Swarm optimization for adaptive phase measurements with low visibility

Alexander J. F. Hayes* and Dominic W. Berry
Department of Physics and Astronomy, Macquarie University, Sydney, New South Wales 2109, Australia
(Received 18 November 2013; published 27 January 2014)

Adaptive feedback normally provides the greatest accuracy for optical phase measurements. New advances in nitrogen-vacancy-center technology have enabled magnetometry via individual spin measurements, which are similar to optical phase measurements but with low visibility. The adaptive measurements that previously worked well with high-visibility optical interferometry break down and give poor results for nitrogen-vacancy-center measurements. We use advanced search techniques based on swarm optimization to design better adaptive measurements that can provide improved measurement accuracy with low-visibility interferometry, with applications in nitrogen-vacancy-center magnetometry.

DOI: 10.1103/PhysRevA.89.013838 PACS number(s): 42.50.Dv, 81.05.ug, 07.55.Ge, 42.50.St

I. INTRODUCTION

An important feature of quantum mechanics is that it imposes fundamental limits on how accurately physical quantities can be measured. On the other hand, by taking advantage of the features of quantum mechanics, one can perform measurements that are far more accurate than would otherwise be possible. Many types of precision measurement use a form of interferometry. One particular example is optical interferometry, which is used to measure distance via changes in phase. In this case, if one uses N independent photons, then the accuracy of the phase measurement scales as $1/\sqrt{N}$. On the other hand, if one takes advantage of quantum mechanics by using N photons in a special entangled state, then the accuracy can scale as $1/N$, yielding vastly higher accuracy for large N [1–3].

A commonly considered entangled state is the NOON state, where N photons are in a superposition of being all in one arm of the interferometer or the other [4–6]. These states have the advantage that they provide accuracy scaling as $1/N$, with the drawback that the phase needs to be initially known to this accuracy; otherwise, the measurement is ambiguous. One way of interpreting the accuracy is that m measurements with m copies of this state will yield accuracy $1/(N\sqrt{m})$. An alternative approach is to combine measurements using NOON states with many different values of N. The measurements with smaller values of N are used to resolve the ambiguity in the measurements with larger values of N. This is the approach used in Refs. [7,8]. Instead of considering NOON states, one can instead consider multiple passes through a phase shift [9].

Another area of interferometry is that using transitions in atomic or solid-state systems. This can be used for time standards or frequency measurement or probing physical quantities that affect the frequency. The particular application we consider here is the use of a single nitrogen-vacancy (NV) center in diamond to probe the magnetic field at the nanoscale [10–15]. Experiments in such systems have demonstrated sensitivity of ∼3 nT [11] and spatial resolution of ∼5 nm [12]. The NV center has excellent properties for magnetic sensing because it can be individually addressed and maintains spin coherence for a significant period of time at room temperature.

A fundamental difficulty with these measurement techniques is that the measurement signal has periodic modulation. The measurement time is either restricted to half an oscillation period, or the magnetic field range must be known accurately in advance. This is essentially the same problem that occurs when performing phase measurements with NOON states, and it is therefore natural to apply the same measurement schemes as were developed for NOON states. This is what was proposed in Ref. [16], and it was experimentally demonstrated in Refs. [17–20]. For this system the interpretation is that the measurements provide a high dynamic range, rather than a quantum improvement as in the case of NOON states.

In optical interferometry, it has been found that adaptive measurements are typically the most accurate [7,9,21–25]. That is, information from early parts of the measurement are used to adjust how the measurement is performed. In fact, for measurements on a single mode, nonadaptive measurements are unable to achieve better than $1/\sqrt{N}$ scaling even for highly nonclassical states [26]. In contrast, for measurements on multiple time modes, such as measurements on multiple NOON states, nonadaptive measurements can yield the same accuracy scaling as adaptive measurements [8,27].

An important difference between optical measurements and measurements with NV centers is that the NV-center measurements have lower visibility. There is both a reduced initial visibility, and the visibility decreases exponentially with interaction time due to decoherence. In the experiment in Ref. [17] the initial visibility is only about 80%, whereas visibilities in optical interferometry are often 98% or better [9]. It turns out that when the visibility is this low, the adaptive measurement schemes that have been developed in previous work become very inaccurate [16]. That is why nonadaptive measurements were used in Refs. [17,18]. Even the nonadaptive measurements have somewhat poor performance for low visibility. It should be possible for some adaptive method to provide improved performance because adaptive measurements are more general than nonadaptive measurements.

The adaptive techniques that were considered for NV-center magnetometry before were those based on the technique from Refs. [21,22]. This technique only optimizes the measurement locally, in the sense that it minimizes the

*alexander.hayes@uqconnect.edu.au

variance after the next detection. An alternative technique is to use global optimization to minimize the final variance at the end of the measurement. Such an optimization is far more challenging, but particle swarm optimization techniques have been found to be effective for single-time-mode adaptive measurements [28,29]. Particle swarm optimization has also been used to design nonadaptive measurements for the related task of Hamiltonian parameter estimation [30]. An alternative technique for measurements with multiple NOON states of different sizes was proposed by Cappellaro [31]. Here we combine particle swarm optimization techniques with the method of Cappellaro to find adaptive measurements that provide improved performance for NV-center magnetometry.

II. RAMSEY INTERFEROMETRY

In this section we summarize Ramsey interferometry and how it is used with NV centers. The probe state is prepared in the superposition state $(|0\rangle + |1\rangle)/\sqrt{2}$. Given that the energy-level splitting is ΔE, the state evolves over time t to $(|0\rangle + e^{-it\Delta E/\hbar}|1\rangle)/\sqrt{2}$. In the case we are interested in, the energy-level splitting is due to the different spin states and is proportional to the magnetic field. Therefore the state is $(|0\rangle + e^{-2it\gamma B}|1\rangle)/\sqrt{2}$, where γ is the gyromagnetic ratio. For readout, the state is measured in the basis $(|0\rangle \pm |1\rangle)/\sqrt{2}$. The probabilities of the measurement results are then

$$P(\pm|B) = \tfrac{1}{2}[1 \pm \cos(2t\gamma B)]. \tag{1}$$

The negatively charged NV center, denoted NV^-, has an energy-level diagram as shown in Fig. 1. There are two sets of energy levels, ^{3}A and ^{3}E, labeled by the irreducible representations of the symmetry group of the defect center. There are three allowable electronic spin states for each, with $m_s = 0$ and ± 1. The electronic spin states $m_s = \pm 1$ for ^{3}A were used in Ref. [18]. In the case of ^{14}N, the energy level $m_s = 0$ for ^{3}A is further split into nuclear energy levels $m_I = 0$ and ± 1. The nuclear spin states $m_I = 0$ and -1 were used in Ref. [17]. These were then mapped to electronic spin states for readout.

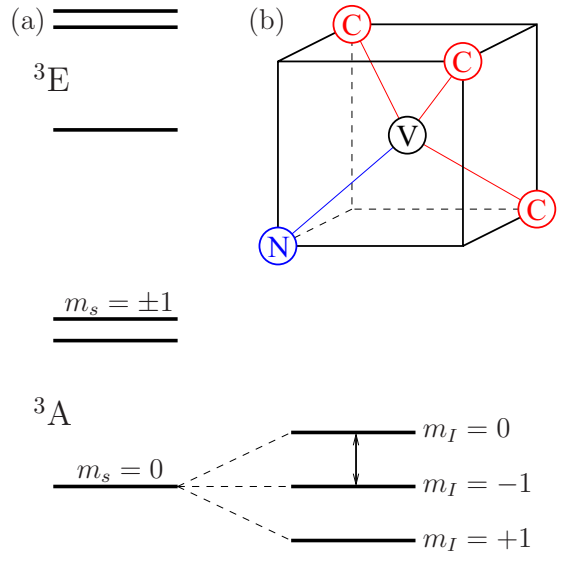

FIG. 1. (Color online) The energy-level structure of an NV center. (a) There are two sets of energy levels, ^{3}A and ^{3}E. We consider the electronic spin state $m_s = 0$ of ^{3}A, which has the further hyperfine splitting into $m_I = 0$ and ± 1. (b) The structure of the NV center, consisting of a substitutional nitrogen impurity N (blue circle) adjacent to a vacancy V (black circle), tetrahedrally arranged with respect to each other and to the nearby carbon atoms C (red circles).

In order to obtain the superposition state, Ref. [17] initialized the system in the $m_I = 0$ state, then used a $\pi/2$ pulse to obtain a superposition of $m_I = 0$ and -1. After allowing the system to accumulate a phase depending on the magnetic field, another $\pi/2$ pulse was used; then the system was measured in the $m_I = 0$ and -1 basis. The advanced phase estimation algorithms developed in optics depend on using a controllable phase that is either controlled in a predetermined way (in the nonadaptive case) or based on the results of measurements (in the adaptive case). In optics this controlled phase can be implemented by putting a phase modulator in one arm of an interferometer. In the interferometry experiment in Ref. [17] this controlled phase was implemented by changing the phase of the second $\pi/2$ pulse. By changing the phase of the pulse, the system could be rotated around the x or y axis of the Bloch sphere.

In general, consider a rotation of the form

$$R(\theta) = \exp[i(\sigma_x \sin\theta + \sigma_y \cos\theta)\pi/4]. \tag{2}$$

Performing this rotation followed by a measurement yields the probabilities

$$P(\pm|B) = \tfrac{1}{2}[1 \pm \cos(2t\gamma B - \theta)]. \tag{3}$$

Hence this technique provides a controlled phase that is mathematically equivalent to that used in optics.

For the phase measurement technique of Refs. [16–18], the times used are multiples of a base time τ. It is therefore convenient to define a phase

$$\phi := 2\gamma B\tau. \tag{4}$$

The phase shifts will be of the form $2^k\phi$. No multiples of ϕ smaller than 1 will be considered, so ϕ is measured in the range $(-\pi,\pi]$. The corresponding range for measurement of B is $(-B_{\max}, B_{\max}]$, where

$$B_{\max} := \frac{\pi}{2\gamma\tau}. \tag{5}$$

An important feature of Ramsey interferometry is the visibility of the interference. It starts significantly less than 100%, then exponentially decays with time due to decoherence. Taking this into account, the probability distribution for the measurement results is [16]

$$P(\pm|\phi) = \tfrac{1}{2}\left[1 \pm f_d e^{-2^k\tau/T_2}\cos(2^k\phi - \theta)\right], \tag{6}$$

where f_d is a factor representing the visibility of the measurement and T_2 is the transverse spin coherence time. The time T_2 indicates the decay rate of spin coherence in the system and sets an effective limit on the maximum duration of measurements on the system.

If one were to just use all measurements with the same interaction time τ, then for a *total* interaction time T the phase

variance would be no less than $1/N$, where $N := T/\tau$. This corresponds to an uncertainty in B lower bounded as

$$\Delta B \geqslant \frac{1}{2\gamma\tau\sqrt{N}}. \tag{7}$$

We can reduce the variance by increasing τ, but that also reduces the range of the measurement. The dynamic range (the ratio of the maximum magnetic field that can be detected to the uncertainty) is upper bounded as

$$\frac{B_{\max}}{\Delta B} \leqslant \pi\sqrt{N}. \tag{8}$$

In contrast, when we use multiple interaction times, the phase variance is lower bounded as approximately $(\pi/N)^2$ [the exact lower bound is in Eq. (15)]. The corresponding upper bound to the dynamic range is then

$$\frac{B_{\max}}{\Delta B} \lesssim N. \tag{9}$$

That is, we potentially obtain a square improvement in the dynamic range by using measurement techniques with multiple interaction times. This is analogous to the improvement from the standard quantum limit to the Heisenberg limit with NOON states.

III. MEASUREMENT TECHNIQUES AND ANALYSIS

A. Bayesian estimation

In order to determine estimates of the phase (and thereby the magnetic field), it is convenient to use Bayesian estimation to calculate the probability distribution for the system phase based on successive measurement results [21,22]. This enables one to calculate the exact phase variance for a given measurement scheme [21,22]. It is assumed that the magnetic field is initially unknown, other than being confined to the range $(-B_{\max}, B_{\max}]$. This means that there is no initial knowledge of the phase, so the prior distribution is

$$P(\phi|\vec{u}_0) = \frac{1}{2\pi}. \tag{10}$$

Here the notation that is used is that the successive measurement results are u_1, u_2, and so forth, and a vector of n measurement results is $\vec{u}_n := (u_1, \ldots, u_n)$. The initial vector of zero length before any measurement results are taken is $\vec{u}_0$.

Each interferometric measurement provides additional information about the phase ϕ and therefore the magnetic field. This information is quantified by using Bayes' rule to update the probability distribution as

$$P(\phi|\vec{u}_n) \propto P(u_n|\phi)P(\phi|\vec{u}_{n-1}), \tag{11}$$

where u_n is the outcome of the most recent measurement. Note that the denominator of the Bayesian function has not been explicitly given here as it is independent of the measurement outcome and serves only as a normalizing factor for the probability distribution. The formula to use for $P(u_n|\phi)$ is given in Eq. (6).

B. The Fourier series representation

As the Bayesian probability distribution is a product of sinusoids, it is convenient to use the Fourier series of the probability distribution function to represent it. We write the general form of the probability distribution as

$$P(\phi) = \sum_{w=-\infty}^{\infty} b_w e^{iw\phi}. \tag{12}$$

Note that, in practice, the sum need not be taken to infinity because the probability distribution is a product of a limited number of sinusoids.

Combining the representation above with Eq. (6), it is possible to derive the simple update rule for the coefficients of the Fourier series

$$b_w^n = \frac{1}{2}b_w^{n-1} + \frac{u_n V}{4}b_{w-2^k}^{n-1}e^{-i\theta} + \frac{u_n V}{4}b_{w+2^k}^{n-1}e^{i\theta}, \tag{13}$$

where $V := f_d e^{-2^k\tau/T_2}$. Note that this formula is only applicable when the system is restricted to integer multiples of the interaction time, which will be the case for the metrology methods considered within this paper. Numeric simulations of the phase estimation process may be performed by tracking the coefficients of the probability distribution and updating them for each measurement using the formula given above.

To estimate the accuracy of the phase measurement, it is convenient to use the Holevo variance [32]

$$V_H := \frac{1}{|\langle e^{i(\hat{\phi}-\phi)}\rangle|^2} - 1, \tag{14}$$

where $\hat{\phi}$ is the estimate of the phase. To account for biased estimates of the phase, one can use $\langle\cos(\hat{\phi}-\phi)\rangle$ in place of $\langle e^{i(\hat{\phi}-\phi)}\rangle$. The Holevo variance is close to the usual variance for narrowly peaked distributions. For phase an advantage of the Holevo variance is that it is naturally modulo 2π. That is, a phase close to $-\pi$ is regarded as also close to π. In the case where the phase is used for measuring the magnetic field, this is a little problematic because an estimate near $-B_{\max}$ would be regarded as accurate if the actual magnetic field is near $B_{\max}$.

To avoid the problem of having a large error in the estimate of the field near $\pm B_{\max}$, one can consider an initial probability distribution for the magnetic field that is in the range $(-B'_{\max}, B'_{\max}]$, where $B'_{\max} = B_{\max} - c\Delta B$. Here c is a constant chosen such that the probability of the error (in the magnetic field estimate) being larger than $c\Delta B$ is negligible. Then the actual dynamic range would be $B'_{\max}/\Delta B$, but this only differs from $B_{\max}/\Delta B$ by a constant. We do not consider that correction in this work because we consider a large dynamic range where it is negligible.

Another feature of the Holevo variance is that it means only one coefficient of the Fourier series for the probability distribution is important. First, the optimal phase estimate to use is $\hat{\phi} = \arg(b_{-1})$ [22]. Second, the Holevo variance can be determined exactly by summing the value of $|b_{-1}|$ over all combinations of measurement results. See Ref. [27] for an explanation of how this is done. It is therefore unnecessary to track all nonzero coefficients, and we can restrict ourselves to recording those that contribute to the final value of b_{-1}. This considerably simplifies the calculation. The calculation can be further simplified by noting that the probability distribution is real, so $b_{-w} = b_w^*$, and only half the coefficients need be stored.

The lower bound to the Holevo variance is [21,33]

$$V_H \geqslant \tan^2\left[\frac{\pi}{N+2}\right]. \tag{15}$$

This bound is slightly less than $(\pi/N)^2$, but it approaches $(\pi/N)^2$ in the limit of large N. The lower bound $1/N$ given in Sec. II for measurements with equal interaction times was also in terms of V_H. In that case it is not a tight lower bound, and there does not appear to be a known analytic expression for the tight bound.

C. Phase measurement protocols

There are several different aspects of the phase estimation procedure that can be controlled with the aim of achieving the best possible estimate. One can obtain improved performance with entangled states; for example, the equivalent of optical NOON states would be N entangled NV centers. Entangled NV centers have been demonstrated [34] but are not yet at the stage where they can effectively be used for magnetometry. For this reason we consider the improvement of the choice of interaction time and controlled phase θ.

The procedure to perform the phase is to begin with the longest viable interaction time and then systematically reduce it by a factor of 2, performing multiple repetitions for each interaction time. In the following we use the terminology "detection" for the individual measurements, to distinguish them from the overall measurement based on the combination of these individual detections. The initial interaction time is $2^K\tau$, and the sequence of multipliers is $2^K, 2^{K-1}, \ldots, 2^1, 2^0$.

The number of detections used for the longest interaction time is denoted G, and each time the interaction time is halved, the number of detections is increased by F. Therefore, for interaction time $2^k\tau$, the number of detections is

$$M_k = G + F(K-k). \tag{16}$$

The total interaction time used is then [16]

$$T = \tau[G(2^{K+1}-1) + F(2^{K+1}-2-K)]. \tag{17}$$

This sequence with varying numbers of detections was introduced in Ref. [8] and used in Refs. [16–18,20].

Although it was found to be possible to take $F = 0$ for adaptive measurements [9], for nonadaptive measurements it is necessary to take $F > 0$ [8]. Reference [16] found that the nonadaptive measurements gave better results than the adaptive technique of Ref. [9] for poor visibility. Nevertheless, if the initial visibility was too low, the nonadaptive measurements were still quite poor. The experiments in Refs. [17,18] addressed this by increasing the number of detections for each interaction time. Reference [17] used $G = 36$ and $F = 8$, whereas Ref. [18] used $G = F = 9$.

The key difference in our approach is how we adjust the controlled phase θ based on the measurement results. In general, the strategy for choosing the controlled phase based on measurement results can be described by a binary decision tree, as in Fig. 2. The approach used in Ref. [28] was to choose the size of the step dependent on the detection result and the number of previous detections.

A more common approach is to update the controlled phase based on the estimate of the system phase. This approach

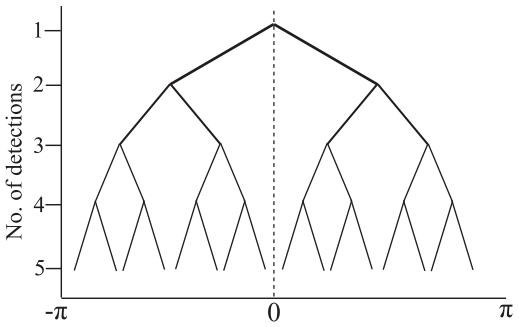

FIG. 2. An illustration of a basic symmetric protocol decision tree. The horizontal axis indicates the phase offset from its initial setting.

is essentially an adaptive homodyne measurement, similar to Refs. [24,35]. Here, after detections with interaction times down to $2^k\tau$, the phase information has only been obtained modulo $2\pi/2^k$. This means that the only nonzero Fourier coefficients b_w are those where $w \mod 2^k = 0$. This is advantageous for computation because it means that only those coefficients need be recorded. However, it means that one cannot use b_{-1} for an estimate of the phase (unless $k = 0$). Instead, one can use b_{-2^k} for an estimate of the phase modulo $2\pi/2^k$; that is, it is an estimate of $2^k\phi$. Taking θ to be an estimate of $2^k\phi$ plus $\pi/2$ yields the point where the probabilities in Eq. (6) are most sensitive to ϕ. One can also choose the controlled phase to minimize the phase variance after the next detection [9,21,22,27].

Nevertheless, this adaptive approach performs very poorly with reduced visibility of the interference. Recently, a variation was proposed by Cappellaro [31] for this type of phase estimation, in which the controlled phase is updated only after each change of interaction time. That is, when the interaction time is changed from $2^k\tau$ to $2^{k-1}\tau$, the controlled phase is taken to be

$$\theta = \tfrac{1}{2}\arg(b_{-2^k}). \tag{18}$$

An estimate of $2^k\phi$ is obtained as $\arg(b_{-2^k})$. This is then divided by 2 to obtain an estimate of $2^{k-1}\phi$, which is needed because one has $2^{k-1}\phi - \theta$ in the argument of the cosine in Eq. (6).

Note that there is ambiguity modulo π in the estimate of $2^{k-1}\phi$, but this is unimportant because adding π yields equivalent results. Specifically, adding π exchanges the probabilities for detection results $u = +1$ and $u = -1$. Hence, if π is added and the detection result is $u = -1$, the following calculations are the same as if π were not added and the detection result were $u = +1$. This means that the same variance must be obtained regardless of whether π is added.

Formula (18) does not have $\pi/2$ in it, so it does not give the point where the probabilities are most sensitive to ϕ. However, it turns out that when $\pi/2$ is not used, this is the phase that minimizes the variance after the next detection (after the change in interaction time). This can be verified using the formula in Ref. [22]. Note that this phase is locally optimal, in that it minimizes the variance after the next detection, but is not (necessarily) globally optimal because it need not minimize the variance at the end of the measurement.

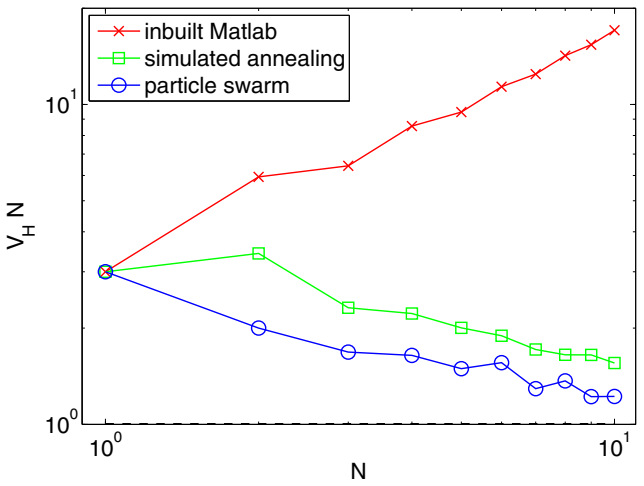

FIG. 3. (Color online) A performance comparison of different optimization methods for an example simulation. The methods shown are Matlab's inbuilt nonlinear programming suite, a simulated annealing approach, and the particle swarm algorithm. Each was given the same runtime for a phase estimation simulation task.

IV. OPTIMIZATION ALGORITHMS AND TECHNIQUES

To find an improved adaptive protocol for this type of phase estimation system, we applied numerical optimization to minimize the variance estimated by repeated simulations. Several different optimization algorithms were tested, including nonlinear programming methods and simulated annealing; however, we found that the procedure which gave the best result in a reasonable amount of time was particle swarm optimization (PSO), which was also used in [28,29].

A comparison of some example results from the three different techniques is shown in Fig. 3. These calculations were for all the same interaction time τ. Because N is the total interaction time divided by τ, here it is the number of detections. For this example T_2 was taken to be infinite for simplicity. Approximately the same calculation times were allowed in each case. Matlab's inbuilt optimization routine produced the largest variances. Simulated annealing produced somewhat improved results, and PSO produced the smallest variance.

The PSO algorithm uses a group of interacting "particles" to search the phase space of the problem. Each particle has a velocity which is updated with semirandom adjustments that are dependent on the best values found by both the individual particle and the entire group. The formula for the velocity update is

$$\vec{v}' = \chi[\vec{v} + c_g r_g(\vec{x}_g - \vec{x}) + c_l r_l(\vec{x}_l - \vec{x})], \qquad (19)$$

where $\vec{x}$ is the particle position, $\vec{v}$ is the particle velocity, and $\vec{v}'$ is the updated velocity. The variables $\vec{x}_g$ and $\vec{x}_l$ are the currently known global and local optimums, respectively, χ is an overall damping factor to ensure convergence, c_g and c_l are weightings for the current global and local optimums, respectively, and r_g and r_l are random numbers that are chosen uniformly in the interval [0,1) at each step. The quantities χ, c_g, and c_l are constant throughout the procedure; we used $\chi = 0.729$ and $c_g = c_l = 2.05$. These are the weighting values recommended by Kennedy and Eberhart [36], the developers of the algorithm, and we found these to perform relatively well.

The space consisted of the phase increments after each detection. That is, after each detection there was an increment in the controlled phase θ that depended on the number of the stage and the detection result but not on prior results. The position is the set of these phase increments. The initial positions were chosen uniformly at random in the space. The initial velocities were chosen uniformly at random, with the maximum velocity corresponding to half the size of the space per iteration. The maximum velocity tended to act as a soft bound on the area covered, as the particles would usually stay within an area of dimensions that were roughly 3 times the maximum velocity.

We used fixed boundaries since we were examining a domain equivalent to one full phase rotation. The phase space had reflective boundaries to prevent a trapping effect near the edges. The simulations used ten particles; we tested higher numbers of particles, but that did not improve the performance. The maximum number of iterations was set at 300 to allow the particles to fully converge.

V. SIMULATION AND OPTIMIZATION OF MEASUREMENT PROTOCOLS

We ran simulations to compare the performance of several different protocols. In each case, we used a sample size of at least 2^{16} (65 536), which is intended to be sufficiently large to ensure accurate results while still being tractable for the computation equipment we used. In order to reduce the random variation just due to the sampling, we used the same set of random numbers each time. At the end we used a different set of random numbers to obtain the final estimate of the variance. This is to ensure that the small variance was not just an artifact of the particular random numbers used.

For our optimized adaptive stepping, the optimization program allowed for asymmetric steps, meaning that the size and direction of a step after a detection outcome are independent of the step resulting from the alternative outcome. This also means that we were required to optimize over twice as many variables.

A further consideration was the number of detections for each interaction time. For most of the results, we used the values $G = 6$ and $F = 2$, which were found to be sufficiently large to give improved scaling at high visibility while also being useful for comparison with results from other work, such as [16]. Smaller detection numbers were found to perform poorly for all protocols.

The results are shown in Figs. 4 and 5, with panels (a) and (b) corresponding to initial visibilities of $f_d = 95\%$ and $f_d = 85\%$, respectively. In Figs. 4 and 5 we took $T_2 = 10^3\tau$. In Fig. 4, the results for nonadaptive measurements are shown, as used in Ref. [16]. For initial visibility of $f_d = 95\%$ the product $V_H N$ decreases with N, whereas for $f_d = 85\%$ the product $V_H N$ slightly increases with N. These results are similar to those in Ref. [16] (see Fig. 3 of that work).

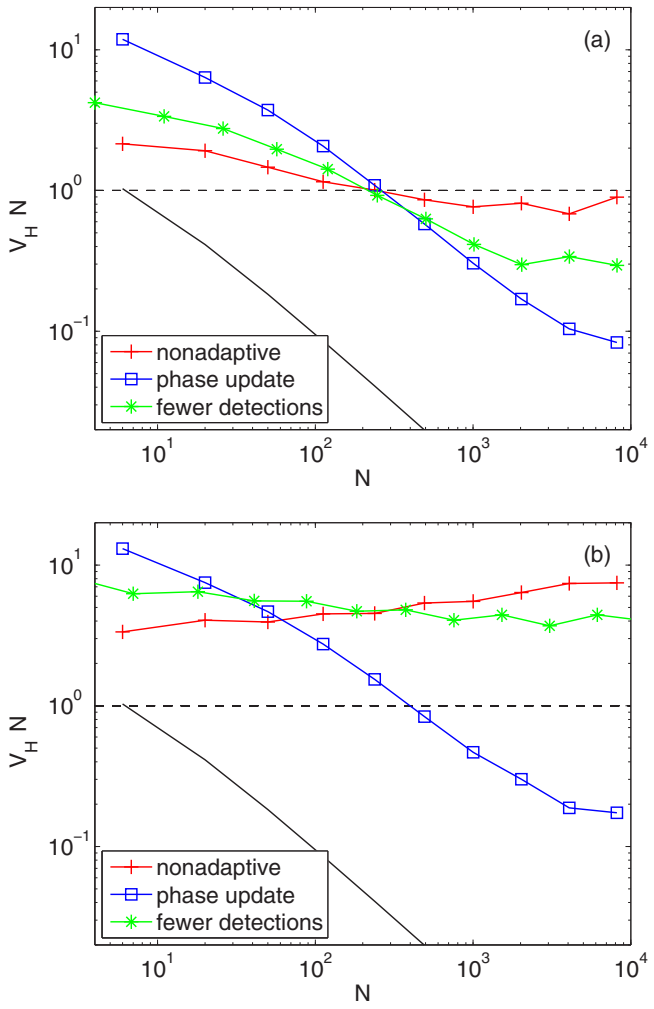

FIG. 4. (Color online) A comparison of variances for different methods vs N (the ratio of the total interaction time to τ). Shown are lines for a nonadaptive decision-tree protocol, a phase update protocol based on Cappellaro's work, and a phase update protocol with a reduced number of detections ($G = 2$, $F = 1$). The limit for a single interaction time is shown as the dashed line, and the limit for multiple interaction times is shown as a solid line. (a) and (b) have initial visibilities of $f_d = 95\%$ and $f_d = 85\%$, respectively. All protocols shown other than the reduced-detection phase update protocol use $G = 6$ and $F = 2$.

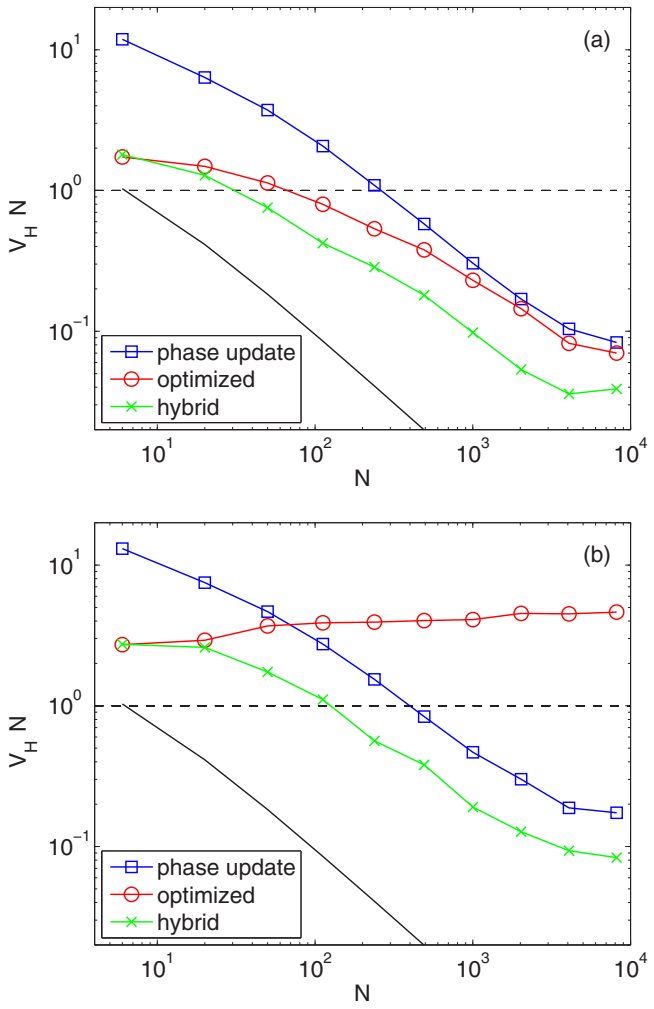

FIG. 5. (Color online) A comparison of variances for different methods vs N (the ratio of the total interaction time to τ). Shown are lines for a phase update protocol based on Cappellaro's work, an optimized decision-tree protocol, and a hybrid protocol combining an optimized decision-tree with phase updates. The limit for a single interaction time is shown as the dashed line, and the limit for multiple interaction times is shown as a solid line. (a) and (b) have initial visibilities of $f_d = 95\%$ and $f_d = 85\%$, respectively.

Note that V_H is inversely proportional to the square of the dynamic range; that is,

$$V_H = \pi^2 \left(\frac{\Delta B}{B_{\max}} \right)^2 . \tag{20}$$

This means that the limit to the dynamic range for equal measurement times (8) corresponds to $V_H N$ being constant. Similarly, the ultimate limit to the dynamic range in (9) corresponds to $V_H N$ scaling as $1/N$. These limits are shown in Figs. 4 and 5 for reference. It can be seen that for the nonadaptive measurement the results are quite similar to what could be obtained for equal interaction times.

An improved result is obtained by only updating the controlled phase each time the interaction time is reduced, according to the Cappellaro approach discussed above. This yields a result that beats the equal-interaction-time limit for visibilities of both 95% and 85%. It has a scaling quite similar to the ultimate limit, although it is a significant distance above the line (about a factor of 10). The alternative values $G = 2$, $F = 1$, which yield fewer detections for each interaction time, give much higher variances.

The phase variances obtained using the PSO method to choose the controlled phases are shown in Fig. 5. This improves on the phase update based on Cappellaro's technique for a visibility of 95% but not for 85% visibility. To improve the performance, we tested a combined protocol using both optimized adaptive steps and a phase update based on the Bayesian distribution at each change in interaction time. This method outperformed all others for both visibilities examined.

To be more specific about how this hybrid protocol works, whenever the interaction time is changed, the controlled phase is updated according to the Cappellaro formula (18).

Otherwise, the controlled phase is adjusted according to step sizes that we search for numerically. This technique should provide a variance that is no larger than Cappellaro's technique because Cappellaro's technique is obtained if the other step sizes are zero. Optimizing the other step sizes should only decrease the variance (unless there was poor convergence of the optimization).

It might at first be expected that the hybrid protocol should perform no better than the optimized protocol. However, the Cappellaro formula depends on all previous detection results, whereas the optimization provides a time step depending on only the most recent detection result. This means that the protocol using only optimized adaptive steps will not be able to provide the same controlled phases as the hybrid protocol. It is likely that this is the reason for the improved performance of the hybrid protocol, although it is also possible that there was poor convergence of the optimization for the protocol using only optimized adaptive steps. On the other hand, the hybrid protocol is not guaranteed to give better performance. This is because the Cappellaro formula is not globally optimal; that is, it does not minimize the final phase variance at the end of the measurement.

Although this hybrid protocol shows scaling similar to the fundamental limit for moderate values of N, it has an upturn for the largest values of N. This is similar to the result found in Ref. [16] and is again due to the fixed coherence time T_2. The largest value of N shown is 8164, which corresponds to the longest interaction time being 512τ. In comparison, the coherence time was $T_2 = 10^3\tau$, so the upturn occurs where the longest interaction time is comparable to T_2. This is similar to the result in Ref. [16]. Hence, although our technique can provide improved performance for reduced *initial* visibility, it is still limited by the coherence time, which is a fundamental limitation.

VI. CONCLUSIONS

Adaptive measurements usually give more accurate results than nonadaptive measurements, but previous adaptive techniques gave poor results for measurements with reduced visibility. The nonadaptive measurements gave better results, but when the initial visibility is too small, even nonadaptive measurements perform poorly. A partially adaptive scheme was proposed by Cappellaro, which provided some improvement [31].

Here we have used PSO to find a hybrid technique that uses the adaptive step from Cappellaro as well as optimizing the other adaptive steps in the measurement. This technique substantially improves on the technique of Cappellaro, as well as the nonadaptive technique and the earlier adaptive technique from Ref. [9]. As a result, this hybrid technique should give a substantially improved result when applied to NV-center magnetometry.

This numerical optimization is relatively time-consuming, but this would not cause a problem for experiments because it can be performed in advance. That is, the numerical optimization provides a lookup table, and during the experiment one would call up this lookup table, which can be performed very rapidly. It would be necessary to calculate the adaptive step of Cappellaro (18) in real time, but that can also be performed very rapidly.

An alternative approach is to use PSO to search for an adaptive technique without also using the Cappellaro update step. It was found that this gives poorer results than the hybrid technique. For an initial visibility of 95%, it still gave an improvement over the Cappellaro technique, but it performed very poorly for 85% visibility.

For these results we have not optimized over the number of detections for each interaction time. There is a possibility that modifying the number of detections for each interaction time may yield a result that is further improved. Further work remains to be done to determine the best numbers of detections to use and the dependence of this number on the visibility.

ACKNOWLEDGMENTS

We would like to acknowledge helpful discussions with Jason Twamley and Alexandr Sergeevich. D.W.B. is funded by an ARC Future Fellowship (FT100100761).

[1] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[2] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[3] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[4] B. C. Sanders, Phys. Rev. A **40**, 2417 (1989).
[5] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).
[6] H. Lee, P. Kok, and J. P. Dowling, J. Mod. Opt. **49**, 2325 (2002).
[7] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde, Nat. Photon. **5**, 43 (2011).
[8] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, New J. Phys. **11**, 073023 (2009).
[9] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature (London) **450**, 393 (2007).
[10] J. M. Taylor, P. Cappellaro, L. Childress, L. Jiang, D. Budker, P. R. Hemmer, A. Yacoby, R. Walsworth, and M. D. Lukin, Nat. Phys. **4**, 810 (2008).
[11] 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, Nature (London) **455**, 644 (2008).
[12] 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, Nature (London) **455**, 648 (2008).
[13] C. L. Degen, Appl. Phys. Lett. **92**, 243111 (2008).

[14] S. Steinert, F. Dolde, P. Neumann, A. Aird, B. Naydenov, G. Balasubramanian, F. Jelezko, and J. Wrachtrup, Rev. Sci. Instrum. **81**, 043705 (2010).
[15] P. Maletinsky, S. Hong, M. S. Grinolds, B. Hausmann, M. D. Lukin, R. L. Walsworth, M. Loncar, and A. Yacoby, Nat. Nanotechnol. **7**, 320 (2012).
[16] R. S. Said, D. W. Berry, and J. Twamley, Phys. Rev. B **83**, 125410 (2011).
[17] G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, Nat. Nanotechnol. **7**, 105 (2012).
[18] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, Nat. Nanotechnol. **7**, 109 (2012).
[19] T. Häberle, D. Schmid-Lorch, K. Karrai, F. Reinhard, and J. Wrachtrup, Phys. Rev. Lett. **111**, 170801 (2013).
[20] N. M. Nusran and M. V. G. Dutt, Phys. Rev. B. **88**, 220410(R) (2013).
[21] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett. **85**, 5098 (2000).
[22] D. W. Berry, H. M. Wiseman, and J. K. Breslin, Phys. Rev. A **63**, 053804 (2001).
[23] D. W. Berry and H. M. Wiseman, Phys. Rev. A **63**, 013813 (2000).
[24] H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, Science **337**, 1514 (2012).
[25] N. Spagnolo, L. Aparo, C. Vitelli, A. Crespi, R. Ramponi, R. Osellame, P. Mataloni, and F. Sciarrino, Sci. Rep. **2**, 862 (2012).
[26] H. M. Wiseman and R. B. Killip, Phys. Rev. A **56**, 944 (1997).
[27] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A **80**, 052114 (2009).
[28] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **104**, 063603 (2010).
[29] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **107**, 233601 (2011).
[30] A. Sergeevich and S. D. Bartlett, in *Proceedings of 2012 IEEE Conference on Evolutionary Computation (CEC 2012)* (IEEE, Piscataway, NJ, 2012), p. 3092.
[31] P. Cappellaro, Phys. Rev. A **85**, 030301 (2012).
[32] A. S. Holevo, in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes*, edited by L. Accardi, A. Frigerio, and V. Gorini, Springer Lecture Notes in Mathematics Vol. 1055 (Springer, Berlin, 1984), p. 153.
[33] A. Luis and J. Peřina, Phys. Rev. A **54**, 4564 (1996).
[34] F. Dolde, I. Jakobi, B. Naydenov, N. Zhao, S. Pezzagna, C. Trautmann, J. Meijer, P. Neumann, F. Jelezko, and J. Wrachtrup, Nat. Phys. **9**, 139 (2013).
[35] T. A. Wheatley, D. W. Berry, H. Yonezawa, D. Nakane, H. Arao, D. T. Pope, T. C. Ralph, H. M. Wiseman, A. Furusawa, and E. H. Huntington, Phys. Rev. Lett. **104**, 093601 (2010).
[36] J. Kennedy and R. C. Eberhart, *Swarm Intelligence* (Morgan Kaufmann Publishers, San Francisco, 2001).