MFL precision scaling will tend to \( \eta^2 \propto 1/\tau_{\text{max}} T \). We highlight this behaviour, using averaged sequences from the whole set \( \alpha_1 \), in Fig. S1b, plotting the scaling in \( \sigma(B_{\text{est}}) \) alongside with the ratio \( \tau/\tau_{\text{max}} \). When the plateau in \( \sigma(B_{\text{est}}) \) occurs, we correspondingly observe that a typical run of MFL starts suggesting \( \tau \geq \tau_{\text{max}} \), before saturating as the uncertainty converges.

Note, this artefact deriving from the artificial choice of a maximum time budget \( \tau_{\text{max}} \) is equivalent to the phenomenon exhibited in correspondence of dephasing noise, reducing the contrast from experimental Ramsey fringes, like in the data reported in Fig. 4. To prove it, we show in Fig. S1c the same performance for the dataset \( \alpha_2 \), where \( \tau_{\text{max}} = 100 \mu s \approx 1.6 T^*_2 \) (estimated from a least-squares fit), and \( B = 6 \mu T \) to have approximately the same number of periods in the corresponding Ramsey fringe, as in dataset \( \alpha_1 \) (refer to Fig. 4). For MFL to deal properly with decaying data, in this analysis we remove any majority voting scheme from the data processing, and at each epoch the corresponding datum \( E \) is probabilistically extracted from the experimentally estimated likelihood (see Methods). This justifies the slowdown in the scaling of \( \sigma(B_{\text{est}}) \), as each data-point is now affected by the same amount of binomial noise that would occur in a set-up with the same readout fidelity, but single-shot measurements. In other words, the additional information acquired about the likelihood \( \mathcal{L}(B; \tau) \) by combining \( M \gg 1 \) sequences for each measurement is partially removed from the inference process by the bipartite \( E \) sampling. For this case, we observe the plateau in \( \sigma(B_{\text{est}}) \) occurs when the adaptive choice of the phase accumulation time saturates in average to \( \tau \approx T^*_2 \) (though a single run will oscillate around this value, as emphasised by the behaviour for a single run also reported in Fig. S1c). Similarly, also the scaling in precision plateaus when \( \tau \approx T^*_2 \) (not reported for brevity), slowing towards \( \eta^2 \propto 1/T^*_2 \). A formal discussion of this saturation is performed in the following section.

**PRECISION BOUNDS AND SENSITIVITY**

In assessing the performance of MFL, it is helpful to compare the uncertainties achieved with those achieved by FFT, using the same datasets. We begin by considering the Cramér–Rao bound [42]. Suppose that our procedure implements a function \( \hat{B}(\text{data}) \) of the entire data record \( \text{data} \) that estimates the true magnetic field \( B \). After, we want to minimise the squared error \( L = (\hat{B}(\text{data}) - B)^2 \) as much as possible. If \( \hat{B} \) as the average of \( \hat{B}(\text{data}) - B \) over all possible data records is zero, then we state that our procedure is unbiased. In this case, the Cramér–Rao bound provides, that on average over all data records one finds [43],

\[
\mathbb{E}[L] \geq \frac{1}{\sum_i \gamma_i^2 \tau_i^2},
\]

(S4)

where \( \tau_i \) is the evolution time used at the \( i \)th step of the MFL procedure. We stress that this inequality holds only on average; after all, we might be “lucky” with the estimate that we assign to any particular data record.
FIG. S2. **Key performances of MFL as the average number of photons collected $n_{\text{phot}}$ is increased**, estimated via 1000 independent runs of MFL, with an underlying lossless model (see Eq. 3 in Methods). **a** Scaling of the median error estimate for a selection of cases, as reported in the colour-coding legend. **b** Final median error achieved by MFL for all the cases considered, along with a power-law fit (dashed line). Error bars indicate 68.27 % percentile ranges. **c** Scaling of the quadratic losses for the same representative runs as in a. **d** Comparison of the final estimates for the Ramsey frequency provided for representative $n_{\text{phot}}$ cases, by FFT and QHL methods, respectively in solid and dash-dotted lines. In a&c, shaded areas indicate the credible intervals corresponding to 68.27 % percentile ranges.

The right-hand side of this inequality is derived using the Fisher information $I$ for a single measurement,

$$I(B, \tau) := \mathbb{E}_E \left[ (\partial_B \log P(E|\bar{x}, \tau))^2 \right]$$

$$= \tau^2 \gamma^2.$$  \hfill (S5)

The Fisher information for an experiment consisting of multiple independent measurements is given by the sum of the Fisher informations for each measurement, giving the Cramér–Rao bound (Eq. S4).

Let $T$ be the total phase accumulation time used for a single “run” of a magnetometry procedure; in our case, $T = \sum_i \tau_i$. By the above argument, $L$ can then scale no better than $T^{-2}$, corresponding to consolidating our phase accumulation into a single measurement. This observation is sometimes referred to as the **Heisenberg limit** for magnetometry.

At the other extreme, suppose that we have a total time budget of $T$, that we are able to spend on a magnetometry experiment, such that we can consider repeating a given procedure $N = T/\tau$ times. The factor of $N$ then factors out of the Cramér–Rao bound, giving

$$\mathbb{E}[L] \geq \frac{1}{N \sum_i \gamma_i^2 \tau_i^2} \propto \frac{1}{T}.$$  \hfill (S7)

The observation that $L \propto 1/T$ is sometimes referred to as the **standard quantum limit**, in the case that we repeat a magnetometry procedure for $N$ independent iterations. Indeed, we can use this observation to motivate a general figure of merit for the time budget of the a given magnetometry procedure. Assume that the Fisher information for a given procedure is $I = NI_0$, where $I_0(\tau)$ is the Fisher information for a single repetition using phase accumulation time $\tau$. Then it follows $I = T(I_0/\tau)$. Next, we define $\eta(T) := \sqrt{I_0(T)/\tau}$ as the **sensitivity** of the proposed magnetometry procedure.
Using this definition, we can then restate the standard quantum limit as the statement that \( \eta^2(T) \) is constant in \( T \). That is, a magnetometry procedure bound by the standard quantum limit gains no advantage from phase accumulation time beyond that conferred by repeating the entire procedure for \( R \) independent runs. By contrast, a Heisenberg limited magnetometry procedure has a sensitivity which scales as \( \eta^2(T) \propto 1/T \), indicating that an additional advantage can possibly be gained by using longer phase accumulation times.

So far we have considered the case in which \( T^*_2 \gg T \), such that we can approximate the dynamics of our magnetometry experiment as dephasing-free. The dichotomy between the Heisenberg and standard quantum limit scalings, however, is changed by dephasing such that we have to consider the definition of the sensitivity in the dephasing-limited case. In particular, Ref. [43] derived that the Fisher information for \( T^*_2 \)-limited magnetometry is given by

\[
\eta \left( \frac{\csc^2(B\gamma t)}{\gamma^2 t^2 (e^{\Gamma t} - 1)^2} \right),
\]

where \( \Gamma := 1/T^*_2 \) is used to represent dephasing in frequency units, in analogy with \( \gamma B \). We note, that unlike the Fisher information describing the noiseless case, the bound Equation S8 for the dephasing-limited case is not independent of the true value of \( B \). Thus, to determine the achievable sensitivity in the case of dephasing-limited magnetometry, we must either assume a particular value of the field \( B \) being estimated, or must generalise beyond the Cramér–Rao bound. We choose the latter case in this work, which provides further insight into the trade-off between phase accumulation time and experimental repetitions for \( \Gamma > 0 \).

Specifically, we consider the van Trees inequality (also known as the Bayesian Cramér–Rao bound) [44],

\[
E_{B,E}[L] \geq \frac{1}{J_\pi + E_B[I(B)]},
\]

where \( J_\pi \) describes the error that can be achieved using prior information, and \( E_B \) an expectation value over a distribution of different hypotheses about the field \( B \). We intentionally do not further define \( J_\pi \), as this term depends on the context in which a magnetometry procedure is used, rather than on the magnetometry procedure itself. Moreover, the effect of \( J_\pi \) is minimal in the limit of large experimental data sets, such that ineffectively consists of a correction to the Cramér–Rao bound in the case of finite data records [45].

In analogy to the Fisher information derivation above, the field-averaged Fisher information in the dephasing-limited case gives \( E_B[I(B)] \leq (\tau e^{-\tau\Gamma})^2 \) for a single phase accumulation \( \tau \). Hence, the analogous bound to Equation S4 is given by

\[
E_{B,E}[L] \geq \frac{1}{J_\pi + \sum_i (\gamma_i e^{-\gamma_i \tau\Gamma})^2}.
\]

To derive the sensitivity in the van Trees case, let \( J_0(T) = \argmax_{T} J_0(T,B) \) be the average Fisher information for a dephasing-limited procedure. We can then define the average sensitivity \( \bar{\eta}(T) = J_0(T)/T \) for a total phase accumulation time \( T \) to reformulate the van Trees inequality in a more practical form for our purposes, thus

\[
E_{B,E}[L] \geq \frac{1}{J_\pi + T\bar{\eta}^2(T)}.
\]

Following Equation S10, \( \bar{\eta}^2(T) \) is constant if a fixed phase accumulation time is used, while \( \bar{\eta}^2(T) \propto 1/T \) if \( TT \ll 1 \). The average Fisher information saturates at \( T = T^*_2 \), however, such that the Heisenberg and standard quantum limits coincide as \( T \) approaches \( T^*_2 \). Therefore, the performance observed in Fig. 1a is limited by saturation near \( T = T^*_2 \gtrsim 0.1s \).

**ABSOLUTE PRECISION SCALING**

As discussed in the main text and Methods, using a room temperature set-up can be challenging for the effect of quantum projection noise and readout infidelities. These need to be properly addressed when reduced sequence repetitions lead to a low number of PL photons to be detected when recording a fringe. The results in terms of absolute scaling \( \bar{\eta} \) have already been discussed (see e.g. Fig. 3 in the main paper). Here, we complete those analyses with additional studies. In Fig. S2a&b, we report respectively the scaling and ultimate uncertainty achievable by MFL after 150 epochs for a subset of cases with \( n_{phot} = 1...2475 \). Fig. S2b suggests an approximate \( \propto 1/n_{phot} \) gain in the uncertainty achievable halting the protocol after a fixed number of steps. From Fig. S2c we observe that the choice for \( N \) in this case is motivated by running MFL for enough steps, to observe for all cases the convergence of
the median quadratic losses – i.e. the square error in the parameters’ estimate, here $Q.L. := (B - B_{est})^2$. To estimate the true $B$, we run MFL once over the whole dataset ($M = M^{tot} = 20275$), checking that the result is consistent with FFT.

We remark how the advantages in increasing the $n_{phot}$ used (along with the higher precision scaling that increases from $\tilde{T}^{-0.73}$ for $n_{phot} = 4$ to Heisenberg limited for $n_{phot} \geq 20$) come at the expense of worse final absolute sensitivities achievable by the protocol. This is due to the linear increment of experimental overheads with $n_{phot}$.

The robustness against sources of noise present in the room temperature set-up is emphasized by Fig. S2d, where we plot the estimates obtained by MFL for a similar subset of $n_{phot}$’s analysed. We observe that for $n_{phot} \geq 4$, MFL estimates are all substantially consistent with the result obtained for $M = M^{tot}$, within the estimated uncertainty $\sigma(B_{est})$, and taking into account minor fluctuations in $B$ that might have occurred during the collection of the whole dataset. By contrast, we observe how FFT estimates are completely unreliable at the noise level corresponding to $O(n_{phot}) = 10$.

![Image](image_url)

**FIG. S3.** **Ramsey sets mediated through different numbers of single sequences,** corresponding to the various $n_{phot}$ reported on the axis below. Experimental data (as blue dots) are reported together with a sinusoidal fit obtained from the case $n_{phot} = 2534$ (as red lines). The unbalance towards the 0 measurement outcome is evident in the cases $n_{phot} = 1, 4$. Data-points whose normalisation is higher than 0.5 correspond to Poissonian distributed multi-photon events still present in this case.

**THE ROLE OF NOISE FOR LOW PL PHOTON COUNTS**

Finally, we observe how for $n_{phot} < 4$, the Bayesian process fails due to increased experimental noise and reduced statistics, underestimating both the real $\omega$ and the uncertainty associated with it. For example, this is evident from Fig. S2c, as the $Q.L.$ does not improve with the number of epochs. In particular, losses in the system cause an asymmetry between $\xi_0$ and $\xi_1$, respectively the overall readout fidelities for the states $|0\rangle$ and $|1\rangle$ (i.e. taken all sources of noise and loss into account). From experimental raw data for $n_{phot} = 1$ (see Fig. S3), we observed that if we assume $\xi_0 \sim 1$, then $\xi_1 \simeq 0.54$. This translates in unbalanced output probabilities, that conflict with the underlying assumption made so far of a binomial model for the outcomes $E \in \{0, 1\}$, with probabilities given by the likelihood in Eq. 3. This level of “poisoning” in the assumed model is evidently beyond the CLE noise robustness [41].

In order to prove there is no fundamental limit preventing MFL to provide correct estimates, within uncertainty, given a correct model, we thus modified the likelihood such that:

$$L'(1|\bar{e}, \tau) = \xi L(1|\bar{x}, \tau)$$

(S12)

where $\xi \in [0, 1]$, $L'(0|\bar{e}, \tau) = 1 - L'(1|\bar{e}, \tau)$ and for $\xi = 1$ we recover the usual $L'(0|\bar{e}, \tau) = L(0|\bar{x}; \tau)$ of Eq. 3. In order to estimate $\xi$, we use it as the free parameter in a preliminary CLE run against the same dataset, but assuming $\omega$
known from the inference process with $M = 20275$ (i.e. having $\mathcal{F} = \{\xi\}$). We thus obtain $\xi = 0.72$, and use this as a known parameter when running MFL with $\mathcal{L}'(B)$. In principle, $\xi$ could also be estimated from a multi-parameter inference model.

The result is reported in Fig. S4. Intuitively, measurement outcomes are interpreted as less informative by the inference process, as $E = 0$ might be due to additional losses. This effectively slows down the learning rate per-epoch, but at the same time restores a correct behaviour of MFL.

![FIG. S4. Noise-compensation in the inference process. a, Estimate of the precession frequency $\omega$ from the set $\epsilon_1$, using $M = 8 \Rightarrow n_{\text{phot}} = 1$ average photons collected per step, in peak configuration. In violet the result adopting the usual likelihood in Eq. 3, capable of handling only Poissonian noise in the the data. In blue, results from the modified likelihood Eq. S12, allowing for an extended number of epochs. Shaded areas here represent the median $\sim 68.27\%$ credible interval provided intrinsically by MFL at each epoch, averaged over 1000 runs. b, Estimates for $\omega$, and uncertainties as a Gaussian fit over the learnt posterior, for the two cases in a, along with some other representative runs from Fig. S2, after 150 epochs. The inference process with no model for infidelities in place, and $n_{\text{phot}} = 1$, falls outside of the plotted interval.]

WIDE RANGE OPERABILITY OF MFL METHODS

It is known how in Ramsey experiments, adaptive choices of time can lead not only to scalings beyond the standard quantum limit, but also to improved dynamic ranges for the sensed magnetic field $B$, up to $\omega_{\text{max}}(B)/\sigma(\omega) \leq \pi T/\tau_{\text{min}}$. Given that MFL is adaptive in the choices of the phase accumulation time $\tau$, and we have shown that its precision scaling is Heisenberg limited, it follows naturally that also MFL benefits from the high-dynamic range already reported by previous experiments.

In the main paper, we already showed applicability of MFL for cases in the dataset $\epsilon_3$, with $B \in [8.3, 550] \mu T$ (see Fig. 2 in the main text). Here we complement this study with an additional case ($\alpha_2$) exhibiting $B_0 = 713 \mu T$. In the case of this dataset, equivalently to $\alpha_1$, $M \sim 20000$ single sequences were collected and averaged from the experimental set-up, so it was reasonable to adopt a majority voting scheme to use the additional information in the data. We stress that such high intensities of $B$ tend to make least-squares fit procedures with no initial guess of the parameters fail.

The results in terms of $\sigma(B_{\text{est}})$ and precision $\eta^2$ scalings are reported in Fig. S5. We observe how after 250 epochs, the difference in the final uncertainties provided by MFL is $|\sigma(B_{\alpha_1}) - \sigma(B_{\alpha_2})| \approx 10^{-3} \mu T$. It can thus be considered approximately independent of the strength of the magnetic field. Also the precision scaling is the same, within error, of the one observed for the lower field in $\alpha_2$.

TRACKING

In the main text, we tested against experimental data the tracking capabilities of the MFL protocol. In Fig. 4 we reported the results in the case where the magnetic field intensity $B$ is synthetically altered stepwise, at random times, in a fashion completely equivalent to a stochastic time-dependent Poisson process $P(t)$. Such random, abrupt variations in the magnetic field might for example reproduce applicative scenarios such as the raster scanning of a surface embedding magnetic nanoparticles. A sketch of a possible experimental set-up is provided in Fig. S6a. The modifications to the standard CLE inference process required by this particularly demanding tracking scenario are summarised as pseudocode in Algorithm 1. The modifications to the standard inference process adopted amount to detect changes in the sensed parameter, that completely invalidate the current posterior, and thus suggest a reset
of the prior as the most effective update step. Without triggering such reset events, huge stepwise changes would otherwise require a long time for MFL to react, because of the little support provided by the prior to the new value.

In Fig. S6b, we show a simulated performance of MFL in a representative run with time-varying $B$. The figure exemplifies the decrease in the rate of failure events, as the frequency of the oscillating signal is decreased with time. We loosely define failure events, all those $\tau_i$ at which the quadratic loss of a single run $Q.L.(\tau_i) \gg \sum_R Q.L.(\tau_i,r)/R$, the mean performance achievable by the protocol, estimated across $R \gg 1$ independent runs. We modify synthetically the magnetic field in the simulations as $B(\tau) = B_0 + b \cos(\nu \tau)$ with $b < B_0$, equivalently to Fig. 4c of the main text, but in this case we chirp the oscillating frequency for each run, and thus $\nu(\tau) = \nu_0 - k \tau$, with $\nu_0$ and $k$ constants. We notice how points where the second derivative of the oscillating magnetic signal is highest are those where failure events tend to occur.

Finally, in Fig. S6c, we display the performance expected for MFL when tracking a brownian-like varying signal. Here the ‘true’ signal is numerically simulated according to an Ornstein-Uhlenbeck process, similarly to the theoretical analysis in [34].
Algorithm 1: MFL algorithm with stepwise change detection

Input : An initial prior distribution $\pi_{ini}$ over models.
Input (additional) : $r_{\text{resample}} \triangleright$ a rate parameter estimating how many epochs occur before RESAMPLE is called
Input (additional) : $p_{\text{reset}} \triangleright$ parameter adjusting the frequency of posterior-reset events

function \text{EstimateAdaptive} (n, $\pi_{ini}$, N, a (the resampling parameter), $t_{\text{resample}}$ (the resample threshold), Optimiz\textunderscore Util, $n_{\text{guesses}}$, \text{GuessExperiment}):

\begin{align*}
    w_i &\leftarrow 1/n \\
    l_{\text{evres}} &\leftarrow 0 \\
    l_{\text{evstp}} &\leftarrow 0 \\
    \text{initialize } \bar{x} = [\bar{x}_1, ..., \bar{x}_N] \\
    \text{for } i_{\exp} \in 1...N \text{ do:}
    \begin{align*}
        &... \\
        \text{if } \sum_i w_i^2 < N t_{\text{resample}}: \triangleright \text{ if the effective sample size is below the threshold} \\
        \text{if } |i_{\exp} - l_{\text{evres}}| \geq r_{\text{resample}} \text{ OR } |i_{\exp} - l_{\text{evstp}}| \leq p_{\text{reset}}: \triangleright \text{ resample as usual} \\
        \{w_i\}, \{x_i\} &\leftarrow \text{RESAMPLE} (\{w_i\}, \{x_i\}, a) \\
        l_{\text{evres}} &\leftarrow i_{\exp} \triangleright \text{ store last resampling event} \\
    \text{else:} \triangleright \text{ reset the procedure} \\
        \pi &\leftarrow \pi_{ini} \\
        w_i &\leftarrow 1/n \\
        \text{draw each } \omega_i \text{ independently from } \pi_{ini} \\
        l_{\text{evstp}} &\leftarrow i_{\exp} \triangleright \text{ store last reset event} \\
        \text{continue from } i_{\exp} + 1 \\
    \text{end if}
    \end{align*}
    \text{end if}
\text{...} \\
\bar{x}_i &\leftarrow \text{MEAN} (\{w_i\}, \{x_i\}) \triangleright \text{ append the new estimate from the mean}
\text{end for}
end function

Output: $\bar{x}$, storing the instantaneous values of the unknown parameter