Sequential Bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers

Sergey Dushenko[1,2], Kapildeb Ambal[1-3], Robert D. McMichael[1]

[1]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

[2]Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD 20742, USA

[3]Department of Physics, Wichita State University, Wichita, KS 67260, USA

Published November 16, 2020
Physical Review Applied 14, 154036 (2020)
DOI: 10.1103/PhysRevApplied.14.054036

ABSTRACT

In magnetometry using optically detected magnetic resonance of nitrogen vacancy (NV^-) centers, we demonstrate more than one order-of-magnitude speed up with sequential Bayesian experiment design as compared with conventional frequency-swept measurements. The NV^- center is an excellent platform for magnetometry with potential spatial resolution down to few nanometers and demonstrated single-defect sensitivity down to $nT/Hz^{1/2}$. The NV^- center is a quantum defect with spin $S = 1$ and coherence time up to several milliseconds at room temperature. Zeeman splitting of the NV^- energy levels allows detection of the magnetic field via photoluminescence. We compare conventional NV^- center photoluminescence measurements that use pre-determined sweeps of the microwave frequency with measurements using a Bayesian inference methodology. In sequential Bayesian experiment design, the settings with maximum utility are chosen for each measurement in real time based on the accumulated experimental data. Using this method, we observe an order of magnitude decrease in the NV^- magnetometry measurement time necessary to achieve a set precision.

INTRODUCTION

This study focuses on magnetometry using optically detected magnetic resonance of NV^- centers. The ability to optically prepare and manipulate spin states, along with a long spin lifetime and robustness to the environment made NV^- centers a promising platform for application in various areas. A few prominent examples include quantum computing [1], cryptography [2] and memory [3,4]; bio-compatible markers [5] and drug delivery [6]; mechanical [7], temperature [8,9], electric [10] and magnetic sensors [11–13]. The concept of NV^- center magnetometry [14] was experimentally demonstrated in 2008 in two independent studies by Balasubramanian *et al*. [11] and Maze *et al*. [12], followed by hundreds of other studies [15].

Magnetometry-based imaging using NV^- centers promises several advantages over the existing magnetic imaging and scanning techniques. The NV^- center does not carry a significant magnetic moment, making it a non-invasive technique, unlike magnetic force microscopy (MFM) which can suffer from the interaction between sample and the magnetic tip. Magneto-optic Kerr effect (MOKE) microscopy is limited by optical resolution and is mostly suitable only for studying materials with a strong Kerr effect. In contrast, NV^- center magnetometry spatial resolution is ultimately only limited by the distance between NV^- center and the sample, which can be less than ten nanometers [16]. Superconducting quantum interference device (SQUID) magnetometry provides unrivaled sensitivity but requires cryogenic temperatures, and has low spatial resolution, though attempts at miniaturizing the technology are in progress [17]. NV^- center magnetometry can operate in a broad range of temperatures, including room temperature and above. These advantages make NV^- center an excellent platform for magnetometry [15,18,19] with potential spatial resolution down to few nanometers and demonstrated sensitivity down to $nT/Hz^{1/2}$ [20,21].

Recent research efforts have been directed at increasing the speed and precision of NV^- center magnetometry measurements. Some of these research efforts summon help from additional hardware to achieve the goal. By modulating the microwave frequency that drives spin-state transitions of the NV^- center and by demodulating the photoluminescence signal using lock-in amplifiers, significant gains in signal-to-noise ratio and measurement speed have been achieved [20,22–24]. However, such an approach generally requires a high photoluminescence signal by simultaneous measurement of multiple NV^- centers, which sacrifices the spatial resolution. Another approach that uses specialized hardware is using the differential photon rate meter that can track photoluminescence signal even at low photon count rate, though it does not significantly improve signal-to-noise ratio [25]. In addition to "hardware" approaches, sophisticated algorithms—"software" approaches—have also shown promise. Simulations have showed that neural networks improve NV^- center readout fidelity [26]. Sequential Bayesian experiment design [27] is another promising machine learning "software" approach. Theoretical studies have discussed how Bayesian methodology [28–31] can be used in determining the unknown parameters of a quantum system [32–36], and magnetometry in particular [37–40]. Encouragingly, in recent experimental studies Bayesian methodology has proven to be advantageous in quantum Hamiltonian learning [41] and measurements of pulsed Ramsey magnetometry using NV^- centers [42,43]. In this study, we show how combining sequential Bayesian experiment design with conventional optically detected magnetic resonance NV^- center magnetometry leads to better measurement strategies. In particular, we carry out experiments that compare using a conventional—swept-frequency NV^- center magnetometry protocol—with the measurements that incorporate sequential Bayesian experiment design.

BACKGROUND

Many of the useful properties of NV^- centers hinge on the fact that their photoluminescence depends on their spin state. The NV^- center is created when two adjacent carbon atoms in a diamond lattice are substituted with a vacancy and a negatively charged nitrogen atom, forming a spin $S = 1$ quantum defect (Fig. 1(a), see Supplemental Material [44] section S.1 for more details). Photon absorption moves the NV^- center from the ground state to the excited state, while preserving its spin projection m_S (Fig. 1(b)) [45,46]. Eventually, the center relaxes back to the ground state, but the relaxation process is spin dependent. An excited state with $m_S = 0$ mostly relaxes back to the ground state with $m_S = 0$ by emitting a red photon. In contrast, the excited state with $m_S = \pm 1$, can relax by two mechanisms: either

back to the ground state with $m_S = \pm 1$ by emitting a red photon, or to any m_S through a dark state, without emitting a visible photon (detailed energy level structure of NV^- center can be found in Supplemental Material S.1). Hence, photoluminescence of NV^- centers under laser excitation is brighter if the center is initially in $m_S = 0$ and dimmer if it is in $m_S = \pm 1$ states. This phenomenon allows optical read-out of the spin state by monitoring the photoluminescence rate. Additionally, the ground state with $m_S = 0$ of the NV^- center can be prepared by continuous illumination that cycles NV^- centers through ground state—excited state—ground state transitions. Since the $m_S = \pm 1$ state can transition to the $m_S = 0$, but no reverse transition is available, eventually, the center ends up in $m_S = 0$ with high probability. In all, the spin-dependent optical relaxation allows the spin state to be both initialized and read out.

The spin state of the NV^- center can also be controlled with microwaves. When the microwave photon energy matches the energy difference between the ground levels with spin projection $m_S = 0$ and the $m_S = \pm 1$ spin state, transitions occur. The microwave energies at this resonance conditions are given by

$$E_{\mathrm{MW}} = hf_{\mathrm{MW}} = hD_{\mathrm{GS}} + g\mu_{\mathrm{B}}\Delta m_S B + m_I A_{\mathrm{GS}}^{\mathrm{HF}}, \tag{1}$$

where $h \approx 6.62\times10^{-34}$ J/Hz is the Planck constant, f_{MW} is the microwave frequency, $D_{\mathrm{GS}} \approx 2.87$ GHz is the zero-field splitting, $g \approx 2$ is the electron g-factor inside the diamond lattice, $\mu_{\mathrm{B}} \approx 9.27$ J·T^{-1} is the Bohr magneton, Δm_S is the spin projection difference between the final and initial ground states, B is the applied magnetic field, m_I is the nuclear spin projection (preserved in the transition), and $A_{\mathrm{GS}}^{\mathrm{HF}}$ is the energy correction due to the hyperfine interaction of the ground state levels with ^{14}N nucleus (spin $I = 1$). Note that strain-induced splitting of the energy levels in diamond should also be considered when measuring small magnetic fields below 1 mT.

Optically detected magnetic resonance [47,48] is observed as a reduction in photoluminescence. Constant illumination populates the $m_S = 0$ state, and dips in the photon count are observed when microwaves induce transitions to the $m_S = \pm 1$ states. One can extract value of the external magnetic field B from the frequencies of the dips in the photoluminescence spectrum that correspond to the frequencies when NV^- center transitions to $m_S = +1$ and $m_S = -1$ states (Fig. 1(c)). This technique is a basis of NV^- magnetometry.

The resonance frequencies described in (1) yield a model for the normalized photon count signal ($y = \{\mu\}$) that is a combination of three Lorentzian curves, one for each of the ^{14}N nuclear I_z states in the hyperfine interaction-split spectrum of the NV^- center:

$$\mu = 1 - \frac{a \cdot k_{\mathrm{NP}}}{(f - f_B - \Delta f_{\mathrm{HF}})^2 + \Omega^2} - \frac{a}{(f - f_B)^2 + \Omega^2} - \frac{a/k_{\mathrm{NP}}}{(f - f_B + \Delta f_{\mathrm{HF}})^2 + \Omega^2}. \tag{2}$$

Here f_B is the center resonance frequency that corresponds to the NV^- center transition from $\{m_S = 0, m_I = 0\}$ to $\{m_S = +1, m_I = 0\}$ state, $\Delta f_{\mathrm{HF}} = A_{\mathrm{GS}}^{\mathrm{HF}}/h$ is the hyperfine splitting, a is an overall contrast factor, Ω is a linewidth, and k_{NP} characterizes the nuclear polarization. The coupling between NV^- center electrons and the nitrogen nucleus spin (naturally abundant ^{14}N, $I = 1$) leads to the weak spin transfer of constant polarization of the electron spin to nucleus. However, the nitrogen nucleus is not fully polarized in the presence of slight misalignment of the external magnetic field with the axis of the NV^- center [49,50]. This leads to the splitting of the NV^- center transitions into three photoluminescence dips of different amplitudes corresponding to $m_I = -1$, 0 and +1, which are separated in frequency by the hyperfine splitting

Δf_{HF} (Fig. 1(c)). For every measurement with microwave excitation, a reference photon count with microwaves switched off is used as a normalizing factor. Throughout this paper, we treat the excitation frequency f as the lone experimental setting design $d = \{f\}$ and the five parameters $\boldsymbol{\theta} = \{f_B, \Delta f_{\mathrm{HF}}, a, \Omega, k_{\mathrm{NP}}\}$ as unknowns.

We use the triple-resonance spectrum described by (2) to compare the effectiveness of measurement protocols. The goal of the experiment is to determine the center resonance frequency f_B. The external magnetic field in NV$^-$ magnetometry is given by the equation $|B| = (h/g\mu_{\mathrm{B}}) \cdot (f_B - D_{\mathrm{GS}})$, where $g\mu_{\mathrm{B}}/h \approx 28$ MHz/mT is the combination of the physical constants. The search range for the signal frequency was from 3040 MHz to 3200 MHz, which corresponds the magnetic field in the range from 6 mT to 12 mT. The generated electromagnet field was set to $B \approx 8.32$ mT (picked by a random number generator) for the results shown in this paper, corresponding to the NV$^-$ resonance frequency $f_B \approx 3103$ MHz. The field was treated as an unknown in the measurements and data analysis.

In the conventional NV$^-$ magnetometry measurements the photoluminescence of the sample was monitored while scanning the microwave frequency from 3040 MHz to 3200 MHz with 20 kHz step. Hence, each frequency scan consisted of 8000 normalized photoluminescence measurements.

The sequential Bayesian experiment design measurements iterated over a three-step cycle comprising a setting choice (design) from the allowed microwave frequencies, measurement, and data analysis via Bayesian inference. Here, we provide an overview of the process, and direct the interested reader to the Supplemental Material [44] (sections S.2 and S.3) and the references [27,34,51,52] for more detailed descriptions.

Bayesian methods treat the unknown parameters $\boldsymbol{\theta}$ as random variables with a probability distribution $p(\boldsymbol{\theta})$. In this application, $\boldsymbol{\theta} = \{f_B, \Delta f_{\mathrm{HF}}, a, \Omega, k_{\mathrm{NP}}\}$ are the parameters of the model function given in Eq. (2). After *n* iterations, the parameters are described by a conditional distribution $p(\boldsymbol{\theta}|\boldsymbol{y_n}, \boldsymbol{d}_n)$ given accumulated measurement results $\boldsymbol{y}_n = \{y_1, y_2, \dots, y_n\}$ obtained at frequency settings (designs) $\boldsymbol{d}_n = \{d_{1,} d_2, \dots, d_n\}$.

In the $n + 1^{\mathrm{th}}$ iteration, the experiment design step uses the parameter distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_n, \boldsymbol{d}_n)$, to inform the choice of a setting design d_{n+1} for the next measurement. The algorithm models a distribution of measurement predictions for each possible design and then predicts the average improvement in the parameter distribution that would result from the predicted data. "Improvement" is quantified as a predicted change in the information entropy of the parameter distribution and it is expressed as a utility function $U(d)$ [53,54]. The derivation of $U(d)$ produces a qualitatively intuitive result: it does the most good to "pin down" the measurement results where they are sensitive to parameter variations. The new setting d_{n+1} is selected to maximize $U(d)$.

After the setting d_{n+1} is used to obtain the measurement result y_{n+1} these values are used to refine the parameter distribution. Using Bayesian inference,

$$p(\boldsymbol{\theta}|\boldsymbol{y_{n+1}}, \boldsymbol{d}_{n+1}) \propto p(y_{n+1}|\boldsymbol{\theta}, d_{n+1}) p(\boldsymbol{\theta}|\boldsymbol{y_n}, \boldsymbol{d}_n) \tag{3}$$

Where $p(y_{n+1}|\boldsymbol{\theta}, d_{n+1})$ is the *likelihood*, the probability of observing the measured value y_{n+1} calculated for arbitrary parameter values $\boldsymbol{\theta}$ given the frequency setting d_{n+1}. With increasing iteration number, the parameter distribution typically narrows, reflecting increasingly precise estimates of the parameter values.

In each iteration, the sequential Bayesian experiment design algorithm makes an informed setting decision and incorporates new data to inform the next decision. On a qualitative level, the Bayesian method formalizes an intuitive approach of making rough initial measurements to guide later runs, but the Bayesian method offers additional advantages. Bayesian inference incorporates new data, allowing for semicontinuous monitoring of "fitting" statistics, and result-based stopping criteria. The utility function provides a non-heuristic, flexible, data-based method for setting decisions. These advantages are especially important for situations where automation is required, speed is essential, or measurement data is expensive.

Software and documentation for sequential Bayesian experiment design is provided online [55].

EXPERIMENTAL DETAILS

In this study, we used a commercially available single crystal diamond grown by chemical vapor deposition (CVD). Sample size was 3.0 mm $\times$ 3.0 mm $\times$ 0.3 mm, with {100} top surface orientation and surface roughness below 30 nm. The diamond (type IIa) had nitrogen concentration below 1 ppm and boron concentration below 0.05 ppm according to the manufacturer. The sample was mounted on top of the 50 mm long microstrip line, which was used to supply microwaves to manipulate spin state of the NV^- center. The microstrip line with the sample was placed in an electromagnet between pincer-shaped poles that were oriented to align with the [111] direction of the diamond lattice ($\arcsin\sqrt{2/3} \approx 54.7°$ from the vertical). In this arrangement, the magnetic field is pointing along one of the four possible orientations of NV^- center axes (vector connecting nitrogen atom to the vacancy site).

A green laser with 520 nm wavelength was used to optically excite NV^- center. The 0.7 numerical aperture (NA) objective of a custom-built confocal microscope was located above the sample to focus laser excitation inside the diamond and to collect fluorescence from the NV^- center. A dichroic beamsplitter with the edge at 650 nm was used to separate excitation laser light from the collected fluorescence. After further wavelength selection with 647 nm long-pass filters, the collected fluorescence was coupled into a multimode fiber and directed to the photon detector. For each data point, a 50 ms photon count with the microwaves on was divided by a subsequent 50 ms reference count with microwaves off. The excitation using green laser light was on continuously. Only 10 mW of microwave power (at the source) and 225 μW of laser power (before the objective) were sent to the sample. The laser power was set using the linear polarizer and the half-wave plate. The combination of laser power, microwave power and counting time produced measurements with a signal-to-noise ratio on the order of 1. Such experimental setup showcases ability of sequential Bayesian experiment design to locate and measure complex multiple-peak signal even in extremely noisy data, and shows its broad dynamical range for sensitivity.

RESULTS AND DISCUSSION

First, we report the results of the conventional NV^- magnetometry measurements. Figure 1(d) shows the photoluminescence data measured in one frequency scan. Dips in the photoluminescence spectrum corresponding to optically detected magnetic resonance are visible with a signal-to-noise ratio on the order of one. We follow the conventional approach to improve the signal-to-noise, which is to re-measure the same scanning range and average the data in the scans. Figures 1(e) – 1(g) show averaged

data for increasing numbers of scans. The signal-to-noise ratio improves as the inverse square root of the number of the averaged scans.

To gauge the evolution of parameter uncertainty as a function of scan number, we "fit" the averaged data using Bayesian inference to determine mean values and standard deviations from the parameter distribution. To allow direct comparison, we used the same algorithm for Bayesian inference as in the sequential design data below. Like the overall signal-to-noise ratio, the standard deviation of the resonance frequency also follows an inverse square root dependence on the total number of the scans (Fig. 1(h)).

Photoluminescence data of the NV^- magnetometry measurements using sequential Bayesian experiment design are shown in Figs. 1(i) – 1(l). Here the data are plotted without averaging. While initial frequency sampling roams across the whole allowed frequency range (Figs. 1(i) and 1(j)), the later measurements almost exclusively focus on the signal location near the resonance dips where the photoluminescence value is lower (Figs. 1(k) and 1(l)). The standard deviation σ_f of the center resonance frequency f_B is plotted as a function of the number of measurements in Fig. 1(m). The standard deviation drops by three orders of magnitude within the first two hundred measurements.

We plot evolution of the probability distribution $p(\boldsymbol{\theta})$ of the signal frequency f_B and hyperfine splitting Δf_{HF} parameters in Figure 2. The probability distribution is implemented using sequential Monte Carlo where the probability density in parameter space is represented by the density of points and by a weight factor attached to each point. After each measurement, the weights are recalculated using Bayesian inference. Fig. 2(a) shows the initial, *prior* distribution, which consists of 10 000 points distributed through the parameter space with equal weights of 10^{-4} (Fig. 2(a)). The sum of all weights is normalized to 1.

Fig. 2(b) plots the probability distribution after the first measurement, which yielded μ_1 = 1.014 for the normalized photon count at $f_1 = 3154.26$ MHz. Since the resonances are dips in the photon count, values of $\mu > 1$ reduce the likelihood that the resonances are located near the measurement frequency f_1. To highlight this effect, distribution points with weights $w < 10^{-4}$ are colored cyan and weights $w \geq 10^{-4}$ are red. After several cycles of measurements and updating the weights, a resampling algorithm redistributes points, allowing high-weight points to survive, multiply, and diffuse slightly while low-weight points face a greater probability of elimination (see section S.4 of Supplemental Material). Resampling allows the computational resources to be focused on high-probability regions of parameter space without completely abandoning low-probability regions. The effects of resampling are visible in Fig. 2(d) and later panels with the higher concentration of points near 3090 MHz. After the first two hundred measurements, the $p(f_B)$ distribution has effectively contracted from spanning over the range of 150 MHz to less than 1 MHz (Figs. 2(k) and 2(l)). Interestingly, redistribution of the weights also allows probability distribution to diffuse beyond the initial boundary conditions. For example, initial weights occupy Δf_{HF} parameter space from 1 MHz to 3 MHz (Figs. 2(a) – 2(c)), but after 100 measurements, resampling steps have allowed the probability distribution to span Δf_{HF} parameter space from 0.5 MHz to 4 MHz. This diffusion allows slow convergence to values outside the prior distribution—i.e., in the areas where the experimenter does not expect to find final parameters' values—which is helpful in cases when experimenter does not have an accurate initial estimate for parameter.

The evolution of the NV^- magnetometry measurements using sequential Bayesian experiment design is in sharp contrast with the evolution of the conventional NV^- magnetometry measurements. The standard deviation of the signal frequency using sequential Bayesian experiment design follows a typical pattern displayed in Figure 1(m). After an initial period of broad sampling of parameter space, the algorithm focuses measurements near the resonance frequencies (Fig. 3(a)) and the probability distribution $p(f_B)$ contracts rapidly. After this contraction, the standard deviation of f_B decreases as the inverse square root of the total number of the measurements n (Fig. 1(m)). In contrast, the standard deviation of the signal frequency in the swept-frequency measurements does not go through such rapid contraction phase and follows an inverse square root of n scaling from the beginning (Fig. 1(h)).

The difference in the measurement strategies can be clearly seen in the photoluminescence data for the first thousand measurements. Sequential Bayesian experiment design has already narrowed down the probability distribution $p(f_B)$ for the signal frequency, and most of the measurements are taken at the signal position—the location of the three hyperfine-split dips (Figs. 3(a) orange solid circles, 3(c) and 3(d)). In contrast, the frequency sweep in the conventional measurements has not even reached the frequency where the signal is located, and all 1000 data points were spent on measuring the background (Figs. 3(a) purple solid circles and 3(b)). After 24 000 measurements (3 full range conventional sweep scans), only 3 measurements were performed at each frequency at the signal location by the conventional NV^- magnetometry (Fig. 3(g)), compared with peak of 214 measurements per frequency for sequential Bayesian experiment design measurements (Fig. 3(j)). This concentration of measurements results in a standard deviation of the averaged Bayesian measurement (Fig 3(i) that is an order of magnitude smaller than in the conventional measurement (Fig. 3(f)).

An interesting behavior of the utility function $U(\boldsymbol{d}=\{f\})$ can be seen in Fig. 3(j). In the central, m_I = 0 photoluminescence dip area most of the measurements are concentrated near its center (frequency f_B) while at the outer dips located at $f_B - \Delta f_{HF}$ and $f_B + \Delta f_{HF}$, measurements are concentrated on the sides of the dips, producing double-peak structures in the distribution of the measurements (Fig. 3(j)). In simulations and measurements on single-dip resonances, similar focus on the sides of dips is typical behavior, and it is consistent with the high sensitivity of the sides of the dip model to the resonance frequency parameter. On the other hand, the central concentration of measurements that we observe at the central dip in Fig. 3(j) would be atypical behavior for single resonances. We speculate that this behavior stems from the triple-resonance model's (2) implicit assumption that the center resonance lies at the midpoint between the outer resonances.

The "smart" measurement strategy of taking data into account on the fly—instead of waiting until the end of the experiment—allows the NV^- magnetometry based on sequential Bayesian experiment design to dramatically outperform conventional NV^- magnetometry. For example, to achieve the precision of σ_f = 5.5 10^{-3} MHz standard deviation of the signal frequency, the conventional sweep-based NV^- magnetometry requires 10^6 measurements, while the NV^- magnetometry based on sequential Bayesian experiment design requires only 24 350 measurements to achieve the same precision. Using the ratio between $1/\sqrt{n}$ scaling of the standard deviations of the signal frequencies for two methods (Fig. 4), sequential Bayesian experiment design magnetometry was determined to be 45 times faster than the conventional measurement approach.

Up to this point, we have compared measurement protocols on the basis of the number of measurements, but "wall-clock" time may be a more relevant basis for comparison, since sequential

Bayesian experiment design comes with an added cost of computational time. Photons from NV^- centers are counted for 100 ms at each data point (50 ms with microwaves on, followed by 50 ms with microwaves off). In the conventional protocol, the average time spent on measuring one data point is 150 ms. The additional 50 ms time is spent on communication between the devices, saving data etc. Using sequential Bayesian experiment design, the average time spent on measuring one data point is 204 ms, a 36 % (54 ms) increase in measurement time compared with the conventional setup. The additional time represents the added computational cost of Bayesian inference and utility calculations for each measurement. The computation time depends on computer hardware and programming methods. Here we report results using a single processor core of an ordinary PC programmed in Python using the Numpy package (see S.4 of Supplemental Material). Compiled code and parallel computation offer avenues for significant reductions in computation time [56,57]. The cost of an additional processor (several hundred dollars) is also negligible compared with the cost of the other hardware typically used in the NV^- magnetometry experiments. However, in the light of the 4400 % speedup, the associated additional Bayesian computation time (36 % longer measurement time) is negligible, even when performed on the ordinary processor and without using parallel threads.

In the NV^- measurements that we have carried out using sequential Bayesian experiment design, we always observe more than one order of magnitude speedup. The amount of speedup depends on experimental setup, signal, set of parameters and settings, and reaches close to two orders of magnitude for some of the experiments that we have carried out. A big factor that influences the speedup is the fraction of settings space occupied by the signal, compared to the whole space spanned by the settings d (scanning or sensing range). In the experiment described in this paper, signal occupies roughly 10 % of the whole scanning range (16 MHz out of the 160 MHz frequency range: 8 MHz is occupied by the dips and 4 MHz on each side by their shoulders). This value can be much smaller in magnetometers/sensors with broad sensing range, which will lead to even larger speedups. However, a focus on the measurements with maximum utility function allows sequential Bayesian experiment design to be beneficial even for measurements where signal is present throughout the whole settings space d (see section S.5 of the Supplemental Material for more details). As a rule of thumb, the more time an experimental procedure spends on measuring data with low utility function values (for example, areas away from the signal or areas with small signal-to-noise ratio), the more beneficial will be implementation of the measurements using sequential Bayesian experiment design. Sequential Bayesian experiment design can be particularly useful for maturing NV^- center magnetometry technology and moving it into the market. Scanning magnetometers or compact in-the-field sensors need to obtain data as fast as possible. Sequential Bayesian experiment design can be used as a much faster alternative to the numerous averaging scans. It can also be combined with other approaches that improve sensitivity, such as magnetometry using complicated pulse sequences. While the current study focused on NV^- center magnetometry using sequential Bayesian experiment design, the reported methods—and corresponding speedups—are directly transferable into other areas beyond NV^- centers magnetometry.

CONCLUSION

In this study, we report more than order-of-magnitude speedup of NV^- magnetometry using sequential Bayesian experimental design, compared with the conventional NV^- magnetometry. The large gain in the speed/precision of the NV^- center magnetometry using sequential Bayesian experiment design demonstrated in this study is readily translatable to other applications beyond magnetometry and

experiments with the NV^- centers. The developed optbayesexpt software that was used to carry out sequential Bayesian experiment design measurements is available online for public use free of charge.

ACKNOWLEDGEMENTS

S.D. and K.A. acknowledge support under the Cooperative Research Agreement between the University of Maryland and the National Institute of Standards and Technology Physical Measurement Laboratory, Award 70NANB14H209, through the University of Maryland. The authors thank Adam Pintar for many helpful discussions.

REFERENCES

[1] Y. Wu, Y. Wang, X. Qin, X. Rong, and J. Du, A programmable two-qubit solid-state quantum processor under ambient conditions, NPJ Quantum Inf. **5**, 9 (2019).

[2] A. Beveratos, R. Brouri, T. Gacoin, A. Villing, J.-P. Poizat, and P. Grangier, Single photon quantum cryptography, Phys. Rev. Lett. **89**, 187901 (2002).

[3] G. D. Fuchs, G. Burkard, P. V. Klimov, and D. D. Awschalom, A quantum memory intrinsic to single nitrogen–vacancy centres in diamond, Nat. Phys. **7**, 789 (2011).

[4] W. L. Yang, Z. Q. Yin, Y. Hu, M. Feng, and J. F. Du, High-fidelity quantum memory using nitrogen-vacancy center ensemble for hybrid quantum computation, Phys. Rev. A **84**, 010301 (2011).

[5] V. Vaijayanthimala and H.-C. Chang, Functionalized fluorescent nanodiamonds for biomedical applications, Nanomedicine **4**, 47 (2009).

[6] I. Badea and R. Kaur, Nanodiamonds as novel nanomaterials for biomedical applications: drug delivery and imaging systems, Int. J. Nanomedicine **8**, 203 (2013).

[7] S. Kolkowitz, A. C. Bleszynski Jayich, Q. P. Unterreithmeier, S. D. Bennett, P. Rabl, J. G. E. Harris, and M. D. Lukin, Coherent sensing of a mechanical resonator with a single-spin qubit, Science **335**, 1603 (2012).

[8] G. Kucsko, P. C. Maurer, N. Y. Yao, M. Kubo, H. J. Noh, P. K. Lo, H. Park, and M. D. Lukin, Nanometre-scale thermometry in a living cell, Nature **500**, 54 (2013).

[9] P. Neumann, I. Jakobi, F. Dolde, C. Burk, R. Reuter, G. Waldherr, J. Honert, T. Wolf, A. Brunner, J. H. Shim, D. Suter, H. Sumiya, J. Isoya, and J. Wrachtrup, High-precision nanoscale temperature sensing using single defects in diamond, Nano Lett. **13**, 2738 (2013).

[10] F. Dolde, H. Fedder, M. W. Doherty, T. Nöbauer, F. Rempp, G. Balasubramanian, T. Wolf, F. Reinhard, L. C. L. Hollenberg, F. Jelezko, and J. Wrachtrup, Electric-field sensing using single diamond spins, Nat. Phys. **7**, 459 (2011).

[11] 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**, 648 (2008).

[12] J. R. Maze, P. L. Stanwix, J. S. Hodges, S. Hong, J. M. Taylor, P. Cappellaro, L. Jiang, M. V. G. Dutt, E. Togan, A. S. Zibrov, A. Yacoby, R. L. Walsworth, and M. D. Lukin, Nanoscale magnetic sensing with an individual electronic spin in diamond, Nature **455**, 644 (2008).

[13] C. Du, T. van der Sar, T. X. Zhou, P. Upadhyaya, F. Casola, H. Zhang, M. C. Onbasli, C. A. Ross, R. L. Walsworth, Y. Tserkovnyak, and A. Yacoby, Control and local measurement of the spin chemical potential in a magnetic insulator, Science **357**, 195 (2017).

[14] 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).

[15] L. Rondin, J.-P. Tetienne, T. Hingant, J.-F. Roch, P. Maletinsky, and V. Jacques, Magnetometry with nitrogen-vacancy defects in diamond, Reports Prog. Phys. **77**, 056503 (2014).

[16] P. Maletinsky, S. Hong, M. S. Grinolds, B. Hausmann, M. D. Lukin, R. L. Walsworth, M. Loncar, and A. Yacoby, A robust scanning diamond sensor for nanoscale imaging with single nitrogen-vacancy centres, Nat. Nanotechnol. **7**, 320 (2012).

[17] P. Reith and H. Hilgenkamp, Analysing magnetism using scanning squid microscopy, Rev. Sci. Instrum. **88**, 123706 (2017).

[18] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, Rev. Mod. Phys. **89**, 035002 (2017).

[19] M. W. Mitchell and S. Palacios Alvarez, Colloquium : quantum limits to the energy resolution of magnetic field sensors, Rev. Mod. Phys. **92**, 021001 (2020).

[20] J. M. Schloss, J. F. Barry, M. J. Turner, and R. L. Walsworth, Simultaneous broadband vector magnetometry using solid-state spins, Phys. Rev. Appl. **10**, 034044 (2018).

[21] P. Balasubramanian, C. Osterkamp, Y. Chen, X. Chen, T. Teraji, E. Wu, B. Naydenov, and F. Jelezko, Dc magnetometry with engineered nitrogen-vacancy spin ensembles in diamond, Nano Lett. **19**, 6681 (2019).

[22] C. S. Shin, C. E. Avalos, M. C. Butler, D. R. Trease, S. J. Seltzer, J. Peter Mustonen, D. J. Kennedy, V. M. Acosta, D. Budker, A. Pines, and V. S. Bajaj, Room-temperature operation of a radiofrequency diamond magnetometer near the shot-noise limit, J. Appl. Phys. **112**, 124519 (2012).

[23] H. A. R. El-Ella, S. Ahmadi, A. M. Wojciechowski, A. Huck, and U. L. Andersen, Optimised frequency modulation for continuous-wave optical magnetic resonance sensing using nitrogen-vacancy ensembles, Opt. Express **25**, 14809 (2017).

[24] H. Clevenson, L. M. Pham, C. Teale, K. Johnson, D. Englund, and D. Braje, Robust high-dynamic-range vector magnetometry with nitrogen-vacancy centers in diamond, Appl. Phys. Lett. **112**, 252406 (2018).

[25] K. Ambal and R. D. McMichael, A differential rate meter for real-time peak tracking in optically detected magnetic resonance at low photon count rates, Rev. Sci. Instrum. **90**, 023907 (2019).

[26] G. Liu, M. Chen, Y.-X. Liu, D. Layden, and P. Cappellaro, Repetitive readout enhanced by machine learning, Mach. Learn. Sci. Technol. **1**, 015003 (2020).

[27] K. Chaloner and I. Verdinelli, Bayesian experimental design: a review, Stat. Sci. **10**, 273 (1995).

[28] T. Bayes, An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. London **53**, 370 (1763).

[29] P. S. Laplace, Memoire sur la probabilite des causes par les evenemens, Mem. Math. Phys. **6**, 621 (1774).

[30] P. S. Laplace, Memoir on the probability of the causes of events, Stat. Sci. **1**, 364 (1986).

[31] E. T. Jaynes, *Probability Theory* (Cambridge University Press, 2003).

[32] H. Mabuchi, Dynamical identification of open quantum systems, Quantum Semiclassical Opt. J. Eur. Opt. Soc. Part B **8**, 1103 (1996).

[33] J. Gambetta and H. M. Wiseman, State and dynamical parameter estimation for open quantum systems, Phys. Rev. A **64**, 042105 (2001).

[34] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online hamiltonian learning, New J. Phys. **14**, 103013 (2012).

[35] S. Gammelmark and K. Mølmer, Bayesian parameter inference from continuously monitored quantum systems, Phys. Rev. A **87**, 032115 (2013).

[36] E. Scerri, E. M. Gauger, and C. Bonato, Extending qubit coherence by adaptive quantum environment learning, New J. Phys. **22**, (2020).

[37] A. Negretti and K. Mølmer, Estimation of classical parameters via continuous probing of complementary quantum observables, New J. Phys. **15**, (2013).

[38] 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).

[39] I. Schwartz, J. Rosskopf, S. Schmitt, B. Tratzmiller, Q. Chen, L. P. McGuinness, F. Jelezko, and M. B. Plenio, Blueprint for nanoscale nmr, Sci. Rep. **9**, 6938 (2019).

[40] C. Bonato and D. W. Berry, Adaptive tracking of a time-varying field with a quantum sensor, Phys. Rev. A **95**, 1 (2017).

[41] 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).

[42] 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).

[43] R. Santagati, A. A. Gentile, S. Knauer, S. Schmitt, S. Paesani, C. Granade, N. Wiebe, C. Osterkamp, L. P. McGuinness, J. Wang, M. G. Thompson, J. G. Rarity, F. Jelezko, and A. Laing, Magnetic-field learning using a single electronic spin in diamond with one-photon readout at room temperature, Phys. Rev. X **9**, 021019 (2019).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.14.054036 for additional details and discussion on structure and physics of NV^- center, sequential Bayesian experiment design, implementation of probability distributions, specifications of the computational hardware used for sequential Bayesian experiment design and speedup of the sequential Bayesian experiment design.

[45] A. Gruber, A. Drabenstedt, C. Tietz, L. Fleury, J. Wrachtrup, and C. von Borczyskowski, Scanning confocal optical microscopy and magnetic resonance on single defect centers, Science **276**, 2012 (1997).

[46] G. Davies and M. F. Hamer, Optical studies of the 1.945 eV vibronic band in diamond, Proc. R. Soc. London. A. Math. Phys. Sci. **348**, 285 (1976).

[47] J. Kohler, J. A. J. M. Disselhorst, M. C. J. M. Donckers, E. J. J. Groenen, J. Schmidt, and W. E. Moerner, Magnetic resonance of a single molecular spin, Nature **363**, 242 (1993).

[48] J. Wrachtrup, C. von Borczyskowski, J. Bernard, M. Orrit, and R. Brown, Optical detection of magnetic resonance in a single molecule, Nature **363**, 244 (1993).

[49] V. Jacques, P. Neumann, J. Beck, M. Markham, D. Twitchen, J. Meijer, F. Kaiser, G. Balasubramanian,

F. Jelezko, and J. Wrachtrup, Dynamic polarization of single nuclear spins by optical pumping of nitrogen-vacancy color centers in diamond at room temperature, Phys. Rev. Lett. **102**, 057403 (2009).

[50] R. Fischer, A. Jarmola, P. Kehayias, and D. Budker, Optical polarization of nuclear ensembles in diamond, Phys. Rev. B **87**, 125207 (2013).

[51] D. V. Lindley, On a measure of the information provided by an experiment, Ann. Math. Stat. **27**, 986 (1956).

[52] X. Huan and Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, J. Comput. Phys. **232**, 288 (2013).

[53] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[54] S. Kullback, *Information Theory and Statistics* (Dover Publications, 1968).

[55] R.D. McMichael, Optimal Bayesian Experiment Design Software [Online] (2020). https://github.com/usnistgov/optbayesexpt, R.D. McMichael, Optimal Bayesian Experiment Design Documentation [Online] (2020). https://pages.nist.gov/optbayesexpt

[56] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt, A review of modern computational algorithms for Bayesian optimal design, Int. Stat. Rev. **84**, 128 (2016).

[57] M. A. Nicely and B. E. Wells, Improved parallel resampling methods for particle filtering, IEEE Access **7**, 47593 (2019).

FIGURES

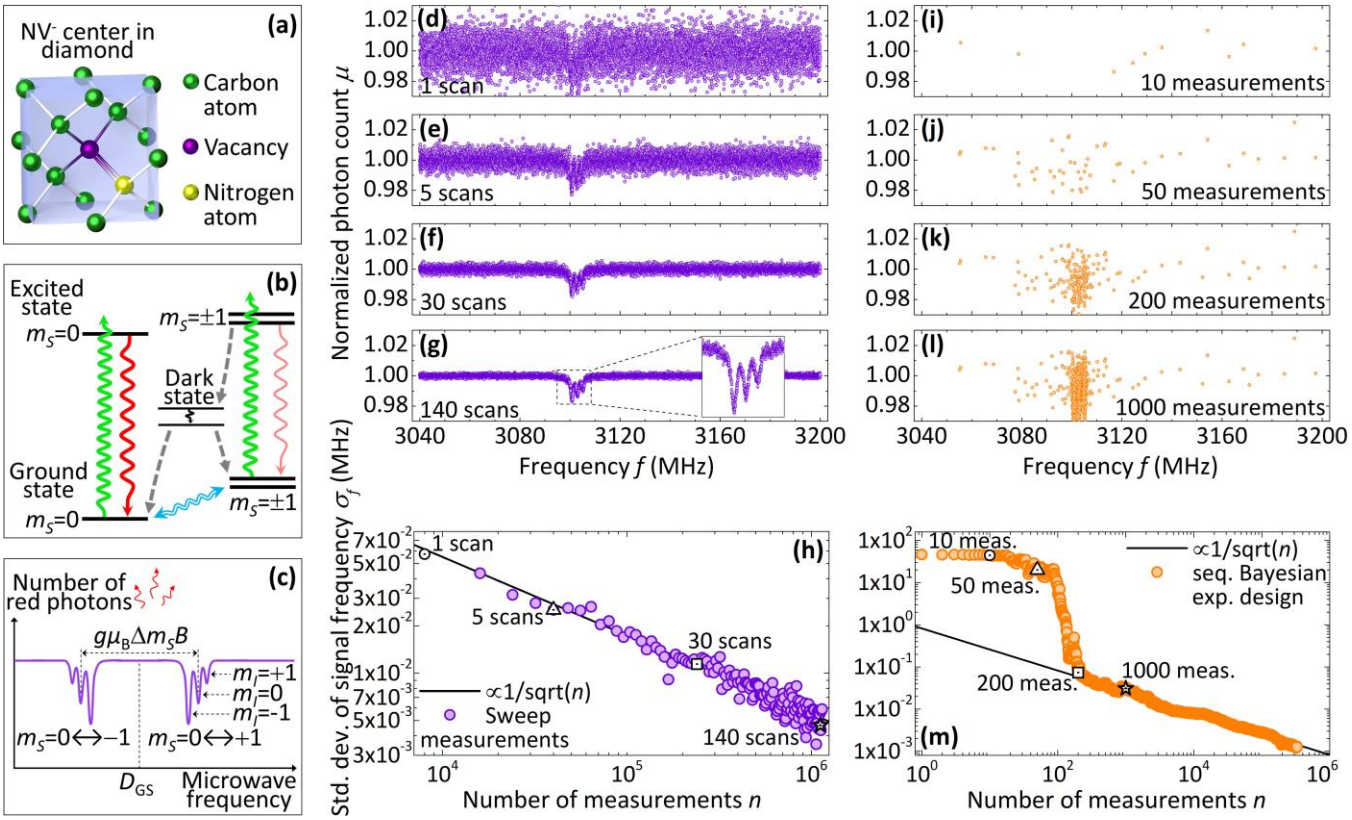

FIG. 1. (a) Crystal structure of the NV- center inside diamond lattice. Green spheres denote carbon atoms, yellow sphere is a nitrogen atom, purple sphere is a vacancy. Each white line corresponds to an sp3 bond created by a pair of electrons. (b) Schematic structure of the transitions between energy levels of the NV- center. NV- center in the ground state can be excited by the laser light (green arrows—transitions due to the absorbed photons); the process preserves spin projection m_S. From the excited state NV- center can relax back to the ground state by emitting red photon (m_S = ±1 or m_S = 0 excited states; red arrow—transitions due to the emitted photons), or non-radiatively relax through the dark state (only m_S = ±1 excited states; dashed gray arrow). Transition between the states with m_S = ±1 and m_S = 0 can be induced by microwaves (blue arrow). (c) Schematics of the photoluminescence spectrum of the NV- center under application of microwave irradiation and the external magnetic field B. The six dips are present due to the Zeeman splitting and hyperfine interaction. (d) – (g) panels show the averaged data from (d) 1 scan, (e) 5 scans, (f) 30 scans, (g) 140 scans (inset shows enlarged signal area) of the conventional NV- magnetometry using photoluminescence detection under sweeping of the microwave frequency. Magnetic field is calculated using the position of the signal (central dip) in the photoluminescence spectrum. (h) Dependence of the standard deviation of the signal frequency f_B on the number of photoluminescence measurements. Each solid purple circle corresponds to a unique number of averaged frequency sweep scans; each scan consists of 8000 measured data points. Black symbols correspond to the data from panels (d) – (g). Black solid line shows inverse square root scaling. Note the logarithmic scale. (i) – (l) panels show the data from (i) 10, (j) 50, (k) 200, (l) 1000 photoluminescence measurements of the NV- magnetometry using sequential Bayesian experiment design. (m) Dependence of the standard deviation of the signal frequency on the number of photoluminescence measurements. Each solid orange circle corresponds to

a unique number of photoluminescence measurements. Black symbols correspond to the data from panels (i) – (l). Black solid line shows inverse square root scaling. Note the logarithmic scale.

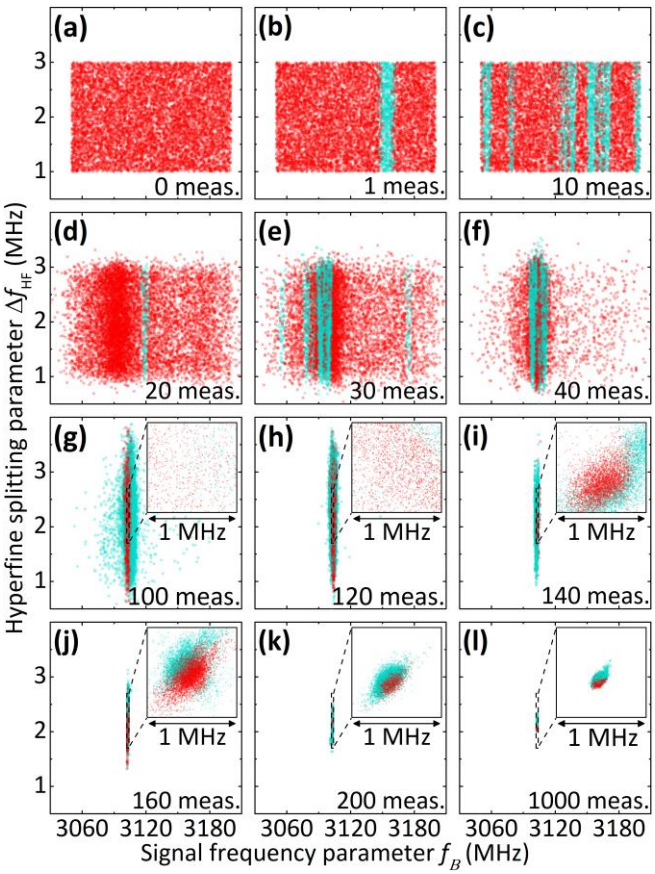

FIG. 2. Dependence of the probability distributions for signal frequency and hyperfine splitting parameters on the number of the measurements in NV^- magnetometry using sequential Bayesian experiment design. Panels shows probability distributions after (a) 0, (b) 1, (c) 10, (d) 20, (e) 30, (f) 40, (g) 100, (h) 120, (i) 140, (j) 160, (k) 200, (l) 1000 measurements. Each probability distribution consists of 10 000 points in parameter space with weights adding up to 1. Color represents weight: $< 10^{-4}$—cyan, $\geq 10^{-4}$—red. Insets show zoomed-in area of the probability distributions. All insets have the same size (1 MHz $\times$ 1 MHz), and span the same parameter space [(3102.5 MHz, 3103.5 MHz); (1.7 MHz, 2.7 MHz)].

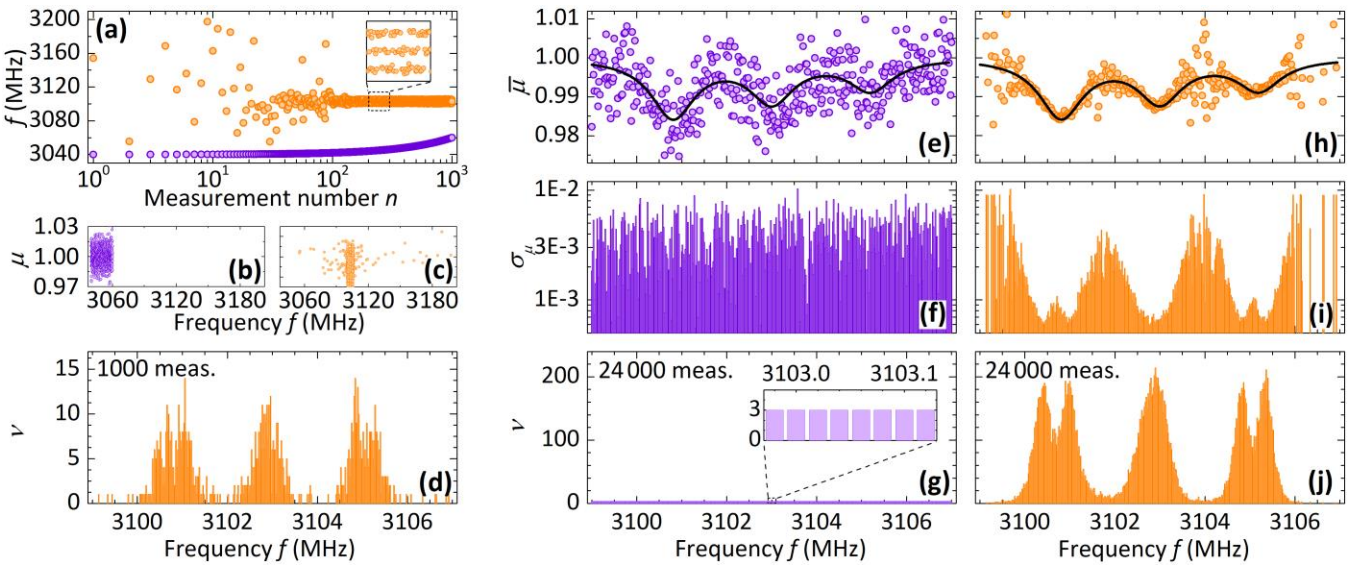

FIG. 3. (a) Dependence of the measurement frequency on the measurement number for the conventional NV^- magnetometry microwave frequency sweep scan (purple solid circles) and the NV^- magnetometry using sequential Bayesian experiment design (orange solid circles). Inset shows zoomed-in view of the area enclosed by the dashed rectangle. Photoluminescence data for the first 1000 measurements of (b) the conventional NV^- magnetometry microwave frequency sweep scan, and (c) NV^- magnetometry using sequential Bayesian experiment design. (d) Distribution of the measurement frequency for the first 1000 measurements of the NV^- magnetometry using sequential Bayesian experiment design. (e, h) Average normalized photon count $\bar{\mu}$, (f , i) standard deviation of the normalized photon count σ_μ and (g, j) number of measurements $\nu(f)$ dependence on the measurement frequency for the first 24 000 measurements. (e, f, g) correspond to data from the conventional NV^- magnetometry scan (purple); (h, i, j) correspond to data from the NV^- magnetometry using sequential Bayesian experiment design (orange). Black solid line (panels (e, h)) shows fitting using function μ of *all* the measured data: 140 scans (1 120 000 measurements) of the conventional NV^- magnetometry and 330 000 measurements of the NV^- magnetometry using sequential Bayesian experiment design. Inset in panel (g) provides a zoomed-in view of the data.

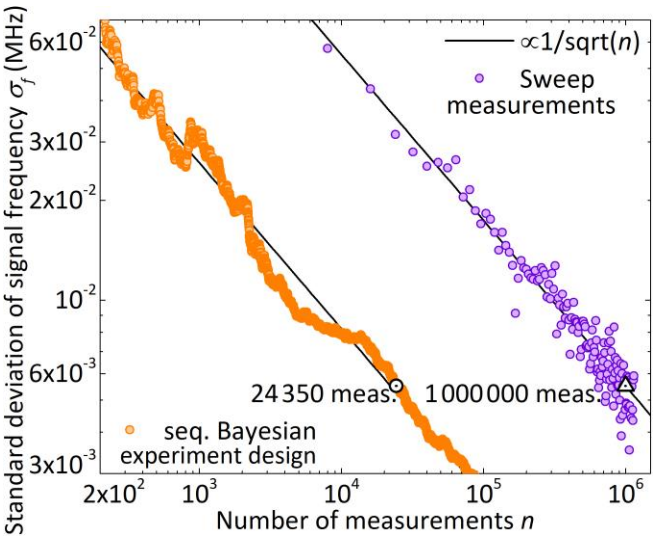

FIG. 4. Dependence of the standard deviation of the signal frequency on the number of photoluminescence measurements. Each orange filled circle corresponds to a unique number of photoluminescence measurements using sequential Bayesian experiment design. Each purple filled circle corresponds to a unique number of averaged frequency sweep scans; each scan consists of 8000 measured photoluminescence data points. Black symbols correspond to equal standard deviation of the signal frequency for sequential Bayesian experiment design (black circle) and conventional sweep measurement (black triangle). Black solid lines show inverse square root scaling. Note the logarithmic scale.

SUPPLEMENTAL MATERIAL

Sequential Bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers

Sergey Dushenko[1,2], Kapildeb Ambal[1-3], Robert D. McMichael[1]

[1]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

[2]Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD 20742, USA

[3]Department of Physics, Wichita State University, Wichita, KS 67260, USA

https://journals.aps.org/prapplied/supplemental/10.1103/PhysRevApplied.14.054036/NU10281_Dushenko_Supplemental_Material_10_14_2020.pdf

Supplemental to Physical Review Applied 14, 154036 (2020)

CONTENT

S.1 ADDITIONAL DETAILS ON STRUCTURE AND PHYSICS OF NV$^-$ CENTER

The NV$^-$ center is a quantum defect that is created when two adjacent carbon atoms in a diamond lattice are substituted with a vacancy and a negatively charged nitrogen atom. This nitrogen-vacancy system in the diamond lattice has six electrons (five electrons from the nitrogen atom and surrounding carbon atoms, plus one additional electron from the lattice), four of which are located on the energy levels inside the energy bandgap of diamond. In the ground and photoexcited states of the NV$^-$ center, two electrons are unpaired, resulting in a total spin of the system adding up to 1, and creating a spin S = 1 quantum defect. NV$^-$ center can be excited using photons in the wavelength range of 480 nm to 637 nm [1], due to 2.6 eV (477 nm) energy difference between the ground state and the bottom of the conduction band and a 1.95 eV energy difference between the ground and excited states (637 nm zero-phonon line) [2]. Figure S1 shows the energy level structure of the NV- center. The NV$^-$ center is not susceptible to photobleaching.

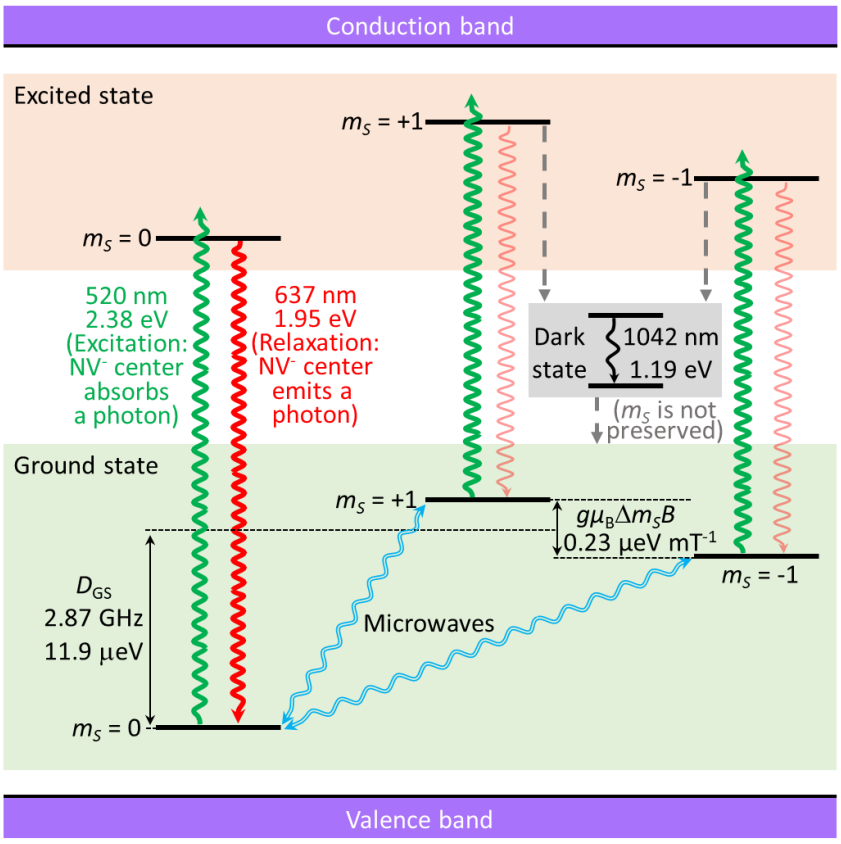

Fig. S1. Schematic structure of the energy levels of NV$^-$ center (not to scale). NV$^-$ center in the ground state can be excited by the green laser (green arrow); the process preserves spin projection. From the excited state NV$^-$ center can relax back to the ground state by emitting a red photon (m_s = ±1 or m_s = 0 excited states), or non-radiatively relaxing through the dark state (only m_s = ±1 excited states). Transition between the states with m_s = ±1 and m_s = 0 can be induced by the microwaves (blue arrows).

S.2 INTRODUCTION TO SEQUENTIAL BAYESIAN EXPERIMENT DESIGN

Sequential design divides measurement runs into a sequence of design-measure-analyze cycles, providing each design decision with information gleaned from all the data that have accumulated up to that point (Fig. S2). In contrast, the traditional method involves a preprogrammed series of settings as the design, a period of measurement, and finally analysis that yields useful information at the very end of the process.

Here, we set the data collection aside and focus on the design and analysis processes in sequential Bayesian experiment design. The methods outlined here have been described previously by numerous authors [3,4]. After a brief introduction to Bayes' theorem (section S.2.1), we describe the application of Bayes' theorem to the analysis process (section S.2.2), and then follow with the use of Bayes' theorem and information theory in making design decisions (sections S.2.3 and S.2.4).

S.2.1 Bayes' Theorem

The ideas of Bayesian methodology were first proposed by Bayes [5] and then independently rediscovered and developed much further by Laplace [6,7]. Bayes' theorem is built upon concepts of probability distributions like $p(A)$, describing the probability of A, conditional probabilities like $p(A|B)$ describing the probability of A given B, and joint probabilities like $p(A,B)$ describing the probability of both A and B. Bayes' theorem follows from the fact that the joint probability can be expressed in terms of a conditional probability: $p(A,B) = p(A|B) \cdot p(B)$, the joint probability of both A and B is the conditional probability of A given B times the probability of B. But it is equally true that, $p(A,B) = p(B|A) \cdot p(A)$. The combination of these last two equations yields Bayes' theorem:

$$p(A|B) = \frac{p(B|A)}{p(B)} \cdot p(A). \qquad [S1]$$

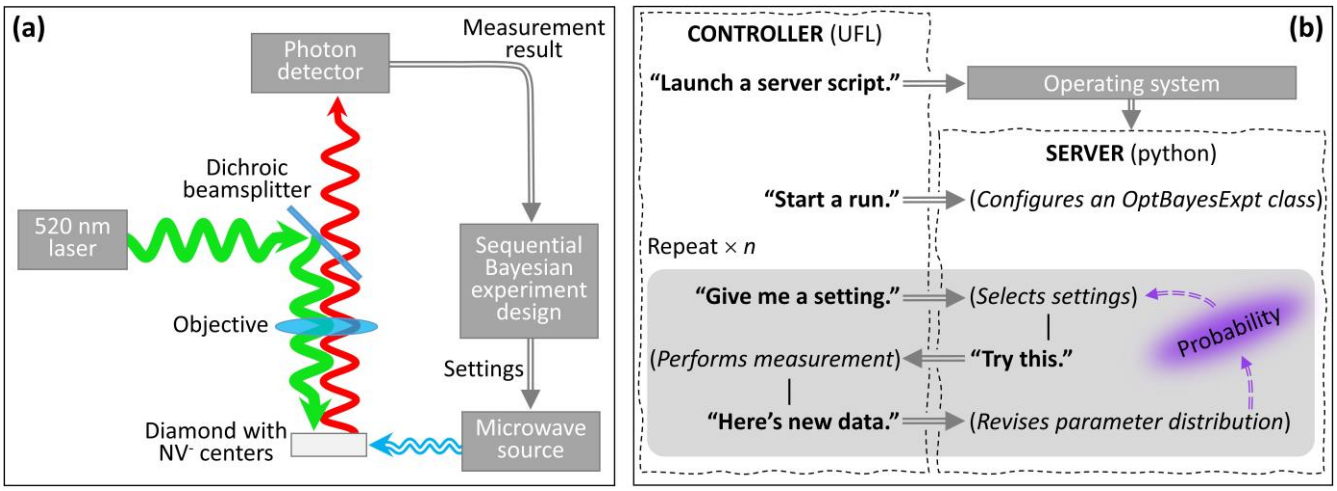

FIG. S2. (a) Schematic layout of the NV^- magnetometry experiment using sequential Bayesian experiment design. (b) Block diagram of the sequential Bayesian experiment design algorithm. UFL stands for user's favorite language.

S.2.2 Bayesian Inference

In the present context, we are interested in the probability distribution $p(\boldsymbol{\theta})$ over parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$ of a model function $f(\boldsymbol{\theta}, \boldsymbol{d})$ that also depends on experimental settings (designs) $\boldsymbol{d}$. Thinking about model parameters as random variables with a distribution to be determined may seem very different from thinking about the same parameters as fixed numbers with values to be determined and uncertainty due to noise. However, the familiar notation $x = \bar{x} \pm \sigma$ is shorthand notation for a Gaussian distribution. Using Bayes theorem to determine the distribution of parameters $\boldsymbol{\theta}$ (A in [S1]) given measurement data $\boldsymbol{y}_n = \{y_1, y_2, \dots, y_n\}$ (B in [S1]) accumulated after n measurements made using settings $\boldsymbol{d}_n = \{d_1, d_2, \dots, d_n\}$, substitution yields

$$p_n(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta}|\boldsymbol{y}_n, \boldsymbol{d}_n) = \frac{p(\boldsymbol{y}_n|\boldsymbol{\theta}, \boldsymbol{d}_n)}{p(\boldsymbol{y}_n|\boldsymbol{d}_n)} p_0(\boldsymbol{\theta}). \quad \text{[S2]}$$

In Bayesian lingo, $p_0(\boldsymbol{\theta})$ is the *prior,* the distribution of parameter values before measurement data are considered. The *posterior,* $p(\boldsymbol{\theta}|\boldsymbol{y}_n, \boldsymbol{d}_n)$ is the parameter distribution given the collected data. The numerator is called the *likelihood,* and the denominator is the *evidence*.

With each additional measurement, the parameter distribution can be refined using Bayes' theorem. With result y_{n+1} measured using settings d_{n+1},

$$p_{n+1}(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta}|y_{n+1}, d_{n+1,}\boldsymbol{y}_n, \boldsymbol{d}_n) = \frac{p(y_{n+1}|\boldsymbol{\theta}, d_{n+1})}{p(y_{n+1}|d_{n+1})} p(\boldsymbol{\theta}|\boldsymbol{y}_n, \boldsymbol{d}_n). \quad \text{[S3]}$$

In the numerator, the *likelihood* is a function of the parameter variables $\boldsymbol{\theta}$ with constants y_{n+1} and d_{n+1}. It is the probability of getting a measurement result y_{n+1} as a function of $\boldsymbol{\theta}$ when setting design d_{n+1} is used. In the denominator, the *evidence* $p(y_{n+1}|d_{n+1})$ is a constant that maintains normalization $\int p(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$.

To estimate the likelihood function, we must provide a connection between settings, parameters and measurement results. Here that connection is provided by a model function $y = f(\boldsymbol{\theta}, \boldsymbol{d}) + \eta$ where η is a model of experimental noise. The model function is roughly equivalent to the fitting function one would use for least-squares regression.

If the noise η follows a normal (Gaussian) distribution with standard deviation σ, the probability of a measurement yielding y_{n+1}depends on the difference between the measured value and the modeled values as a function of $\boldsymbol{\theta}$:

$$p(y_{n+1}|\boldsymbol{\theta}, d_{n+1}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-[y_{n+1} - f(\boldsymbol{\theta}, d_{n+1})]^2}{2\sigma^2}\right]. \quad \text{[S4]}$$

Qualitatively, some parameter values, say $\boldsymbol{\theta}_a$, will produce model results $f(\theta_a, d_{n+1})$ that are closer to y_{n+1} than will other parameter values $\boldsymbol{\theta_b}$. It follows that the likelihood given in [S4] is greater for $\boldsymbol{\theta}_a$ than for $\boldsymbol{\theta}_b$. In a quantitative way, the likelihood formalizes the notion that $\boldsymbol{\theta}_a$ "explains the data" better than $\boldsymbol{\theta}_b$. Although the model function does not depend on the noise parameter σ, the likelihood does. If σ is treated as an additional parameter, some values of σ will "explain the data" better than others.

S.2.3 Bayesian Experiment design

We now turn to the problem of selecting a design (choosing settings) for a future measurement, preferably making good use of the refined parameter distribution $p_n(\boldsymbol{\theta})$. We will frame the problem in terms of defining a utility function $U(\boldsymbol{d})$ that expresses the predicted benefit of a future measurement made with setting $\boldsymbol{d}$.

First, we look at predicted measurement values and their distributions. The distribution of predicted measurement values for a design $\boldsymbol{d}$ and fixed $\boldsymbol{\theta}$ is

$$p(y|\boldsymbol{\theta},\boldsymbol{d}) = p_\eta(y - f(\boldsymbol{\theta},\boldsymbol{d})) \quad [S5]$$

Where $p_\eta(\cdot)$ is the distribution of measurement noise values. To obtain the full distribution of predicted y values, $p(y|\boldsymbol{d})$, we must also account for the probability distribution of $\boldsymbol{\theta}$ values by integrating over the $\boldsymbol{\theta}$ values weighted by the $p_n(\boldsymbol{\theta})$ distribution.

$$p(y|\boldsymbol{d}) = \int p(y|\boldsymbol{\theta},\boldsymbol{d})p_n(\theta)d\theta. \quad [S6]$$

Next, in order to make decisions about future measurements, we need to quantify the "goodness" of a θ distribution. For this purpose, the information entropy is the conventional measure. For an arbitrary distribution $p(x)$ the information entropy is defined as

$$H = -\int p(x)\ln[p(x)]\,dx. \quad [S7]$$

The change in information entropy of the parameter distribution that would result from a future measurement value y is given by the difference in entropy between the posterior distribution given predicted measurements, $p(\boldsymbol{\theta}|y,\boldsymbol{d})$ and the prior distribution $p(\boldsymbol{\theta})$

$$\Delta H(y|\boldsymbol{d}) = -\int p(\boldsymbol{\theta}|y,\boldsymbol{d})\ln[p(\boldsymbol{\theta}|y,\boldsymbol{d})]d\boldsymbol{\theta} + \int p(\boldsymbol{\theta})\ln[p(\boldsymbol{\theta})]\,d\boldsymbol{\theta}. \quad [S8]$$

The expectation value of ΔH is our utility function,

$$U(\boldsymbol{d}) = \int dy\, p(y|\boldsymbol{d})\Delta H(y|\boldsymbol{d}), \quad [S9]$$

which predicts the mean benefit of a future measurement made using setting $\boldsymbol{d}$.
Combining [S8] and [S9], and using Bayes' theorem:

$$\mathrm{U}(\boldsymbol{d}) = -\iint p(y|\boldsymbol{d})\frac{p(y|\boldsymbol{\theta},\boldsymbol{d})}{p(y|\boldsymbol{d})}\mathrm{p}(\boldsymbol{\theta})\ln[\frac{p(y|\boldsymbol{\theta},\boldsymbol{d})}{p(y|\boldsymbol{d})}\mathrm{p}(\boldsymbol{\theta})]d\boldsymbol{\theta}dy + \int p(\boldsymbol{\theta})\ln[p(\boldsymbol{\theta})]\,d\boldsymbol{\theta}. \quad [S10]$$

Expanding the logarithm,

$$\begin{aligned} U(\boldsymbol{d}) = &-\iint p(y|\boldsymbol{\theta},\boldsymbol{d})p(\boldsymbol{\theta})\ln p(y|\boldsymbol{\theta},\boldsymbol{d})\,d\boldsymbol{\theta}dy - \iint p(y)|\boldsymbol{\theta},\boldsymbol{d})p(\boldsymbol{\theta})\ln p(\boldsymbol{\theta})\,d\boldsymbol{\theta}dy \\ &+\iint p(y|\boldsymbol{\theta},\boldsymbol{d})\mathrm{p}(\boldsymbol{\theta})\ln p(y|\boldsymbol{d})\,d\boldsymbol{\theta}dy + \int p(\boldsymbol{\theta})\ln[p(\boldsymbol{\theta})]\,d\boldsymbol{\theta}. \end{aligned} \quad [S11]$$

The y integral in the 2nd term amounts to 1, so the 2nd and 4th terms cancel, yielding

$$\mathrm{U}(\boldsymbol{d}) = -\int\left\{\int p(y|\boldsymbol{\theta},\boldsymbol{d})\ln p(y|\boldsymbol{\theta},\boldsymbol{d})\,dy\right\}\mathrm{p}(\boldsymbol{\theta})d\boldsymbol{\theta} + \iint p(y|\boldsymbol{d})\ln p(y|\boldsymbol{d})\,dy. \quad [S12]$$

Recalling earlier expressions, $p(y|\boldsymbol{\theta},\boldsymbol{d})$ is essentially the distribution of noise, so the first term is the entropy of the measurement noise distribution, averaged over parameters. The second term is the information entropy of $p(y|\boldsymbol{d})$, the y distribution with only the design given. $p(y|\boldsymbol{d}) = \int p(y|\boldsymbol{\theta},\boldsymbol{d})p(\theta)d\boldsymbol{\theta}$, or more explicitly,

$$p(y|\boldsymbol{d}) = \int p_\eta\big(y - f(\boldsymbol{\theta}, \boldsymbol{d})\big)p(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad \text{[S13]}$$

This expression shows that $p(y|\boldsymbol{d})$ is a convolution of the noise distribution and the distribution of model values due to the parameter distribution. Loosely, [S10] suggests that the highest utility will be made with designs where random draws from the parameter distribution produce the largest variations in the model function results.

S.2.4 Evaluation of Utility Function

The double integrals in the entropy loss [S10] are potentially quite expensive. Both involve integration over both the noise distribution and the parameter distribution. However, there are several factors that relax the need for precise evaluation of the utility.

- If the utility is only used to select designs d where $U(d)$ is large or maximum, precise evaluation is not needed. Only the relative magnitude of $U(d)$ is important for selecting candidate d values.
- Further, precision in selecting d is also non-critical in many cases. All measurements decrease information entropy in expectation, so sloppy selection of d only affects the efficiency of a measurement, not the validity of the measurement results.
- In many common applications, the measurement noise does not depend on model parameters θ, simplifying the first term in [S10].

In view of these factors, the optbayesexpt software adopts two approximations that dramatically reduce the computational cost of evaluating the utility.

For many common distributions, the information entropy has the form $\ln w + C$ where w is a parameter describing the width of the distribution. A much smaller sample is needed to estimate the width of a distribution than to estimate the information entropy. The width of $p(y|d)$ described by the convolution in [S11] is approximated by the standard deviation of the noise distribution σ_η, and the standard deviation σ_θ of the $f(\theta, d)p(\theta)$ distribution, summed in quadrature:

$$H(y|d) \approx \frac{1}{2}\ln(\sigma_\eta^2 + \sigma_\theta^2) \quad \text{[S14]}$$

To ensure smoothness of $U(d)$, the same draws from $p(\theta)$ are used to form and estimate $U^*(d)$ for all values of d. When $f(\theta, d)$ is a smooth function of d for fixed θ, the estimate $U^*(d)$ will also be smooth. A small sample (tens) are drawn from $p(\theta)$ to estimate the width of distributions.

S.3 IMPLEMENTATION OF PROBABILITY DISTRIBUTIONS

We use sequential Monte Carlo (SMC) methods to provide a computer-friendly approximation to analytical probability distributions. The distribution $p(\theta_1, \theta_2, \dots, \theta_k)$ is represented by N samples $\boldsymbol{\theta}_i = \{\theta_{1,i}, \theta_{2,i}, \dots, \theta_{k,i}\}, i = 1, \dots, N$. Each sample can be regarded as the coordinates of a particle in k-dimensional parameter space and the ensemble of particles as a cloud or swarm. Each particle is also assigned a weight w_i, so that the probability density is represented by the weighted density of points in θ space. Computationally, the distribution is implemented by a dimension $N \times k$ array listing the particle coordinates and a length N array listing the corresponding weights.

To incorporate new data y_{n+1} using Bayesian inference, the likelihood of the result $p\big(y_{n+1}\big|\theta_{i,n}, d_{n+1}\big)$ initially modifies the weights, but does not affect the particle coordinates.

$$W_{i,n+1} = p\left(y_{n+1}\middle|\theta_{i,n}, d_{n+1}\right)w_{i,n}$$

$$w_{i,n+1} = W_{i,n+1}/(\Sigma W_{i,n+1})$$

This Bayesian inference process will tend to decrease weights for low-probability regions in θ space to very small values, eventually leaving a small number of particles to represent higher-probability regions. To circumvent this problem, SMC methods typically use a resampling method that effectively reassigns particles into high-probability regions.

After each inference step, the effective number of particles, $N_{\text{eff}} = 1/\Sigma w_i^2$ is calculated. If N_{eff} is less than (typically) half of N, the resampling procedure is executed as follows:

1. N particles are chosen with probability w_i from the current distribution with replacement. Some particles may be chosen more than once, some once, and those that are not chosen are abandoned.
2. To separate particles that were chosen multiple times, each of the particles is given a random displacement that is small compared to the distribution's standard deviation. Then, to compensate for the diffusion that this random displacement produces, all particles are contracted slightly toward the distribution's mean value.
3. Finally, each particle weight is assigned a uniform value $w_i = 1/N$.

S.4 SPECIFICATIONS OF THE COMPUTATIONAL HARDWARE USED FOR SEQUENTIAL BAYESIAN EXPERIMENT DESIGN

All sequential Bayesian experiment design calculations were performed on a single core (one thread) of Intel Xeon Processor E3-1225 v2 @3.20 GHz [8] using the optbayesexpt python package [9]. The wall-clock measurement times are estimated from data file modification times. The data saving was performed eight times more often for sequential Bayesian experiment design compared with the conventional setup (every 1000 points, instead of every scan of 8000 points). Moreover, the file modification times do not account for the additional time required to fit the conventional data. Hence, the 36 % per-measurement slowdown associated with the Bayesian computational time is an upper bound and the difference in total throughput may be smaller.

S.5 SPEEDUP OF THE SEQUENTIAL BAYESIAN EXPERIMENT DESIGN

As discussed in the manuscript, a big factor that influences the speedup of the sequential Bayesian experiment design is the fraction of settings space occupied by the signal, compared with the whole scanning or sensing range. This ratio of signal to sensing range can vary significantly depending on the task at hand, and can be both much smaller than in our study (for example, in magnetometers/sensors with broad sensing range) leading to even larger speedup, or bigger, leading to a smaller speedup. In any case, the Bayesian algorithm is going to outperform the conventional scan-and-average technique. Here, we consider the "worst case" scenario for a gained speedup, when an experimenter can guess the minimal size and location of the scanning range just after one quick scan and adjust the settings appropriately. For the experiment described in the manuscript, the lower bound for such a range would be about 16 MHz, or a tenth of the scanning range used in the manuscript (8 MHz is occupied by the dips, plus at least 4 MHz on each side for the shoulders). Such scanning range would allow the conventional method to collect ten times more measurements in the signal area in the same period of time, and potentially making it up

to ten times faster. Even if one assumes no increase in the Bayesian algorithm speed with smaller range, the sequential Bayesian experiment design will still be 4.5 times faster than the conventional measurement. A milestone of speedup by more than a factor of 2 is very likely to be of practical relevance in any measurement, technology or business. Hence—even for applications where one, in a way, already knows the answer and can guess the location and size of the minimal scanning range—the speedup that can be achieved using sequential Bayesian experiment design surpasses this milestone by a huge margin.

REFERENCES

[1] A. Gruber, A. Drabenstedt, C. Tietz, L. Fleury, J. Wrachtrup, and C. von Borczyskowski, Scanning confocal optical microscopy and magnetic resonance on single defect centers, Science **276**, 2012 (1997).

[2] G. Davies and M. F. Hamer, Optical studies of the 1.945 eV vibronic band in diamond, Proc. R. Soc. London. A. Math. Phys. Sci. **348**, 285 (1976).

[3] K. Chaloner and I. Verdinelli, Bayesian experimental design: a review, Stat. Sci. **10**, 273 (1995).

[4] X. Huan and Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, J. Comput. Phys. **232**, 288 (2013).

[5] T. Bayes, An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. London **53**, 370 (1763).

[6] P. S. Laplace, Memoire sur la probabilite des causes par les evenemens, Mem. Math. Phys. **6**, 621 (1774).

[7] P. S. Laplace, Memoir on the probability of the causes of events, Stat. Sci. **1**, 364 (1986).

[8] Disclosure: certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

[9] R.D. McMichael, Optimal Bayesian Experiment Design Software [Online] (2020). https://github.com/usnistgov/optbayesexpt

R.D. McMichael, Optimal Bayesian Experiment Design Documentation [Online] (2020). https://pages.nist.gov/optbayesexpt