Journal of Applied Physics

RESEARCH ARTICLE | FEBRUARY 21 2025

Weighing unequal parameter importance and measurement expense in adaptive quantum sensing

Special Collection: Defects in Solids for Quantum Technologies

M. Kelley ; R. D. McMichael

Check for updates

J. Appl. Phys. 137, 074401 (2025)
https://doi.org/10.1063/5.0251881

View Online Export Citation

Articles You May Be Interested In

Novel 3D reciprocal space visualization of strain relaxation in InSb on GaAs substrates

J. Vac. Sci. Technol. A (April 2024)

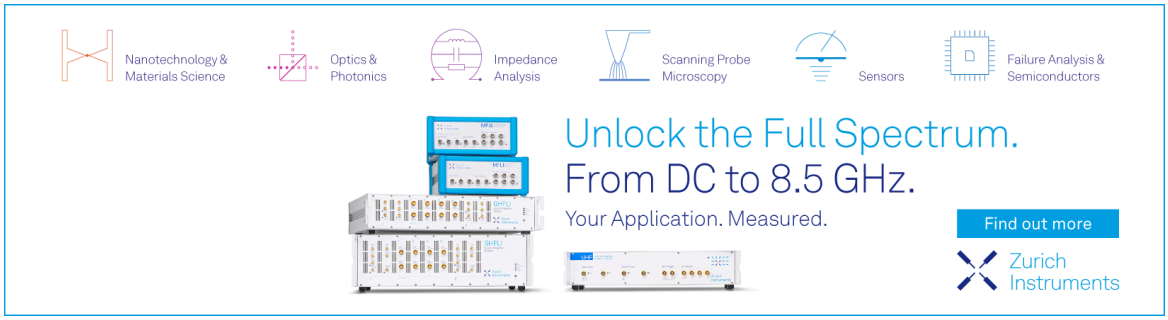
Corrosive behavior of 440C and M50 steels for marine applications

AIP Conference Proceedings (November 2022)

High-coverage surface phases of tellurium on Ni(111)

J. Vac. Sci. Technol. A (July 1987)

02 March 2025 10:26:28



AIP Publishing

Journal of Applied Physics ARTICLE pubs.aip.org/aip/jap

Weighing unequal parameter importance and measurement expense in adaptive quantum sensing

Cite as: J. Appl. Phys. **137**, 074401 (2025); doi: 10.1063/5.0251881
Submitted: 5 December 2024 · Accepted: 30 January 2025 ·
Published Online: 21 February 2025

View Online Export Citation CrossMark

M. Kelley and R. D. McMichael[a)]

AFFILIATIONS

National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

Note: This paper is part of the Special Topic, Defects in Solids for Quantum Technologies.
[a)]**Author to whom correspondence should be addressed:** robert.mcmichael@nist.gov

ABSTRACT

A large class of experiments consists of measuring the parameters of physical models. In these experiments, the goal is to learn about these parameters as accurately and, often, quickly as possible. Adaptive experiment design works by yielding instrument control to Bayesian-based algorithms that alter instrument settings based on potential information gain about the parameters. By actively learning from data in real-time where to measure instead of determining instrument settings *a priori*, striking improvements in experiment efficiency are possible. Here, two new algorithms that improve upon previous implementations of adaptive experiment design are introduced. The first algorithm focuses on learning the model parameters that matter the most. The second algorithm considers the expense of a measurement and prioritizes information that can be gained at a lower cost. We demonstrate the remarkable improvement in efficiency and sensitivity that these algorithms provide for quantum sensing, specifically magnetometry, with nitrogen-vacancy centers in diamond. Most notably, we find an almost fivefold improvement in magnetic field sensitivity.

02 March 2025 10:26:28

© 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/). https://doi.org/10.1063/5.0251881

I. INTRODUCTION

Adaptive experiment design exploits accumulated measurement data to create efficient measurement strategies.[1–7] The impact of adaptive experiment design tends to be most significant in experiments that are slow and noisy. An important example of current interest is quantum sensing with solid-state point defects. Quantum sensing with optically active solid-state point defects is a rapidly developing technology, as these point defects or "color centers" have technical and practical advantages over atomic systems. Nitrogen-vacancy (NV) centers in diamond are one of the most well-studied color centers, with sensitivity to magnetic fields, electric fields, strain, and temperature.[8] Despite the exciting prospects of diamond-based quantum sensors, the conventional readout process, which is based on spin-dependent photon emission, has poor fidelity at room temperature. Even the best setups have readout fidelities, a measure of confidence, of ≈0.01 for NV ensembles, which sets the sensitivity two orders of magnitude below the spin-projection limit.[9] As a result, measurements are repeated many times (upward of 10^4–10^6), leading to time-consuming experiments and limited sensitivity.

Here, we demonstrate improvements in the efficiency and sensitivity of quantum sensing using adaptive experiment design. A core concept in adaptive experiment design is the utility function, which estimates the benefit of a measurement as a function of the proposed instrument setting. Recently, improved algorithms for calculating the utility function, typically a computationally cumbersome task, have been introduced.[2,10] We expand on this work by honing the utility function for specific experimental goals. First, we recognize that some parameters in the experiment are more important than others. We reason that adaptive design strategies can be made more efficient by de-emphasizing instrument settings that inform about nuisance parameters and instead focus resources on determining valuable parameters. Second, we consider the case where each measurement setting has a varying expense, such as time, and minimizing the overall cost of an experiment is desirable.

© Author(s) 2025

Previous works have shown how adaptive methods can improve the sensitivity of NV-based relaxometry[3] as well as improve the sensitivity to the decoherence, dephasing, and decay times.[7] Arshad *et al.*[7] showed that adaptive algorithms (in their case, based on the Fisher information) can be implemented directly on hardware via a microcontroller for very fast computation. Our work here complements and builds upon these works by (1) making fast decisions in software for easy implementation and (2) further honing the adaptive algorithms for the desired outcome.

In this work, we describe utility algorithms adapted to the two aforementioned cases and present testing of these algorithms using both simulations and laboratory experiments. Section II reviews the theoretical background for utility calculations and then presents the adaptations for de-emphasizing nuisance parameters and incorporating economic expenses into setting decisions. Section III describes the laboratory measurement system using NV centers in diamond and the basics of quantum sensing. Section IV contains the main results of the paper, where we compare the behavior of the algorithms across several core quantum sensing experiments. All algorithms used in this work are described in the Appendix.

II. BACKGROUND

A. Adaptive experiment design

The measurement runs of interest are sequences of epochs. Each epoch involves a design or setting selection d, data collection, and interpretation of the resulting data. We use Bayesian methods to interpret the data in terms of the parameters θ of parametric model f,

$$y = f(\theta, d) + \zeta, \tag{1}$$

where ζ is the experimental noise. Bayesian inference operates on the probability distribution of the parameters $P(\theta)$, which is a more general method than considering just the values and uncertainties of the parameters. In a typical measurement, the parameter distribution will become narrower, corresponding to shrinking uncertainties.

The algorithms used in adaptive experiment design rely on finding the setting combination likely to make the most progress toward an experiment goal. The utility function, $U(d')$, is an estimate of the progress to be made as a function of the proposed setting combination, d'. In general, d' can be n-dimensional, but in typical experiments, it may exist as one or two instrument settings, such as power, frequency, or time. An adaptive measurement epoch consists of locating the maxima of the utility and subsequently taking measurements at its argument d', updating the model parameter distributions using the new data, and recalculating the utility. By repeating this process, one makes efficient measurements by maximizing the information gain.[2,10] Shortly, adaptive experiment design uses incoming data to actively make good decisions. In traditional measurements, the instrument settings are determined *a priori*, data points are collected, and then, the data are fit to the model to estimate the model parameters. All experiment optimization relies on prior knowledge in this case.

The utility function can be represented by the difference or distance between two distributions—the current parameter distribution $P(\theta)$ (the prior) and that which is *predicted* given a proposed measurement using setting design d' and yielding outcome y', $P(\theta|y', d')$ (the posterior).

The goal of the design task is to pick a setting that is likely to move us closer to the goal of narrow parameter distributions. It is convenient to quantify "narrowness" with the information entropy of the distribution,

$$H = -\int P(x) \ln P(x) dx. \tag{2}$$

Changes in the information entropy (our "narrowness") are calculated using the Kullback–Leibler divergence:

$$D_{KL}(d', y') = -\int P(\theta|y', d') \ln\left(\frac{P(\theta)}{P(\theta|y', d')}\right) d\theta. \tag{3}$$

The Kullback–Leibler divergence quantifies how much information about $P(\theta)$ is contained in $P(\theta|y', d')$, that is, the mutual information between the two distributions.[11] Note that if the prior and posterior are equivalent, the argument of the logarithm is unity, and the mutual information is zero, meaning no progress would be made with that setting. The setting d' that produces the greatest mutual information (on average) is likely to make the most progress.

In order to rewrite Eq. (3) in a usable form for a utility function, the posterior can be expressed using the Bayes rule as

$$P(\theta|y', d') = \frac{P(y'|\theta, d')}{P(y'|d')} P(\theta). \tag{4}$$

This transformation eliminates the need to calculate posterior distributions. Equation (3) becomes

$$D_{KL}(d', y') = -\int \frac{P(y'|\theta, d')}{P(y')} P_n(\theta) \ln\left(\frac{P(y')}{P(y'|\theta, d')}\right) d\theta. \tag{5}$$

Averaging over potential measurement outcomes, y', yields the utility function,

$$U(d') = -\int D_{KL}(y', d') P(y'|d') dy', \tag{6}$$

where

$$P(y'|d') = \int P(y'|\theta, d') P(\theta) d\theta. \tag{7}$$

Substituting Eq. (7) in Eq. (6),

$$U(d') = \int P(\theta) \int P(y'|\theta, d') \ln P(y'|\theta, d') dy' d\theta - \int P(y'|d') \ln P(y'|d') dy'. \tag{8}$$

In the above expression, $P(y'|d')$ represents the measurement outcome at d', whereas $P(y'|\theta, d')$ represents variations in the outcome due to noise (i.e., if θ and d' are held constant the only

02 March 2025 10:26:28

© Author(s) 2025

variation in outcome will be due to noise). Note that Eq. (8) can be rewritten using Eq. (2) as

$$U(d') = H_{y'|d'}(d') - \int H_{y'|\theta,d'}(\theta, d')P(\theta)d\theta. \tag{9}$$

The first term in Eq. (9) is the information entropy, including experimental noise and the parameter distributions. The second term is the information entropy of the experimental noise, averaged over the parameters, θ. A simple interpretation is that an optimal measurement is where the variation in the outcome y is greatest due to the underlying model parameters with respect to the experimental noise.

Furthermore, if we assume that the model parameters are normally distributed with variance v, then

$$H = \frac{1}{2}\ln(2\pi v) + \frac{1}{2}, \tag{10}$$

and Eq. (9) becomes

$$\begin{aligned} U(d') &= \frac{1}{2}\ln\left(2\pi v_\theta(d') + 2\pi v_\zeta\right) - \frac{1}{2}\ln\left(2\pi v_\zeta\right) \\ &= \frac{1}{2}\ln\left(1 + \frac{v_\theta(d')}{v_\zeta}\right), \end{aligned} \tag{11}$$

where v_ζ is the variance of the experimental noise and $v_\theta(d')$ is the variance of the distribution of the outcomes, $P(y'|d', \zeta = 0)$ without any noise, that is,

$$v_\theta(d') \equiv \mathrm{Var}[P(y|d', \zeta = 0)]. \tag{12}$$

Recall that the variance of a distribution is determined by what is not fixed (i.e., what does *not* appear after "|"). This means that $P(y'|d', \zeta = 0)$ describes the variance in what we measure, y', just due to underlying model parameters, θ, which will be a function of the setting d'. Again, we want select settings, d', that will maximize the utility and make our measurements more efficient. In Eq. (11), it is clear that maximizing the utility comes from taking measurements where the *variance in outcomes due to the parameter distribution* $v_\theta(d')$ is the largest relative to the variance due to noise v_ζ. Simply put, we want to measure where the outcome of the measurement will tell us more about the underlying parameters. This can also be thought of as measuring at a higher "information signal-to-noise ratio (SNR)."

Equation (11) is also the same form as the variance approximation that was demonstrated to compute much faster than Eq. (9), while retaining almost identical behavior.[12] In this work, the variance approximation is also used. The algorithm is outlined in the Appendix.

B. Nuisance parameter algorithm

In practice, not all model parameters are equally important. For instance, the ultimate goal of an experiment may be to ascertain the frequency of an oscillation. However, the background level or even amplitude of such oscillation may not be as important. The above implementation of the utility function treats all of the parameters on equal footing, and as a result, resources are designed to be spent almost equally on learning about each model parameter. If resources can be diverted from learning about less important parameters, even more efficient measurements are possible.

Suppose that for a given experiment, there are important model parameters to infer, still labeled θ. However, there are other model parameters that, while necessary for the model, are physically irrelevant or otherwise unimportant. We call these other "nuisance" parameters, ϕ. Including the nuisance parameters explicitly, the experimental model becomes

$$y = f(\theta, \phi, d) + \zeta. \tag{13}$$

Now, we wish to construct a utility function that de-emphasizes the nuisance parameters, ϕ, such that parameters of interest, θ, are prioritized. We begin by modifying the posterior, to distinguish the nuisance parameters $P(\theta|y', d') \to P(\theta, \phi|y', d')$. Yet, since we are only concerned with the θ parameters, we can marginalize the posterior distribution such that we remove dependence on ϕ,

$$P(\theta|y', d') = \int P(\theta, \phi|y', d')\, d\phi. \tag{14}$$

Therefore, we can rewrite Eq. (6) as

$$\begin{aligned} U^\phi(d') = \int P(y'|d') \iint P(\theta, \phi|y', d')\, d\phi \\ \times \ln\left(\frac{\int P(\theta, \phi|y', d')\, d\phi}{P(\theta)}\right) d\theta dy'. \end{aligned} \tag{15}$$

With an application of the Bayes rule,

$$P(\theta, \phi|y', d') = \frac{P(y'|\theta, \phi, d')}{P(y')} P(\theta, \phi), \tag{16}$$

the utility function becomes

$$\begin{aligned} U^\phi(d') = \int P(y'|d') \int \frac{P(y'|\theta, \phi, d')P(\theta, \phi)}{P(y'|d')}\, d\phi \\ \times \ln\left(\frac{\int \frac{P(y'|\theta, \phi, d')P(\theta, \phi)}{P(y'|d')}\, d\phi}{P(\theta)}\right) d\theta. \end{aligned} \tag{17}$$

After rearranging,

$$\begin{aligned} U^\phi(d') = \iiint P(y'|\theta, \phi, d')P(\theta, \phi)\, d\phi \\ \times \ln\left(\frac{\int P(y'|\theta, \phi, d')P(\theta, \phi)\, d\phi}{P(\theta)P(y'|d')}\right) d\theta dy'. \end{aligned} \tag{18}$$

02 March 2025 10:26:28

© Author(s) 2025

Noting that

$$\int P(y'|\theta, \phi, d')P(\theta, \phi)d\phi = \int P(y', \theta, \phi|d')d\phi = P(y', \theta|d') = P(y'|\theta, d')P(\theta), \quad (19)$$

we can rewrite the utility as

$$U^{\phi}(d') = \int\int P(y'|\theta, d')P(\theta)\ln\left(\frac{P(y'|\theta, d')P(\theta)}{P(\theta)P(y'|d')}\right)d\theta dy'. \quad (20)$$

Finally, rearranging,

$$U^{\phi}(d') = \int P(\theta)\int P(y|\theta, d')\ln\left(\frac{P(y'|\theta, d')}{P(y|d')}\right)d\theta dy', \quad (21)$$

then expanding,

$$U^{\phi}(d') = \int P(\theta)\int P(y'|\theta, d')\ln P(y'|\theta, d')dy'd\theta - \int P(y'|d')\ln P(y'|d')dy'. \quad (22)$$

Comparing Eqs. (8) and (22), it is evident that the explicit form of the utility function is the same whether or not nuisance parameters are marginalized. The absence of the nuisance parameters makes sense as we set out to only consider the information entropy of parameters θ.

In Eq. (8), the leading term involves distributions of outcomes when parameter values are given or fixed. In Eq. (22), the leading term involves model distributions of outcomes y' when θ values are given, but the nuisance parameters, ϕ, are not fixed. In the second term of Eq. (22), the distribution of y' includes the effects of both θ and ϕ parameters.

As before, if we use the approximation that the noise (ζ), θ, and ϕ are all normally distributed, we can re-express Eq. (22) as

$$U^{\phi}(d') = \frac{1}{2}\ln\left(2\pi v_{\theta}(d') + 2\pi v_{\zeta} + 2\pi v_{\phi}\right) - \frac{1}{2}\ln\left(2\pi v_{\zeta} + 2\pi v_{\phi}\right) = \frac{1}{2}\ln\left(\frac{v_{\theta}(d') + v_{\zeta} + v_{\phi}}{v_{\zeta} + v_{\phi}}\right). \quad (23)$$

We wish to emphasize that while the nuisance parameters have been marginalized from the utility function, they will still be inferred in the same process as the other parameters. Additionally, it should be expected that nuisance parameters (ϕ) that are highly correlated with parameters of interest (θ) cannot be fully de-emphasized. That is, throughout the experiment, instrument settings that contain relatively more information about ϕ that, in turn, inform about θ may be chosen. This means that prior knowledge of parameter correlations is not needed; it is "safe" to de-emphasize parameters that are not of interest, no matter the underlying correlations. We should expect potentially fewer gains in efficiency if there are strong correlations in ϕ and θ. In the Appendix, we outline the modified algorithm for the case of nuisance parameters (U^{ϕ}).

C. Expense algorithm

Suppose that for a given experiment, the associated expense of taking a measurement varies with the setting d'. The expense could be time or another type of resource. An example of a time expense is when the setting is a time delay or if changing the setting results in an overhead time (e.g., reprogramming an instrument). Here, we wish to modify the utility function such that measurements that are "cheaper" or faster are prioritized.

We take a direct weighting approach,

$$U^{\varepsilon}(d') = \frac{U(d')}{\varepsilon(d')}, \quad (24)$$

where $\varepsilon(d')$ is the expense function describing the economic cost of a proposed measurement design. Here, we distinguish this as an expense function to avoid conflation with the cost function used in machine learning. This formulation retains two fundamental properties of mutual information: first, when the prior and predicted posterior distributions are the same, the utility equals 0, and second, the utility should be non-negative. Equation (24) also has an intuitive interpretation—selecting a setting is a benefit-to-cost ratio. This approach has some precedence in weighted or "compound" utility functions, which are typically weighted averages of multiple utility functions to create a composite design with two or more distinct objectives.[4,6,13–15] However, unlike these methods, the weight here is a function of the proposed design, d', and is imposed on a singular utility function. We would like to note that other approaches that directly maximize sensitivity, which by definition consider the time expense of a measurement, such as those by Caouette-Mansour *et al.*[3] and Arshad *et al.*[7] have demonstrated great success.

Furthermore, our approach is compatible with any utility function, including the aforementioned nuisance parameter utility or various approximations.[12] In general, it is expected to be preferable to combine the expense function with the nuisance parameter algorithm such that the expense function does not preference learning about unimportant parameters just because they are "less expensive." The algorithm that combines the nuisance parameter and expense approach ($U^{\phi,\varepsilon}$) is included in the Appendix.

III. METHODS

A combination of simulations and experimental measurements was used to study the impact of the adaptive algorithms on NV-based quantum sensing protocols. For the simulations, we study the behavior of the setting selection and track parameter uncertainties for a group of 100 experiment runs, under typical noise levels for the Rabi, Ramsey, Hahn echo, and T_1 sequences. Similarly, for the experimental measurements, we record the setting selection and parameter uncertainties as a function of time and number of measurements for five individual runs per algorithm per sequence. We represent parameter uncertainties as the width of the parameter distributions, which we have approximated to be Gaussian. We will, therefore, report the standard deviations.

02 March 2025 10:26:28

© Author(s) 2025

A. Nitrogen vacancy centers in diamond

The nitrogen-vacancy (NV) center in diamond is formed by a substitutional nitrogen atom paired with a neighboring vacancy. These defects are present at concentrations varying from a few hundredths of a ppb to several ppm depending on the growth method and treatment.[16–18] Quantum sensing can be performed at the single NV center level, as well as with ensembles of various sizes. In general, the achievable spin-state contrast of NV ensembles (a few percent) is much lower than that of a single NV ($\simeq$30%).[19–22]

When an electron is trapped in the defect (NV^-), the NV center becomes a $S = 1$ electronic spin system. The NV center can be initialized into the ground $m_s = 0$ spin state and read out via spin-dependent photo-luminescence under optical pumping with green light. After absorbing a green photon, an electron originating in the $m_s \pm 1$ state is more likely to decay through a secondary spin non-conserving, non-radiative pathway—the inter-system crossing (ISC).[8] This secondary decay pathway forms the basis for the initialization and readout of the NV center. When the green laser light is first turned on, the spin state can be determined by the relative number of red photons emitted. More photons imply $m_s = 0$; fewer photons imply $m = \pm 1$. When the laser is left on for a longer time, the spin-mixing process eventually shuttles most of the electrons into the ground $m_s = 0$ state resulting in an initialized (polarized) spin state.

For the experimental measurements, our lab-built confocal quantum diamond microscope was used. The diamond sample was a chemical vapor deposition (CVD)-grown, single crystal with less than 1 ppm of nitrogen ($\approx$0.5 ppb NV in the negatively charged state).[9,23,24] A bias magnetic field of a few mT was formed by two permanent magnets aligned along one family of NV axes. The sample was optically pumped with approximately 8 mW of 514 nm laser light. The laser light was focused into the sample using a 100×, 0.7 N.A. objective lens ($\approx 2\mu m^3$ confocal volume). The collected photo-luminescence was redirected using a dichroic mirror, passed through a pinhole at the focal plane, and then focused into a fiber-coupled single photon counter. The microwave frequency magnetic field was provided using a $\approx$1 mm outer diameter shorted loop made from the $\approx 100\,\mu$m diameter inner conductor of a stripped coaxial cable. Experiment control was implemented with a programmable transistor–transistor logic (TTL) generator with a 300 MHz time base, and the data acquisition was controlled using a digital time-to-streaming converter. Python was used to control hardware and implement the adaptive algorithms. Measurements at each setting were repeated thousands of times, forming an epoch. After each epoch, the parameter distributions and utility functions were updated.

B. Quantum sensing

This work focuses on the core quantum sensing techniques of the Rabi, Ramsey, Hahn echo, and T_1 relaxometry sequences. We briefly describe these measurements in the rest of the section.

The Rabi sequence gauges the strength of the microwave magnetic fields used to manipulate the spin states. In order to perform other sensing experiments, the $\pi/2$ and π pulses must be calibrated with a Rabi experiment. The Rabi pulse sequence begins with a laser pulse that initializes the spins into the $m_s = 0$ state; then, an on-resonance microwave pulse is applied for a duration of τ before the laser is turned back on to read out the spin state. By scanning over the microwave pulse duration, Rabi oscillations are revealed. The time corresponding to the first minima (i.e., where $m_s = \pm 1$ depending on the resonance used) is the π pulse duration. The $\pi/2$ pulse duration corresponds to the time that it takes to transition halfway to the other eigenstate, which here we have estimated as half the length of the π pulse.

A Ramsey sequence is used for measuring DC or slowly varying magnetic fields.[8] The NV system is initially prepared in the $m_s = 0$ state with a green laser pulse. Then, it is put into a superposition with either the $m_s = \pm 1$ state, $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, using a $\pi/2$ pulse. Once in the superposition, the magnetic field information is encoded in the relative phase accumulated after a time τ, $\frac{1}{\sqrt{2}}(|0\rangle + e^{-i\omega\tau}|1\rangle)$. The population is then mapped back to the $m_s = 0$ and $m_s = \pm 1$ states after time τ using another $\pi/2$ pulse. By scanning over the delay time τ, Ramsey fringes emerge, which are a result of the frequency detuning of the microwave pulses from resonance, plus the contributions of the DC or slowly varying field. The sensitivity of the Ramsey sequence is limited by the dephasing time T_2^*.

The Hahn echo sequence, which measures AC magnetic fields, is similar to the Ramsey sequence but includes an additional π pulse halfway between the initial and final $\pi/2$ pulses. The π pulse rephases the spins, thus making the Hahn echo sequence insensitive to DC fields and additionally, instead limited by the decoherence time, T_2, where $T_2 \geq T_2^*$.[8]

The final type of sensing experiment studied here is relaxometry, which relies on detecting changes in the spin relaxation time. The spin relaxation time, T_1, describes how long the spin population will remain out of thermal equilibrium. T_1 is sensitive to any transverse fluctuations at the spin transition frequency and is, therefore, used in a wide variety of material and biological applications.[8,25–28] These experiments can be performed in a variety of ways[25] but generally rely on initially polarizing the NV center into $m_s = 0$ spin state, and coherently transferring the population to $m_s = \pm 1$ with a π pulse, and then allowing the system to relax for a varied amount of time before reading out the spin state.[29]

All pulse sequences described above are shown in Fig. 1.

IV. RESULTS AND DISCUSSION

A. Rabi

During a Rabi experiment the microwave (MW) pulse duration, τ, is scanned. Most often the Rabi experiment is used to calibrate the microwave pulses for other experiments as the location of the first minima corresponds to the duration of the π pulse. It is also a helpful way of measuring the resonant microwave strength. The contrast measured as a function of MW pulse duration is

$$S_{Rabi}(\tau) = Ae^{-\tau/T_{1\rho}}\cos(\omega\tau) + b, \tag{25}$$

where the decoherence time $T_{1\rho}$ (also known as $T^*_{2(Rabi)}$) is extended compared to the true T_2^* due to continuous driving and, thus, is typically not the main parameter of interest. Therefore the main parameter of interest is the Rabi frequency, ω, and the other three parameters are set as nuisance parameters.

02 March 2025 10:26:28

© Author(s) 2025

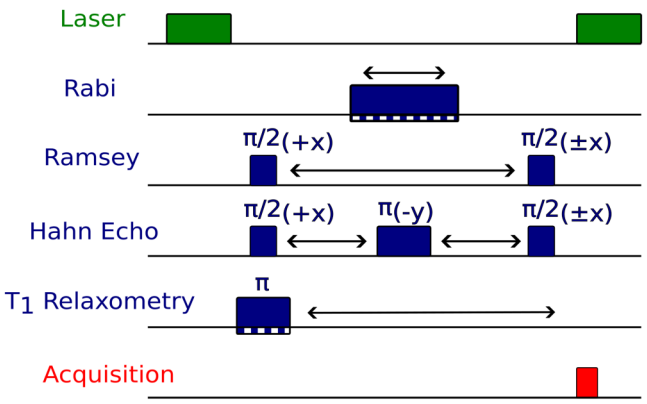


FIG. 1. Pulse sequences used in this work. All sequences use the same initiation and acquisition pulses (green and red, respectively) and the arrows indicate the scanned parameter, either delay or pulse duration. The MW pulses used in each of the four sequences are shown individually in blue which are used for the "signal" in contrast and visibility calculations [contrast = (signal)/(background) visibility = (signal − background)/(signal + background)]. Sequences are repeated either without microwaves, denoted above with a dashed bottom (Rabi, T_1 relaxometry), or with microwave phase cycling, denoted above with $\pm x$ (Ramsey, Hahn echo) for the "background."

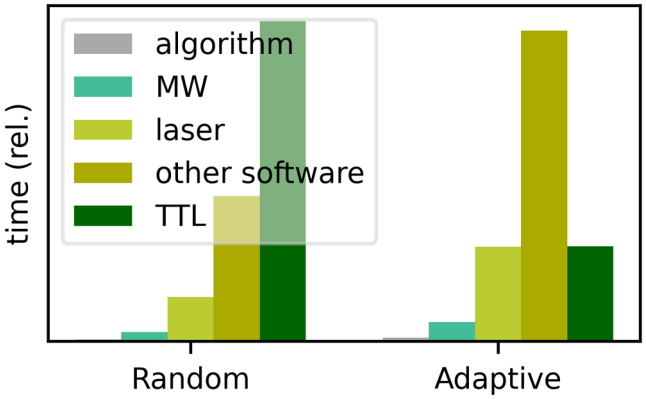


FIG. 2. Representation of the average division of time during a 200 s Rabi experiment (number of averages in an epoch, $N_A = 6000$) between the TTL generator reprogramming time, other software time, the adaptive algorithm overhead, which includes the updating parameter distributions, the laser pulse duration, and setting, here the microwave duration (MW). For "Random" on the left, settings were randomly selected. Note that when randomly selecting settings, we are still updating parameter distributions through the Bayesian inference after each epoch. However, this amount of time is visually indiscernible. The average number of epochs in 200 s was 567. For "Adaptive" on the right, settings were selected using the nuisance parameter and expense algorithm, $U^{\phi,\varepsilon}$. The average number of epochs in 200 s was 1215, just over twice the amount as the "Random." Note that this means that all categories of time besides the TTL have approximately doubled. Similar behavior is seen in the Ramsey sequence.

The MW duration τ (on the order of 10–1000 ns) is much shorter than the initialization and readout time (few μs per measurement) or software time (hundreds of ms per epoch), so each setting "costs" approximately the same. However, the amount of time to reprogram the TTL generator for a new MW duration can be relatively time-consuming (220 of the 320 ms total software overhead, where the utility calculation and parameter inference constitute less than 1 ms). It is, thus, more economical to repeat epochs at the same value of τ to avoid reprogramming the TTL generator. These effects are accounted for using the expense function:

$$\varepsilon_{Rabi}(\tau') = N_A(\tau' + t_{OH}) + t_{SW} - t_{TTL}\delta(\tau - \tau'), \tag{26}$$

where N_A represents the number of averages in an epoch, t_{OH} is the overhead time which includes laser initialization, readout and interpulse delays (5 μs), t_{SW} is the total software time per epoch (320 ms, as measured on our instrument), and t_{TTL} is the time required to reprogram for a new pulse delay (220 ms). The delta function, $\delta(\tau - \tau')$, incorporates a reduced expense if the MW duration is not changed. Figure 2 shows a breakdown of how time is spent during Rabi experiment and how considering the TTL reprogramming time with an expense function leads to a dramatic increase in useful experiment time as opposed to overhead.

Figure 3(a) shows the simulated impact of the standard adaptive (U, purple), nuisance parameter (U^{ϕ}, light blue), and combined nuisance parameter and expense ($U^{\phi,\varepsilon}$, pink) algorithms to randomly selected settings (orange) for a Rabi experiment with an SNR of 1. The combined power of the nuisance parameter and expense algorithm results in the greatest increase in efficiency. After 200 s, the uncertainty in the Rabi frequency is reduced by 43 % using this algorithm vs the randomly selected settings. Note that in this work, we use randomly selected settings to represent non-adaptive measurements without introducing scanning artifacts.

It is expected that the most information about ω is located where $|\delta S_{Rabi}/\delta\omega|$ is at a local maxima and where $\tau \approx T_{1\rho}$.[12] One can see how setting selection is concentrated around these regions by the nuisance parameter algorithm in Figs. 3(b) and 3(c). Figure 3(d) shows how the inclusion of an expense function results in epochs repeated at the same setting as a consequence of the delta function in Eq. (26).

Complementary experiments were performed using the same three algorithms and randomly selected settings. In Fig. 4, each Rabi experiment was repeated five times for each of the four methods and the uncertainty in the Rabi frequency was recorded after each epoch. All methods were initiated with the following priors: $P(f) = \text{uniform}(2, 6)$ MHz, $P(T^*_{2,Rabi}) = \text{uniform}(250, 1500)$ ns, $P(a) = \text{uniform}(0.0075, 0.03)$, and $P(b) = \text{normal}(1, 0.01)$. All experiments were allowed to run for 3 min (wall clock). The final mean uncertainty in the Rabi frequency at 3 min of total experiment time was 0.061, 0.035, 0.030, and 0.022 MHz, for randomly selected, U, U^{ϕ}, and $U^{\phi,\varepsilon}$, respectively. The best-performing adaptive algorithm, $U^{\phi,\varepsilon}$, resulted in a 64% reduction in uncertainty in Rabi frequency at 3 min. In other terms, this adaptive algorithm can significantly reduce experiment time.

B. Ramsey

The Ramsey sequence is used to measure static or very slowly varying magnetic fields.[8,30,31] The Ramsey fringes depend on the

02 March 2025 10:26:28

© Author(s) 2025

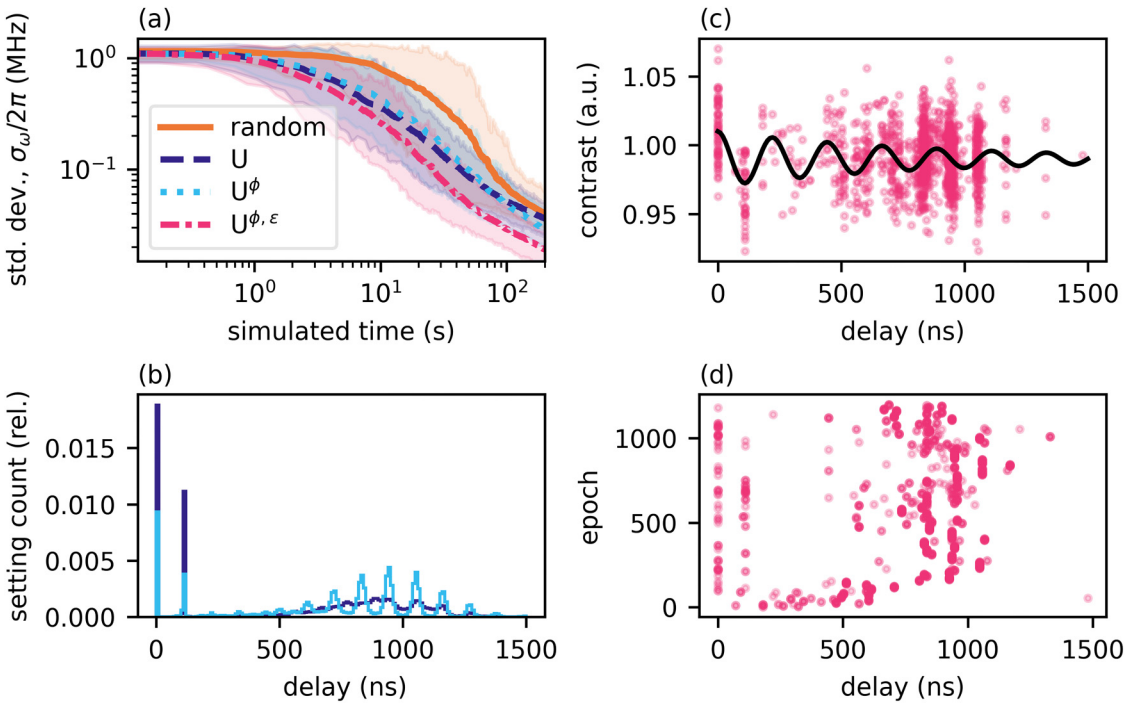


FIG. 3. Simulated comparison of randomly selected settings (orange) vs the U (purple), U^ϕ (light blue), and $U^{\phi,\varepsilon}$ (pink) algorithms for a Rabi experiment with a signal-to-noise ratio (SNR) of 1. For an SNR of 1, an epoch = 6000 total averages, 3000 each for signal and background. Potential settings were chosen from 150 values ranging from 0 to 1500 ns. The prior distribution of the Rabi frequency, $\omega/(2\pi)$, was a uniform distribution ranging from 2 to 6 MHz (true value 4.5 MHz). $T_{1\rho}$ prior distribution was uniform from 250 to 1500 ns (true value 900 ns). (a) The mean standard deviation of the Rabi frequency (f) vs experiment time over 100 individual runs. The shaded area corresponds to values within a 95% confidence interval. (b) The relative setting counts of U and U^ε algorithms. Notice how U^ϕ (light blue) is more selective with the settings. (c) The model function and an example set of measurements using the $U^{\phi,\varepsilon}$ algorithm. (d) The settings selected vs epoch when an expense function is used. The additional expense of changing settings leads to epochs using repeated settings.

02 March 2025 10:26:28

detuning of the microwave field from resonance, $f = f_0 - f_{MW}$. Hyperfine coupling to ^{14}N leads to three resonances split by $|A_\parallel| = 2.16$ MHz,[32]

$$S_{Ramsey}(\tau) = e^{-\tau/T_2^*}(a_1 \cos(2\pi(f + A_\parallel)\tau) + a_2 \cos(2\pi f\tau) + a_3 \cos(2\pi(f - A_\parallel)\tau)) + b. \quad (27)$$

In general, there are two parameters of interest in Eq. (27)—the dephasing time, T_2^*, and the Ramsey frequency, f. The most information about the parameters f and T_2^* will be where $|\delta S_{Ramsey}(\tau)/\delta f|$ and $|\delta S_{Ramsey}(\tau)/\delta T_2^*|$, respectively, are at local maxima near $\tau \approx T_2^*$. This means information about f and T_2^* is "out-of-phase."

Therefore, no improvements are expected with U^ϕ over U if both f and T_2^* are considered equally important for the same experiment. Indeed, this is what we found ($\approx$5% decrease in σ_f and $\approx$5% increase in $\sigma_{T_2^*}$). However, rarely are f and T_2^* considered to be equally important. Typically, T_2^* is used to characterize the sensitivity of the diamond[9,16,18,33] or determine the preferential sensing time, $\tau \approx T_2^*$. The preferential sensing time is important when Ramsey fringes are not mapped out, and instead, small changes around a set point near maximum sensitivity (i.e., $\tau \approx T_2^*$) are measured. The latter approach of using "magnetometry curves" is sometimes used to improve the efficiency as compared to mapping out Ramsey fringes.[8] In the case of mapping out Ramsey fringes for magnetometry, high sensitivity to f is prioritized.

Figures 5 and 6 show the results when optimizing the adaptive algorithm for T_2^* and f, respectively. All adaptive algorithms showed tremendous improvement in efficiency as compared to the randomly selected setting (orange), reaching the inverse square root scaling limit much faster, that is where the sensitivity plateaus in Fig. 6(d). There is a significant improvement in sensitivity by de-emphasizing the nuisance parameters (purple vs light blue). The inclusion of the expense function (pink) appears to further improve the sensitivity, particularly after reaching the inverse square root scaling limit.

Experimental results for the Ramsey sequence are shown in Fig. 7. Each of the four protocols is run five times for 20 min each. The MW pulses were applied 4 MHz off-resonance to simulate magnetic field sensing ($\approx$150 μT). The priors were the same as in the simulations shown in Figs. 5 and 6. The uncertainty of f and the T_2^* is recorded as a function of experiment time. At the end of 20 min, the mean uncertainty of T_2^* was 223, 96, 85, and 70 ns for

© Author(s) 2025

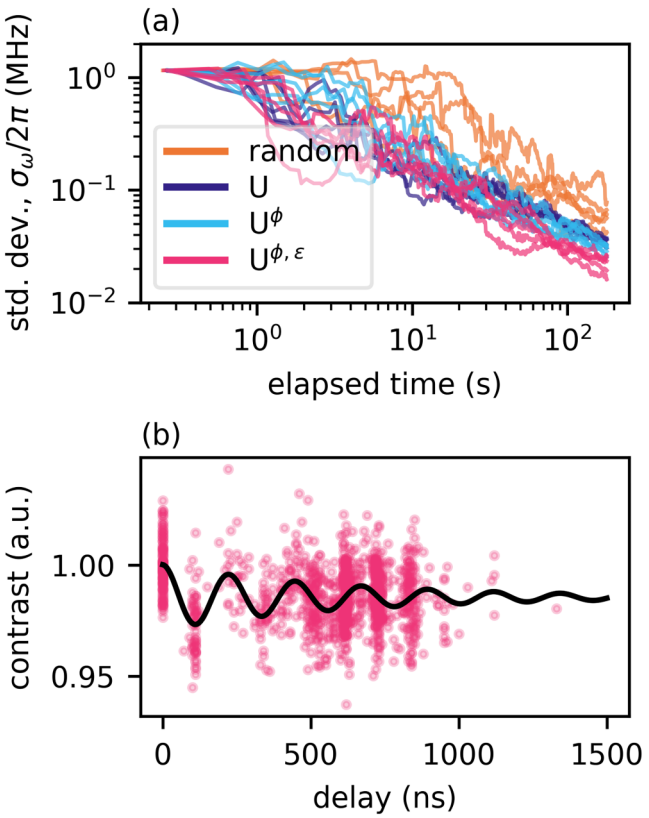


FIG. 4. Experimental data showing how adaptive algorithms are able to learn about the Rabi frequency with fewer measurements even at an SNR of 1. The settings were chosen from 150 values ranging from 0 to 1500 ns and the prior distribution for the Rabi frequency, $\omega/(2\pi)$, was a uniform distribution from 2 to 6 MHz. The prior distribution for $T_{1\rho}$ was uniform from 250 to 1500 ns. (a) Individual traces of uncertainty in the Rabi frequency vs epoch for five experiments with randomly selected settings (orange) and the U (purple), U^ϕ (light blue), and $U^{\phi,\varepsilon}$ (pink) algorithms. (b) Example setting selection with $U^{\phi,\varepsilon}$ and fitted model, which can be compared to Fig. 3(c).

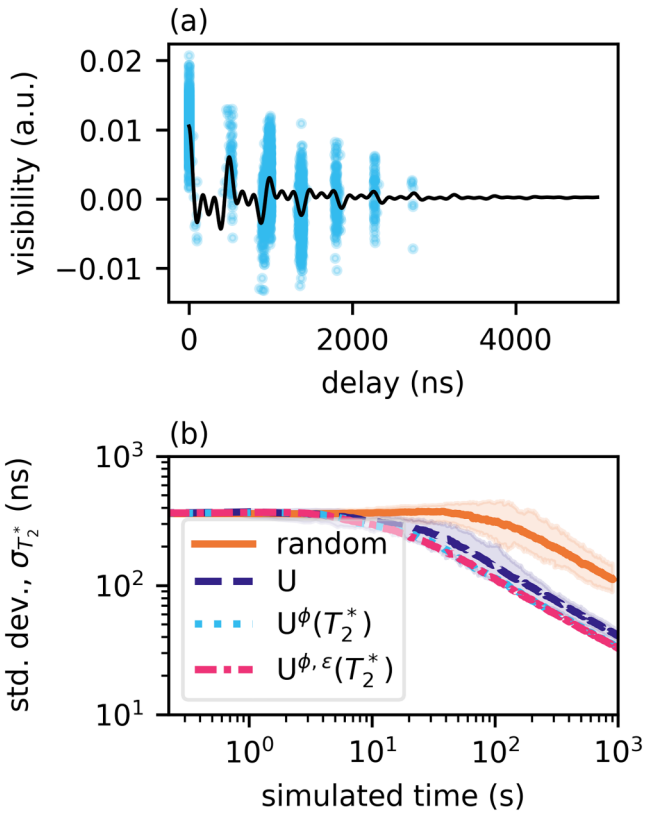


FIG. 5. Simulations using randomly selected settings (orange) and the U (purple), U^ϕ (light blue), and $U^{\phi,\varepsilon}$ (pink) algorithms for a Ramsey experiment with an SNR of 1. In this case, learning the T_2^* was prioritized by de-emphasizing all other model parameters for U^ϕ and $U^{\phi,\varepsilon}$. Potential delay settings were selected from 800 values ranging from 0 to 5000 ns. The prior distribution of f was a uniform distribution ranging from 1.5 to 8 MHz (true value 4 MHz) and prior for T_2^* ranged from 250 to 1500 ns (true value 1000 ns). (a) Example set of measurement using U and the ground truth (solid black curve). (b) The mean standard deviation of T_2^* over 100 individual runs. The bands represent a 95% confidence interval.

the randomly selected settings (orange) and the U (purple), U^ϕ (light blue), and $U^{\phi,\varepsilon}$ (pink) algorithms, respectively. Similarly, the mean uncertainty of f was 0.0203, 0.0086, 0.0060, and 0.0046 MHz, respectively. The last of which represents an almost fivefold improvement in sensitivity as compared to the randomly selected settings.

For magnetometry, the sensitivity is an important criterion that describes the minimum detectable field within a certain amount of experiment time. We define the sensitivity of the measurements as

$$\eta = \sigma\sqrt{t_{total}}, \tag{28}$$

where σ is the uncertainty and t_{total} is the total experiment time.[19,22] We wish to emphasize that the shot noise, spin-projection limited sensitivity of the Ramsey is in general a function of the delay time τ,[8,9]

$$\eta \propto \sqrt{\frac{\tau + t_{OH} + t_{SW}/N_A}{\tau^2}}\left(\frac{1}{e^{-\tau/T_2^*}}\right), \tag{29}$$

where we have included the software time (per epoch) evenly distributed over each individual average. Note that as we are programming our pulse generator as we go, we are explicitly including the programming time in our sensitivity calculations. Software time is typically omitted from theoretical sensitivity considerations. Despite sensitivity's dependence on τ, theoretical limits are typically reported using the minimum of this function where $\tau = T_2^*$. It should, therefore, not be surprising that non-adaptive protocols that map out the Ramsey fringes fail to meet the predicted sensitivity limit (see orange curves in Figs. 6 and 7). Typically, magnetometry curves are used to

02 March 2025 10:26:28

© Author(s) 2025

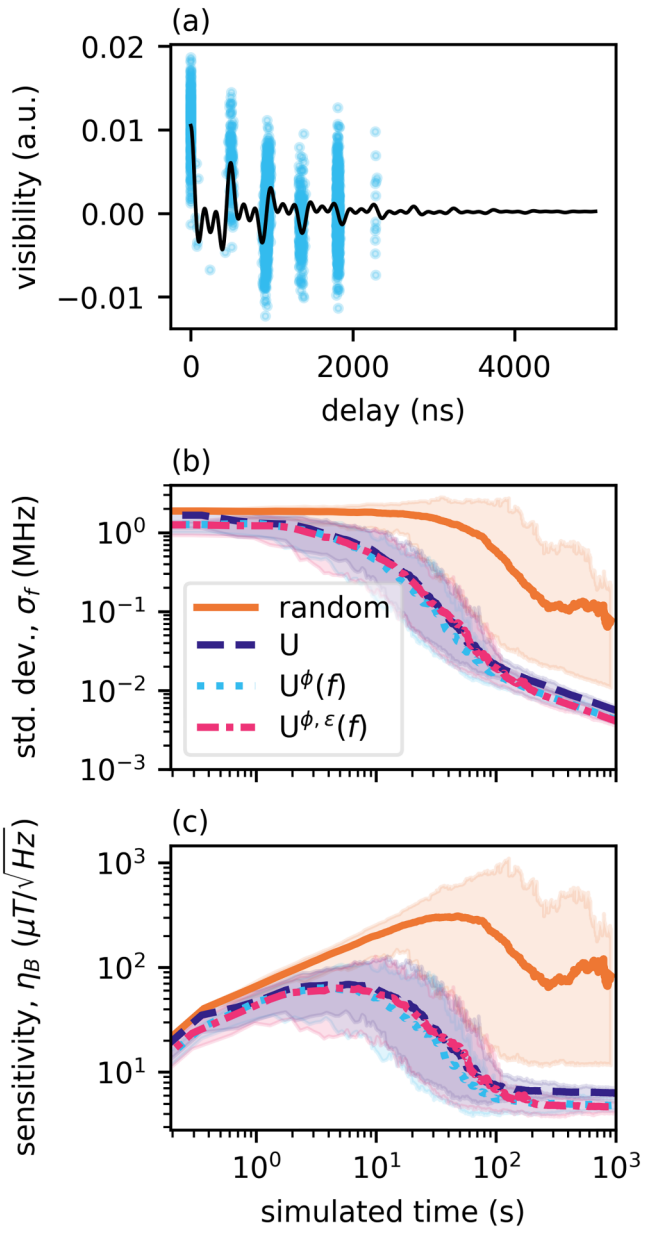


FIG. 6. Same as Fig. 5 but for U^{ε} and $U^{\varepsilon,\phi}$ all parameters except for f were de-emphasized. (a) Example experiment using U^{ε} and the ground truth (solid black curve). Note how the setting selections are slightly different from Fig. 5(a). (b) The uncertainty in the Ramsey frequency as a function of time. (c) The sensitivity to the magnetic field.

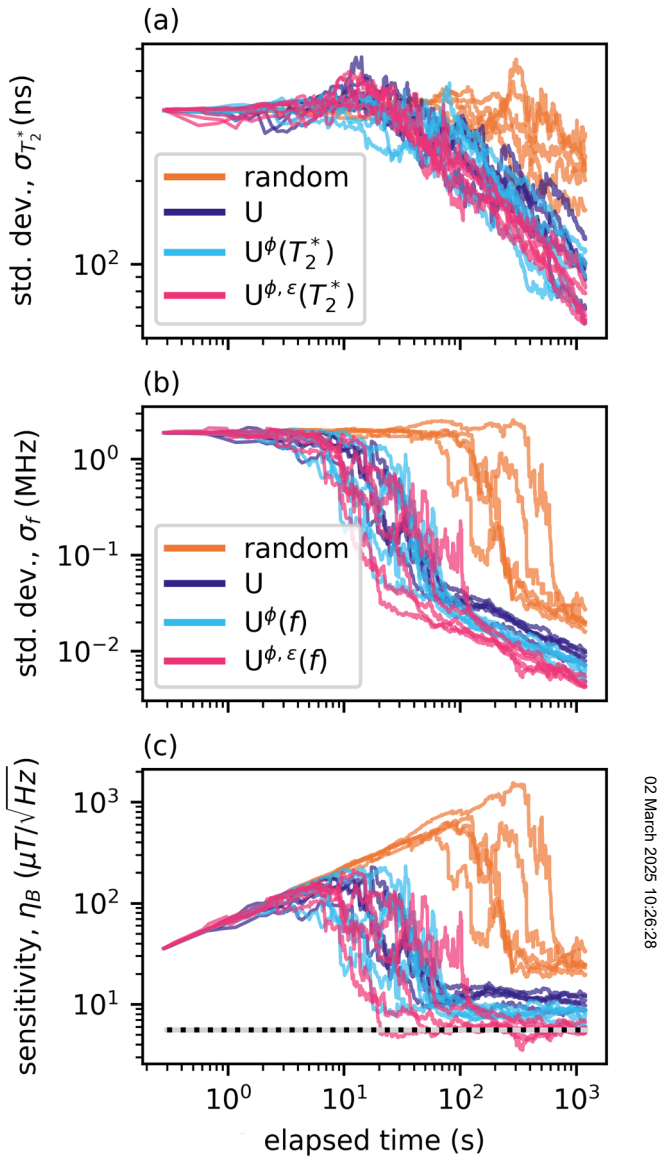


FIG. 7. Experimental uncertainty of (a) T_2^* (measured value 1160 ns) and (b) f as a function of experiment time for the four experiment protocols at an SNR of approximately 1. Each trace represents an individual experiment. (c) Sensitivity achieved with each method. The solid gray line shows the sensitivity limit without the adaptive overhead ($\lesssim$1 ms per epoch), and the dashed black line shows the sensitivity limit, including the adaptive computational overhead. The two limits are indistinguishable. The sensitivity is limited by the low intrinsic NV density of our sample and the limitations of working in the regime where Rabi frequency $\approx$ hyperfine splitting $\approx$ the detuning from resonance.

help reach the minimum sensitivity; however, doing this requires selecting a set point on the side of a Ramsey fringe where $\tau \approx T_2^*$,[8] which requires knowing T_2^* and to a high degree, an accurate estimation of the magnetic field to be measured. The latter requirement derives from the Ramsey being a phase-based measurement and to avoid ambiguity, the resultant phase must be known within a range of $[-\pi/2, \pi/2)$. That is, magnetometry curves have an inherently limited dynamic range.[34] The optimization provided by adaptive algorithms allows us to reach the minimum sensitivity, *without*

02 March 2025 10:26:28

© Author(s) 2025

knowing T_2^* or the magnetic field ahead of time. The ability of our methods to optimize based on incoming data and not prior knowledge or operator expertise is an enormous advantage for the adoption of color centers beyond specialty lab settings.

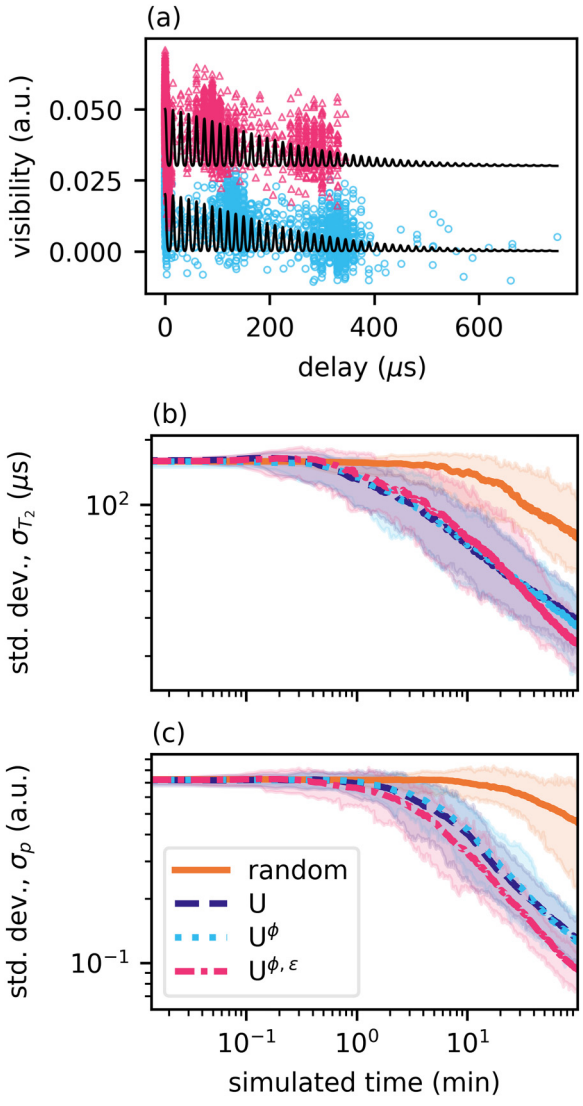


FIG. 8. Simulated comparison of randomly selected settings (orange) and the U (purple), U^{ϕ} (light blue), and $U^{\phi,\varepsilon}$ (pink) algorithms for a Hahn experiment with an SNR of 3. Potential delay settings were selected from 750 values ranging from 0 to 750 μs. The prior distribution of T_2 was a uniform distribution ranging from 350 to 650 μs (true value 500 μs) and the prior for p was a uniform distribution from 0.5 to 3 (true value 1.4). (a) Example set of measurements using the nuisance parameter algorithm with (pink triangles) and without (light blue circles) an expense function and the ground truth (solid black curve). A vertical offset was applied for clarity. (b) The mean standard deviation of T_2 over 100 individual runs. The bands represent a 95% confidence interval (c) Same as (b) but for p.

C. Hahn echo

The Hahn echo sequence adds an additional "refocusing" π pulse in the Ramsey sequence, which serves to rephase any coherence lost due to inhomogeneous contributions from the local spin

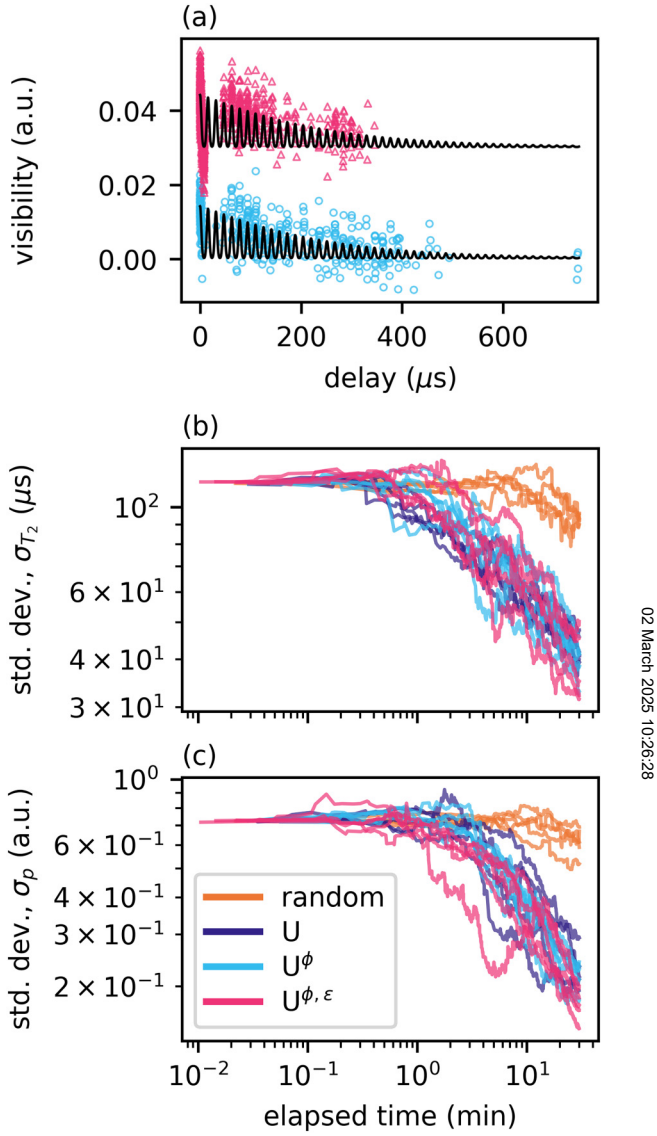


FIG. 9. Experimental results of the Hahn echo sequence with an SNR of 3. Potential delay settings were selected from 750 values ranging from 0 to 750 μs. The prior distribution of the T_2 was a uniform distribution ranging from 350 to 650 μs and the prior for p was a uniform distribution from 0.5 to 3. The measured T_2 was 455 μs with a p of 1.34. (a) Comparison of the settings selected using U^{ϕ} (light blue circles) and $U^{\phi,\varepsilon}$ (pink triangles). The latter was vertically offset for clarity. (b) The uncertainty in T_2 as a function of experiment time for U (purple), U^{ε} (light blue), and $U^{\phi,\varepsilon}$ (pink) adaptive algorithms vs randomly selected settings (orange). Each trace represents an individual experiment run. (c) Same as (b) but for p.

02 March 2025 10:26:28

© Author(s) 2025

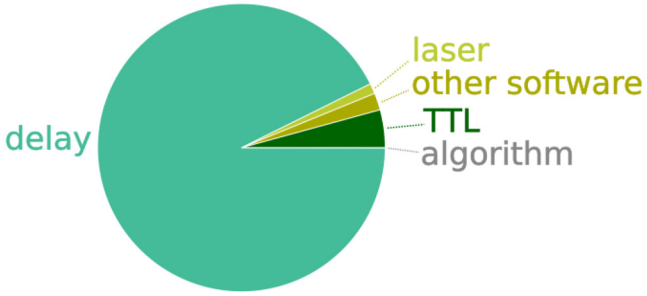


FIG. 10. Breakdown of time spent during a single epoch with a delay of 400 μs where $N_A = 12\,000$. Epoch time is dominated by the delay, which is the field sensing time. This can be compared to the Rabi and Ramsey sequences in Fig. 2 where overhead (software and TTL) dominate.

bath, strain, and magnetic field gradients. Therefore, the decoherence time, T_2, is typically at least an order of magnitude larger than the dephasing time, T_2^*.[8,17,18,31] In general, the Hahn echo takes on the form,

$$S_{Hahn}(\tau) = Ae^{-(2\tau/T_2)^p} \sum_j e^{-((2\tau - jT_R)/T_D)^2} + b, \tag{30}$$

where the delay τ is the time between the $\pi/2$ and π pulses. In Eq. (30), the summation represents the echos due to ^{13}C nuclear spin procession, with revival and decay periods of T_R and T_D, respectively. The revival and decay periods of the echos will be dependent on the bias (DC) magnetic field and may or may not be apparent due to the combination of field strength and the decoherence time.[35] The stretched exponential parameter p encapsulates the different broadening mechanisms and can vary from 0.5 to 3 for ensembles.[9,18,35] Note that T_2 and p limit the minimum detectable magnetic field frequency, and the Rabi frequency limits the maximum detectable frequency.

In AC magnetometry, the echo sequence works by creating a filter centered on $f = 1/(2\tau)$. When scanning over τ, the filter rejects noise outside its bandwidth, therefore extending the coherence time. Simultaneously, the spins become sensitive to magnetic noise within the filter bandwidth. Therefore, if the frequency of the AC magnetic field is unknown, τ is swept, and dips in the signal are correlated to the frequency. Once the frequency is known, the

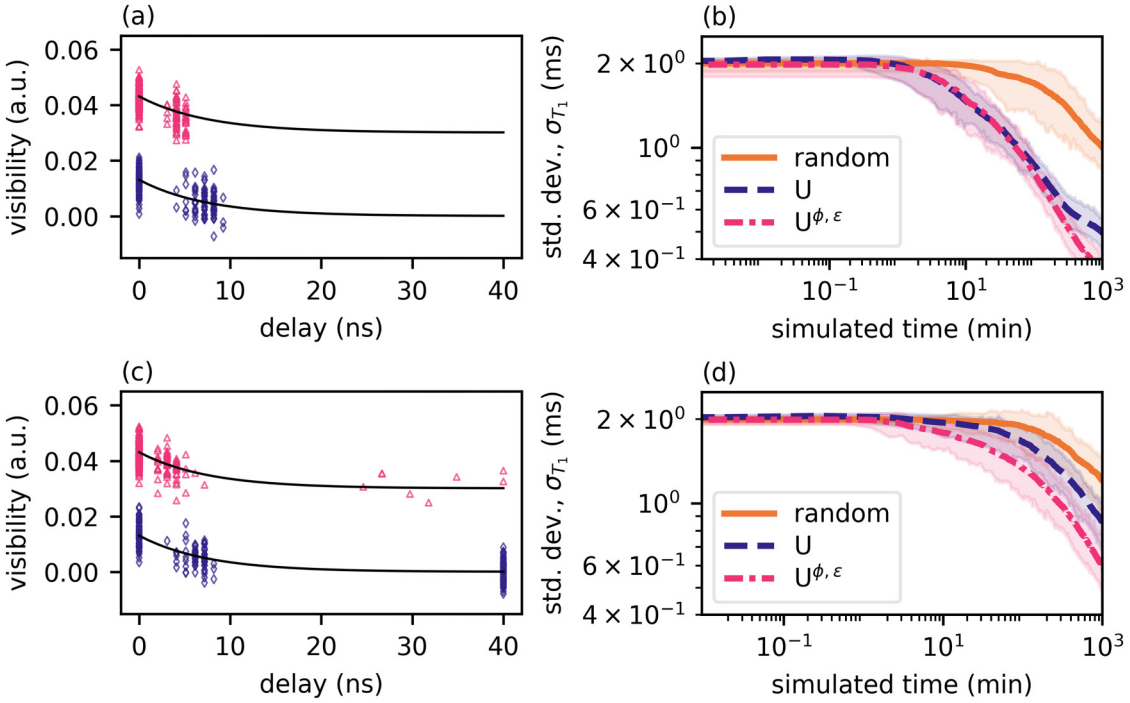


FIG. 11. Comparison of U (purple) and $U^{\phi,\varepsilon}$ (pink) to randomly selected settings (orange) for a T_1 experiment with an SNR of 3. The time was calculated with Eq. (26) with $N_A = 12000$. The potential settings were 40 values ranging from 0 to 40 ms. T_1 prior was uniform from 3 to 10 ms. The true values of the parameters were $T_1 = 7.5$ ms, $A = 0.013$, and $b = 0.0001$. In the top row, the background, b, is assumed to be known (prior width 2.5×10^{-4}), and in the bottom row, it is assumed to be less well known (prior width 2.5×10^{-3}). Settings selected for the first 200 epochs are shown in (a) and (c) where a vertical offset was applied for clarity. Note how in (c) many more settings out at long delays are selected to learn about b. (b) and (d) The uncertainty in T_1 as a function of time. The bands represent a 95% confidence interval.

02 March 2025 10:26:28

© Author(s) 2025

delay is set to $\tau = 1/(2f)$, and changes in the signal can related to the field amplitude.

Similar to the Ramsey, the optimal sensitivity is set by T_2 and p (i.e., $\tau \approx p^{-1/p} T_2^*$), but τ must also be matched to the period (or multiples of the period) of the AC magnetic field. Sequences incorporating dynamic decoupling such as the Carr–Purcell–Meiboom–Gill (CPMG) or the XY8 can extend the coherence time by rejecting more noise by creating a narrower filter.[8,9,31] Optimizing these dynamic decoupling sequences to match the T_2 for particular fields of interest can be time-consuming using non-adaptive methods. Here, we just study the Hahn echo, but the methods can be easily extended to the CPMG and the XY8.

Unlike the Rabi and Ramsey sequences, the delay times in the Hahn echo sequence are long enough that the expense of each measurement can vary significantly. The expense function is now,

$$\varepsilon_{Hahn}(\tau') = N_A(2\tau + t_{OH}) + t_{SW} - t_{TTL}\delta(\tau - \tau'), \tag{31}$$

where the factor of 2 accounts for the delays both before and after the refocusing π pulse. Figure 10 shows the division of measurement time during a single epoch.

The results of the simulations are shown in Fig. 8. Figure 8(a) compares the setting selection of the nuisance parameter algorithm with and without an expense function. For the U^{ϕ} and $U^{\phi,\varepsilon}$, all parameters except T_2 and p were de-emphasized. Figures 8(b) and 8(c) show the uncertainty of T_2 and p as a function of time for all three adaptive algorithms vs randomly selected settings.

Figure 9 shows the experimental behavior of the setting selection and the subsequent uncertainties in the T_2 and p as a function of experiment time for the various algorithms. For an SNR of 3 ($N_A = 12\,000$), all adaptive algorithms show an improvement in measurement efficiency. The most well-performing algorithm, the combined nuisance parameter and expense algorithm, resulted in a 2.4 and 3.6-fold reduction in uncertainty in T_2 and p, respectively, at 30 min of experiment time. The average number of epochs within 30 min of experiment time were 195, 437, 407, and 1280 for the randomly selected, U, U^{ϕ}, and $U^{\phi,\varepsilon}$, respectively. As expected, the expense algorithm leads to the prioritization of shorter delays as can be seen in the location of the pink vs purple data points in Fig. 8 and Fig. 9(a).

Due to the fact that T_2 and p are highly correlated and we elected to consider both as key parameters of interest [i.e., we de-emphasized all other model parameters in Eq. (30)] that meant experiment time was split between learning about T_2 and p. An analysis using partial derivatives shows that information about these parameters is located at different points in the setting space, with information about p at shorter delays. Therefore, we should expect more gains in p with the expense function. If, instead, we de-emphasize p, then $U^{\phi,\varepsilon}$ (pink) should instead produce higher gains for T_2. This is the exact behavior we observed, and we have included these results in the supplementary material.

D. T_1 relaxometry

Typically, T_1 is phonon-limited to several ms ($\lesssim$10 ms) unless the NV centers are subjected to magnetic noise at frequencies near their resonance.[36–39] Therefore, T_1 relaxometry is useful for sensing very high-frequency signals ($\geq$100 MHz) or paramagnetic spins.[8,25–28] As T_1 can be quite long, this sequence poses as the most time-consuming experiment. We assume a simple exponential decay:

$$S_{T_1}(\tau) = Ae^{-\tau/T_1} + b. \tag{32}$$

In Fig. 11, the simulated uncertainty in T_1, as a function of experiment time, is shown for an SNR of 3. U (purple) and $U^{\phi,\varepsilon}$ (pink) algorithms both show a substantial improvement over the randomly selected settings (orange). In the case of a simple exponential decay, all parameters are highly interdependent; that is, to know T_1, we must also know A and b. This means that we should not expect the nuisance parameter approach to be as impactful as in other sequences, such as the Rabi. However, the inclusion of the

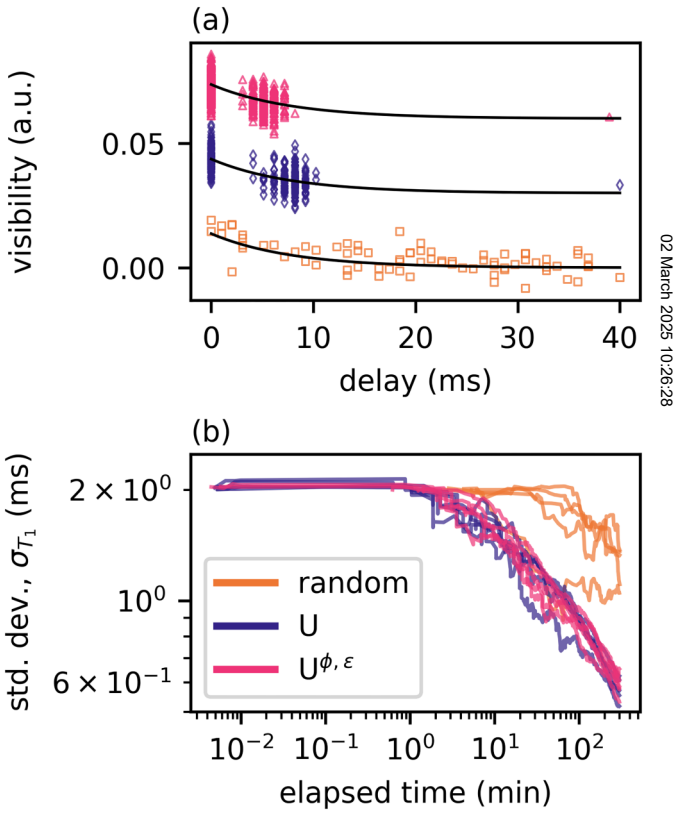


FIG. 12. (a) Comparison of the settings selected by the U (purple diamonds) and $U^{\phi,\varepsilon}$ (pink triangles) to randomly selected settings (orange squares) for a T_1 experiment with an SNR of 3. A vertical offset was applied for clarity. (b) The uncertainty in T_1 as a function of time. Potential delay settings were selected from 40 values ranging from 0 to 40 ms. The prior distribution of the T_1 was a uniform distribution ranging from 3 to 10 ms. The background was assumed to be well known with a normally distributed prior with a standard deviation of 2.5×10^{-4}. The measured T_1 was 7.5 ms.

02 March 2025 10:26:28

© Author(s) 2025

expense function can significantly improve the performance of nuisance parameter algorithm, depending on the parameter priors.

The background level b in Figs. 11(a) and 11(b) is assumed to be relatively well-known and described with a normally distributed prior with a standard deviation of 2.5×10^{-4}. In Figs. 11(c) and 11(d), the background level is assumed to be less well-known and instead described with a normally distributed prior with a standard deviation of 2.5×10^{-3}. In the case where the background is well-known, the similar behavior of the two adaptive algorithms can be understood as in neither case time was spent learning about b. If, instead, the background was not as well-known, the behavior of the two adaptive algorithms differ as U spends more time learning about b out at long delays. The difference can be seen by comparing Figs. 11(a) and 11(c). However, note in Fig. 11(d) that the T_1 uncertainty is still not as low as when the background was assumed to be well known in Fig. 11(b).

The experimental results are shown in Fig. 12. In this case we assume that the background level was well known [same as in Figs. 11(a) and 11(b)]. Neither adaptive algorithm spends much time a long delays [Fig. 12(a)] and one can see that the inclusion of the expense function (pink) creates a slight preference for earlier delays. Both adaptive algorithms perform approximately equally well as expected in this case [Fig. 12(b)].

V. CONCLUSIONS AND OUTLOOK

In this work, we have demonstrated through simulations and experiments how adaptive experiment design can make dramatic improvements in the efficiency, and, thus, sensitivity, of quantum sensing with NV centers in diamond. For optimal adaptive experiments, it is important that the utility function adequately reflects the underlying physical model, real experimental conditions, and ideally, *the experimental goals*. We have introduced two new algorithms that incorporate two types of experimental goals. The nuisance parameter algorithm is able to prioritize learning about model parameters that matter the most. The expense algorithm prioritizes measurements that can be performed faster or at lower economic cost, which can lead to further improvements in efficiency. Using these algorithms, we noted an almost fivefold improvement in magnetic field sensitivity as measured with a Ramsey sequence and reached the shot noise and spin-projection limit without using a magnetometry curve, a method that typically requires initial optimization experiments and imparts limited dynamic range.

Furthermore, these methods are easily transferable to quantum sensing with other color centers, such as those in silicon carbide, that suffer from much lower spin-state contrast as these algorithms operate quite well even at an SNR of 1. They can also be used to increase the sensitivity of emerging readout techniques such as photoelectrically detected magnetic resonance.[13,30] Future work could also consider the use of these algorithms in wide-field quantum sensing where the measurement outcome and the individual parameter space is a $n \times m$ dimensional matrix, corresponding to the pixels of a camera.

Finally, we wish to emphasize that these methods are broadly applicable to other types of experiments beyond quantum sensing, although they are particularly relevant for slow or noisy experiments.

SUPPLEMENTARY MATERIAL

See the supplementary material for the additional simulations of the Hahn echo sequence where the model parameter p is de-emphasized in U^{ϕ} and $U^{\phi,\varepsilon}$ as mentioned in Sec. IV C.

ACKNOWLEDGMENTS

M.K. would like to acknowledge funding support from the National Research Council Postdoctoral Fellowship.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M. Kelley: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Software (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **R. D. McMichael:** Conceptualization (equal); Funding acquisition (supporting); Methodology (supporting); Software (equal); Writing – review & editing (supporting).

DATA AVAILABILITY

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

APPENDIX: PSEUDOCODE

ALGORITHM 1. Standard adaptive algorithm.

Input: Parameter distribution $P(\theta)$
  for N_s parameter samples θ_j drawn from $P(\theta)$ **do**
    for all candidate designs, d_i **do**
      Evaluate model $y_{i,j} \leftarrow f(\theta_j, d_i)$
                ▷ reusing θ_j samples
    end for
  end for
  for all candidate designs d_i **do**
    Variance $v_{\theta,i} \leftarrow Var(y_{i,1} \ldots y_{i,N_s})$ over parameters
    Variance of noise $v_{\eta,i} \leftarrow Var[P(\eta)]$
    Calculate utility $U_i \leftarrow \ln[1 + v_{\theta,i}/v_{\eta,i}]/2$
  end for
  Find max utility $i_{best} \leftarrow Argmax(U_i)$
Output: corresponding setting $d_{i_{best}}$

02 March 2025 10:26:28

© Author(s) 2025

ALGORITHM 2. Nuisance parameter algorithm.

Input: Parameter distribution $P(\theta)$
  for N_s parameter samples θ_j drawn from $P(\theta)$ **do**
    for all candidate designs, d_i **do**
      Evaluate model $y_{i,j} \leftarrow f(\theta_j, \phi_j, d_i)$
                                                ▷ reusing θ_j samples
    end for
  end for
  for all candidate designs d_i **do**
    Variance $v_{\theta,i} \leftarrow Var(y_{i,1} \ldots y_{i,N_s})$ over parameters
    Variance of noise $v_{\eta,i} \leftarrow Var[P(\eta)]$
    Variance of nuisance parameters $v_{\phi,i} \leftarrow Var[P(\phi)]$
    Calculate utility $U_i^{\phi} \leftarrow \ln[(v_{\theta,i} + v_{\eta,i} + v_{\phi,i})/(v_{\eta,i} + v_{\phi,i}]/2$
  end for
  Find max utility $i_{best} \leftarrow Argmax(U_i^{\phi})$
Output: corresponding setting $d_{i_{best}}$

ALGORITHM 3. Nuisance parameter + expense algorithm.

Input: Parameter distribution $P(\theta)$
  for N_s parameter samples θ_j drawn from $P(\theta)$ **do**
    for all candidate designs, d_i **do**
      Evaluate model $y_{i,j} \leftarrow f(\theta_j, \phi_j, d_i)$
                                                ▷ reusing θ_j samples
    end for
  end for
  for all candidate designs d_i **do**
    Variance $v_{\theta,i} \leftarrow Var(y_{i,1} \ldots y_{i,N_s})$ over parameters
    Variance of noise $v_{\eta,i} \leftarrow Var[P(\eta)]$
    Variance of nuisance parameters $v_{\phi,i} \leftarrow Var[P(\phi)]$
    Calculate utility $U_i^{\phi,\varepsilon} \leftarrow \ln[(v_{\theta,i} + v_{\eta,i} + v_{\phi,i})/(v_{\eta,i} + v_{\phi,i}]/(2\varepsilon_i)$
  end for
  Find max utility $i_{best} \leftarrow Argmax(U_i^{\phi})$
Output: corresponding setting $d_{i_{best}}$

02 March 2025 10:26:28

REFERENCES

[1]S. Dushenko, K. Ambal, and R. D. McMichael, "Sequential Bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers," Phys. Rev. Appl. **14**, 054036 (2020).
[2]R. D. McMichael, S. M. Blakley, and S. Dushenko, "Optbayesexpt: Sequential Bayesian experiment design for adaptive measurements," J. Res. Natl. Inst. Stan. **126**, 126002 (2021).
[3]M. Caouette-Mansour, A. Solyom, B. Ruffolo, R. D. McMichael, J. Sankey, and L. Childress, "Robust spin relaxometry with fast adaptive Bayesian estimation," Phys. Rev. Appl. **17**, 064031 (2022).
[4]K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," Stat. Sci. **10**, 273–304 (1995).
[5]X. Huan and Y. M. Marzouk, "Simulation-based optimal Bayesian experimental design for nonlinear systems," J. Comput. Phys. **232**, 288–317 (2013).
[6]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**, 128–154 (2016).
[7]M. J. Arshad, C. Bekker, B. Haylock, K. Skrzypczak, D. White, B. Griffiths, J. Gore, G. W. Morley, P. Salter, J. Smith, I. Zohar, A. Finkler, Y. Altmann, E. M. Gauger, and C. Bonato, "Real-time adaptive estimation of decoherence timescales for a single qubit," Phys. Rev. Appl. **21**, 024026 (2024).
[8]E. V. Levine, M. J. Turner, P. Kehayias, C. A. Hart, N. Langellier, R. Trubko, D. R. Glenn, R. R. Fu, and R. L. Walsworth, "Principles and techniques of the quantum diamond microscope," Nanophotonics **8**, 1945–1973 (2019).
[9]J. F. Barry, J. M. Schloss, E. Bauch, M. J. Turner, C. A. Hart, L. M. Pham, and R. L. Walsworth, "Sensitivity optimization for NV-diamond magnetometry," Rev. Mod. Phys. **92**, 015004 (2020).
[10]R. D. McMichael, S. Dushenko, and S. M. Blakley, "Sequential Bayesian experiment design for adaptive Ramsey sequence measurements," J. Appl. Phys. **130**, 144401 (2021).
[11]C. E. Rodriguez-Carranza and M. H. Loew, "Weighted and deterministic entropy measure for image registration using mutual information," Proc. SPIE **3338**, 155 (1998).
[12]R. D. McMichael and S. M. Blakley, "Simplified algorithms for adaptive experiment design in parameter estimation," Phys. Rev. Appl. **18**, 054001 (2022).
[13]M. Clyde and K. Chaloner, "The equivalence of constrained and weighted designs in multiple objective design problems," J. Am. Stat. Assoc. **91**, 1236–1244 (1996).

© Author(s) 2025

[14]A. B. Antognini and A. Giovagnoli, "Compound optimal allocation for individual and collective ethics in binary clinical trials," Biometrika **97**, 935–946 (2010).
[15]A. Giovagnoli and I. Verdinelli, "Bayesian adaptive randomization with compound utility functions," Stat. Sci. **38**, 52–67 (2023).
[16]R. Rubinas, V. V. Vorobyov, V. V. Soshenko, S. V. Bolshedvorskii, V. N. Sorokin, A. N. Smolyaninov, V. G. Vins, A. P. Yelisseyev, and A. V. Akimov, "Spin properties of NV centers in high-pressure, high-temperature grown diamond," J. Phys. Commun. **2**, 115003 (2018).
[17]P. L. Stanwix, L. M. Pham, J. R. Maze, D. Le Sage, T. K. Yeung, P. Cappellaro, P. R. Hemmer, A. Yacoby, M. D. Lukin, and R. L. Walsworth, "Coherence of nitrogen-vacancy electronic spin ensembles in diamond," Phys. Rev. B **82**, 201201 (2010).
[18]E. Bauch, C. A. Hart, J. M. Schloss, M. J. Turner, J. F. Barry, P. Kehayias, S. Singh, and R. L. Walsworth, "Ultralong dephasing times in solid-state spin ensembles via quantum control," Phys. Rev. X **8**, 031025 (2018).
[19]A. Dréau, M. Lesik, L. Rondin, P. Spinicelli, O. Arcizet, J.-F. Roch, and V. Jacques, "Avoiding power broadening in optically detected magnetic resonance of single NV defects for enhanced dc magnetic field sensitivity," Phys. Rev. B **84**, 195204 (2011).
[20]F. M. Stürner, Y. Liu, P.-O. Colard, M. Markham, and F. Jelezko, "Magnetometry based on the excited-state lifetimes of a single nitrogen-vacancy center in diamond," Appl. Phys. Lett. **119**, 134001 (2021).
[21]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).
[22]D. A. Hopper, H. J. Shulevitz, and L. C. Bassett, "Spin readout techniques of the nitrogen-vacancy center in diamond," Micromachines **9**, 437 (2018).
[23]A. F. L. Poulsen, J. D. Clement, J. L. Webb, R. H. Jensen, L. Troise, K. Berg-Sørensen, A. Huck, and U. L. Andersen, "Optimal control of a nitrogen-vacancy spin ensemble in diamond for sensing in the pulsed domain," Phys. Rev. B **106**, 014202 (2022).
[24]A. M. Edmonds, U. F. S. D'Haenens-Johansson, R. J. Cruddace, M. E. Newton, K.-M. C. Fu, C. Santori, R. G. Beausoleil, D. J. Twitchen, and M. L. Markham, "Production of oriented nitrogen-vacancy color centers in synthetic diamond," Phys. Rev. B **86**, 035201 (2012).
[25]A. Mzyk, A. Sigaeva, and R. Schirhagl, "Relaxometry with nitrogen vacancy (NV) centers in diamond," Acc. Chem. Res. **55**, 3572–3580 (2022).
[26]A. Finco, A. Haykal, R. Tanos, F. Fabre, S. Chouaieb, W. Akhtar, I. Robert-Philip, W. Legrand, F. Ajejas, K. Bouzehouane, N. Reyren, T. Devolder, J.-P. Adam, J.-V. Kim, V. Cros, and V. Jacques, "Imaging non-collinear antiferromagnetic textures via single spin relaxometry," Nat. Commun. **12**, 767 (2021).
[27]D. Schmid-Lorch, T. Häberle, F. Reinhard, A. Zappe, M. Slota, L. Bogani, A. Finkler, and J. Wrachtrup, "Relaxometry and dephasing imaging of superparamagnetic magnetite nanoparticles using a single qubit," Nano Lett. **15**, 4942–4947 (2015).
[28]J.-P. Tetienne, T. Hingant, L. Rondin, A. Cavailles, L. Mayer, G. Dantelle, T. Gacoin, J. Wrachtrup, J.-F. Roch, and V. Jacques, "Spin relaxometry of single nitrogen-vacancy defects in diamond nanocrystals for magnetic noise sensing," Phys. Rev. B **87**, 235436 (2013).
[29]I. Cardoso Barbosa, J. Gutsche, and A. Widera, "Impact of charge conversion on NV-center relaxometry," Phys. Rev. B **108**, 075411 (2023).
[30]C. Degen, F. Reinhard, and P. Cappellaro, "Quantum sensing," Rev. Mod. Phys. **89**, 035002 (2017).
[31]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**, 810–816 (2008).
[32]B. Smeltzer, J. McIntyre, and L. Childress, "Robust control of individual nuclear spins in diamond," Phys. Rev. A **80**, 050302 (2009).
[33]Z.-H. Wang and S. Takahashi, "Spin decoherence and electron spin bath noise of a nitrogen-vacancy center in diamond," Phys. Rev. B **87**, 115122 (2013).
[34]G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, "High-dynamic-range magnetometry with a single nuclear spin in diamond," Nat. Nanotechnol. **7**, 105–108 (2012).
[35]E. Bauch, S. Singh, J. Lee, C. A. Hart, J. M. Schloss, M. J. Turner, J. F. Barry, L. M. Pham, N. Bar-Gill, S. F. Yelin, and R. L. Walsworth, "Decoherence of ensembles of nitrogen-vacancy centers in diamond," Phys. Rev. B **102**, 134210 (2020).
[36]A. Norambuena, E. Muñoz, H. T. Dinani, A. Jarmola, P. Maletinsky, D. Budker, and J. R. Maze, "Spin-lattice relaxation of individual solid-state spins," Phys. Rev. B **97**, 094304 (2018).
[37]M. Mrózek, D. Rudnicki, P. Kehayias, A. Jarmola, D. Budker, and W. Gawlik, "Longitudinal spin relaxation in nitrogen-vacancy ensembles in diamond," EPJ Quantum Technol. **2**, 1–11 (2015).
[38]A. Jarmola, V. M. Acosta, K. Jensen, S. Chemerisov, and D. Budker, "Temperature- and magnetic-field-dependent longitudinal spin relaxation in nitrogen-vacancy ensembles in diamond," Phys. Rev. Lett. **108**, 197601 (2012).
[39]Z. Wang, C. McPherson, R. Kadado, N. Brandt, S. Edwards, W. Casey, and N. Curro, "ac sensing using nitrogen-vacancy centers in a diamond anvil cell up to 6 GPa," Phys. Rev. Appl. **16**, 054014 (2021).
[40]C. T.-K. Lew, V. K. Sewani, N. Iwamoto, T. Ohshima, J. C. McCallum, and B. C. Johnson, "Enhanced magnetometry with an electrically detected spin defect ensemble in silicon carbide," Appl. Phys. Lett. **122**, 234001 (2023).
[41]E. Bourgeois, M. Gulka, and M. Nesladek, "Photoelectric detection and quantum readout of nitrogen-vacancy center spin states in diamond," Adv. Opt. Mater. **8**, 1902132 (2020).

02 March 2025 10:26:28

© Author(s) 2025