Optimal frequency estimation and its application to quantum dots

Ángel Gutiérrez-Rubio,1 Peter Stano,1 and Daniel Loss1,2

1RIKEN Center for Emergent Matter Science, 2-1 Hirosawa, Wako, Saitama, 351-0198 Japan
2Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

(Dated: April 28, 2020)

We address the interaction-time optimization for frequency estimation in a general two-level system. The goal is to track with maximum precision a stochastic perturbation with arbitrary dynamics. Our approach is valid for any figure of merit used to define optimality, and is illustrated for the variance and entropy. For the entropy, we clarify the connection to maximum-likelihood estimation. We devise novel estimation protocols with and without feedback. They outperform common protocols given in the literature. We design a probabilistic self-consistent protocol as a generically optimal estimation without feedback. It can improve current experimental techniques and boost coherence times in quantum computing.

PACS numbers: 03.67.Ac, 89.70.Cf

I. INTRODUCTION

With the steady progress of experimental techniques, estimating a parameter with the precision limited by Heisenberg’s uncertainty principle is beyond theoretical speculation.1 As a result, quantum measurement stands out in fields like interferometry,2 spectroscopy,3 quantum positioning and clock synchronization,4,5 magnetometry,6,7 or two-photon microscopy14 to name a few.

Since phases are the origin of many quantum phenomena, there is abundant literature on their measurement, and also of the related frequencies, in the context of parameter estimation.15,22 The ability to accurately measure phases and frequencies has deep implications. For example, in quantum computing, dephasing and decoherence have to be controlled to make a qubit usable.15,22

Previous works improved phase and frequency estimation through several techniques. Usually, these did not concern apparatus performances, but the theoretical design of measurement protocols. One can highlight quantum tricks like squeezing and entanglement, as often encountered in interferometry; and the adaptation of Kitaev’s algorithm,23,24 to experiments, with modifications to handle systematic errors and the time dependence of the estimated parameter.11,22 Remarkably, the achieved precision surpasses the shot-noise limit in many cases.

However, several relevant questions remain unanswered. To motivate our study, here we summarize some of them. First and foremost, despite the improvements reported in the literature, the path to optimality is still unclear. Often, achieving Heisenberg scaling is conditioned on phenomenological parameter tuning. A good example is the variable number of repetitions made in Kitaev’s algorithm,11,22,23 with optimality subordinated to numerical tests. Ideally and in contrast to this approach, there should be a robust method to find the optimal protocol without heuristic parameters. Secondly, the performance of measurements with feedback11,23,26 versus without feedback11,25 has been discussed for over a decade, opening the possibility of optimization with less resources. At last, optimality is assessed with different figures of merit in the literature: probability theory often deals with maximum-likelihood estimation31 and information theory with the entropy.32–35 However, many works in physics focus on the variance or confidence intervals,11,22 together with versions like the Holevo variance.10,26 In the quest for optimality, it is worth to clarify the relations and possible equivalence between these figures of merit.

This article addresses these matters generally for a two-level system perturbed by dephasing noise. Our main result is the design of protocols with and without feedback that optimize an arbitrary figure of merit. Remarkably, our protocols without feedback are not heuristic, they are efficient and of very general applicability. Numerical simulations compare them to other protocols based on the literature. Ours perform better in terms of the entropy. As we prove, this is a reliable figure of merit due to its equivalence to maximum likelihood. These results clarify the open questions discussed before, improving frequency measurement and coherence times in quantum computing.

Our work is structured as follows. In Sec. II we describe the system and the procedure of Bayesian estimation. Sec. III introduces our figures of merit and the relation to maximum likelihood. In Sec. IV we design the estimation protocols and evaluate them with numerical simulations. Our general conclusions appear in Sec. V.

II. BAYESIAN ESTIMATION ON A TWO-LEVEL SYSTEM

Let \( H(t) = \frac{\hbar}{2} |\Delta + \Gamma(t)| \sigma_z \) be the Hamiltonian of a two-level system. Define \( \Omega(t) = |\Delta + \Gamma(t)| \) as the frequency of the oscillations between the eigenvectors of \( \sigma_z \), denoted by \( |\downarrow\rangle \) and \( |\uparrow\rangle \); let \( \Delta > 0 \) be constant, \( \Gamma(t) \) a time-dependent perturbation and assume \( \langle \Gamma(t) \rangle = 0 \). \( \Gamma(t) \) is the source of dephasing with respect to the average frequency \( \Delta = \langle \Omega(t) \rangle \).

\[\begin{align*}
\text{arXiv:2004.12049v1 [cond-mat.mes-hall] 25 Apr 2020}
\end{align*}\]
This system is particularly interesting because its Bayesian update (defined below) has a closed solution for all diffusion regimes. App. A analyzes this in detail, discussing also a general two-level Hamiltonian and to what extent the results extrapolate. Moreover, our Hamiltonian maps to a prospective qubit, a two-electron double quantum dot in GaAs. In that case, $|\downarrow\rangle$ and $|\uparrow\rangle$ represent the singlet and triplet states, $\Delta$ is a micromagnet gradient and $\Gamma(t)$ is the difference between the projections of the Overhauser field on the micromagnet field.

Our goal is to detect and track the value of $\Omega(t)$ with as much precision as possible. Not only for the sake of its measurement, but also to mitigate dephasing. With estimation of a variable, we mean giving its probability distribution. To track $\Omega(t)$, we estimate $\Omega(t_i)$ at a discrete set of times $t_i$, $i = 1, \ldots, N$. These are called initialization times in the typical experiment described below.

To simplify the notation, $\omega$ denotes $\Omega(t_i)$ irrespective of the index $i$. The context that will accompany $\omega$ will make this notation unambiguous. Uppercase symbols like $\Gamma$ or $\Omega$ represent functions of time and usually omit the time argument.

The estimation of $\omega$ involves two processes: the Bayesian update after each qubit measurement and the diffusion of $\omega$ from then on. In this section, we separately outline both. After that, we explain how to apply them sequentially and repeatedly in a typical experiment. This is all we need to design estimation protocols.

Qubit measurements are the only way to access $\Omega$. To describe the Bayesian update, let us look first at a single initialization $\omega$, say $\omega = \Omega(t_i)$, and then we generalize to a general two-level Hamiltonian and to what extent the results extrapolate. Moreover, our Hamiltonian maps to a prospective qubit, a two-electron double quantum dot in GaAs. In that case, $|\downarrow\rangle$ and $|\uparrow\rangle$ represent the singlet and triplet states, $\Delta$ is a micromagnet gradient and $\Gamma(t)$ is the difference between the projections of the Overhauser field on the micromagnet field.

Knowing $P(m|\omega, \tau)$, we can apply Bayes’ rule to estimate $\omega$. Bayes’ rule updates the prior distribution $P(\omega)$ to the so-called posterior distribution $P(m|\omega, \tau)$ after the interaction time $\tau$ and the outcome $m = \pm 1$,

\[
P(\omega|m, \tau) = \frac{P(m|\omega, \tau)P(\omega)}{P(m|\tau)}.
\] (3)

$P(m|\tau)$ can be calculated as a normalization constant. We emphasize that in these expressions, $\omega$ represents the value of $\Omega(t)$ at the last initialization time.

Now, we analyze how the diffusion of $\Omega(t)$ affects frequency estimation. It already entered the problem through the functional integral of Eq. (1). But it also broadens $P(\omega)$ onto $P(\omega)$ after a time $\delta t$,

\[
P_d(\omega) = \int_{-\infty}^{\infty} d\omega' P(\omega')K(\omega', \omega, \delta t).
\] (4)

$K(\omega', \omega, \delta t)$ is the diffusion kernel that encodes the time dynamics of $\Omega(t)$. For simplicity, we assume that it is Markovian. App. A comments on possible generalizations.

A model for diffusion closes the description of the system. A random walk in the presence of a harmonic potential meets our needs: fluctuations happen around zero with typical deviation $\sigma_\Omega$ in a timescale given by $\kappa$. Its diffusion kernel reads

\[
K(\omega', \omega, \delta t) = \frac{1}{\sqrt{2\pi\sigma_\delta^2}} \exp \left[ -\frac{(\omega' - \omega - e^{-\delta t/\kappa})^2}{2\sigma_\delta^2} \right],
\] (5)

where $\sigma_\delta^2 = \sigma_\Omega^2(1 - e^{-2\delta t/\kappa})$. Subtracting $\Delta$ from $\omega$ and $\omega'$ accounts for a nonzero mean value of the frequency.

The kernel $K(\omega', \omega, \delta t)$ gives the probability that $\omega'$ evolves to $\omega$ after a time $\delta t$. Therefore, it determines $P(\Omega)$ in Eq. (1). Eq. (5) allows us to express Eq. (1) in closed form,

\[
P(m|\omega, \tau) = \frac{1}{2} \left[ 1 + m e^{-\psi(\tau)} \cos \phi(\omega, \tau) \right].
\] (6)

Here,

\[
\psi(\tau) \equiv \sigma_\Omega^2[\tau + \kappa(1 - e^{-\tau/\kappa})(-3 + e^{-\tau/\kappa})], \quad \phi(\omega, \tau) \equiv \Delta \tau + \kappa(\omega - \Delta)(1 - e^{-\tau/\kappa}).
\]

The first-order expansion in $\tau/\kappa$ reduces Eq. (6) to an expression with $\psi(\tau) \simeq 0$ and $\phi(\omega, \tau) \simeq \omega \tau$. This is the approximation of slow diffusion mentioned before. For what follows, it is essential that we do not adopt this approximation and we use the general expression of Eq. (6).

Bayesian update and diffusion are applied alternatingly and repeatedly. They compete in the tracking of $\Omega$, roughly narrowing and smearing the probability distribution $P(\omega)$, respectively. We lay out a typical experiment and pinpoint the changes in $P(\omega)$, see also Fig. 1. First, at time $t_1$ we initialize the system in $|\uparrow\rangle$, and after oscillating for a time $\tau_1$ of our choice, we perform
a projective measurement. With the result \( m_1, \) Eq. (3) and Eq. (6) narrow the probability distribution of the frequency. Eq. (4) diffuses the resulting posterior until a new measurement is performed. We repeat the whole procedure \( N \) times, and after a set of measurements \( M \equiv \{ m_1, \ldots, m_N \} \) corresponding to the interaction times \( T \equiv \{ \tau_1, \ldots, \tau_N \} \), one obtains the posterior probability distribution \( P(\omega|TM) \). As we pointed out before, here \( \omega \) corresponds to \( \Omega(t_N) \), with \( t_N \) the initialization time of the \( N \)th measurement. For the prior \( P(\omega) \) at the beginning of the experiment, we take a Gaussian centered at \( \Delta \) with a dispersion \( \sigma_\Omega \).

Now that we have described the system, we can rephrase our main goal more precisely: we want to find \( T \) for the optimal tracking of \( \Omega \). It only remains to properly define optimality. To do so, we discuss different figures of merit and their relations in the next section.

### III. Variance, Entropy, and Maximum-Likelihood Estimation

In the literature on estimation, terms like optimal or accuracy are associated to different magnitudes that are not necessarily equivalent. On the one hand, the variance is ubiquitous in quantum measurement. It quantifies the error and defines the shot-noise and Heisenberg limits. Some modifications like the Holevo variance are used in the estimation of periodic quantities like phases.

On the other hand, the entropy takes over in the context of information theory due to the Shannon theorem. Adopting this figure of merit is extraordinarily far-reaching. For example, one can build statistical mechanics almost exclusively from entropy maximization. The fundamental binomial and Gaussian distributions also maximize the entropy.

It is possible to justify the use of the entropy by arguments other than the relation to knowledge or information. In this work we show how it emerges naturally from maximum-likelihood estimation. For a protocol \( T \), let \( L_T \) be the log likelihood of estimating the true frequency at \( t = t_N \). And let \( \langle S_T \rangle \) be the mean entropy of the final posteriors \( P(\omega|TM) \) after sampling the whole \( (\Omega, M) \) subspace. This subspace is a horizontal plane in Fig. 1b (left). App. B proves the relation:

\[
L_T \equiv \sum_{\Omega,M} P(\Omega|M(T) \log P(\omega = \Omega(t_N))TM) = -\langle S_T \rangle.
\]

In other words, optimizing the maximum average likelihood also optimizes the entropy. This connects maximum-likelihood estimation with entropy optimization.

Now, we give an intuitive understanding of Eq. (7). Fig. 1 illustrates the case of a single trajectory \( \Omega(t) \) and the three protocols \( T^A, T^B, \) and \( T^C \). The ranking (a) T-B better than (b) T-B and (c) T-B better than (c) is straightforward from the plot on the right. As we prove in App. B, the left-hand side of Eq. (7) is the cumulative application of the same criterion for all possible experiments given a protocol \( T \). Hence the sum over \( \Omega \) and \( M \). On the right-hand side, the resulting quantity is the mean entropy of the final distributions \( P(\omega|TM) \).

A given variance sets an upper bound for the entropy, but the converse is not true. Consequently, the variance and the entropy are related but not equivalent. This raises a fundamental question in parameter estimation: what is preferable, to optimize the variance and get an

**Fig. 1.** (a) Outline of a frequency-estimation experiment for a given protocol \( T = \{ \tau_1, \ldots, \tau_N \} \). The distributions of \( \omega \) prior (posterior) to each measurement appear to the left (right) of the oscillation pictures. The red arrows represent diffusion. (b) The left side depicts the configuration space \( (\Omega, T, M) \) and three experiments (colored dots). The protocols \( T^A \), \( T^B \), and \( T^C \) applied to the same trajectory \( \Omega(t) \) result in the different measurements \( M_A, M_B, \) and \( M_C \), respectively. The final posterior distribution of each experiment appears on the right in the same color. The knowledge of the true final frequency \( \Omega(t_N) \) would clearly rank the performance of the protocols. This is the basis of our discussion on the figures of merit and maximum likelihood.

We conclude this section by describing the configuration space of the problem. It contains the main variables and helps to understand their relations. Every experiment is uniquely represented by the coordinates \( (\Omega, T, M) \). \( \Omega \) and \( M \) reflect two different sources of randomness: one is the stochastic evolution of \( \omega \); the other comes from the quantum probabilistic nature of projective measurement. In contrast, the interaction times \( T = \{ \tau_1, \ldots, \tau_N \} \) are up to our choice. It is within this freedom where we consider optimality. A procedure to choose \( T \) and \( T \) itself, will be called a protocol.
estimation with the least squared error, or to minimize the entropy and thus maximize the likelihood of guessing the variable right? Actually, there is no categorical answer: rejecting either sacrifices optimality in a different manner. One must choose the most suitable figure of merit for a specific purpose.

Nevertheless, the protocol’s likelihood $L_T$, given by $\langle S_T \rangle$, is a natural benchmark for protocols in our configuration space. $\langle S_T \rangle$ stands out versus other in principle valid quantities: for example, instead of considering the average of the entropy $\langle S_T \rangle$, why not take the entropy of the averaged error distributions? With our derivation, we can answer that $\langle S_T \rangle$ has a deeper meaning in terms of likelihood. For this reason, we focus on $\langle S_T \rangle$ to compare protocols from now on, but also discuss errors and the variance for the sake of completeness.

IV. OPTIMAL FREQUENCY ESTIMATION

In this section, we design the protocols that can optimize an arbitrary figure of merit. We illustrate them for the entropy $\langle S_T \rangle$ and the variance. $T^S$ and $T^\sigma$ denote these protocols, respectively. We design these protocols with and without feedback. By feedback, we mean using the outcomes $M_{j-1} \equiv \{m_1, \ldots, m_{j-1}\}$ to set $\tau_j$. This is denoted by an asterisk as in $T^{S*}$. On the contrary, protocols without feedback set all $\tau_j$ before the experiment starts. Some works call these protocols online and offline, respectively.

We limit our protocols to a memoryless choice of interaction times: $\tau_k$ for $k<j$ does not directly influence $\tau_j$, but only indirectly through Bayesian update (explained in detail below). The literature sometimes refers to this as local optimization. Proceeding otherwise (or globally), one would have to sample a space of exponentially growing dimension and thus apply Monte Carlo techniques.

This study is outside the scope of this article and we defer it to future works.

By numerical simulations, we compare our protocols to others inspired by the literature. First, we compare them with a set of linear protocols $T^{\alpha_k} \equiv \tau_k \cdot k = 1, \ldots, K$, with $\tau_j = \alpha_k j$, $j = 1, \ldots, N$ and $\alpha_k$ constant. Second, with the saw-toothed protocol $T^{saw}$ with $\tau_j = \alpha_m \times (j \mod n)$ for some $m \leq K$ and $n < N$.

Our simulations take as a benchmark the double quantum dot from Ref. 11 mentioned in Sec. III. But we make one important modification: we deal with fast diffusion, $\delta t / \kappa \sim 1$. The reason is that this makes our analysis more general. Actually, fast diffusion can be trivially extended to slow diffusion, but not the other way around. Consequently, we take $\kappa \approx 3.1 \times 10^{-2}$s, three orders of magnitude smaller than the typical value for the Overhauser field. For the rest of the parameters, we adopt $\sigma_\Delta = 10$ MHz, $\Delta = 30$ MHz, $N \approx 150$ and $\delta t \simeq 15$ms. 1000 simulated experiments are run for each protocol $T$ to sample the configuration space.

A. Protocols with feedback

In this section, we design a protocol with feedback $T^{X*}$ to optimize an arbitrary figure of merit $X$. We illustrate it for the entropy, and we show the results also for the variance.

The protocols with feedback set $\tau_j$ as follows:

Before measurement $j \leq N$, we have the past outcomes $M_{j-1} \equiv \{m_1, \ldots, m_{j-1}\}$ for the interaction times $T_{j-1}^{S*} \equiv \{\tau_1^{*}, \ldots, \tau_{j-1}^{*}\}$, and thus $P(\omega|T_{j-1}^{S*}M_{j-1})$. We choose the optimal interaction time $\tau_j^*$ that minimizes the expected value of the entropy right before measurement $j+1$,

$$S_j = \sum_{m_{j+1} \pm 1} P(m_j|T_j^{S*}M_{j-1})S[P^d(\omega|T_j^{S*}M_j)].$$

Here, the first factor can be calculated as $\int_0^\infty d\omega P(m_j|\omega)P^{d}(\omega|T_j^{S*}M_{j-1})$. $S[P(\omega)]$ is the entropy of the probability distribution $P(\omega)$. In this way, by minimizing $S_j$ at each measurement $j$, we aim at reducing $\langle S_T \rangle$ as much as possible. As mentioned before, we note that $T_{j-1}^{S*}$ only influences $\tau_j^*$ through $P(\omega|T_{j-1}^{S*}M_{j-1})$. Besides this, the values $\tau_1^{*}, \ldots, \tau_{j-1}^{*}$ do not determine $\tau_j^*$.

The upper inset of Fig. 2 illustrates the calculation of $\tau_j^*$ with Eq. (8). An analytical calculation is not possible, but here are two qualitative remarks. $\tau_j$ has a global minimum. For too short $\tau_j$, the Bayesian update yields a smeared posterior. On the contrary, for $\tau_j$ too long, the posterior shows multiple narrow peaks. Fig. 2(c) clearly shows these cases, cf. its discussion below. The optimal $\tau_j^*$ lies in between. For much greater $\tau_j$ values, a high damping $\psi(\tau)$ in Eq. (6) makes the measurement less informative and eventually the entropy saturates to a constant value. $\tau_j$ shows many local minima and maxima. These correspond to the oscillations of $\cos \int_t^{t+\tau} dt \Omega(t)$ in Eq. (4). We interpret them as follows. A projective measurement is most informative when the two binary outcomes $m = \pm 1$ are equally probable. And it is least informative when there is only one possible outcome. Changing $\tau_j$ alternates between these cases and makes the entropy oscillate.

Now, let us evaluate the accuracy of our protocols with feedback. This analysis includes (i) $\langle S_T \rangle$ in Fig. 2(a); (ii) the lower inset of Fig. 2(a), with the so-called odds or ratios of likelihoods $\exp[(\langle S_T \rangle - \langle S_{T^\sigma} \rangle)]$, $T^\sigma \in \{T^S, T^{saw}, T^{saw}, T^{lin}, \ldots, T^{lin,k}\}$ (Eq. 7 and $T^\sigma$ defined below); and (iii) Fig. 2(b), with the average error at $t = t_N$ with respect to the true value. (i) and (ii) mainly concern the entropy and likelihood, related through Eq. 7. (iii) is linked to the variance.

One concludes the following. First of all, the protocols with feedback $T^{S*}$ and $T^{saw}$ beat all other protocols. They reach a lower value of $\langle S_T \rangle$ after fewer measurements and indefinitely sustain it. $T^{lin,k}$ eventually give worse results, see also the multiple peaks in Fig. 2(b). The same happens with $T^{saw}$, cf. Sec. IVB. In contrast, pro
Despite their good performance, protocols with feedback require costly calculations. The diffusion to compute the curve $S_j$ is the most limiting. For this reason, they cannot keep up with fast measurement rates. Without feedback, one can trade this cost for a reasonable compromise in precision. We analyze to what extent this is excellent. They work for arbitrary diffusion speed and automatically handle perturbations like experimental errors: after all, they optimize the figure of merit regardless of the circumstances. Consequently, our protocols can be directly applied to very general systems. This is in contrast to previous works. For example, Refs. 11 and 25 modify Kitaev’s algorithm by repeating each $\tau_j$. The number of repetitions is the parameter tuned numerically. This approach seeks optimality with a single parameter, but why not use more? In order to sustain the maximum precision, when should one restart that protocol? How to modify it when experimental errors happen? Our protocols do not raise this kind of questions.
B. Protocols without feedback

Protocols without feedback set \( T = \{ \tau_1, \ldots, \tau_N \} \) before the experiment starts. Examples in the literature are \( T_{\text{raw}}^{\text{II}} \), defined earlier as \( \tau_j = \sigma_m(j \mod n) \), or Kitaev's protocol: \( \tau_j = 2^{N-j}\tau_0 \), \( j = 1, \ldots, N^{\text{II}} \). In our protocols without feedback, instead of giving a deterministic expression for \( \tau_j \), we take a different and novel approach. We generate each \( \tau_j \) from a distribution \( \hat{P}(\tau_j) \) (notice the dependence of \( \hat{P} \) on \( j \), according to the usual notation \( \tau_j \) in probability theory). In other words, we pick \( \tau_j \) randomly with probability \( \hat{P}(\tau_j) \) before the experiment starts. Thus the label probabilistic from now on. But what is the \( \hat{P}(\tau_j) \) that makes the protocol optimal? Now, we present the self-consistent method to construct it. As we discuss later on, not only does self-consistency guarantee optimality, but also universal applicability. In fact, the construction of the set of distributions \( \hat{P}(\tau_j) \) means adapting to the particular features of the system.

An outline of our probabilistic self-consistent protocol appears in Fig. 3b. The construction proceeds iteratively. In the iteration \( i \), pick the values \( \{ \tau_1, \ldots, \tau_N \} \) randomly with probabilities given by the initial distributions \( \{ P_1(\tau_1), \ldots, P_1(\tau_N) \} \), respectively (for \( i = 1 \), take \( \tau_j \) arbitrarily). Next, while an experiment is running (or more often, being simulated) with that protocol (see blue in Fig. 2a), compute also the optimal \( \tau_j^* \) for each \( j \) with Eq. (8) (red in Fig. 3b). Once enough data have been collected, construct the distributions (by histograms or other parametrization) \( \{ P_{i+1}(\tau_1), \ldots, P_{i+1}(\tau_N) \} \) out of those optimal \( \tau_j^* \). These distributions feed the next iteration, \( i+1 \). Repeat the process until reaching the limit

\[
\hat{P}(\tau_j) = \lim_{i \to \infty} P_i(\tau_j). \tag{9}
\]

As we show below, only a few repetitions are needed.

The convergence to this limit is critical to apply our method. Let us prove its existence and uniqueness. For this assume \( \hat{P}(\tau_k) = \lim_{i \to \infty} P_i(\tau_k) \) exists and is unique for \( k = 1, \ldots, j \), and let \( T_j = \{ \tau_1, \ldots, \tau_j \} \) be a protocol generated by \( \hat{P}(\tau_k) \). \( M_j, T_j \) uniquely determines \( \tau_j^* \) through Eq. (8), and thus \( \hat{P}(\tau_j) \) is the distribution of \( \tau_j^* \) after sampling the whole subspace \( (M_j, T_j) \).

Moreover, \( \hat{P}(\tau_1) = \lim_{i \to \infty} P_i(\tau_1) = \delta(\tau_1 - \tau_1^*) \) trivially exists because \( \tau_1^* \) only depends on the prior \( \hat{P}(\omega) \). This completes the proof. Note that this reasoning does not impose anything on the protocol \( T \). For this reason, our method can optimize general protocols without feedback as we discuss below.

Our method can only be applied in practice if \( P_i(\tau_j) \) converges to \( \hat{P}(\tau_j) \) fast. This is proved in Fig. 4 convergence is reached after less than 8 iterations. Additionally, an adjustment on the calculation of \( \hat{P}(\tau_j) \) allows us to speed up the method. It consists of applying the self-consistent method described above but only for the distributions \( \{ \hat{P}(\tau_j) \mid j \in r\mathbb{N}, 1 < r < N \} \). We take \( r = 15 \). The remaining distributions \( \{ \hat{P}(\tau_k) \mid k \notin r\mathbb{N} \} \) are linearly interpolated.

With this method, we generate the protocols \( T^S \) and \( T^* \). Let us analyze what precision they achieve. Fig. 2 shows that they perform similarly to \( T^S \) and \( T^* \), but there are two main differences. First, as expected, the minimum \( \langle S_T \rangle \) from \( T^S \) and \( T^* \) is slightly above the minimum from \( T^S \) and \( T^* \). The odds quantify the difference in terms of likelihood, stabilizing at \( \sim 1.5 \). Second and more strikingly, several linear protocols surpass \( T^S \) and \( T^* \) during the first half of the experiment.

This does not mean that linear protocols are better. The reason is that \( T^S \) and \( T^* \) eventually outperform \( T_{\text{raw}}^{\text{II}} \). This happens when \( \tau_j \) in \( T_{\text{lin}}^{\text{II}} \) becomes too large, producing multiple peaks in Fig. 2a (right). But then the question is: can the protocols \( T_{\text{lin}}^{\text{II}} \) be modified to sustain the minimum \( \langle S_T \rangle \) they get, and therefore beat \( T^S \) and \( T^* \)? This is what we aim at with \( T_{\text{raw}}^{\text{II}} \). It restarts a linear protocol when it reaches the minimum \( \langle S_T \rangle \), expecting to maintain that value from then on. However, Fig. 2 shows that rather than keeping \( \langle S_T \rangle \) constant, \( T_{\text{raw}}^{\text{II}} \) makes it oscillate. \( T_{\text{raw}}^{\text{II}} \) periodically recovers the minimum \( \langle S_T \rangle \), but on average \( T^S \) and \( T^* \) perform better. We expect this same behavior for any other protocol without feedback. We conjecture that the protocol generated by Eq. (9) is the ultimate protocol without feedback to sustain the maximum precision indefinitely.

Now, we examine the limit distributions \( \hat{P}(\tau_j) \), plotted for all \( j \) in Fig. 2b. As expected, they share the main qualitative features discussed in Sec. IV A. Slight differences are that \( \hat{P}(\tau_j) \) are more smeared, and that the replicated plateaus display more weight. Once again, the important point is that optimality does not depend on imposed features or heuristic parameters. The true value of self-consistency is this robust and automatic tuning. As we already pointed out in Sec. IV A, it works under very general conditions, even against measurement errors or wide variations of the noise-dynamics parameters.

We conclude by pointing out the broad applicability of our probabilistic self-consistent protocol. Not only can it optimize the memoryless protocols we focused on, but it would apply to more general cases. This is a direct consequence of the general proof of existence and uniqueness we gave after Eq. (9). For example, we can use self-consistency to improve any protocol without feedback in the literature. Let the protocol be originally \( \tau_j = \tau(j) \), with \( \tau(j) \) a certain function, and let us improve it to \( \tau^* \) within the constraint \( \tau^* \in [\tau(j) - L(j), \tau(j) + L(j)] \). \( L(j) \) must be large enough for this interval to contain at least one local minimum in the figure of merit, see Fig. 2a (top inset). We choose the optimal \( \tau^*_j \) as one of those minima. The rest of our method stays the same, and so we generate a protocol without feedback. Remarkably, this combines self-consistency with the memory provided by the function \( \tau(j) \). Modifications like this would respond to self-consistency rather than to the heuristic parameters used in the literature. We leave these optimizations as a continuation to this work.

In sum, this section proves that our self-consistent
method (i) yields good precision, slightly less than protocols with feedback; (ii) keeps that precision stable in time, beating other protocols without feedback and preserving more coherence; and (iii) can improve any other protocol without feedback. For these reasons, we propose our self-consistent probabilistic protocol as a robust way to optimize generic protocols without feedback. This is the main result of this article.

Despite being memoryless, the protocols with feedback outperform all others we tried, mostly inspired by the literature, and moreover indefinitely sustain the maximum precision attained. But most remarkably, we have designed a protocol without feedback that performs almost identically and (i) can also optimize generic protocols for any figure of merit and does not require heuristic parameters; (ii) can be applied to experiments with arbitrarily fast measurement rates; (iii) is robust in general circumstances, including measurement errors; and (iv) is numerically feasible with a few self-consistent iterations.

In sum, we have designed and tested a versatile protocol that can significantly improve the precision in parameter estimation. Its implementation in generic quantum dots would result in the increase of the coherence times of potential qubits.

VI. ACKNOWLEDGMENTS

Á. Gutiérrez-Rubio acknowledges the project CREST JST (JPMJCR1675) through the Japanese Science and Technology Agency. This work was supported by the Swiss National Science Foundation and NCCR QSIT.

Appendix A: Generalizations of the system

In this appendix, we give some guidelines to generalize our study. We analyze how the design of protocols for frequency estimation should change (i) for a general two-level Hamiltonian and (ii) under measurement errors. Finally, we briefly point out further possible extensions.

We focus first on a general two-level system. Let $H(t) = H_0 + H'(t)$, with

$$H_0 = \hbar \Delta / 2 \hat{n} \cdot \hat{\sigma}, \quad H'(t) = [\hbar \Gamma(t)/2] \sigma_x,$$

(A1)

$\Delta > 0$ constant, $\hat{n} = n_x \hat{x} + n_y \hat{y} + n_z \hat{z}$ a unit vector, and $\Gamma(t)$ a stochastic variable. We assume $\langle \Gamma(t) \rangle = 0$. 

---

FIG. 3. Outline of the self-consistent method to obtain $P(\tau_j)$, depicting iteration $i$ and measurement $j$. Blue corresponds to the application of the protocol generated with $P_i(\tau_j)$. Red only gathers the information to construct $P_{i+1}(\tau_j)$. The plot of $S_j$ shows that the red $\tau_j^*$ is better than the blue $\tau_j$. This information will be used in the next iteration $i+1$.

FIG. 4. Waiting-time distributions for the measurement $j$ displayed on the top left corners. The results for the iterations $i = 1, \ldots, 8$ are shown with increasing opacity, and the seed for $\tau_j$ with a dotted line. The limit distributions $P(\tau_j)$, filled with a yellow shadow as a guide for the eye, are the sections of the density plots in Fig. 2c. Blue (green) corresponds again to the variance (entropy).

V. CONCLUSIONS

We have analyzed the optimization of frequency estimation in the presence of dephasing sources with great generality, including generic two-level systems, arbitrarily fast diffusion, protocols with and without feedback and two archetypal figures of merit, the squared error and entropy, in turn related to maximum likelihood.
The eigenstates of $\sigma_z$ are denoted by $|\uparrow\rangle$ and $|\downarrow\rangle$ and define our space for projective measurements. Further, we define the frequency $\Omega(t) = |\Delta \hat{n} + \Gamma(t) \hat{x}|$, where $\Gamma(t)$ is the source of dephasing with respect to the average $\Delta = (\Omega(t))$. The goal is to estimate $\Gamma(t)$ or, equivalently, $\Omega(t)$. One can map this model to a variety of systems where dephasing has different origins. For example, with $\hat{n} = \hat{z}$, $H(t)$ maps to the double dot discussed in Sec. [11] or to holes with Ising-like interactions [25].

Importantly, almost all the expressions and protocols in the main text remain valid for the general Hamiltonian $H(t)$ of Eq. (A1). But there is one important change. Now, the commutator of the Hamiltonian with itself at different times is different from zero:

$$[H(t), H(t')] = 2i\Delta[(\Gamma(t) - \Gamma(t'))(n_x \sigma_y + n_y \sigma_x)].$$

Therefore, in Eq. (4) we must insert

$$P(m|\Omega, \tau) = |\langle m | T \exp \left[ -i \frac{\hbar}{2} \int_{t_i}^{t_i+\tau} dt H(t) \right] |\uparrow\rangle|^2$$

and we cannot drop the time-ordering operator $T$. Consequently, there is no simple expression for $P(m|\Omega, \tau)$ like Eq. (2), and no closed expression for $P(m|\omega, \tau)$ like Eq. (6). In sum, Bayesian update is now a difficult task.

A workaround is to approximate $\Gamma(t)$, namely take $P(H|\omega) \propto \exp (-\beta H(t))$ for $t \in [t_i, t_i + \tau]$. As we discuss in the main text, this approximation is valid only for negligible diffusion during interaction times $\tau$. In this case,

$$P(m|\omega, \tau) \approx \frac{1}{2} \left[ 1 + m \left( \cos^2 \beta + \sin^2 \beta \cos(\omega \tau) \right) \right],$$

with $\omega = \Omega(t_i)$ and $\cos \beta = \hat{n} \cdot \hat{z}$. Always within the regime of slow diffusion, this expression extends our protocols to general two-level systems. One should use it instead of Eq. (6), but the rest of our analysis remains the same.

Now, we extend our study with measurement errors. Let $\eta_1$ ($\eta_2$) be the probability to make an error when measuring $|\uparrow\rangle$ ($|\downarrow\rangle$). Define $\mu = \eta_1 - \eta_2$ and $\nu = 1 - \eta_1 - \eta_2$.

In the main text, Eq. (2) turns into

$$P(m|\Omega, \tau) = \frac{1}{2} \left( 1 + m \left[ \mu + \nu \cos \int_{t_i}^{t_i+\tau} dt \Omega(t) \right] \right),$$

and Eq. (4) changes accordingly. For the general Hamiltonian of Eq. (A1), Eq. (A2) takes the form

$$P(m|\omega, \tau) \approx \frac{1}{2} \left[ 1 + m \left[ \mu + \nu \cos^2 \beta + \sin^2 \beta \cos(\omega \tau) \right] \right].$$

This expression shows the equivalence between measurement errors on the one hand, and a nonzero $\cos \beta = \hat{n} \cdot \hat{z}$ (coming from $\sigma_y$ and $\sigma_z$ terms in the Hamiltonian) on the other. Either of them separately, or both together, yield $r + s \cos(\omega t)$ inside the square brackets, with $r + s = 1$. Their effect is to slow down the Bayesian update. Actually, they prevent zeros in $P(m|\omega, \tau)$. Such zeros are desirable because they discard frequencies when applying Bayes’ rule, see Eq. (3).

One can also think of further generalizations. For example, the perturbation in Eq. (A1) could be $H'(t) = (h/2) \sum_i \Gamma_i(t) \sigma_i$, with $i$ running over $\{x, y, z\}$. This would require us to estimate three stochastic variables instead of one. The Bayesian formalism, although more involved, would stay the same. At last, realistic noise often has a more complicated kernel than Eq. (5). Similarly, the diffusion of probability distributions through Eq. (4) does not cover the most general case. Using ARMA models seems a viable option to simulate such noise. We leave these studies as a possible continuation to our work.

**Appendix B: Proof of Eq. (7)**

Log likelihood is a relevant quantity encountered in estimation and hypothesis testing. In this appendix, we discuss its optimization on general grounds first. Afterwards we prove Eq. (7).

Let $\mathcal{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses or models that presumably govern a given phenomenon, and $D = \{x_1, \ldots, x_n\}$ contain all the data we have about it. According to maximum-likelihood estimation, the hypothesis $H_j$ that best fits $D$ is the one maximizing $P(H_j|D)$. Remarkably, this ranks the elements in $\mathcal{H}$ according only to the available evidence $D$.

Assume that no hypothesis is preferred over any other, namely take $P(H)$ constant for all $H \in \mathcal{H}$. By Bayes’ rule, i.e., $P(H|D) = P(D|H)P(H)/P(D)$, the maximum-likelihood estimation of $H$ is equivalent to the optimization of $P(D|H)$. Thus, for logically independent data, for which $P(x_1, \ldots, x_N|H) = \prod_{i=1}^N P(x_i|H)$, we have

$$\log P(H|D) = \sum_{i=1}^N \log P(x_i|H) + \text{const.} \quad (B1)$$

We highlight two important aspects of this result. First of all, the performance or ranking of a general hypothesis, model or method $H$ is computed by testing it with the true data values $x_i$. And secondly, the log likelihood is an additive function of the data.

Let us now introduce the log likelihood in our context. Consider first a single experiment: we have a given trajectory $\Omega$, a protocol $T_A$ and the measurements $M_A$. The final frequency has the true (and unknown in a real experiment) value $\omega = \Omega(t_N)$ and our estimation of $\omega$ is given by $P(\omega|T_A M_A)$. The log likelihood of this single experiment is $\log P(\omega = \Omega(t_N)|T_A M_A)$. This value quantifies, for this experiment we are looking at, how good the estimation with $T_A$ is. If another protocol $T_B$ had been used on the same $\Omega$, yielding the measurements $M_B$, we would have the log likelihood $\log P(\omega = \Omega(t_N)|T_B M_B)$. By comparing $\log P(\omega = \Omega(t_N)|T_A M_A)$ and $\log P(\omega = \Omega(t_N)|T_B M_B)$, one could decide whether $T_A$ or $T_B$ es-
estimates ω better for that particular experiment. Fig. illustrates this trivial comparison.

The additivity of likelihoods in Eq. (B1) allows us to extend this reasoning to multiple experiments. Ideally, if one could sample the whole configuration space (Ω, M) for a given protocol T, the total log likelihood (averaged over the number of experiments) would add up to

\[ L_T = \sum_{\Omega, M} P(\Omega M | T) \log P(\omega = \Omega(t_N) | TM) \]  \hspace{1cm} (B2)

Here, \( P(\Omega M | T) \) represents the probability to encounter (Ω, M) given T. Then, in terms of maximum likelihood, \( L_T \) is the figure of merit that evaluates how well the protocol T estimates the frequency.

Now, let us prove that for a given protocol T, \( L_T \) equals the mean entropy \( \langle S_T \rangle \) of all the posteriors \( P(\omega | TM) \).

Departing from Eq. (B2),

\[
L_T = \sum_{\Omega, M} P(\Omega M | T) \log P(\omega = \Omega(t_N) | TM) \\
= \sum_{\Omega, M} P(M | T) P(\Omega | TM) \log P(\omega = \Omega(t_N) | TM) \\
= \sum_{\Omega(t_N), M} P(M | T) P(\Omega(t_N) | TM) \times \log P(\omega = \Omega(t_N) | TM) \\
= - \sum_{M} P(M | T) S[P(\omega | TM)] = - \langle S_T \rangle. \]  \hspace{1cm} (q.e.d.)

Remarkably, this result stems from (i) the application of elementary probability rules and (ii) from the assumption \( P(\Omega(t_N) | TM) = P(\omega = \Omega(t_N) | TM) \), namely that the sampling probability \( P \) in the configuration space is correctly predicted by our Bayesian estimation \( P(\omega | TM) \). Condition (ii) basically relies on the agreement between the true diffusion dynamics of \( \Omega(t) \) and the kernel, see Eq. (4). We obviously simulate our experiments under this condition. Therefore, our theorem applies to our data. In an actual experimental setup, however, this requirement should be confirmed by a procedure external to ours.

We conclude with a remark on the practical use of our result, Eq. (7). Eq. (B2) defines the log likelihood \( L_T \) in our context. But its computation in a real experiment is impossible in that form: as we pointed out, \( P(\omega = \Omega(t_N) | TM) \) cannot be calculated because the true frequency \( \Omega(t_N) \) is unknown. With Eq. (7), however, we can tackle this problem: the knowledge of the true frequencies \( \Omega(t_N) \) is not necessary if we can calculate the average entropy of the posteriors \( P(\omega | TM) \), denoted by \( \langle S_T \rangle \). Our result is valuable from a theoretical point of view, mainly discussed in Sec. IV, but it is also a practical tool to calculate log likelihoods, as we do throughout Sec. IV.

\[ ^{1} \text{V. Giovannetti, S. Lloyd, and L. Maccone, Science 306, 1330 (2004).} \]
\[ ^{2} \text{C. L. Degen, F. Reinhard, and P. Cappellaro, Rev. Mod. Phys. 89, 035002 (2017).} \]
\[ ^{3} \text{M. O. Scully and J. P. Dowling, Phys. Rev. A 48, 3186 (1993).} \]
\[ ^{4} \text{J. Jacobson, G. Björk, and Y. Yamamoto, Appl. Phys. B Laser Opt. 60, 187 (1995).} \]
\[ ^{5} \text{B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature 450, 393 (2007).} \]
\[ ^{6} \text{Z. Y. Ou, Phys. Rev. Lett. 77, 2352 (1996).} \]
\[ ^{7} \text{J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A 54, R4649 (1996).} \]
\[ ^{8} \text{V. Giovannetti, S. Lloyd, and L. Maccone, Nature 412, 417 (2001).} \]
\[ ^{9} \text{V. Giovannetti, S. Lloyd, L. Maccone, and F. N. Wong, Phys. Rev. Lett. 87, 117902 (2001).} \]
\[ ^{10} \text{M. Mullan and E. Knill, Phys. Rev. A 90, 042310 (2014).} \]
\[ ^{11} \text{H. T. Dinani, D. W. Berry, R. Gonzalez, J. R. Maze, and C. Bonato, Phys. Rev. B 99, 125413 (2019).} \]
\[ ^{12} \text{R. S. Said, D. W. Berry, and J. Twamley, Phys. Rev. B 83, 125410 (2011).} \]
\[ ^{13} \text{S. Danilin, A. V. Lebedev, A. Vepsäläinen, G. B. Lesovik, G. Blatter, and G. S. Paraoanu, npj Quantum Inf. 4, 29 (2018).} \]
\[ ^{14} \text{H. B. Fei, B. M. Jost, S. Popescu, B. E. Saleh, and M. C. Teich, Phys. Rev. Lett. 78, 1679 (1997).} \]
\[ ^{15} \text{Z. Y. Ou, Phys. Rev. A 55, 2598 (1997).} \]
\[ ^{16} \text{H. M. Wiseman and R. B. Killip, Phys. Rev. A 56, 944 (1997).} \]
This expression is straightforward from elementary probability rules:

\[
P(m|\omega, \tau) = P(m, \int D\Omega|\omega, \tau) = \int D\Omega P(m, \Omega|\omega, \tau)
\]

Here, the functional integral in \(\Omega\) runs over all arbitrary functions of \(t\) and the integral in \(\Omega\), as we defined in the main text, over all functions with \(\Omega(t_0) = \omega\). Throughout this article, we follow the propositional notation common in probability theory arguments inside \(P\) represent logical statements. Accordingly, sums or integrals act as the logical ‘or’ operator.

E. T. Jaynes, *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst (Cambridge University Press, 2003).

32. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).

33. D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys. Rev. A **61**, 032306 (2000).

34. C. Ferrie, C. E. Granade, and D. G. Cory, AIP Conf. Proc. **1443**, 165 (2012).

35. F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).

36. D. W. Berry, H. M. Wiseman, and J. K. Breslin, Phys. Rev. A **63**, 053804 (2001).

37. A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Phys. Rev. A **84**, 052315 (2011).

38. This expression is straightforward from elementary probability rules:

\[
P(m|\omega, \tau) = P(m, \int D\Omega|\omega, \tau) = \int D\Omega P(m, \Omega|\omega, \tau)
\]

Here, the functional integral in \(\Omega\) runs over all arbitrary functions of \(t\) and the integral in \(\Omega\), as we defined in the main text, over all functions with \(\Omega(t_0) = \omega\). Throughout this article, we follow the propositional notation common in probability theory arguments inside \(P\) represent logical statements. Accordingly, sums or integrals act as the logical ‘or’ operator.

39. G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

40. K. Rabinstein, V. A. Sverdlov, and D. V. Averin, J. Exp. Theor. Phys. Lett. **79**, 646 (2004).

41. H. W. Bode and C. E. Shannon, Proc. IRE **38**, 417 (1950).

42. C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, Nat. Nanotechnol. **11**, 247 (2016).

43. E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

44. J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, J. Phys. A Math. Theor. **49**, 143001 (2016).

45. A. Doucet, S. Godsill, and C. Andrieu, Stat. Comput. **10**, 197 (2000).

46. T. Ruster, H. Kaufmann, M. A. Luda, V. Kaushal, C. T. Schmiegelow, F. Schmidt-Kaler, and U. G. Poschinger, Phys. Rev. X **7**, 031050 (2017).

47. The probability distributions of two different variables \(x\) and \(y\) are usually represented by \(P(x)\) and \(P(y)\), although \(P(x = a)\) does not necessarily equal \(P(y = a)\). The same applies to our distributions \(\hat{P}(\tau_j), j = 1, \ldots, N\).

48. J. Fischer, W. A. Coish, D. V. Bulaev, and D. Loss, Phys. Rev. B **78**, 155329 (2008).