

# Applications of model-aware reinforcement learning in Bayesian quantum metrology

Federico Belliardo *NEST, Scuola Normale Superiore, I-56126 Pisa, Italy*Fabio Zoratti *Scuola Normale Superiore, I-56126 Pisa, Italy*

Vittorio Giovannetti

*NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy*

(Received 12 March 2024; accepted 22 May 2024; published 14 June 2024)

An important practical problem in the field of quantum metrology and sensors is to find the optimal sequences of controls for the quantum probe that realize optimal adaptive estimation. In Belliardo *et al.* [[arXiv:2312.16985](#)], we solved this problem in general by introducing a procedure capable of optimizing a wide range of tasks in quantum metrology and estimation by combining model-aware reinforcement learning with Bayesian inference. We take a model-based approach to the optimization where the physics describing the system is explicitly taken into account in the training through automatic differentiation. In this follow-up paper we present some applications of the framework. The first family of examples concerns the estimation of magnetic fields, hyperfine interactions, and decoherence times for electronic spins in diamond. For these examples, we perform multiple Ramsey measurements on the spin. The second family of applications concerns the estimation of phases and coherent states on photonic circuits, without squeezing elements, where the bosonic lines are measured by photon counters. This exposition showcases the broad applicability of the method, which has been implemented in the `qsensoropt` library released on PyPI, which can be installed with `pip`.

DOI: [10.1103/PhysRevA.109.062609](https://doi.org/10.1103/PhysRevA.109.062609)

## I. INTRODUCTION

In recent years, the intersection of quantum mechanics and machine learning has become a focal point of exploration, with a particular emphasis on leveraging quantum technologies for various applications. This convergence holds immense potential for mutual enhancement across both domains. Quantum technologies, notably quantum computers, offer unique capabilities to tackle classical machine-learning challenges, such as classification and sampling, utilizing both classical and quantum data sources [1–3].

Conversely, traditional machine-learning methodologies can enhance quantum information tasks, including state preparation [4–7], optimal quantum feedback [8], error correction [9], device calibration [10–13], characterization [14], and quantum tomography [15–17]. This work aligns with the latter category, employing model-aware reinforcement learning (RL) to derive optimized adaptive and nonadaptive control strategies for quantum metrology and estimation tasks [8,18–20].

Specifically, we address the challenge of optimal experimental design [21], a task already explored using machine-learning techniques [22–26], which we explore in the quantum realm, finding performances going beyond the current state of the art [27]. Estimation processes in this context involve numerous nondifferentiable steps, such as simulating measurements and resampling from posterior distributions, posing challenges to the application of model-aware RL. To

overcome these obstacles, our approach incorporates techniques like importance sampling, adding log-likelihood to the loss [8], the reparametrization trick, and the Scibior and Wood correction [28].

The methodology involves identifying tunable parameters in a given physical platform and metrological task, allowing an agent (which can be a neural network, a decision tree, or a list of trainable controls) to learn an optimal policy through gradient descent optimization. We have abstracted and packaged this procedure into the `qsensoropt` library, which is available on GitLab [29], together with the online documentation of the classes and the examples of this paper [30]. The library can also be installed with `pip` on every machine without the need for cloning the environment, being it available on PyPI. This is a versatile tool for finding adaptive policies to optimize the precision of quantum sensor, which we think will be of much use for the quantum information community. By exploiting the classes and function preprogrammed in this library, users can implement the quantum-mechanical model of their sensors, and optimize an adaptive or nonadaptive policy, which can be later exported and implemented in the experiments.

Demonstrating the broad applicability of our approach, we optimized examples on the nitrogen-vacancy (NV) center platform for various metrology tasks, including DC and AC magnetometry, decoherence estimation, and hyperfine coupling characterization. In the realm of photonic circuits, we explored applications such as multiphase discrimination, the

agnostic Dolinar receiver [31], and coherent states classification. Our findings showcase the superiority of model-aware RL over traditional control strategies, even outperforming model-free RL in multiple scenarios. Through this work we lay the tools to accelerate the search for optimal control policies in quantum sensing and metrology, potentially expediting the widespread industrial application of this quantum technology.

The existing literature encompasses various works addressing challenges akin to our approach, which we have categorized into four classes according to their relation with our work. The first class encompasses the competitor approaches for optimization in quantum metrology [32–36]. These lack coverage of non-greedy or adaptive policies of Bayesian estimation. The second class contains those optimal control algorithms based on the optimization of the Fisher information [37–45], which again often lack coverage of Bayesian estimation, adaptivity through the use of neural networks or can be applied only to NV centers. The third class design a theoretical approach to optimal control in quantum sensing, but lack an implementation [20,23,24,46,47]. The fourth and last class applies variational quantum circuits to specific metrological tasks [48–54].

In this work, we give an overview of the main components of the framework that allows quantum metrology tasks to be optimized with machine learning, with several examples of interest where we achieved results that are comparable to or better than the current state of the art for each task. However, we refer to Ref. [55] for a complete description of the method and a review of the literature. See also Ref. [56] for a three-pages explanation of the theory with a single example. Although we have also implemented the optimization of frequentist estimation, based on the Fisher information, the results present in this paper concern exclusively the domain of Bayesian estimation.

### A. Review of the framework

This section is meant to be a quick review of the various components of the qsensoropt framework used throughout the paper. For a rigorous treatment of the theory behind, please refer to Ref. [55]. See also Sec. ID for a more detailed review of the quantum model of a sensor and how it is used in Bayesian estimation with different degrees of control. This section defines a common mathematical schematization to quantum metrology, in which we fit all the examples of this paper. If an optimization problem can be expressed in terms of this scheme, then it can be optimized with the qsensoropt library.

*The sensor's model.* Some elements of the Bayesian inference process in quantum metrology are universal, in the sense that they do not depend on the particular task. These are the Bayesian filtering, the repeated execution of the measurements, and the training of the agent. Together with other general aspects of the optimization routine, these are implemented in the library. To apply the framework the user needs to code the model of the specific quantum sensor of interest. This model should simulate the stochastic outcome extraction of the measurements and evaluate the probability of observing a specific outcome for the particular quantum system at hand.

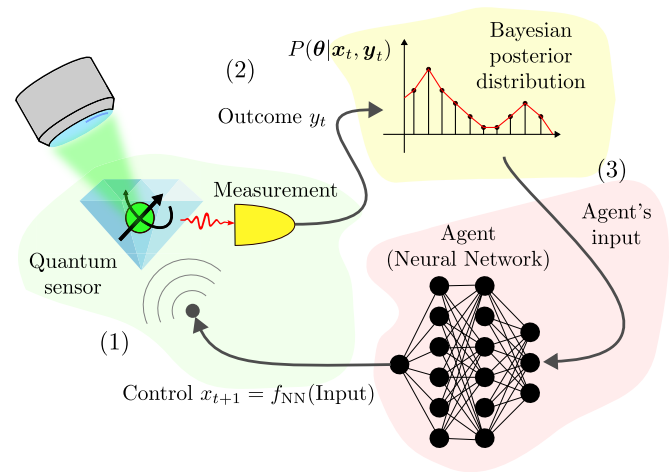


FIG. 1. This general scheme illustrates the information flow within the measurement loop for a quantum metrology task where the quantum probe is an NV center. Refer to Sec. II A for a brief description of the physics of such systems. The environment we aim to study interacts with the quantum probe and encodes it with the unknown variables  $\theta$ . This probe is then measured using a tunable instrument (1). The outcome of this measurement provides us with information about the probe's state and in turn about the environment's variables. This information is used to update the posterior Bayesian distribution on  $\theta$  (2). Some summary information derived from the posterior is then input into an agent that decides the new control parameters for the measurement in the next iteration of the loop (3). This control is then realized through the electronics of the experiment. In this picture, the agent is a neural network.

In this description, it should be stated which parameters are controls, that can be tuned by the experimenter, and which are the target of the estimation. These last ones, which will be indicated collectively by the symbol  $\theta$ , are unknowns related to the environment of the probe, which are codified during the probe's evolution, or they are parameters of the state of the probe if it has been encoded in a distant laboratory, on which we have no control. This encoding is formalized later within the object  $\Phi_\theta$ , defined in Sec. ID.

*The measurements loop.* A quantum metrology or estimation task usually involves many measurements indexed by  $t$ , through which the information on the target parameters is accumulated step by step in the Bayesian posterior. The sequence of measurements is organized in a “measurement loop,” which is represented in Fig. 1 and it involves three fundamental steps. An iteration of the loop starts with the simulation of the evolution and the measurement of the sensor's probe, all done thanks to the model implemented by the user. We indicate with  $y_t$  the outcome of the  $t$ th measurement and with  $\mathbf{y}_t := (y_0, y_1, \dots, y_t)$  the list of outcomes until the  $t$ th iteration (included). The second step in the loop's iteration is the processing of the outcome, which allows the internal representation of the Bayesian posterior to be updated. This is done via a particle filter, which uses an ensemble of points in the parameter space (called particles) and weights to approximate the posterior, see Refs. [55,57] for more details. The initial distribution is the prior  $\pi(\theta)$ , which is uniform over some interval in every application presented in this paper.

At the third and last step of the iteration, some information on the state of the estimation and on the posterior is computed and fed to an agent which evaluates the controls for the next measurement on the probe, i.e., for the next iteration of the loop. We indicate with  $x_t$  the controls of the  $t + 1$ st measurement and with  $\mathbf{x}_t := (x_0, x_1, \dots, x_t)$  the list of controls until iteration  $t$  (included). This agent doing the controlling will typically be a neural network (NN). This agent can leverage knowledge from past measurements to optimize the estimation task's overall performance, which makes the control strategy adaptive and non-greedy. The posterior on  $\theta$  at the  $t$ th iteration of the loop is conditioned on the full list of previous controls and outcomes, and it will be indicated with  $P(\theta|\mathbf{x}_t, \mathbf{y}_t)$ . Typically, during the training, a number  $B > 1$  of simulations of the estimation are executed in parallel with the same agent in a batch; we call  $B$  the batch size. From this batch, the statistical properties of the strategy produced by the agent are estimated, like a precision figure of merit specified by the user, e.g., the mean square error or the error probability.

*The precision-resources paradigm.* Each iteration of the measurement loop consumes a certain amount of a specific "resource" such as the estimation time or the available probes. Once these resources are depleted, the measurement loop concludes. For each application, the resource must be specified, and it is as important as the definition of the precision figure of merit. The goal of the machine-learning framework is to optimize the precision for a fixed maximum amount of consumed resources. In the simplest case, the resource is just the number of measurements executed in the whole estimation, which we indicate with  $M_{\max}$ , which is also the number of total iterations of the measurement loop.

*Overview of the training process.* After the measurement loop ends, an estimator  $\hat{\theta}$  for the parameters  $\theta$  is computed from the Bayesian posterior. The user-defined precision metric serves as the training loss to be minimized through a stochastic gradient descent procedure, which acts on the trainable variables of the agent, e.g., the biases and weights in the case of a NN, which are indicated collectively with the symbol  $\lambda$ . The gradient of the loss is computed via automatic differentiation, where the derivatives flow through the sampling of stochastic variables, like the measurement outcomes, and through the physical model of the sensor. This characterizes the training as model-aware policy gradient reinforcement learning and special precautions are taken to compute an unbiased estimator of the gradient [8,55]. Each training step involves a complete execution of the measurement loop until the resources are depleted. After multiple training steps, in the order of  $O(10^4)$ , the precision reaches a plateau and the training is stopped. At this point, the agent has been optimized. In all the applications of this paper Adam [58] has been used as an optimizer. The learning rate is set to decrease at each training step, with an inverse square root decay law, as specified in Ref. [59]. The initial learning rate  $\alpha_0$  has been chosen optimally for each simulation, through trial and error.

## B. Specifics of each experiment

From the description of this machine-learning approach, it is clear that in order to apply this technique on a particular sensor four pieces of information must be specified and

implemented in the code by the user. These are listed in the following:

(1) The sensor's model, represented by the code that defines the probability of each potential outcome in a measurement. This model is typically derived from the Born rule applied to the physics of the sensor, necessitating the specification of a division between control and target parameters. Further insights on how to schematize the quantum mechanics behind this aspect are detailed in Sec. ID.

(2) The nature and the amount of resources consumed in the metrological task.

(3) The input to the agent, which must be a function of the posterior distribution, and in general of the measurement outcomes for the implementation of an adaptive control policy.

(4) The error figure of merit used to gauge the precision of the estimation, such as the mean square error.

## C. Selection of the agent

In all the adaptive experiments the NN has by default five hidden layers with 64 neurons each, and the activation function is tanh. For the nonadaptive strategy, the controls for each measurement step do not depend on the Bayesian posterior. In this case, the agent can still be a NN with reduced input, or instead, we directly optimize the values of the controls in the training by setting  $\lambda := \mathbf{x}_t$ . In Sec. IV B we experimented with using a decision tree to compute the controls. This means that the posterior is bypassed and the adaptive controls are directly selected based on the list of past measurement outcomes  $\mathbf{y}_t$ . More precisely the outcome selects the tree branch at each node.

## D. Dynamical model

In this section, we present a rigorous characterization of the measurement loop of Fig. 1. To start with, let us observe that the whole procedure can be represented as an ordered sequence of concatenated events describing the information acquired by the agent at the various stages and the associated evolution of the probe state (see top panel of Fig. 2). In particular, the event  $t$  takes as input the state of the probe at the beginning of the  $t$ th measurement loop (orange), the control value  $x_t$  synthesized by the agent in the previous event (red line), the complete list of the outcomes  $\mathbf{y}_{t-1} = (y_0, y_1, \dots, y_{t-1})$  of the measurements performed up to that point, as well as a possible new encoding of the parameters  $\theta$  into the probe state. As shown in the bottom panel of Fig. 2, a characterization of the  $t$ th event is obtained by assigning the following three elements:

(1) a completely positive, trace preserving (LCPT) linear map [60,61]  $\Phi_{\theta}^{(t, x_t)}$  describing the physical process which encodes the parameters  $\theta$  into the state of the probe at the beginning of the event (red box of the figure);

(2) a collection of operators  $\{M_{y_t}(x_t)\}_{y_t}$  fulfilling the normalization condition  $\sum_{y_t} E_{y_t}(x_t) = \mathbb{1}$ ,  $E_{y_t}(x_t) := M_{y_t}^{\dagger}(x_t)M_{y_t}(x_t)$  describing the measurement process (black box);

(3) a control unit (white box).

This is effectively a different way to divide the iteration, more centered on the quantum-mechanical evolution

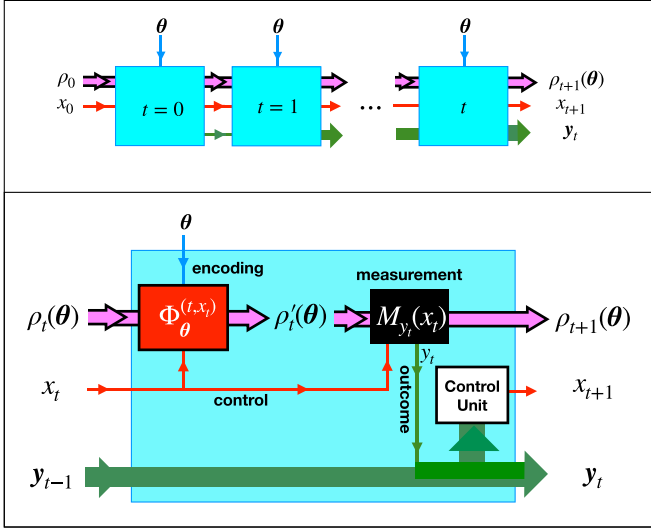


FIG. 2. (top panel) The measurement loop of Fig. 1 represented as an ordered sequence of concatenated events (cyan boxes). The violet arrows describe the temporal evolution of the probe state, the red lines the controls, and the green arrow the measurement outcomes that accumulate along the way. (bottom panel) Schematic description of the processes involved in the  $t$ th event: first the input state  $\rho_t(\theta)$  of the probe undergoes to a possible new encoding stage of the parameters  $\theta$  via the action of the LCPT map  $\Phi_\theta^{(t, x_t)}$ ; then the transformed state  $\rho'_t(\theta)$  gets measured producing the outcome  $y_t$  and the probe emerges as the conditional state  $\rho_{t+1}(\theta)$ . As indicated by the figure all these operations are in part determined by the inputted control  $x_t$ . The new value of the parameter  $x_{t+1}$  is determined by the control unit element elaborating the newly acquired outcome  $y_t$ . To connect with the scheme in Fig. 1 observe that the control unit here encompasses both the particle filter updating the Bayesian distribution and the neural network.

with respect to Fig. 1 and the previous section, which was more centered on the machine-learning side and the flow of information. It is important to connect this two different representation, and this can be done by observing that the first two element in the above list, i.e., the LCPT map and the measurement, are part of the sensor's model, while the Bayesian update and the NN are part of the last element, i.e., the control unit. The inclusion of the index  $t$  in the definition of  $\Phi_\theta^{(t, x_t)}$  allows for the analysis of nonuniform models where the encoding mechanism varies along the measurement loop. For instance, setting  $\Phi_\theta^{(t, x_t)}$  as the identity map  $\mathbb{1}$  for  $t \geq 1$  represents models where  $\theta$  is imprinted on the probe only at the beginning of the first measurement loop. The explicit inclusion of the parameter  $x_t$  in the definition of  $\{M_{y_t}(x_t)\}_{y_t}$  and  $\Phi_\theta^{(t, x_t)}$  accommodates instead models where the agent exerts control over the measurement and possibly on the encoding mechanism (e.g., determining factors such as the duration of the probe's interaction with the external system that is responsible for the effect). Define now  $\rho_t(\theta) := \rho(\theta, \mathbf{x}_{t-1}, \mathbf{y}_{t-1})$  the density matrix of the probe at the beginning of the  $t$ th event [i.e., the state that emerges as output from  $(t-1)$ st event]. According to the model, such configuration is first evolved via  $\Phi_\theta^{(t, x_t)}$  through the mapping

$$\rho_t(\theta) \mapsto \rho'_t(\theta) := \Phi_\theta^{(t, x_t)}[\rho_t(\theta)], \quad (1)$$

producing the state  $\rho'_t(\theta) := \rho'(\theta, \mathbf{x}_t, \mathbf{y}_{t-1})$  that undergoes to the measurement process defined by the operators  $\{M_{y_t}(x_t)\}_{y_t}$ . The probability of outcome  $y_t$  follows the Born rule:

$$P(y_t | \theta, \mathbf{x}_t, \mathbf{y}_{t-1}) := \text{Tr}[E_{y_t}(x_t) \rho'_t(\theta)]. \quad (2)$$

The conditional probe state  $\rho_{t+1}(\theta) := \rho(\theta, \mathbf{x}_t, \mathbf{y}_t)$  after measurement is determined instead by the formula

$$\rho_{t+1}(\theta) := \frac{M_{y_t}(x_t) \rho'_t(\theta) M_{y_t}^\dagger(x_t)}{P(y_t | \theta, \mathbf{x}_t, \mathbf{y}_{t-1})}. \quad (3)$$

Equations (1) and (3) define the update of the input probe density matrix from the event  $t$  to event  $t+1$ . The update of the control parameter  $x_{t+1}$  is instead determined by the classical data processing unit of the model that uses  $y_t$  as input information. An essential ingredient of such a procedure is the Bayesian inference formula,

$$P(\theta | \mathbf{x}_t, \mathbf{y}_t) := \frac{P(y_t | \theta, \mathbf{x}_t, \mathbf{y}_{t-1}) P(\theta | \mathbf{x}_{t-1}, \mathbf{y}_{t-1})}{\sum_{\theta'} P(y_t | \theta', \mathbf{x}_t, \mathbf{y}_{t-1}) P(\theta' | \mathbf{x}_{t-1}, \mathbf{y}_{t-1})}, \quad (4)$$

which allows one to update the posterior probability  $P(\theta | \mathbf{x}_{t-1}, \mathbf{y}_{t-1})$  acquired at the beginning of the  $t$  event to the posterior probability  $P(\theta | \mathbf{x}_t, \mathbf{y}_t)$  of the next event.

## II. PLATFORMS FOR QUANTUM METROLOGY

### A. Experiments on the nitrogen-vacancy center

The first family of examples presented in Sec. III are applications on the platform of single electronic spins in diamond. The nitrogen-vacancy (NV) center is a point defect of the diamond crystal that allows for the initialization, detection, and manipulation of its electronic spin. It exhibits an exceptionally long quantum coherence time, maintaining this property even at room temperature. Consequently, it has found applications in areas such as magnetometry, thermometry, and stress sensing [62–65]. The  $e^-$  spin of the NV center is a two-level system, and the Ramsey measurement, upon which the examples in this paper are based, involves the initialization of such spin in the coherent state  $|\psi\rangle := (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$  with a  $\pi/2$  microwave (MW) pulse, where  $|\uparrow\rangle$  ( $|\downarrow\rangle$ ) is the spin-up (spin-down) state. The spin is then left to freely evolve and interact with the environment in order to encode the target parameters. The duration  $\tau$  of the time interval of free evolution is the control parameter in this platform, changed from measurement to measurement. A second  $\pi/2$  MW pulse closes the encoding stage. When the NV center is excited with green light, according to the state of the  $e^-$  spin, it has different probabilities of decaying through a radiative or a nonradiative path, which means that the number of photoluminescence photons emitted is different for the two states of the spin. This mechanism allows a reliable direct measurement of the spin even at room temperature. We can also tune an extra phase  $\varphi$  of the spin evolution through the MW pulse. There are two possible choices for the resources on this platform, either we fix the maximum number of Ramsey measurements  $M_{\max}$ , or we fix the maximum total free evolution time  $T = \sum_{t=0}^{M_{\max}-1} \tau_t$ , which is the sum of the evolution times in each Ramsey measurement. In this second case, the measurement number  $M$  is a stochastic variable. We demonstrated the applicability of our machine-learning methods for single and

multiparameter metrology on various estimation tasks on NV centers including both DC [27] and AC magnetometry, decoherence estimation [66], and the characterization of the hyperfine coupling with a  $^{13}\text{C}$  nucleus [67].

### B. Experiments with photonic circuits

The second family of examples we study is based on photonic circuits [68–72]. The systems we simulated can all be realized with lasers as photon sources and passive elements, like phase plates and beam splitters (BS), together with number resolving photon counters. We assume to have programmable elements, which means that we can dynamically change the values of the phase imprinted by a plate and of the transmissivity of a BS on a timescale faster than the time interval between the measurements. This is necessary if we want to have adaptivity in the estimation. In the examples of this paper, we range from using a single bosonic mode system to controlling ten modes. The resources for this class of experiments are either the number of input states or the number of photons in a signal. We studied multiphase estimation, the agnostic Dolinar receiver [31], and coherent states classification, both in the case where the states are classically known and in the case they must be learned from a quantum training set. We have avoided the use of active elements in all these experiments, which would generate single or multimode squeezing. We also avoided the use of other nonclassical states of light like Fock states, which would require an implementation of the physics of the sensor that goes beyond that of Gaussian systems. This is a limitation only of the presented applications, and nothing prevents our framework from being useful also for those systems.

## III. APPLICATIONS ON THE NITROGEN-VACANCY CENTER PLATFORM

All the NV center applications share the same input to the agent, independently of the nature or the number  $d$  of parameters to estimate. The construction of this input is presented in Sec. III A. Similarly all these application share also the same loss, see Sec. III B.

### A. Input to the neural network

The input to the NN is obtained by concatenating, at each iteration of the measurement loop, the estimators for the unknown parameters, their standard deviations, their correlation matrix, the total number of consumed resources up to that point, and the index of the measurement iteration. All these variables are rescaled to make them fit in the  $[-1, 1]$  interval, which makes them more suitable to be the inputs of a NN. More precisely the input at the  $t$ th measurement step is composed of

(1) the mean of the posterior given by

$$\hat{\theta}_t := \int \theta P(\theta | x_t, y_t), \quad (5)$$

normalized to lay in the interval  $[-1, 1]$ , which is possible since the prior is uniform with known extrema. These inputs are  $d$  scalars, where  $d$  is the number of parameters to estimate, and will be indicated with the symbol  $\tilde{\theta}_t$  in the following.

(2) standard deviations around the mean for each parameter computed from the Bayesian posterior distribution. Given the covariance matrix  $\Sigma_t$  defined as

$$\Sigma_t := \int (\theta - \hat{\theta}_t)(\theta - \hat{\theta}_t)^\top P(\theta | x_t, y_t) d\theta, \quad (6)$$

the next  $d$  inputs to the NN are given by the vector  $\tilde{\sigma}_t$ , with entries

$$\tilde{\sigma}_{t,j} := -\frac{2}{10} \ln \sqrt{\Sigma_{t,jj}} - 1, \quad (7)$$

being  $\sqrt{\Sigma_{t,jj}}$  the said standard deviations. This time, since we do not know in advance the admissible values for the covariance matrix, we cannot cast the standard deviation exactly in  $[-1, 1]$ , but we can do it approximately for standard deviations in the range  $(10^{-5}, 1)$ , through the above formula. These inputs are  $d$  scalars.

(3) the correlation matrix  $\chi_t$  between the parameters, computed as

$$\chi_{t,ij} := \frac{\Sigma_{t,ij}}{\sqrt{\Sigma_{t,ii}\Sigma_{t,jj}}}. \quad (8)$$

This matrix does not need the normalization since its entries are already in the interval  $[-1, 1]$ . The matrix  $\chi_t$  is flattened and each entry is added to the input of the NN. These are  $d^2$  further scalars.

(4) the iteration index of the measurement loop  $t$  normalized in  $[-1, 1]$ , according to the maximum number of measurement steps  $M_{\max}$ , fixed beforehand. This input is a single scalar, indicated with  $\tilde{t} := 2t/M_{\max} - 1$ .

(5) the amount of consumed resources  $R_t$  normalized in  $[-1, 1]$ , according to the maximum amount of resources,  $R$ , also fixed beforehand. This is a single scalar indicated with  $\tilde{R}_t := 2R_t/R - 1$ . For the NV center applications the resources are either the number of measurements, i.e.,  $R_t := t$  and  $R := M_{\max}$ , or the total free evolution time, i.e.,  $R_t := \sum_{k=0}^t \tau_k$  and  $R := T_{\max}$ .

The total length of the NN input as a function of the number of parameters is  $n_t := d^2 + 2d + 2$ . If the NN is used for the nonadaptive strategy it receives in input only the two scalars  $\tilde{R}_t$  and  $\tilde{t}$ , this is the case for the decoherence estimation in Sec. III E.

### B. Definition of the precision figure of merit

The error on the estimation task for the NV center is

$$\mathcal{L}(\lambda) := \text{tr}[G \cdot K(\lambda)], \quad (9)$$

where  $K(\lambda)$  is the mean error matrix of the estimator  $\hat{\theta}$  on the batch of  $B$  parallel simulations, i.e.,

$$K(\lambda) := \sum_{k=1}^B (\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top. \quad (10)$$

$G \geq 0$  is the weight matrix used to obtain a scalar error for the multiparameter metrological problem, and  $\lambda$  are the trainable variables of the agent. The estimator  $\hat{\theta}$  is the mean of the posterior defined in Eq. (5) but  $K(\lambda)$  ought not to be confused with the covariance  $\Sigma_t$  defined in Eq. (6). The latter refers to a single estimation, the former to multiples; in other words  $K(\lambda)$  is the empirical dispersion of the estimator, computed through

multiple experiments, while  $\Sigma_t$  is the uncertainty on  $\hat{\theta}_t$  of a single estimation. The mean error in Eq. (9) can be expanded as

$$\mathcal{L}(\lambda) = \sum_{k=1}^B \text{tr}[G \cdot (\hat{\theta}_k - \theta_k)(\hat{\theta}_k - \theta_k)^\top], \quad (11)$$

from which it is clear that the error  $\ell(\hat{\theta}_k, \theta_k)$  for each single estimation in the batch should be

$$\ell(\hat{\theta}_k, \theta_k) := \text{tr}[G \cdot (\hat{\theta}_k - \theta_k)(\hat{\theta}_k - \theta_k)^\top], \quad (12)$$

where the subscript  $k$  is the index of the simulation inside the batch. For those examples with a fixed maximum amount of free evolution time  $T_{\max}$  we use the cumulative loss as a figure of merit for the precision in the training:

$$\mathcal{L}_{\text{cum}}(\lambda) := \frac{1}{M_{\max}B} \sum_{t=0}^{M_{\max}-1} \sum_{k=1}^B \frac{\ell(\hat{\theta}_{k,t}, \theta_k)}{\eta(\theta_k, T_{t,k})}, \quad (13)$$

with normalizing factor  $\eta(\theta_k, T_k)$  given by

$$\eta(\theta_k, T_{t,k}) := \min \left( \sum_{i=1}^d G_{ii} \frac{(b_i - a_i)}{12}, \frac{1}{T_{t,k}} \right), \quad (14)$$

where the sum is over the  $d$  parameters  $\theta$ , being  $(a_i, b_i)$  the extrema of the uniform prior on the  $i$ th parameter, and  $G_{ii}$  the  $i$ th diagonal entry of the weight matrix. The quantity  $T_{t,k}$  is the total elapsed evolution time, i.e.,

$$T_{t,k} := \sum_{m=0}^t \tau_{m,k}, \quad (15)$$

which plays the role of the amount of consumed resources and can be different across the batch of estimations. The quantity  $\tau_{t,k}$  is the evolution time at the iteration index  $t$ , for the  $k$ th instance of the estimation in the batch. The figure of merit  $\mathcal{L}_{\text{cum}}$  is designed to take into account also the precision of the intermediate results, and not only the error of  $\hat{\theta}$  at the end of the estimation. For those examples referring to a measurement-limited estimation we use the logarithm loss, i.e.,

$$\mathcal{L}_{\log}(\lambda) := \frac{1}{M_{\max}} \sum_{t=0}^{M_{\max}-1} \ln \left[ \frac{1}{B} \sum_{k=1}^B \ell(\hat{\theta}_{k,t}, \theta_k) \right], \quad (16)$$

which is a version of the cumulative loss that does not require the normalizing factor  $\eta$  and has been found to work better for measurement-limited estimations.

## C. DC magnetometry with time and phase control

### 1. Description of the task

The NV center electron spin is sensitive to magnetic fields; for example, static fields determine the electron Larmor frequency, which can be measured as an accumulated phase by a Ramsey experiment. The spin projection measurement that follows has a binary outcome according to the selected spin state. Indicating with  $\pm 1$  these two outcomes their probabilities are

$$p(\pm 1 | \omega, T_2, \tau) := \frac{1}{2} \pm \frac{1}{2} e^{-\tau/T_2} \cos(\omega\tau + \varphi). \quad (17)$$

This is the quantum sensor's model that has to be hard coded for the RL framework to be applicable. We can easily identify the theoretical description of the model using the convention defined in Sec. ID. The initial state for the system is the state  $|+\rangle\langle+|$ , while the physical map  $\Phi_\theta^{(t,x_t)}$  is the same at each step and can be divided in the usual evolution of a spin under the action of a magnetic field, with Hamiltonian

$$\hat{H} := \frac{\hbar\omega}{2} \hat{\sigma}_z, \quad (18)$$

followed by a  $\varphi$  phase rotation and a dephasing of the state. The phase  $\omega := \gamma B$  represents the unknown precession frequency to be estimated, which is proportional to the static magnetic field  $B$  with  $\gamma \simeq 28$  MHz/mT. The unitary component of the evolution is

$$\hat{U}(\tau, \varphi) := \exp \left[ -i \left( \frac{\hat{H}\tau}{\hbar} + \frac{\varphi}{2} \sigma_z \right) \right], \quad (19)$$

and we define the action on the state as

$$\mathcal{U}_{\tau,\varphi}(\hat{\rho}) := \hat{U}(\tau, \varphi) \hat{\rho} \hat{U}(\tau, \varphi)^\dagger. \quad (20)$$

The action of the dephasing integrated over the evolution time  $\tau$  can be described by a depolarizing dissipative term that is written in the form of an LCPT map:

$$\Phi_\tau(\hat{\rho}) := \hat{\rho} e^{-\tau/T_2} + \frac{\mathbb{1}}{2} (1 - e^{-\tau/T_2}), \quad (21)$$

so that the map  $\Phi_\theta^{(t,x_t)}$  defined in the breakdown of Sec. ID is

$$\Phi_\theta^{(t,x_t)} := \Phi_\tau \circ \mathcal{U}_{\tau,\varphi_t}, \quad (22)$$

with  $x_t := (\tau_t, \varphi_t)$ . The free evolution time  $\tau$  and the phase  $\varphi$  are controlled by the trainable agent, while  $\omega$  is the unknown parameter to be estimated. The parameter  $T_2$  denotes the transverse relaxation time, serving as the timescale for the dephasing induced by magnetic noise. Mostly this is caused by the  $^{13}\text{C}$  in the diamond lattice. In some of the examples of this section we perform multiparameter estimation, indeed the transverse relaxation time  $T_2$  may or may not be an unknown in the estimation. The prior on the frequency  $\omega$  is uniform in  $(0, 1)$  MHz. In the simulations with unknown  $T_2^{-1}$ , a narrow prior, uniform in  $T_2^{-1} \in (0.09, 0.11)$  MHz, was chosen. The optimization of the NV center as a magnetometer has been extensively studied in the literature with analytical tools [73,74], with numerics [75–87], and with machine learning [27,88,89], but with the model-aware RL implemented by qsensoropt, we were able to outperform these works and redefine the state of the art for the performances of DC magnetometry with NV centers.

### 2. Discussion of the results

We conducted multiple estimations summarized in Fig. 3, where we compared the performances of the optimized adaptive (NN) and nonadaptive strategies against the particle guess heuristic (PGH) [90], which is a commonly referenced strategy in the literature. Additionally, we introduced a variant of the  $\sigma^{-1}$  strategy [74], named  $\sigma^{-1} \& T^{-1}$ , which accounts for the finite coherence time. According to the  $\sigma^{-1} \& T^{-1}$  strategy, the evolution time  $\tau_t$  is computed from the covariance

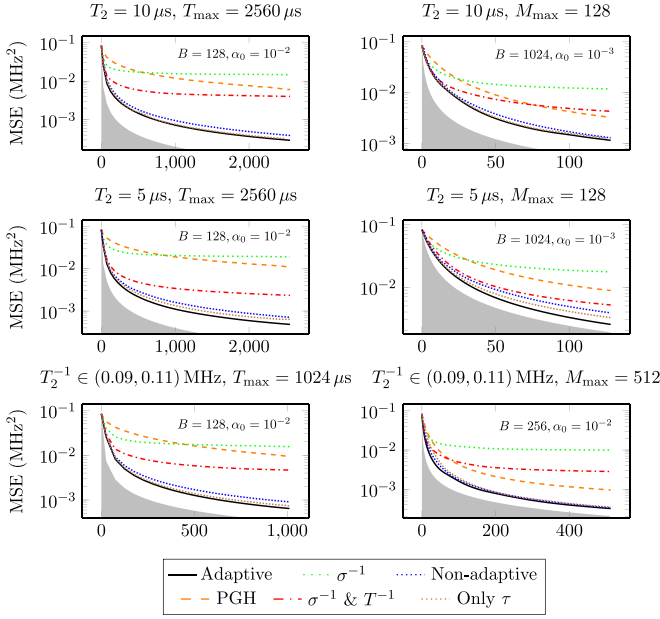


FIG. 3. The performances of the optimized adaptive and non-adaptive strategies are reported in this plots for DC magnetometry on NV center with time and phase control, together with the same strategies commonly used in the literature discussed in Sec. III C. Along with the NN optimized to control  $\tau$  and  $\varphi$ , the performances of the NN trained to control only the evolution time  $\tau$  are reported as the dotted brown line. The first two lines of plots have a fixed  $T_2^{-1}$ , with only the precession frequency  $\omega$  being estimated. In contrast, the last line refers to the estimation of both  $T_2^{-1}$  and  $\omega$  simultaneously, i.e.,  $\theta = (\omega, T_2^{-1})$ . In both cases  $G = \mathbb{1}$ . The shaded gray areas indicate the Bayesian Cramér-Rao bound, which is the ultimate precision bound computed from the Fisher information, which can be found in Appendix A 3. For these estimations, we have used  $N = 480$  particles for the particle filter. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the “adaptive” strategy.

matrix  $\Sigma_t$  of the current posterior distribution as

$$\tau_t = [\text{tr}(\Sigma_t)^{\frac{1}{2}} + \widehat{T}_2^{-1}]^{-1}, \quad (23)$$

while the standard  $\sigma^{-1}$  strategy prescribes  $\tau_t = \text{tr}(\Sigma_t)^{-\frac{1}{2}}$ . In the case of a fixed  $T_2^{-1}$ , this value was used instead of its estimator  $\widehat{T}_2^{-1}$ . For computing the controls of the PGH strategy, two particles  $\theta_1$  and  $\theta_2$  are drawn from the Bayesian posterior distribution; the evolution time is then computed as

$$\tau_t = (\|\theta_1 - \theta_2\|_2 + \varepsilon)^{-1}, \quad (24)$$

with  $\varepsilon := 10^{-5}$  MHz. As a function of the normalized input to the NN, the controls  $\tau$  and  $\varphi$  are obtained from

$$\begin{pmatrix} \tau_t \\ \varphi_t \end{pmatrix} = \begin{pmatrix} h \\ \pi \end{pmatrix} |f_{\text{NN}}(\tilde{\theta}_t, \tilde{\sigma}_t, \chi_t, \tilde{R}_t, \tilde{r})| + \begin{pmatrix} 1 \mu\text{s} \\ 0 \end{pmatrix}, \quad (25)$$

where  $h$ , which stands for “height,” is a prefactor that is chosen to be appropriate for each simulation, and should be of the order of magnitude of the expected optimal  $\tau_t$ . The NN is the function  $f_{\text{NN}}$ . The prefactors  $h$  for our experiments appear in Table I. For the time-limited estimations these prefactors are chosen to be roughly in accordance with the maximum value

TABLE I. Prefactor  $h$  appearing in Eq. (25) for DC magnetometry on an NV center.

	Time	Measurement
$T_2 = \infty$	$\frac{T_{\text{max}}}{20}$	$\lceil 2\sqrt{M_{\text{max}}} \rceil \mu\text{s}$
$T_2 < \infty$	$\max\{\frac{T_{\text{max}}}{20}, T_2\}$	$\max\{\lceil 2\sqrt{M_{\text{max}}} \rceil \mu\text{s}, T_2\}$
$T_2^{-1} \in (a, b)$	$\max\{\frac{T_{\text{max}}}{20}, a^{-1}\}$	$\max\{\lceil 2\sqrt{M_{\text{max}}} \rceil \mu\text{s}, a^{-1}\}$

of the control predicted by the analysis in Refs. [59,91], for the measurement-limited case see Appendix C. The constant shift added in Eq. (25) and the absolute value are necessary to keep  $\tau_t$  strictly positive, which is helpful for the convergence of the training. The NN has been pretrained to reproduce a linear ramp for the control time  $\tau$ , starting from  $1 \mu\text{s}$  and reaching  $\tau = h$  at the end of the estimation. This behavior has been suggested to us by the analytical results exposed in Appendix C and in Ref. [59]. The same initial controls are used for the nonadaptive strategy. Besides the optimized adaptive strategy with time and phase controls we reported also, under the name “Only  $\tau$ ,” the optimized precisions achieved through controlling the time  $\tau_t$  only, also reported in Ref. [55]. From the plots in Fig. 3 we conclude that there is only a very small advantage in controlling the phase for  $T_2 = 10 \mu\text{s}$ , if there is any at all. Similarly, with  $T_2 = \infty$ ,  $100 \mu\text{s}$  no advantage has been found, although the plots have not been reported. For  $T_2 = 10 \mu\text{s}$  and  $T_{\text{max}} = 2560 \mu\text{s}$  the phase control had converged to the constant  $\varphi = \pi$ , and more training could not take it out of this minimum. From the plot for  $T_2 = 5 \mu\text{s}$  we see that the advantage of controlling the phase  $\varphi$  grows as the coherence time becomes smaller. The results of the simulations with  $T_2^{-1} \in (0.09, 0.11)$  MHz are very similar to that with  $T_2 = 10 \mu\text{s}$  because of the relatively narrow prior on  $T_2^{-1}$ .

### 3. Future directions

Typically, in the applications, the meaningful resource is the total time required for the estimation. This does not coincide however with  $T = \sum_{t=0}^{M_{\text{max}}-1} \tau_t$ , because the initialization of the NV center, the readout, and the data processing all take time. This overhead time is proportional to the number of measurements, so, in a real experiment, we expect the actual resource to be a combination of the evolution time  $T$  and the number of measurements,  $M$ . In future work, the role of the higher moments of the Bayesian posterior distribution in the determination of the controls should be explored. In particular, it should be understood if with

$$\begin{pmatrix} \tau_t \\ \varphi_t \end{pmatrix} = \begin{pmatrix} h \\ \pi \end{pmatrix} |f_{\text{NN}}(\tilde{\theta}_t, \tilde{\sigma}_t, \chi_t, \tilde{\gamma}_1, \tilde{\mu}_4, \tilde{\mu}_5, \dots, \tilde{R}_t, \tilde{r})| + \begin{pmatrix} 1 \mu\text{s} \\ 0 \end{pmatrix}, \quad (26)$$

more precision from the adaptivity can be achieved. In this formula  $\tilde{\gamma}_1$  is the skewness of the posterior and  $\tilde{\mu}_i$  are its higher moments. A further improvement of this work would be to implement nonsingle-shot readout of the NV center state, since at room-temperature it is the only way to measure the spin, as done by Zohar *et al.* [82]. The decoherence model

we have used in this example and in all the others of the section works well for surface NV center, while for bulk centers a better model is

$$p(\pm 1|\omega, T_2, \tau) := \frac{1}{2} \pm \frac{1}{2} e^{-(\tau/T_2)^2} \cos(\omega\tau + \varphi). \quad (27)$$

The majority of the theoretical works, especially those papers involving ML [27], refer to Eq. (17), and we wanted to be consistent with this trend in order to facilitate the comparisons. A further study could be treating the exponent of the dephasing term as a nuisance parameter.

## D. AC magnetometer

### 1. Description of the task

In this section, we study the estimation of the intensity of an oscillating magnetic field of known frequency with an NV center used as an AC magnetometer. The NV center spin precesses in the magnetic field of intensity  $B$  and frequency  $\omega$ , then its state is observed. The model for the binary outcome of the Ramsey measurement is

$$p(\pm 1|\Omega, T_2, \tau) := \frac{1}{2} \pm \frac{1}{2} e^{-\tau/T_2} \cos\left[\frac{\Omega}{\omega} \sin(\omega\tau)\right]. \quad (28)$$

These probabilities can be found with a physical setup very similar to the one described in Sec. III C, where we need to change the Hamiltonian of the system, which now oscillates with frequency  $\omega$ , i.e.,

$$\hat{H} := \frac{\hbar\Omega}{2} \cos(\omega\tau) \hat{\sigma}_z, \quad (29)$$

in this situation we also neglect the controllable phase  $\varphi$ , i.e., we set  $\varphi = 0$  for all the measurement steps. The evolution time  $\tau$  is controlled by the trainable agent, while  $\Omega := \gamma B$  is the unknown parameter to be estimated. The parameter  $T_2^{-1}$  may or may not be an unknown in the estimation; in all cases  $G = \mathbb{1}$ . The prior on  $\Omega$  is uniform in  $(0, 1)$  MHz for all the examples. The formula for  $\tau_t$  is

$$\tau_t = (1 \mu\text{s}) \times |f_{\text{NN}}(\tilde{\theta}_t, \tilde{\sigma}_t, \tilde{\chi}_t, \tilde{R}_t, \tilde{t})| + 1 \mu\text{s}. \quad (30)$$

The weights and biases of the NN are randomly initialized, and so are the controls for the nonadaptive strategy.

### 2. Discussion of the results

The results of the strategy optimization for this model are reported in Fig. 4. Remarkably the time-limited estimations with long coherence time can saturate the bound set by the Fisher information. The model used in this example is formally equivalent to Eq. (17) with  $\varphi = 0$ , where the adaptive strategy gives only small advantages, at difference with the results obtained here for AC magnetometry. This is because the AC model maps to the DC one in a very different region of parameters with respect to the region we have explored in Sec. III C.

### 3. Future directions

In AC magnetometry the technique of dynamical decoupling is often used to improve the sensitivity with respect to  $\Omega$  and to increase the coherence time  $T_2$ . This consists of a series of  $\pi$  pulses that reverse the sign of the accumulated phase.

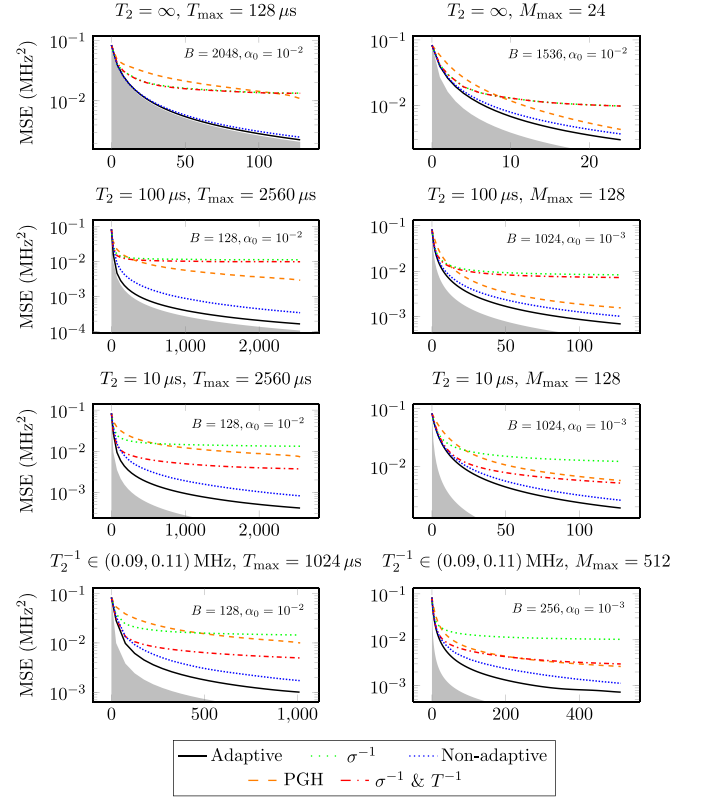


FIG. 4. This figure refers to the problem of AC magnetometry on an NV center with known frequency  $\omega = 0.2$  MHz; see Eq. (28) for the model. The mean-square error (MSE) of the optimized NN and static strategies are plotted, together with the performances of the  $\sigma^{-1}$ , the  $\sigma^{-1} & T^{-1}$ , and the PGH strategies, discussed in Sec. III C. On the left, the precisions of the time-limited simulations are reported, while the plots on the right refer to the measurement-limited estimation. For all the plots  $G = \mathbb{1}$ . The shaded gray areas indicate the Bayesian Cramér-Rao bound, which is the ultimate precision bound computed from the Fisher information, which can be found in Appendix A 4. For these estimations, we used  $N = 480$  particles for the particle filter. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the adaptive strategy.

Given  $t_i$ ,  $0 < i < L + 1$ , the times at which the instantaneous  $\pi$  pulse are applied, the outcome probabilities for a Ramsey measurement on a noiseless NV center is

$$p(\pm 1|\Omega, T_2, \tau) := \frac{1}{2} \pm \frac{1}{2} \cos\left[\frac{\Omega}{\omega} \sum_{i=1}^{L+1} (-1)^i [\sin(\omega t_i) - \sin(\omega t_{i-1})]\right], \quad (31)$$

with  $L$  being the number of pulses, and  $t_0$  and  $t_{L+1}$  being respectively the initialization and the measurement times. In the current simulations we have not used  $\pi$  pulses, but optimizing their application is the natural extension of this example, left for future works. For controlling the pulses there are three possibilities:

(1) The interval between all the pulses is fixed to  $\tau = t_{i+1} - t_i$ , which is produced by a NN together with the number



of pulses  $L$ . In this case, we have two controls, one of them  $L$  being discrete.

(2) The controls are the  $L$  time intervals  $\tau_i = t_i - t_{i-1}$ . The number of pulses is fixed but they can be made ineffective with  $\tau_i = 0$ .

(3) The control is the free evolution time  $\tau$  together with a boolean variable, which tells whether a pulse or a measurement has to be applied after the free evolution. This would require a stateful model for the NV center.

The problem of estimating the magnetic field knowing the frequency is complementary to the protocol for the optimal discrimination of frequencies [73], which has been applied to distinguish two chemical species in a sample. In this work the authors put forward an optimal strategy for the discrimination of two frequencies  $\omega$  and  $\omega + \Delta\omega$ , knowing the intensity of the field. If  $\Delta\omega \ll \omega$  and the field intensity is unknown, then a two stage approach to the problem is possible. We first estimate the intensity of the field as done in this example, while considering the frequency  $\omega$  to be fixed and known, then we proceed with the optimal frequency discrimination based on the intensity just estimated. The error probability of the second stage depends on the precision of the first. Given a maximum total time for the discrimination, the assignment of time to the first and the second stages can be optimized to minimize the final error. It is important for the first stage to have low sensitivity to variations in the frequency, and for  $\Delta\omega \simeq 0$  we expect this two stage protocol to be close to optimality. For optimizing the frequency discrimination with the field intensity as a nuisance parameter in a fully integrated protocol the introduction of  $\pi$  pulses is necessary, as they have been used in Ref. [73]. The natural extension of frequency discrimination is the estimation of both the intensity and the frequency of the magnetic field optimally, both starting from a broad prior. All these improvements are left for future work.

## E. Decoherence estimation

### 1. Description of the task

In this example, we study an NV center subject to variable decoherence of the dephasing type, that we want to characterize. The model for the binary outcome of the Ramsey measurement is

$$p(\pm 1|T, \beta, \tau) := \frac{1 \pm e^{-(\tau/T)^\beta}}{2}, \quad (32)$$

that can be obtained with a dynamical model similar to the one described in Sec. III C, where we need to set the magnetic field  $B$  or equivalently the frequency  $\omega$  to zero and change slightly the noise model, using

$$\Phi_\tau(\hat{\rho}) := \hat{\rho} e^{-(\tau/T)^\beta} + \frac{\mathbb{1}}{2} [1 - e^{-(\tau/T)^\beta}]. \quad (33)$$

The evolution time  $\tau$  is controlled by the trainable agent, while  $T^{-1}$  and  $\beta$  are the two unknown characteristic parameters of the dephasing noise, i.e., respectively the transverse relaxation time and the decay exponent. In the applications, the coherence time of the noise encodes some useful information about the environment, like the concentration of radicals in a biological sample or the transport property of a material.

The priors on  $T^{-1}$  and  $\beta$  are uniform in (0.01, 0.1) MHz and (1.5, 4.0), respectively.

## 2. Discussion of the results

In this section, we compare the results of various strategies to control  $\tau$ , which are the

(1) optimized adaptive strategy with a trained NN, which outputs the control according to

$$\tau_t = (100 \mu\text{s}) \times |f_{\text{NN}}(\tilde{\theta}_t, \tilde{\sigma}_t, \chi_t, \tilde{R}_t, \tilde{t})| + 1 \mu\text{s}. \quad (34)$$

(2) static strategy implemented with a NN that receives in input only  $\tilde{R}_t$  and  $\tilde{t}$ , i.e.,

$$\tau_t = (100 \mu\text{s}) \times |f_{\text{NN}}(\tilde{R}_t, \tilde{t})| + 1 \mu\text{s}. \quad (35)$$

(3) random strategy, where the inverse of the evolution time  $\tau^{-1}$  is chosen randomly and uniformly in the support of the prior for  $T^{-1}$ , i.e., (0.01, 0.1) MHz.

(4) inverse time strategy. In this case, the evolution time is  $\tau := \alpha_M^{1/\beta} \widehat{T}^{-1}$  for the measurement-limited estimation and  $\tau := \alpha_T^{1/\beta} \widehat{T}^{-1}$  for the time-limited estimations, where  $\alpha_M = 0.79681$  and  $\alpha_T = 0.43711$ . These numerical coefficients come from the optimization of the Fisher information reported in Appendix A 5. For the simulations where  $\beta$  is treated as a nuisance parameter,  $\alpha_M$  is used also for the time-limited estimation. This strategy not only depends on the current estimate  $\widehat{T}^{-1}$  for the decoherence time but also on the estimator for the decay coefficient  $\widehat{\beta}$ , and it is, of course, adaptive.

Notice that, differently from the other NV center examples, the nonadaptive strategy for this application has been implemented not as a list of controls for each individual measurement but as a NN. The results of the strategy optimization for this model are reported in Fig. 5. Both the adaptive and nonadaptive strategies have been pretrained to reproduce a linear ramp that reaches the maximum  $\tau = 100 \mu\text{s}$  at the end of the estimation. From these simulations, we conclude that there is no advantage in using an NN instead of the strategy that optimizes the Fisher information, except when we are interested in the estimation of  $\beta$ .

## 3. Future directions

Identifying the decay exponent has applications in distinguishing the type of NV center, surface of bulk, that we are dealing with. Such a problem would be formulated in terms of a discrimination task between  $\beta = 1, 2$ , with  $T$  as the nuisance parameter.

## F. Hyperfine coupling estimation

### 1. Description of the task

In this example, we study the measurement of an NV center strongly coupled to a  $^{13}\text{C}$  nuclear spin in the diamond lattice. Such nuclear spin is not hidden in the spin bath of nuclei that are responsible for the dephasing noise, instead, it causes a relatively large split of the energy levels of the NV center according to the hyperfine interaction strength and that can be measured in an experiment. The precession frequency of the NV center in a magnetic field is determined by the state of the nuclear spin. In the experiment of Joas *et al.* [67] multiple

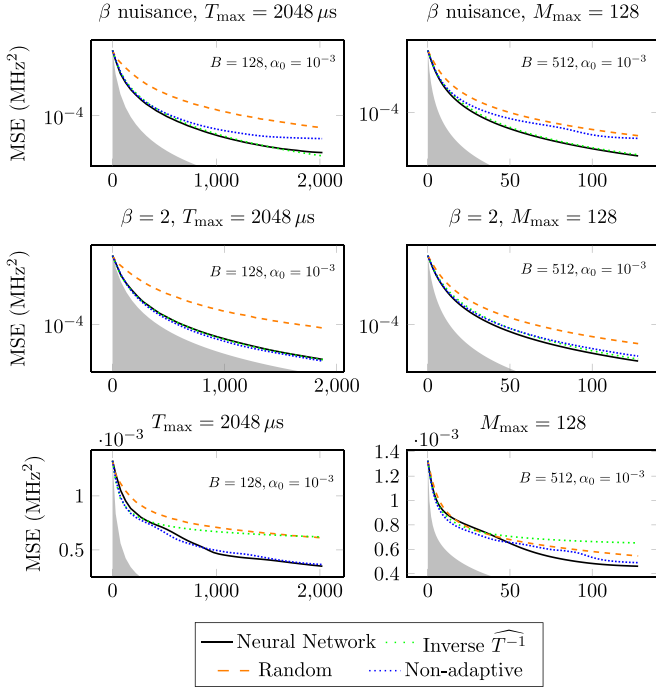


FIG. 5. In these plots the MSE for the decoherence estimation on an NV center is reported as a function of the consumed resources, with the time-limited estimations on the left and the measurement-limited ones on the right. In the first line of plots, the decaying exponent  $\beta$  is unknown and it is treated as a nuisance parameter. Therefore, only the precision on the inverse decay time appears in the loss. The plots on the second line refer to the case where  $\beta$  is known and fixed ( $\beta = 2$ ), and only  $T^{-1}$  is estimated. In the third line of plots  $\beta$  and  $T^{-1}$  are unknown and are both parameters of interest. In this case, the weight matrix is chosen to be  $G = \text{diag}(1, 1/800 \text{ MHz}^2)$ , to compensate for the different order of magnitude of the values of the parameters  $\beta$  and  $T^{-1}$ . The shaded gray areas indicate the Bayesian Cramér-Rao bound, which is the ultimate precision bound computed from the Fisher information, see Appendix A 5 for details. For these estimations, we have used  $N = 2048$  particles for the particle filter. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the “neural network” strategy.

incoherent nuclear-spin flips happen during the read out, so that the nuclear spin is in each eigenstate approximately half of the time. This motivates the choice for model probability of the Ramsey measurement:

$$p(\pm 1 | \omega_0, \omega_1, T_2, \tau) := \frac{1}{2} \pm \frac{1}{4} e^{-\tau/T_2} [\cos(\omega_0 \tau) + \cos(\omega_1 \tau)], \quad (36)$$

which we take directly from Ref. [67]. In such a model,  $\omega_0$  and  $\omega_1$  are the two precession frequencies to be estimated, split by the hyperfine interaction,  $T_2$  is the coherence time, while  $\tau$  and  $\varphi$  are the controls, that are respectively the evolution time and the phase. This model is completely symmetric under permutation of the two precession frequencies, therefore only those weight matrices  $G$  that are permutationally invariant in the two parameters should be considered for this estimation. The prior on  $(\omega_0, \omega_1)$  is the uniform distribution over the triangle in the  $(\omega_0, \omega_1)$  plane identified by the points  $(0, 0)$  MHz,  $(0, 1)$  MHz, and  $(1, 1)$  MHz, since we have decided to cancel

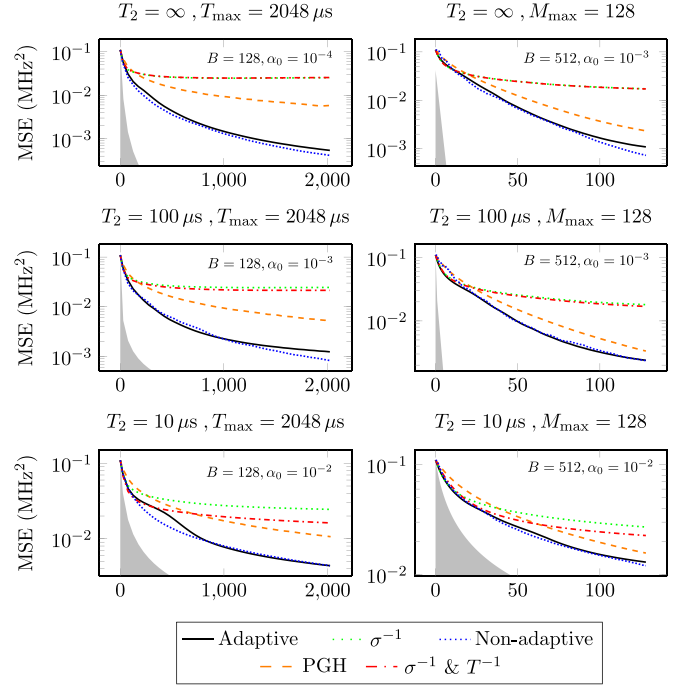


FIG. 6. Sum of the mean square errors on the frequencies  $\omega_0$  and  $\omega_1$  for the optimized adaptive (NN) and nonadaptive strategies (static) compared with other common strategies used in the literature [67] and described in Sec. III C. The shaded gray areas in the above plot indicate the Bayesian Cramér-Rao bound, which is the ultimate precision bound computed from the Fisher information, see Appendix A 6 for details. For these estimations, we have used  $N = 4096$  particles for the particle filter. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the adaptive strategy.

this permutation symmetry by fixing  $\omega_1 > \omega_0$ . An important observation to be made is that the Fisher information matrix (FI)  $I(\tau, \varphi)$  for  $\omega_0$  and  $\omega_1$  of the model in Eq. (36) is singular and remains singular even for multiple measurements with different  $\tau$  and  $\varphi$ . This means that the Fisher information is of no use in the optimization of the strategy. The controls  $\tau$  and  $\varphi$  are computed according to Eq. (25) with the prefactor

$$h = \min\left(40, \frac{T_2}{2}\right). \quad (37)$$

The means square errors of the two frequencies are weighted equally with  $G = \mathbb{1}$ . The frequency difference is the component of the hyperfine interaction parallel to the NV center quantization axis, i.e.,  $A_{||} = |\omega_1 - \omega_0|$ .

## 2. Discussion of the results

Besides the NN and the static strategies, the performances of the particle guess heuristic (PGH) and of the  $\sigma^{-1}$  strategies have been tested. The results are reported in Fig. 6. The NN is pretrained to reproduce a linear ramp for the control time, which reaches its maximum  $\tau = h$  at the end of the estimation, while the phase control is random. Similarly, the nonadaptive strategy is initially a linear ramp in  $\tau$  and is random in  $\varphi$ . From these simulations, there seems to be no significant advantage in using an adaptive strategy for the simultaneous

estimation of the two precession frequencies, neither for large nor for small coherence times  $T_2$ . Nevertheless optimizing with model-aware reinforcement learning still gives us an advantage with respect to other simpler approaches.

### 3. Future directions

There is also another way of expressing the outcome probability of the measurement: instead of defining the model in terms of  $\omega_0$  and  $\omega_1$  we could define it in terms of the frequencies sum  $\Sigma = \omega_0 + \omega_1$  and difference  $\Delta = \omega_1 - \omega_0$ , thus writing

$$p(\pm 1 | \Sigma, \Delta, T_2, \tau) := \frac{1}{2} \pm \frac{1}{4} e^{-\tau/T_2} \left[ \cos\left(\frac{\Sigma + \Delta}{2} \tau\right) + \cos\left(\frac{\Sigma - \Delta}{2} \tau\right) \right]. \quad (38)$$

In this form there is no permutational invariance in  $\Sigma$  and  $\Delta$ , instead, the model is invariant under the transformation  $\Delta \rightarrow -\Delta$ . Using this expression, we should choose a prior having  $\Delta > 0$ , which means  $\omega_1 > \omega_0$ , and we should impose the positivity of the frequencies by requiring  $\Sigma > \Delta$  in the prior. Since we are interested in the difference  $\Delta$ , the sum of the frequencies  $\Sigma$  would be treated as a nuisance parameter. The absence of an advantage of the adaptive strategy over the nonadaptive one is probably also due to the hyper-simplified information passed to the NN. We are in fact approximating a complex two-dimensional (2D) posterior, with many peaks and valleys with a Gaussian. A better approach would be to train an autoencoder to compress the information contained in the posterior and pass it to the NN. The autoencoder will be trained to compress the class of distributions that are produced by the likelihood of the Sec. III F. In future work, we plan to extend the estimation to the detection of multiple nuclear spins surrounding the NV center. For the estimation of  $n$  frequencies in the model

$$p(\pm 1 | \{\omega_i\}_{i=0}^n, T_2, \tau) := \frac{1}{2} \pm \frac{1}{2n} e^{-\tau/T_2} \sum_{i=0}^{n-1} \cos(\omega_i \tau), \quad (39)$$

an appropriate precision figure of merit would be

$$\ell(\hat{\theta}, \theta) := \min_{\pi \in S_n} \sum_{i=0}^{n-1} (\hat{\omega}_i - \omega_i)^2, \quad (40)$$

with  $\theta = (\omega_0, \omega_1, \dots, \omega_{n-1})$  and  $\hat{\theta} = (\hat{\omega}_0, \hat{\omega}_1, \dots, \hat{\omega}_{n-1})$ , and  $S_n$  being the permutation group of  $n$  elements. If we impose the condition  $\omega_i \leq \omega_{i+1}$ , then we can get rid of the minimization in the permutation. In this case, it is interesting to notice that the  $n$ -dimensional volume of the parameter space for a uniform prior on  $\omega_i$  which is null outside of  $(a, b)$  is reduced by a factor  $n!$  due to the symmetry of the parameters, i.e.,

$$\mathcal{V}_n \leq \frac{(b-a)^n}{n!}. \quad (41)$$

This may reduce the number of particles necessary to represent the Bayesian posterior and make the optimization of this application accessible.

## IV. APPLICATIONS ON THE PHOTONIC CIRCUITS PLATFORM

In this section, we report all the information and data related to the examples on the photonic platform, i.e., the agnostic Dolinar receiver, the classification of known and unknown coherent states, and multiphase discrimination in a four-arm interferometer.

### A. Agnostic Dolinar receiver

#### 1. Description of the task

In Ref. [55] we already discussed the Dolinar receiver, and in this section we briefly review the physics of this system and the task we are trying to solve. The goal consists of distinguishing between the two coherent states  $|\alpha\rangle$  and  $|\pm\alpha\rangle$  with  $\alpha$  unknown using a single copy of the signal  $|\pm\alpha\rangle$ . The Dolinar receiver, known for its good performance, traditionally uses a local oscillator (LO) synchronized with the sender's laser. Recent studies introduced an alternative, the agnostic Dolinar receiver, eliminating the need for the LO. Instead, the new device utilizes  $n$  reference states  $|\alpha\rangle$  sent alongside the signal. In this setup, classical knowledge about  $|\alpha\rangle$  is absent, treating it as an unknown parameter. The device, as shown in Fig. 7, utilizes  $|\alpha\rangle^{\otimes n}$  to discriminate the sign of the signal. The signal is combined with a reference state on a programmable BS with transmissivity  $\theta_i$ . The result of the photon counting at each BS is used to update the Bayesian posterior on  $\alpha$  and the signal's sign. A NN determines the angle  $\theta_{i+1}$  for the next BS. This task involves continuous (the signal's amplitude  $\alpha \in \mathbb{R}$ ) and discrete (signal sign) parameters. The performance of the receiver is evaluated based on the error probability in the classification of the signal's sign, while the amplitude  $\alpha$  is a nuisance parameter. With this schematization, we can identify the parameters of the dynamical model defined in Sec. I D. Differently from all the variations of the NV center studied until now, in this case, there is no dynamical evolution, meaning that the LCPT map  $\Phi_{\theta}^{t,x_r}$  is always equal to the identity. In this case, all the control is in the collection of operators  $\{\hat{M}_{y_i}(x_i)\}$ , which means that the initial state of the probe is already encoded, as it comes from a different laboratory on which we do not have any control. The input state for this procedure is the pure state  $|\alpha\rangle^{\otimes n} \otimes |\pm\alpha\rangle$ . At the  $k$ th step, the measurement operators  $\hat{M}_{y_k}(x_k)$  can be written as the product of a BS operation, implemented by  $\hat{U}^{(k,n+1)}(\theta_k)$ , that mixes the  $(n+1)$ st state, namely, the one to discriminate, with the reference state in position  $k$ , followed by a measurement on the  $k$ th component using photon counting. In equations,

$$\hat{M}_{m_k}(\theta_k) = |m_k\rangle_k \langle m_k| \hat{U}^{(k,n+1)}(\theta_k), \quad (42)$$

where we revised slightly the notation to make it closer to the standard one for a beam splitter, using  $\theta_k$  instead of  $x_k$  as the mixing angle, and using  $m_k$  instead of  $y_k$  as the integer number that represents how many photons have been counted. This formula holds for all the steps except the last one, where two photon-counting operators are present, instead of one.

#### 2. Input to the neural network

The input to the NN is the concatenation of the following eight scalar values:

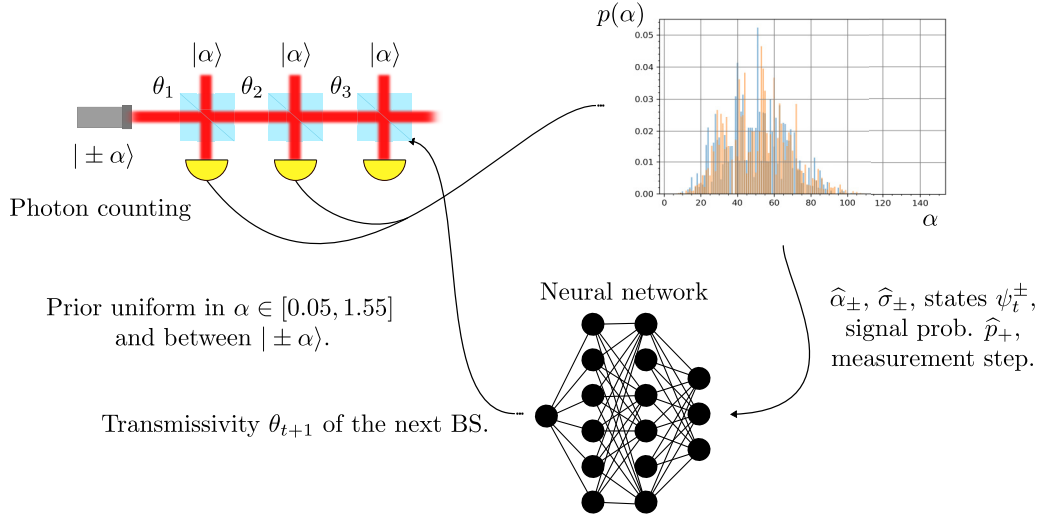


FIG. 7. Representation of the measurement loop for the Dolinar receiver. The signal state  $|\pm\alpha\rangle$  is mixed with the reference states and measured with photon counting. The posterior distribution on  $\alpha$  has two components, corresponding to the signs  $s = \pm 1$ . The intensity of the intermediate signal  $\psi_i^\pm$ , the mean and the variance of the posterior, are fed to the NN that computes the next transmissivity.

- (1) the signal intensity  $\psi_i^+$  after the  $i$ th measurements, assuming that the initial state was  $|\alpha\rangle$ ;
- (2) the posterior estimator for the initial signal intensity  $\hat{\alpha}_+$ , assuming the signal's sign was  $s = +$ ;
- (3) the variance of the posterior distribution  $\hat{\sigma}_+$  for the initial signal intensity, assuming the sign was  $s = +$ ;
- (4) the signal intensity  $\psi_i^-$  after the  $i$ th measurements, assuming that the initial state was  $|\alpha\rangle$ ;
- (5) the posterior estimator for the initial signal intensity  $\hat{\alpha}_-$ , assuming the signal's sign was  $s = -$ ;
- (6) the variance of the posterior distribution  $\hat{\sigma}_-$  for the initial signal intensity, assuming the sign was  $s = -$ ;
- (7) the posterior probability  $\hat{p}_+$  that the original signal had a  $+$  sign;
- (8) the index of the current measurement step, normalized in the interval  $[-1, +1]$  with respect to the total number of measurements, indicated with  $\tilde{r} := 2i/M_{\max} - 1$ .

The reflectivity  $\theta_i$  is produced by the NN through the following formula:

$$\theta_i = f_{\text{NN}}(\psi_i^+, \hat{\alpha}_+, \hat{\sigma}_+, \psi_i^-, \hat{\alpha}_-, \hat{\sigma}_-, \hat{p}_+, \tilde{r}) - \pi \times (n_{\text{Phot}} \bmod 2). \quad (43)$$

The symbol  $n_{\text{Phot}}$  indicates the total number of photons measured up to the point where the NN is called, which controls the addition of an extra phase  $-\pi$  to the transmissivity. The same mechanism is implemented for the nonadaptive strategy.

### 3. Definition of the precision figure of merit

The loss in the agnostic Dolinar receiver measures the error in guessing the sign of the signal  $|\pm\alpha\rangle$ , after it has been completely measured. The lower bound on the error probability given  $n$  copies of the reference states is  $p_H(\alpha)$ , and can be found in Appendix B 1. We have implemented two possible policies for guessing the sign  $s$ , that we indicate respectively

with  $\hat{s}_{\text{Bayes}}$  and  $\hat{s}_{\text{Parity}}$ . They are, respectively,

$$\hat{s}_{\text{Bayes}}(\hat{p}_+) := \begin{cases} +1 & \text{if } \hat{p}_+ > 0.5 \\ -1 & \text{if } \hat{p}_+ \leq 0.5, \end{cases} \quad (44)$$

and

$$\hat{s}_{\text{Parity}}(n_{\text{Phot}}) := \begin{cases} +1 & \text{if } n_{\text{Phot}} \bmod 2 = 0 \\ -1 & \text{if } n_{\text{Phot}} \bmod 2 = 1. \end{cases} \quad (45)$$

The Kronecker  $\delta$  function

$$\delta(x, y) := \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y, \end{cases} \quad (46)$$

is necessary to introduce the losses, which are

- (1) loss=0:  $\ell(\hat{p}_+, s) := 1 - \delta(\hat{s}_{\text{Bayes}}, s)$ ;
- (2) loss=1:  $\ell(\hat{p}_+, s) := 1 - \hat{p}_s$ ;
- (3) loss=2:  $\ell(n_{\text{Phot}}, s) := 1 - \delta(\hat{s}_{\text{Parity}}, s)$ ;
- (4) loss=3:  $\ell(\hat{p}_+, s, \alpha) := 1 - \delta(\hat{s}_{\text{Bayes}}, s) - p_H(\alpha)$ ;
- (5) loss=4:  $\ell(\hat{p}_+, s, \alpha) := 1 - \hat{p}_s - p_H(\alpha)$ ;
- (6) loss=5:  $\ell(n_{\text{Phot}}, s, \alpha) := 1 - \delta(\hat{s}_{\text{Parity}}, s) - p_H(\alpha)$ ;
- (7) loss=6:  $\ell(\hat{p}_+, s, \alpha) := \frac{1 - \delta(\hat{s}_{\text{Bayes}}, s)}{p_H(\alpha)}$ ;
- (8) loss=7:  $\ell(\hat{p}_+, s, \alpha) := \frac{1 - \hat{p}_s}{p_H(\alpha)}$ ;
- (9) loss=8:  $\ell(n_{\text{Phot}}, s, \alpha) := \frac{1 - \delta(\hat{s}_{\text{Parity}}, s)}{p_H(\alpha)}$ .

All these losses, when averaged, converge to the probability of a wrong classification. It is not obvious *a priori* which loss is the best one.

### 4. Discussion of the results

The simulation results are presented in Fig. 8. While in Ref. [55] we trained the optimized strategies with loss=3 in this paper we used loss=6, which seems however to deliver slightly worst results. In both cases, the plots represent loss=0, averaged over many executions of the estimation task.

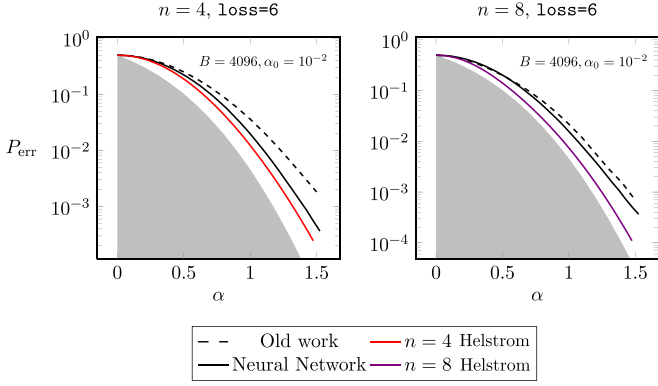


FIG. 8. Comparison of error probabilities for various strategies with different numbers of copies of  $|\alpha\rangle$ , specifically  $n = 4$  and  $n = 8$ , for  $\text{loss}=6$ . The shaded gray area is the region excluded by the Helstrom bound [92,93], which is the lowest error probability theoretically achievable when assuming having an infinite number of reference states ( $n = \infty$ ) at disposal. The solid red and violet lines are the Helstrom bound calculated for a finite number of copies of  $|\alpha\rangle$  [31], respectively  $n = 4$  and  $n = 8$ . For the details on the computation of the Helstrom bound see Appendix B 1. The black dashed line showcases the lowest error found in the old work [31], without machine learning, while the black solid line is the performance achieved using the NN. The performances of the optimal nonadaptive strategies have not been reported since they cannot rival those of the NN. For both the training and the performance evaluation we used  $N = 512$  particles. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the neural network strategy.

**B. Quantum machine-learning classification of states**

In this example, we put forward a quantum-machine-learning-based (QML-based) classifier able to distinguish between three coherent states  $|\alpha_0\rangle, |\alpha_1\rangle, |\alpha_2\rangle$ , with  $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{C}$ , given  $n$  copies of each of them, which constitute the quantum training set. The signal state is a single copy of a coherent state, promised to be one of the three training states, i.e.,  $|\alpha_s\rangle$  with  $s = 0, 1, 2$ . The priors on the components of the training states  $|\alpha_0\rangle, |\alpha_1\rangle, |\alpha_2\rangle$  are uniform in the harmonic-oscillator

phase space in a square of side  $2\alpha$  centered at the origin, that is,

$$\alpha_i^R := \text{Re}(\alpha_i) \in (-\alpha, +\alpha), \quad \alpha_i^I := \text{Im}(\alpha_i) \in (-\alpha, +\alpha), \tag{47}$$

for  $i = 1, 2, 3$ . Also the prior on the signal classes  $s = 0, 1, 2$  is uniform. What makes this task difficult is that the complex numbers  $\alpha_0, \alpha_1, \alpha_2$  are not precisely known, not at least beyond the information coming from the prior. There are therefore seven parameters to be estimated: the real and imaginary components of these three complex amplitudes (six continuous parameters in total), plus the class of the signal  $s$ , which is a discrete parameter. The device that performs this task is made of programmable BS only, of which we control the transmissivity  $\theta$  and the phase  $\phi$ , and is represented in Fig. 9. The resource consumed in a single execution of the discrimination task is the average number of photons in the device, i.e.,

$$N_{\text{ph}} := n(|\alpha_0|^2 + |\alpha_1|^2 + |\alpha_2|^2) + |\alpha_s|^2. \tag{48}$$

In this application, we experimented also with using a ternary decision tree as an agent instead of a NN. The device presented in this section can also be interpreted as a generalization of the agnostic Dolinar receiver of Sec. IV A.

**1. Input to the agent**

Let us define the posterior probabilities of the signal being respectively  $s = 0, 1, 2$  as  $\hat{p}_0, \hat{p}_1$ , and  $\hat{p}_2$ . This distribution is computed as the marginal of the Bayesian posterior by integrating over the continuous variables, and can be computed on the particle filter. In case the agent is a NN, then it receives as input the following 19 scalars:

(1) the average state of the signal after the  $t$ th photon counting measurements. These are two scalar values, i.e., the real and imaginary parts of the coherent state, and will be indicated collectively as  $\hat{\psi}_t$ . The average is taken over the posterior distribution.

(2) the mean posterior estimator for the real and imaginary components of the complex amplitudes of the three

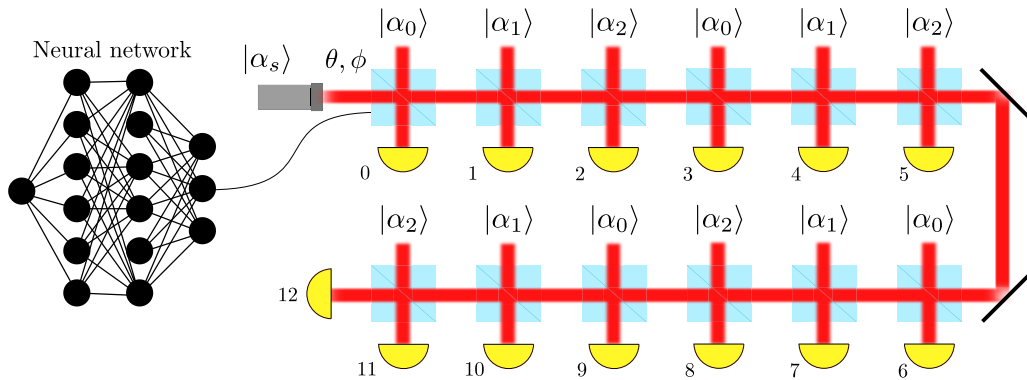


FIG. 9. Schematic representation of the quantum machine-learning discrimination device controlled by the NN. The training set states enter sequentially the ports of the BS from above, where they are mixed with the signal. The outcome of the photon counting measurement is then used to update the posterior distribution on the values of  $\alpha_0, \alpha_1, \alpha_2$ , and  $s$ . The order in which the measurements are performed is indicated near the symbols for the photon counters, from which we deduce  $M = 13$ . All the outcomes of the previous measurements contribute to the determination of the next controls through the PF and the NN.

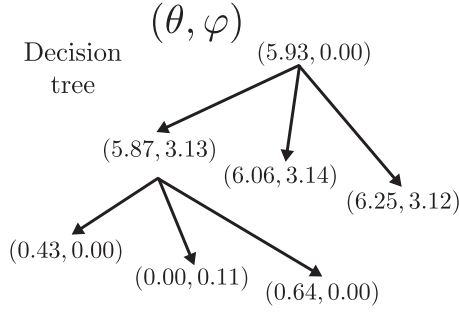


FIG. 10. Example of a decision tree used in the quantum machine-learning classifier. A tuple of controls  $(\theta, \varphi)$  is associated with each node in the tree. The value  $\tilde{y}_t$  determines the path chosen in the tree at each node in an experiment.

training states at the  $t$ th step, collected in the tuple  $\hat{\alpha}_t := (\hat{\alpha}_0^R, \hat{\alpha}_0^I, \hat{\alpha}_1^R, \hat{\alpha}_1^I, \hat{\alpha}_2^R, \hat{\alpha}_2^I)$ .

(3) the standard deviations from the mean of the amplitudes of the reference states, collected in the tuple  $\hat{\sigma}_t := (\hat{\sigma}_0^R, \hat{\sigma}_0^I, \hat{\sigma}_1^R, \hat{\sigma}_1^I, \hat{\sigma}_2^R, \hat{\sigma}_2^I)$ . The scalars passed to the NN are actually  $-\frac{1}{10} \ln \hat{\sigma}_t$ .

(4) the posterior probability for the initial state to be  $s = 0$ , normalized in  $[-1, +1]$ , i.e.,  $\tilde{p}_0 := 2\hat{p}_0 - 1$ .

(5) the posterior probability for the initial state to be  $s = 1$ , normalized in  $[-1, +1]$ , i.e.,  $\tilde{p}_1 := 2\hat{p}_1 - 1$ .

(6) the posterior probability for the initial state to be  $s = 2$ , normalized in  $[-1, +1]$ , i.e.,  $\tilde{p}_2 := 2\hat{p}_2 - 1$ .

(7) the index of the current photon counting measurement normalized in  $[-1, +1]$ , indicated with  $\tilde{t} := 2t/M_{\max} - 1$ .

(8)  $\tilde{t} = t \bmod 3 - 1$ . This tells us whether the current photon counting is performed by mixing the signal with  $|\alpha_0\rangle$ ,  $|\alpha_1\rangle$ , or  $|\alpha_2\rangle$ .

The two controls for the beam splitter are obtained directly from the NN as

$$\begin{pmatrix} \theta_t \\ \varphi_t \end{pmatrix} = f_{\text{NN}}\left(\hat{\psi}_t, \hat{\alpha}_t, -\frac{1}{10} \ln \hat{\sigma}_t, \tilde{p}_0, \tilde{p}_1, \tilde{p}_2, \tilde{t}, \tilde{t}\right). \quad (49)$$

Now we focus on the case the agent is a ternary decision tree, see Fig. 10. In this case, the input is based directly on the string of outcomes  $\mathbf{y}_t$  instead of going through the Bayesian posterior distribution. The outcome  $y_t$  of each measurement, i.e., the number of photon observed in the photon counter, is classified into one of the three classes according to the relation with the mean number of photons  $\lfloor \alpha^2 \rfloor$ . This is realized by defining the variable  $\tilde{y}_t$ , i.e.,

$$\tilde{y}_t := \begin{cases} 0 & \text{if } y_t \leq \lfloor \alpha^2 \rfloor - 1 \\ 1 & \text{if } y_t = \lfloor \alpha^2 \rfloor \\ 2 & \text{if } y_t \geq \lfloor \alpha^2 \rfloor + 1. \end{cases} \quad (50)$$

The infinite possibilities for the outcomes get reduced to only three classes without losing too much information. The modified outcomes up to the  $t$ th measurement are used to decide which branch to follow at each node of a ternary decision tree and are collected in the tuple  $\tilde{\mathbf{y}}_t := (\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_t)$ . The path in the tree (and the whole controls trajectory) is completely identified by  $\tilde{\mathbf{y}}_t$ . At each node there are a couple of controls  $(\theta, \varphi)$  to be used in the next measurement, which are the trainable variables of the agent and are returned in the call

to the strategy, i.e.,

$$\begin{pmatrix} \theta_t \\ \varphi_t \end{pmatrix} = f_{\text{Tree}}(\tilde{\mathbf{y}}_t, t). \quad (51)$$

where  $f_{\text{Tree}}$  represents the decision tree.

## 2. Definition of the precision figure of merit

The precision is the estimated error probability of a wrong classification. The estimator for the signal class is

$$\hat{s}(\hat{\mathbf{p}}) := \operatorname{argmax}(\hat{p}_0, \hat{p}_1, \hat{p}_2), \quad (52)$$

where  $\hat{\mathbf{p}} := (\hat{p}_0, \hat{p}_1, \hat{p}_2)$ , and the loss for each instance of the task reads

$$\ell(\hat{\mathbf{p}}, s) := 1 - \delta(\hat{s}, s), \quad (53)$$

which is one if an error is made and zero if the guess is right. Averaged on many estimations, it converges to the error probability for the classification task.

## 3. Discussion of the results

We have trained the QML classifier for  $\alpha = 0.75$  and  $\alpha = 1.00$ . All the estimations have been performed for  $n = 4$ . We studied the precision of the discrimination device in the quantum regime, i.e., with a small number of photons, since for  $\alpha \gg 1$  the coherent states are perfectly distinguishable. With  $\alpha = 0.75$ , the signal contains on average 0.5 photons, we should therefore expect relatively large discrimination errors even with optimal strategies. Nevertheless, the point of the optimization is to extract every last bit of information from the device, even in a regime where the errors are relatively frequent. The performances of the NN and the decision tree are compared with that of the optimized nonadaptive strategy and the precision of a nonoptimized strategy, for which each photon counter receiver the same fraction of the signal  $|\alpha_s\rangle$ . The ultimate precision limit is represented by the pretty good measurement (PGM), discussed in Appendix B. This last cannot be realized with linear optics and photon counting and assumes a perfect *a priori* knowledge of the training states  $|\alpha_0\rangle$ ,  $|\alpha_1\rangle$ , and  $|\alpha_2\rangle$  which we do not have. To make a more fair comparison, we assumed the PGM consumes  $2n$  training states where this QML device consumes only  $n$ . The  $2n$  copies are intended to be sufficient to perfectly identify  $\alpha_i$ , in order to later perform the corresponding PGM on the signal  $|\alpha_s\rangle$ . Still, the PGM error probability is not expected to be achievable. The performances of the QML classifier for different controls are reported in Fig. 11. There is a gap between the performances of the nonadaptive controls and the two adaptive strategies (NN and decision tree). This gap starts from zero and grows with the average number of photons. This is intuitive since with very few photons we do not expect the adaptive measurements to be better than the nonadaptive one, because almost always the measurement outcomes of the photon counters will be zero, which does not give any information about the training states. On the other hand, with many photons there is more space for learning. This can be also understood considering the number of possible trajectories of the Bayesian posterior during the estimation, which is only one for  $\alpha = 0$  (no update is made, the outcomes are always zero), and grows exponentially with the number of photons.

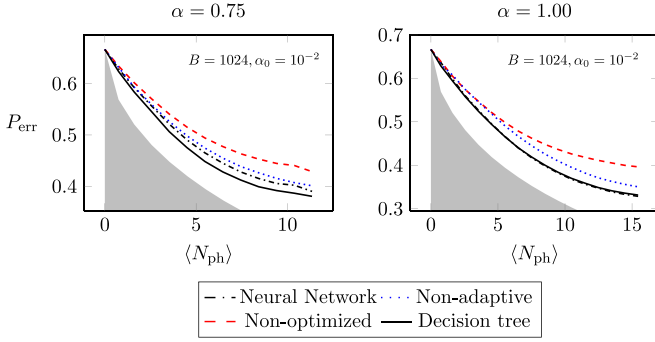


FIG. 11. Error probabilities as a function of the average number of photons in the QML device for the classification of coherent states, reported for different strategies. Both neural network and decision tree are optimized adaptive strategies. The “nonoptimized” strategy refers to a device where the phase imprinted by the BS are all zero and the reflectivity is chosen, so that the same fraction of the signal reaches each one of the photon counters. The shaded gray areas indicate the performances of the pretty good measurement (PGM) computed for  $2n$  copies of each training state, see Appendix B 4 for details. This is not the ultimate achievable precision for the classification problem, but it is a reasonable reference value not achievable with linear optics. Both the weights of the NN and the controls in the nonadaptive strategies have been initialized randomly. For these estimations, we have used  $N = 512$  particles to represent the posterior. In each plot, we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the neural network strategy.

More possible trajectories means having potentially different controls on each one of them in order to improve the precision, which is by definition adaptivity. In Fig. 12 we report some examples of trajectories.

In our experiments, we observed that the convergence of the NN to the optimal strategy is slow and does not always happen (see the case  $\alpha = 0.75$  in Fig. 11), while the decision tree is shown to be a superior control for this device. In general, for a problem with many parameters and a relatively

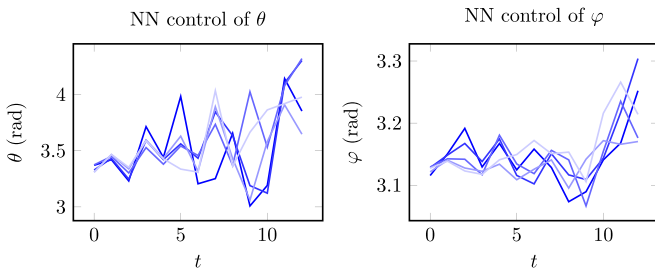


FIG. 12. Five trajectories for the neural network controls  $\theta$  and  $\phi$  are shown as a function of the measurement index  $t$ . The color shade indicates the five different executions of the experiment. The parameters  $\alpha_0, \alpha_1, \alpha_2$  has been selected at random within the prior interval, like the signal class  $s = 0, 1, 2$  for these five instances estimation. There is a tendency for the control trajectories to diverge in time (this can be seen well in the plot for  $\theta$ ), since having already done some measurements means that the controls can be tailored to the particular instance of the task. This behavior is trained in the NN but is hard-coded in the decision tree strategy.

small space of possible outcome trajectories, the decision tree is superior to the NN. On the other hand, with few parameters and a large space of outcome trajectories (like in the examples on NV centers), the decision tree is unusable in the form we presented here and the NN is superior. For a large number of average photons ( $N_{\text{ph}}$ ) the error probability curves tend to saturate. This is because even with infinite  $n$ , and therefore perfect knowledge of the training states, a single copy of the signal cannot be classified unambiguously. In this application, we have used  $N = 512$  particles to represent the posterior distribution, which conventional wisdom says are too few for the estimation of seven parameters. We have observed though that using  $N = 1024$  particles does not improve significantly the precision. Notice that in this device we are performing the simultaneous estimation of multiple quantum incompatible parameters [94], that are the real and imaginary components of the states amplitudes, whose corresponding generators  $x$  and  $p$  do not commute. We have also tried to implement the discrimination on a parallel device, instead of the serial one represented in Fig. 9. In this case, the original signal is first equally split on three wires, i.e.,  $|\alpha_s\rangle \rightarrow |\frac{\alpha_s}{\sqrt{3}}\rangle^{\otimes 3}$ , and on each line the signal is mixed with the  $n$  copies of the same training state. The performances of this parallel device, however, turned out to be always worst then the serial one and are not reported here.

#### 4. Future directions

The controls from the decision tree have the advantage of being faster to compute with respect to those of the NN but storing the tree requires more memory. A future direction of research could be to trim the decision tree after the training by keeping only those paths that correspond to the most probable trajectories.

### C. Multiphase discrimination on a photonic circuit

#### 1. Description of the task

In this section, we examine an example of multiphase estimation. We consider an interferometer consisting of four arms [95], with two balanced quarters serving as the opening and closing elements, respectively. A quarter generalizes a beam splitter to have four inputs and four outputs. After passing through the closing quarter, all wires are measured using photon counters. Three of the four arms experience phase shifts with unknown phases  $\varphi_i$  for  $i = 0, 1, 2$ . Immediately after they are subjected to three other phase shifts  $c_i \in [0, 2\pi)$ , which are the controls in our experiment, managed by the agent. In this interferometer, represented in Fig. 13, we simultaneously test each line to determine the presence or absence of a phase shift, which can only assume the two values  $\varphi_i \in \{0, 1\}$  rad. In some practical applications, these phase shifts could result from the presence on the line of a specific object, a chemical species, or it could represent a piece of encoded information that we want to retrieve (think of an optical memory device). This photonic application can also be considered as a very rudimentary form of adaptive imaging with few photons. During the experiment a total of  $M_{\text{max}}$  identical copies of an input state are fed sequentially to the interferometer and measured independently on the interferometer’s end. Based on

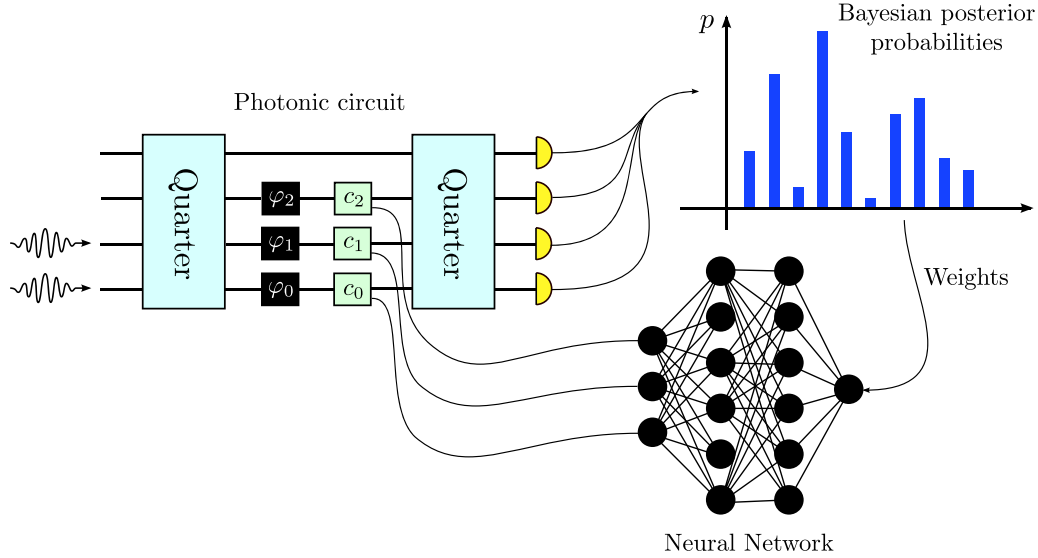


FIG. 13. Four modes interferometer for the simultaneous estimation of three phases. The interferometer has two balanced quarter as opening and closing elements, which are the generalization of the balanced BS for four modes in input and four modes in output. The three unknown phases are called  $\varphi_0, \varphi_1, \varphi_2 \in \{0, 1\}$  rad, while the control phases are  $c_0, c_1, c_2$ . The input of the NN are the eight probabilities constituting the posterior distribution on the values of the three phases.

the outcomes of the four photon counting measurements, the correct values for the tuple  $\boldsymbol{\varphi} := (\varphi_0, \varphi_1, \varphi_2)$  is inferred using Bayesian inference. The agent is an NN that outputs the tuple of control phases  $\boldsymbol{c} := (c_0, c_1, c_2)$  to be applied in the next round of measurements. The schematization of the dynamical model defined in Sec. ID can be applied also in this case. The LCPT map  $\Phi_{\boldsymbol{\varphi}}^{(t, x_t)}$  is unitary and thus can be written in the form

$$\Phi_{\boldsymbol{\varphi}}^{(t, x_t)}(\hat{\rho}) := \hat{U}_{\boldsymbol{\varphi}} \hat{\rho} \hat{U}_{\boldsymbol{\varphi}}^\dagger, \quad (54)$$

where we have introduced the unitary operator  $\hat{U}_{\boldsymbol{\varphi}}(\boldsymbol{c}_t)$ , which is a simple composition of the unitary matrices representing the quarter and the phase shifts shown in Fig. 13 on each bosonic line:

$$\hat{U}_{\boldsymbol{\varphi}} := \hat{U}(\varphi_2) \hat{U}(\varphi_1) \hat{U}(\varphi_0) \hat{U}_{\text{quarter}}. \quad (55)$$

The measurement operators are similar, and can be written as a product of unitaries implementing the controlled phase shifts  $c_i$ , then the same quarter unitary operator and, lastly, a series of photon-counting measurements on all the lines, which can be written as projectors on the base that diagonalize the number operators for each of the lines. Notice that the control  $\boldsymbol{c}$  can be inserted as a part of the dynamical model  $\Phi_{\boldsymbol{\varphi}}^{(t, x_t)}$  as well as a part of the measurement.

## 2. Input to the neural network

The input to the NN is build from the concatenation of the following 10 scalar quantities:

- (1) the Bayesian posterior probabilities for the hypotheses on the values of the phases  $\varphi_i$ , i.e.,  $\{w_j\}_{j=1}^8$  with  $\sum_{j=1}^8 w_j = 1$ .
- (2) the index of the measurement step normalized in  $[-1, +1]$ , i.e.,  $\tilde{t} := 2t/M_{\text{max}} - 1$ , where  $M_{\text{max}}$  is the number of measurements. A single use of the interferometer, in which we sent an input state and read all the four photon counters, is treated as a single measurement.

- (3) the normalized average number of photons consumed in all the previous measurements, indicated with  $\tilde{N}_{\text{ph}}$ .

The three controls phases  $c_i$  are computed as

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = 2\pi f_{\text{NN}}(\{w_j\}_{j=1}^8, \tilde{t}, \tilde{N}_{\text{ph}}). \quad (56)$$

## 3. Definition of precision figure of merit

On each estimation in the batch the error is zero if the correct phases are guessed correctly and one if the guess is wrong. Let us indicate with the symbol  $h \in \{1, 2, \dots, 8\}$  the index of the correct value, then the estimator for this discrete parameter is

$$\hat{h}(\{w_j\}_{j=1}^8) := \text{argmax}(w_1, w_2, \dots, w_8), \quad (57)$$

and the loss is given by

$$\ell(\hat{h}, h) := 1 - \delta(\hat{h}, h), \quad (58)$$

which, averaged on the batch, is the probability of a wrong classification.

## 4. Discussion of the results

In our simulations the input state always takes the form of a product of coherent states on the four bosonic wires, i.e.,  $|\psi_{\text{in}}\rangle = |\alpha_0 \alpha_1 \alpha_2 \alpha_3\rangle$ . We investigate the performance of both optimized adaptive and nonadaptive strategies for the following four inputs having different average number of photons  $\langle n \rangle$ :

- (1)  $|\psi_{\text{in}}\rangle = |1000\rangle$ :  $\alpha_0 = 1, \alpha_1 = \alpha_2 = \alpha_3 = 0, \langle n \rangle = 1$ ;
- (2)  $|\psi_{\text{in}}\rangle = |1100\rangle$ :  $\alpha_0 = \alpha_1 = 1, \alpha_2 = \alpha_3 = 0, \langle n \rangle = 2$ ;
- (3)  $|\psi_{\text{in}}\rangle = |1110\rangle$ :  $\alpha_0 = \alpha_1 = \alpha_2 = 1, \alpha_3 = 0, \langle n \rangle = 3$ ;
- (4)  $|\psi_{\text{in}}\rangle = |1111\rangle$ :  $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1, \langle n \rangle = 4$ .

The total number of measurements for each estimation remains fixed to  $M_{\text{max}} = 32$  in all the plots of Fig. 14, however,



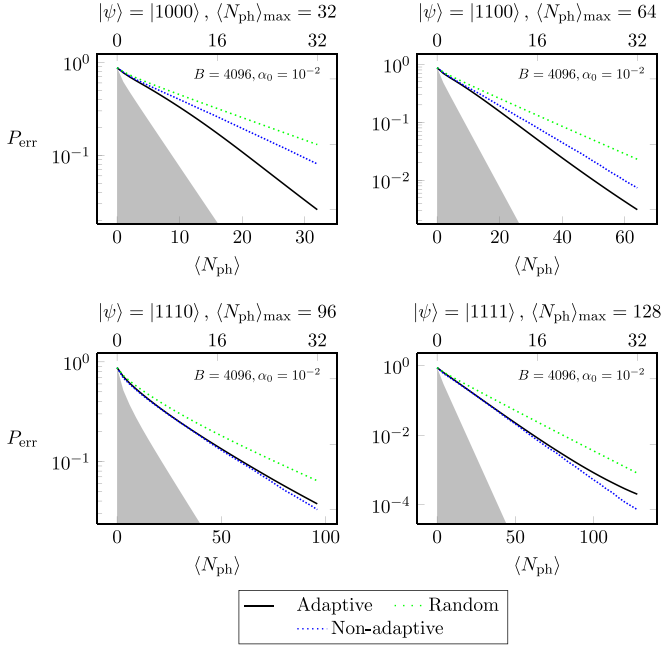


FIG. 14. These plots display the probability of incorrectly guessing the value of the unknown phases ( $\varphi_0, \varphi_1, \varphi_2$ ) as a function of the average number of photons consumed during the estimation, shown on the lower  $x$  axis. The title of each plot indicates the input state and the maximum average number of photons  $\langle N_{\text{ph}} \rangle_{\text{max}}$  used for this values discrimination task. The upper  $x$  axis represents the number of measurements  $t$ , i.e., the number of input states  $|\psi\rangle$  utilized. The optimized adaptive and nonadaptive strategies are compared with the random strategy, where the controls  $c_i$  are randomly selected uniformly in the interval  $[0, 2\pi)$ . The shaded gray areas indicate the performances of the pretty good measurement (PGM), computed for multiple copies of the encoded states, see Appendix B 3 for the details. This is not the ultimate precision bound regarding this discrimination problems, but it is a reasonable reference value not achievable with linear optics. In each plot we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the adaptive strategy.

the maximum number of photons is given by  $\langle N_{\text{ph}} \rangle_{\text{max}} := \langle n \rangle M_{\text{max}}$ , and varies according to the input state. From Fig. 14, we infer that for the input states  $|\psi_{\text{in}}\rangle = |1000\rangle, |1100\rangle$ , the adaptive strategy offers some advantage over the nonadaptive one, while for states with a higher number of photons the nonadaptive strategy is optimal. The most efficient input, in terms of the number of photons consumed to achieve a specific error probability, is  $|\psi_{\text{in}}\rangle = |1000\rangle$ . The state with three photons does not perform as well as the other inputs. Regarding the damage to the sample or the energy consumed, both of which are proportional to the total number of photons, it is preferable to conduct multiple measurements, each involving fewer photons, rather than fewer measurements with a larger number of photons. Thus, minimizing energy consumption or potential sample damage requires extending the total measurement time, which may become impractical beyond a certain limit. For a set number of measurements, and hence a fixed estimation time, states with a higher photon number generally outperform those with fewer photons, with  $|\psi_{\text{in}}\rangle = |1110\rangle$  being an exception. The observation that either the states

with high or low photon count are optimal, depending on the resources we consider, underlines the importance of agreeing on the nature of the expensive resource while discussing metrology. For the states  $|\psi\rangle = |1111\rangle$  the NN is not able to perfectly reproduce the performances of the optimal nonadaptive strategy, most probably because it converges to a local minimum of the loss during the training. Both the NN and the nonadaptive strategies have been randomly initialized in  $[0, 2\pi)$  before the training.

## D. Nonadaptive linear classifier of coherent states

### 1. Description of the task

In the example of this section the goal is to classify a coherent state  $|\alpha_s\rangle$ , which we are promised is one of the states  $\{|\alpha_i\rangle\}_{i=1}^{d-1}$ , which are classically known, at difference with Sec. IV B. For doing this we employ a network of beam splitters and phase plates (BS network) that receives in input the signal  $|\alpha_s\rangle$  to be classified. The other inputs of the network are in order  $|\alpha_0\rangle, |\alpha_1\rangle, \dots, |\alpha_{d-1}\rangle$ , that is, the reference states corresponding to the possible values of signal, which are fixed for each execution of the task. It follows that the network, represented in Fig. 15, must have  $d + 1$  bosonic wires. The output of the network is measured with individual photon counters on each wire. The signal  $|\alpha_s\rangle$  will be classified according to the photon-counter outcomes. If more than one signal state is available multiple BS networks can be stacked on one another and the classification will be based on the measurements from all layers. Each layer can have a different network. This is the only example in which we do not use a NN, that means we do not have adaptivity. Instead, we directly optimized the parametrized BS. For the same reason, it is very simple to describe the dynamical model with the framework defined in Sec. ID. There is no encoding of the state, meaning that the LCPT map  $\Phi_{\theta}^{(t, x_r)}$  is the identity, and the measurement operators can be written as a product of the unitary matrix implementing the BS network followed by projectors on the Fock states, similarly to what has been done for the Dolinar receiver in Sec. IV A. We now briefly comment on the parametrization of the BS network. The initial state of the light is Gaussian, just like the linear circuit is a Gaussian operation [96]. We can keep track of the Gaussian state of  $d + 1$  modes by following the evolution of the displacement  $\mathbf{r} \in \mathbb{R}^{2(d+1)}$  and of the covariance matrix  $\Sigma \in \mathbb{R}^{2(d+1)} \times \mathbb{R}^{2(d+1)}$ , which in general is

$$\mathbf{r} \rightarrow \mathbf{r}' := S\mathbf{r} + \mathbf{d}, \quad \text{and} \quad \Sigma \rightarrow \Sigma' := S\Sigma S^T, \quad (59)$$

where  $S$  is a real symplectic matrix. In this context  $\Sigma$  is not the covariance matrix of the posterior distribution but the covariance of Gaussian state of the modes. For a BS network, which is made of passive elements, we have no additional displacement, i.e.,  $\mathbf{d} = 0$ . Furthermore the energy is conserved, which means  $\text{tr}(\Sigma') = \text{tr}(\Sigma S^T S)$ , that implies  $S^T S = \mathbb{1}$ , since this condition must hold for every  $\Sigma$ . We have arrived at the results, that the action of a BS network is represented by a real symplectic and orthogonal matrix  $S$ . These two conditions are equivalent to the matrix  $S$  being in the form

$$S = \begin{pmatrix} \text{Re}U & \text{Im}U \\ -\text{Im}U & \text{Re}U \end{pmatrix}, \quad \text{with } U \in U(d + 1). \quad (60)$$

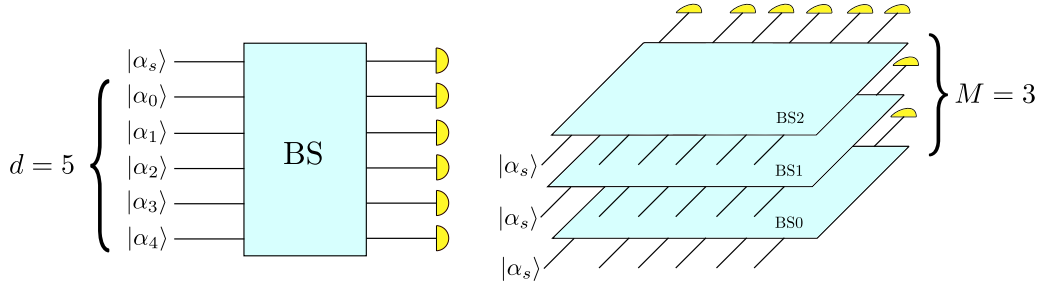


FIG. 15. On the left we represent the programmable interferometer with the BS network, the input being the signal state  $|\alpha_s\rangle$  and the fixed reference states  $|\alpha_i\rangle$  with  $i = 0, 1, \dots, d - 1$ . On the right we have stacked  $M$  layers of interferometers, each receiving a copy of the signal. The classification is done on the basis of the outcomes of the photodetectors.

The unitary matrix  $U$  can then be parametrized as  $U = e^{i(A+A^\dagger)}$ , where  $A \in GL(d+1, \mathbb{C})$  is a complex matrix, whose entries are the trainable variables of the agent, i.e., the BS network in this case. Despite not being an adaptive experiment, this is still a Bayesian estimation, since we start from a uniform prior over the hypotheses and apply the Bayes' rule to incorporate the outcomes of the photon counters in the posterior.

### 2. Definition of the precision figure of merit

After all the measurements have been performed we obtain the discrete Bayesian posterior probabilities on the possible states, i.e.,  $\{\hat{p}_i\}_{i=0}^{d-1}$ , from which we define the estimator for the signal's class

$$\hat{s} = \operatorname{argmax}(\hat{p}_0, \hat{p}_1, \dots, \hat{p}_{d-1}), \quad (61)$$

and the loss for an instance of the task

$$\ell(\hat{s}, s) := 1 - \delta(\hat{s}, s), \quad (62)$$

which averaged over a batch of tests converges to the error probability of classification.

### 3. Discussion of the results

The linear classifier has been trained and tested for some symmetric configurations of the  $\alpha_i$ . In particular the roots of the unit for  $d = 3, 5, 7, 9$  have been chosen. The results are reported in Fig. 16. For a symmetric configuration of states the PGM is optimal, but it cannot be achieved with linear optics, see Appendixes B 2 and B 5 for details. The optimality of the PGM means that a single layer BS network cannot achieve its performances, however this raises the interesting question of how many layers  $M$  and copies of the signal  $|\alpha_s\rangle$  are necessary to match or surpass the error probability of the PGM. Three copies of the signal are always sufficient in these examples to match the PGM error probability. Beside that, the performances of the trained BS networks are also compared with that of randomly extracted BS networks. An interesting observation is that the advantage over the random strategy that the trained network has decreases as the number of hypotheses grows. For  $d \rightarrow \infty$  we expect to have no gain. An adaptive strategy may be able to restore the gap with the untrained BS networks.

## V. CONCLUSIONS

Our findings indicate that model-aware RL outperforms traditional control strategies in multiple scenarios, surpassing even model-free RL. The fact that many problems can be solved by minimal changes to the examples could pave the way for researchers to expedite the search for optimal controls in quantum sensors, potentially accelerating the advent of their widespread industrial application. With this work, we believe we have solved the problem of optimal controls for quantum sensors that are simple enough to be efficiently simulated. Since the sensor's model must be simulated on a

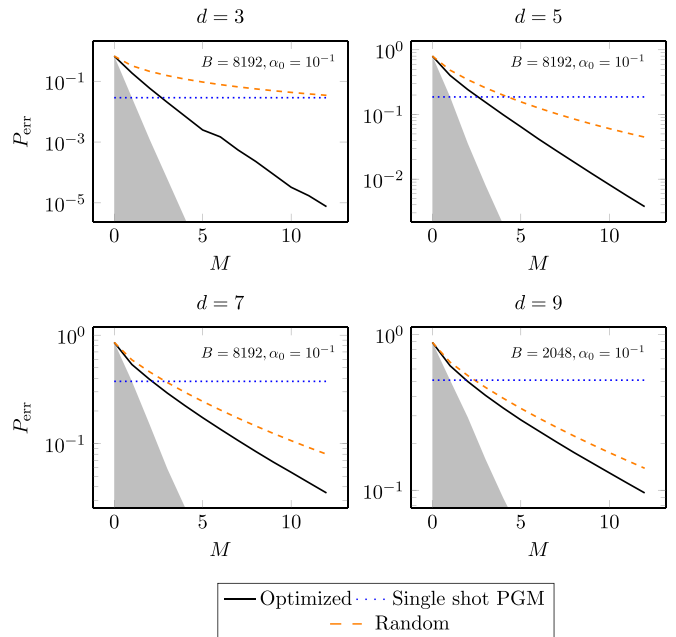


FIG. 16. Probability of error in the classification of a coherent state as a function of the number  $M$  of copies of the signal  $|\alpha_s\rangle$  used, which is also the number of layers of BS networks used in the classifier. The dashed orange line is the average performance of randomly extracted BS networks. The dotted blue line is the optimal error probability for a single copy of the signal, while the shaded gray area indicates the optimal error probability for  $M$  copies of the signal, both computed with the PGM. More details can be found in Appendix B 5. In each plot we reported also the batch size  $B$  and the initial learning rate  $\alpha_0$  used in the simulations for the “optimized” strategy.

computer with the overhead of automatic differentiation, we do not expect to be able to scale this procedure to sensors more complex than a handful of qubits in the near future. This limitation excludes the application of model-aware RL to domains where complex entangled states are used. However, given their limited experimental use at the moment, this constraint on efficient simulability may not significantly impact the utility of model-aware RL. It is notable that the most relevant platform for quantum metrology today is the single spin in diamond, which is straightforward to simulate. While large quantum systems remain inaccessible, we believe that scaling the procedure to a large number of unknown parameters is feasible. Although the number of particles and, consequently, the memory requirements must increase for a growing number of parameters, we have already successfully optimized a seven-parameter problem in Sec. IV B. We believe it is possible to find even more efficient representations of the Bayesian posterior distribution than the particle filter, which will enable us to optimize problems with tens of parameters. While we acknowledge that experimentally relevant problems can only rarely be approached analytically, the results obtained in this way can often be valuable in setting up and guiding more powerful numerical optimization, just as the analytical study of the Fisher information in Appendix A and of phase estimation in Ref. [59] and Appendix C have been useful in setting the initial state of the neural network and of the adaptive controls for the NV centers, and in defining the normalization of the network’s output.

**ACKNOWLEDGMENTS**

We gratefully acknowledge the computational resources of the Center for High Performance Computing (CHPC) at SNS. F.B. thanks F. Marquardt for useful discussions. We acknowledge financial support by MUR (Ministero dell’Istruzione, dell’Università e della Ricerca) through the following projects: PNRR MUR project PE0000023-NQSTI, PRIN 2017 Taming complexity via Quantum Strategies: a Hybrid Integrated Photonic approach (QU-SHIP) Id. 2017SRN-BRK. Lastly, this work was supported by the Open Access Publishing Fund of the Scuola Normale Superiore.

**APPENDIX A: LOWER BOUNDS FOR NITROGEN-VACANCY CENTERS**

In this section, we apply the Bayesian Cramér-Rao bound to the estimation of various parameters on the NV center platform. This bounds will be based on the Fisher information [21]. Consider a stochastic variable  $y$ , which is extracted from a probability distribution  $p(y|\theta)$ , where  $\theta$  is a parameter we want to estimate. This is a model for an experiment leading to a stochastic outcome. The information on  $\theta$  available from the knowledge of  $y$  can be measured by the Fisher information (FI), defined as

$$I(\theta) := \mathbb{E}_y \left[ \left( \frac{\partial \ln p(y|\theta)}{\partial \theta} \right)^2 \right], \tag{A1}$$

where the expectation value is taken over the distribution  $p(y|\theta)$ . There is also a multiparameter version of the FI, called

the Fisher information matrix (FI matrix), defined as

$$I_{ij}(\theta) := \mathbb{E}_y \left[ \frac{\partial \ln p(y|\theta)}{\partial \theta_i} \frac{\partial \ln p(y|\theta)}{\partial \theta_j} \right]. \tag{A2}$$

If the experiment allows us to be controlled through the parameter  $x$ , then the outcome probability is  $p(y|x, \theta)$  and the FI inherits such dependence, i.e., we write  $I(\theta|x)$ . In this paper the control parameter  $x$  is computed from a strategy  $h$ , which stands for “heuristic” and which could be the particle guess heuristic or a neural network, for example. In this case, we indicate it explicitly in symbol for the control  $x_h$ .

**1. Bayesian Cramér-Rao bound**

Given  $\theta$  a single parameter to estimate, we call  $I(\theta|x_h)$  the Fisher information of a sequence of measurements with controls  $\mathbf{x}_h = (x_0^h, x_1^h, \dots, x_{M-1}^h)$ , which are computed from a strategy  $h$ . The quantity  $I(\theta|x_h)$ , together with the Fisher information of the prior  $\pi(\theta)$ , i.e.,  $I(\pi)$ , defines a lower bound on the precision  $\Delta^2 \hat{\theta}$  of whatever estimator  $\hat{\theta}$ , which contains the expectation value of  $I(\theta|x_h)$  on  $\pi(\theta)$  and is optimized on the strategy  $h$ . This lower bounds reads

$$\Delta^2 \hat{\theta} \geq \frac{1}{\sup_h \mathbb{E}_\theta [I(\theta|x_h)] + I(\pi)}. \tag{A3}$$

This definition appears in the work of Fiderer *et al.* [27]. For the NV center the controls are the evolution time  $\tau$  and the phase  $\varphi$ , this last, however, does not play any role in the computation of the lower bound and will be omitted in the following. For the multiparameter bound we avoid the complication of introducing the Fisher information matrix and consider instead the sum of the bounds for two independent estimations, thus writing instead

$$\text{tr}[G \cdot (\hat{\theta} - \theta)^\top (\hat{\theta} - \theta)] \geq \sum_{k=1}^d \frac{1}{\sup_h \mathbb{E}_{\theta_k} [I(\theta_k|\tau_h)] + I(\pi_k)}. \tag{A4}$$

The Fisher information of a sequence of measurements is always additive, even if the quantum probe is only measured weakly, but in dealing with projective measurements, as it is the case for an NV center, the advantage is that the measurements are uncorrelated, and the same expression for the Fisher information applies to all of them, independently of the results of the previous measurements, i.e.,

$$I(\theta|\tau) = \sum_{\tau=1}^M I(\theta|\tau_\tau) \leq M \sup_{\tau} I(\theta|\tau), \tag{A5}$$

where  $M$  is the total number of measurements. The optimization of the single measurement FI gives directly the precision bound for the measurement-limited estimation:

$$\begin{aligned} \Delta^2 \hat{\theta} &\geq \frac{1}{\sup_h \mathbb{E}_\theta [I(\theta|\tau_h)] + I(\pi)} \\ &\geq \frac{1}{M \mathbb{E}_\theta [\sup_{\tau} I(\theta|\tau)] + I(\pi)}. \end{aligned} \tag{A6}$$

TABLE II. Lower bounds for the precision of the frequency and decoherence time estimation in DC magnetometry on an NV center.

	Time		Measurement	
$T_2 = \infty$	$\frac{1}{T^2 + I(\omega)}$	(A10)	$\frac{2^{-2(M+1)}}{3} \text{ MHz}^2$	(A11)
$T_2 < \infty$	$\frac{1}{0.5T T_2 + I(\omega)}$	(A12)	$\max \left\{ \frac{2^{-2(M+1)}}{3} \text{ MHz}^2, \frac{1}{\mu M T_2^2 + I(\omega)} \right\}$	(A13)
$T_2^{-1} \in (a, b)$	$\frac{1}{0.5T \mathbb{E}[T_2] + I(\omega)} + \frac{1}{0.5T \mathbb{E}[T_2] + I(T_2^{-1})}$	(A14)	$\frac{1}{\mu M \mathbb{E}[T_2^2] + I(\omega)} + \frac{1}{\mu M \mathbb{E}[T_2^2] + I(T_2^{-1})}$	(A15)

If the total evolution time is the limiting resource, then, the expression for the total FI is

$$I(\theta|\tau) = T \sum_{i=1}^M \frac{\tau_i}{T} \left[ \frac{I(\theta|\tau_i)}{\tau_i} \right] \leq T \sup_{\tau} \frac{I(\theta|\tau)}{\tau}, \quad (\text{A7})$$

with  $\sum_{i=1}^M \tau_i = T$ . In this expression the total FI is the weighted sum of the renormalized FI of each measurement, i.e.,  $I(\theta|\tau_i)/\tau_i$ , and can be manifestly upper bounded by concentrating all the weights on the supremum of the renormalized FI. This gives the lower bound for the precision of the time-limited estimation:

$$\begin{aligned} \Delta^2 \hat{\theta} &\geq \frac{1}{\sup_h \mathbb{E}_{\theta} [I(\theta|\tau_h)] + I(\pi)} \\ &\geq \frac{1}{T \mathbb{E}_{\theta} [\sup_{\tau} \frac{I(\theta|\tau)}{\tau}] + I(\pi)}. \end{aligned} \quad (\text{A8})$$

In the following we apply this general observations to the derivation of the numerical bounds for DC and AC magnetometry, for decoherence estimation, and for the measurement of the parallel hyperfine coupling.

## 2. Evaluation of Fisher information

Since the measurement outcome in the NV center is binary, we can compute the Fisher information for a parameter  $\theta$ , given the control  $\tau$ , as

$$\begin{aligned} I(\theta|\tau) &= \mathbb{E} \left[ \left( \frac{\partial \ln p(\pm 1|\theta, \tau)}{\partial \omega} \right)^2 \right] \\ &= \frac{(\frac{\partial p}{\partial \theta})^2}{p(1-p)} = \frac{(\frac{\partial p}{\partial \theta})^2}{\frac{1}{4} - (p - \frac{1}{2})^2}, \end{aligned} \quad (\text{A9})$$

where we have used the definition in Eq. (A1) and where  $p := p(+1|\theta, \tau)$ . For example, for a decoherence-free estimation of the precession frequency  $\omega$  we have  $p := \cos^2(\omega\tau/2)$ , from which  $\partial p/\partial \theta = \tau \sin(\omega\tau/2) \cos(\omega\tau/2)$ , and finally  $I(\omega|\tau) = \tau^2$ .

## 3. DC magnetometry

The lower bound on the estimation of the frequency  $\omega$  and of the inverse of the decoherence time  $T_2^{-1}$  are reported in Table II and are represented in Fig. 3. The left column contains the bound for a finite number of measurements  $M$ , while right column refers to the estimation with a fixed total evolution time  $T$ . The first row refers to the estimation of  $\omega$  with perfect coherence, the second row refers to the estimation of  $\omega$  with a finite and known  $T_2$ , while the last row refers to the simultaneous estimation of  $\omega$  and  $T_2^{-1}$  treated on equal footing, i.e.,  $G = \mathbb{1}$ . The symbols  $I(\omega)$  and  $I(T_2^{-1})$  indicate the FI of the prior of the precession frequency and of the decoherence time respectively. The numerical values of the quantities appearing in the table, for  $\omega \in (0, 1)$  MHz and  $T_2^{-1} \in (0.09, 0.11)$  MHz, are  $\mu = 0.1619$ ,  $\mathbb{E}[T_2^2] = 101.01 \mu\text{s}^2$ ,  $\mathbb{E}[T_2] = 10.0336 \mu\text{s}$ ,  $I(\omega) = 12 \mu\text{s}^2$ ,  $I(T_2^{-1}) = 3 \times 10^4 \mu\text{s}^2$ . In the following we derive these bounds.

(1) The Fisher information for the precession frequency  $\omega$  is given by  $I(\omega|\tau) = \tau^2$ , so that  $\sup_{\tau} I(\omega|\tau) = \infty$  and the analysis based on the Cramér-Rao bound does not give a useful bound. Eq. (A11) in Table II can be found by observing that each measurement gives at most one bit of information about the value of  $\omega$ , because it has a binary outcome [27]. This bound is applied to all the measurement-limited estimations in the table in addition to the one coming from the Fisher information.

(2) With a finite decoherence time  $T_2 < \infty$  the FI for the frequency  $\omega$  is

$$I(\omega|\tau, T_2) = \frac{1}{\mathcal{D}} \tau^2 e^{-\frac{2\tau}{T_2}} \cos^2 \left( \frac{\omega\tau}{2} \right) \sin^2 \left( \frac{\omega\tau}{2} \right), \quad (\text{A16})$$

with

$$\begin{aligned} \mathcal{D} &:= \left[ e^{-\frac{\tau}{T_2}} \cos^2 \left( \frac{\omega\tau}{2} \right) + \frac{1 - e^{-\frac{\tau}{T_2}}}{2} \right] \\ &\times \left[ e^{-\frac{\tau}{T_2}} \sin^2 \left( \frac{\omega\tau}{2} \right) + \frac{1 - e^{-\frac{\tau}{T_2}}}{2} \right] \end{aligned} \quad (\text{A17})$$

which, by defining  $C := \cos^2(\omega\tau/2)$ , can be bounded in the following way:

$$I(\omega|\tau, T_2) = \frac{\tau^2 e^{-\frac{2x}{T_2}} C(1-C)}{\left[\frac{1}{4} - e^{-\frac{2x}{T_2}} \left(C - \frac{1}{2}\right)^2\right]} \quad (\text{A18})$$

$$\leq \frac{\tau^2 e^{-\frac{2x}{T_2}}}{1 - e^{-\frac{2x}{T_2}}} = T_2^2 \frac{x^2 e^{-2x}}{1 - e^{-2x}}, \quad (\text{A19})$$

where  $x = \tau/T_2$ . The maximization in  $x \in \mathbb{R}_+$  gives  $\sup_{\tau} I(\omega|\tau, T_2) = \mu T_2^2$  with  $\mu = 0.1619$ . Inserting this expression in Eq. (A6) gives the first term in the maximum of Eq. (A13), the second term was explained in the previous point.

(3) We now turn to the estimation of an unknown  $T_2^{-1}$ , with prior uniform in  $(0.09, 0.11)$ , alongside  $\omega$ . The total MSE is  $\Delta^2 \widehat{\omega} + \Delta^2 T_2^{-1}$ . The supremum of the FI for  $\omega$  is always  $\sup_{\tau} I(\omega|\tau, T_2) = \mu T_2^2$ , but this time the expectation value  $\mathbb{E}_{\theta}$  is not trivial, i.e.,  $\mathbb{E}_{\theta}[\sup_{\tau} I(\omega|\tau, T_2)] = \mu \mathbb{E}[T_2^2]$ . The derivative of the probability with respect to  $T_2^{-1}$ , used to compute the FI is

$$\frac{\partial p}{\partial T_2^{-1}} = -\tau e^{-\frac{x}{T_2}} \left[ \cos^2\left(\frac{\omega\tau}{2}\right) - \frac{1}{2} \right]. \quad (\text{A20})$$

With the same notation for the cosine we write the FI for the inverse of the decoherence time as

$$I(T_2^{-1}|\omega, \tau) = \frac{\tau^2 e^{-\frac{2x}{T_2}} \left(C - \frac{1}{2}\right)^2}{\left[\frac{1}{4} - e^{-\frac{2x}{T_2}} \left(C - \frac{1}{2}\right)^2\right]} \leq T_2^2 \frac{x^2 e^{-2x}}{1 - e^{-2x}}. \quad (\text{A21})$$

The maximization of this expression gives  $\sup_{\tau} I(T_2^{-1}|\omega, \tau) = \mu T_2^2$ , which is the same expression of  $I(\omega|T_2, \tau)$ , that again has a nontrivial expectation value on the parameters. Putting everything together and adding the prior information on  $T_2^{-1}$  gives Eq. (A15).

(4) Regarding the time-constrained lower bounds, for  $T_2 = \infty$ , the total FI is maximized by performing a single measurement of time duration  $\tau = T$ , which gives Eq. (A10), through the application of Eq. (A8).

(5) For  $T_2 < \infty$  we have to maximize the normalized FI in  $x \in \mathbb{R}_+$ , i.e.,

$$\frac{I(\omega|\tau, T_2)}{\tau} \leq \frac{\tau e^{-\frac{2x}{T_2}}}{1 - e^{-\frac{2x}{T_2}}} \leq T_2 \frac{x e^{-2x}}{1 - e^{-2x}} \leq \frac{T T_2}{2}, \quad (\text{A22})$$

from which Eq. (A12) follows from Eq. (A8).

(6) In the last case we have to estimate both  $\omega$  and  $T_2^{-1} \in (0.09, 0.11)$  MHz. Regarding the estimation of  $\omega$  the Fisher information is always given by Eq. (A22), only that this time the expectation value on  $T_2$  is nontrivial, and produces the first addend of Eq. (A14). For the estimation of  $T_2^{-1}$  we have a similar expression for the normalized FI:

$$\frac{I(T_2^{-1}|\tau, \omega)}{\tau} \leq \frac{\tau e^{-\frac{2x}{T_2}}}{1 - e^{-\frac{2x}{T_2}}} \leq T_2 \frac{x e^{-2x}}{1 - e^{-2x}} \leq \frac{T T_2}{2}, \quad (\text{A23})$$

which similarly needs the expectation value appearing in Eq. (A8) to give the second piece of Eq. (A14).

#### 4. AC magnetometry

The lower bound on the precision of the frequency  $\Omega$  and of the inverse of the decoherence time  $T_2^{-1}$ , are reported in Table III and are represented in Fig. 4 of Sec. III D. The meaning of the rows and columns here are the same as in Section A3. The symbols  $I(\Omega) = 12 \mu\text{s}^2$  and  $I(T_2^{-1}) = 3 \times 10^4 \mu\text{s}^2$  indicate respectively the FI of the prior of the frequency and of the inverse of the decoherence time. In the following we derive each of these bounds.

(1) The decoherence-free FI for  $\Omega$  in the measurement-limited case is obtained from Eq. (A9) through the derivative

$$\frac{\partial p}{\partial \Omega} = \frac{\sin(\omega\tau)}{\omega} \cos\left[\frac{B}{2\omega} \sin(\omega\tau)\right] \sin\left[\frac{B}{2\omega} \sin(\omega\tau)\right] \quad (\text{A30})$$

and reads

$$I(\Omega|\tau) = \frac{\sin^2 \omega\tau}{\omega^2} \leq \frac{1}{\omega^2}. \quad (\text{A31})$$

Inserting this expression in Eq. (A8) gives immediately the first part of Eq. (A25) in Table III. The second operand of the max come from analysis on the number of bits [27], which we already discussed in the previous section.

(2) For the time-limited estimation of  $\Omega$  we have to maximize the normalized FI in  $x = \omega\tau \in \mathbb{R}_+$ :

$$\frac{I(\Omega|\tau)}{\tau} = \frac{1}{\omega} \frac{\sin^2(\omega\tau)}{\omega\tau} = \frac{1}{\omega} \frac{\sin^2 x}{x} \leq \frac{\gamma}{\omega}, \quad (\text{A32})$$

where  $\gamma = 0.724611$ . By inserting this result into Eq. (A8) we get Eq. (A24).

(3) For the case  $T_2 < \infty$  we do not compute any new bound, instead we use the one for the decoherence-free estimation, which must be valid also for the noisy model. Accordingly, Eq. (A27) is equal to Eq. (A25) and Eq. (A26) is equal to Eq. (A24).

(4) For the simultaneous estimation of  $\Omega$  and of  $T_2^{-1}$  we use the same bound computed for DC magnetometry for the precision on  $T_2^{-1}$  and obtain Eqs. (A29) and (A28) for the measurement-limited and time-limited estimation, respectively.

#### 5. Decoherence estimation

In this section, we report the bounds based on the FI for the estimation of the inverse of the decoherence time  $T_2^{-1}$  and the exponent  $\beta$  on the NV center platform, which are reported in Fig. 5 of Sec. III E. The numerical coefficients appearing in Table IV are  $\mu = 0.16190$ ,  $\delta = 0.24429$ ,  $\chi = 0.23966$ ,  $\varepsilon = 0.20687$ ,  $\eta = 0.10582$ ,  $\psi = 2.43013$ ,  $\mathbb{E}[T_2^2] = 10^3 \mu\text{s}^2$ ,  $\mathbb{E}[T_2] = 25.5848 \mu\text{s}$ ,  $\mathbb{E}[T_2^{-1}] = 0.0450 \text{ MHz}^2$ ,  $\mathbb{E}[\beta^2] = 8.08332$ ,  $\mathbb{E}[\beta^{-2}] = 0.16666$ ,  $I(T^{-1}) = 1481.48 \mu\text{s}^2$ , and  $I(\beta) = 1.92$ . We proceed in deriving point by point the lower bounds of the above table from the Fisher information.

(1) The derivative of the model probability in  $T_2^{-1}$  is given by

$$\frac{\partial p}{\partial T_2^{-1}} = -\frac{1}{2} e^{-\left(\frac{\tau}{T_2}\right)^{\beta}} \beta \tau \left(\frac{\tau}{T_2}\right)^{\beta-1}, \quad (\text{A39})$$

TABLE III. Lower bounds on the estimation precision of the magnetic field intensity and decoherence time in AC magnetometry on an NV center.

	Time		Measurement	
$T_2 = \infty$	$\frac{1}{\frac{\gamma T}{\omega} + I(\Omega)}$	(A24)	$\max \left\{ \frac{2^{-2(M+1)}}{3} \text{MHz}^2, \frac{1}{\frac{M}{\omega^2} + I(\Omega)} \right\}$	(A25)
$T_2 < \infty$	$\frac{1}{\frac{\gamma T}{\omega} + I(\Omega)}$	(A26)	$\max \left\{ \frac{2^{-2(M+1)}}{3} \text{MHz}^2, \frac{1}{\frac{M}{\omega^2} + I(\Omega)} \right\}$	(A27)
$T_2^{-1} \in (a, b)$	$\frac{1}{\frac{M}{\omega^2} + I(\Omega)} + \frac{1}{0.5T\mathbb{E}[T_2] + I(T_2^{-1})}$	(A28)	$\frac{1}{\frac{\gamma T}{\omega} + I(\Omega)} + \frac{1}{\mu M\mathbb{E}[T_2^2] + I(T_2^{-1})}$	(A29)

from which we obtain the FI from Eq. (A9):

$$I(T_2^{-1}|\tau, \beta) = \beta^2 T_2^2 \frac{e^{-2\left(\frac{\tau}{T_2}\right)^\beta} \left(\frac{\tau}{T_2}\right)^{2\beta}}{1 - e^{-2\left(\frac{\tau}{T_2}\right)^\beta}}. \quad (\text{A40})$$

By defining  $x = (\tau/T_2)^\beta \in \mathbb{R}_+$  we rewrite the above expression as

$$I(T_2^{-1}|\tau, \beta) = \beta^2 T_2^2 \frac{e^{-2x} x^2}{1 - e^{-2x}} \leq \mu \beta^2 T_2^2, \quad (\text{A41})$$

which inserted into Eq. (A6), by taking the expectation value on the parameters  $\beta$  and  $T_2$ , gives the expression in Eq. (A34).

(2) For  $\beta = 2$  we can avoid the expectation value of  $\beta^2$ , which is replaced by the factor 4 in Eq. (A34) of Table IV.

(3) For the simultaneous estimation of  $T_2^{-1}$  and  $\beta$  we need an expression for the Fisher information on  $\beta$ , which can be obtained from the derivative

$$\frac{\partial p}{\partial \beta} = -\frac{1}{2} \left(\frac{\tau}{T_2}\right)^\beta e^{-\left(\frac{\tau}{T_2}\right)^\beta} \ln \left(\frac{\tau}{T_2}\right), \quad (\text{A42})$$

and reads

$$\begin{aligned} I(\beta|\tau, T_2) &= \frac{1}{\beta^2} \frac{\left(\frac{\tau}{T_2}\right)^{2\beta} e^{-2\left(\frac{\tau}{T_2}\right)^\beta} \log^2 \left(\frac{\tau}{T_2}\right)^\beta}{1 - e^{-2\left(\frac{\tau}{T_2}\right)^\beta}} \\ &\leq \frac{1}{\beta^2} \frac{x^2 e^{-2x} \log^2 x}{1 - e^{-2x}} \leq \frac{\chi}{\beta^2}, \end{aligned} \quad (\text{A43})$$

where  $x = (\tau/T_2)^\beta \in \mathbb{R}_+$ . The coefficient  $\chi$  comes from the maximization of the function in  $x$  only. Given the result  $\sup_x I(\beta|\tau, T_2) = \chi \beta^{-2}$ , we insert this expression into Eq. (A6), obtaining the second term of Eq. (A38).

(4) Regarding the time-limited estimation, the relevant quantity to optimize is the renormalized FI again. For the estimation of  $T_2^{-1}$  this reads

$$\begin{aligned} \frac{I(T_2^{-1}|\tau, \beta)}{\tau} &= \beta^2 T T_2 \frac{e^{-2x} x^{2-\frac{1}{\beta}}}{1 - e^{-2x}} \\ &\leq \beta^2 T T_2 \sup_{x, \beta} \frac{e^{-2x} x^{2-\frac{1}{\beta}}}{1 - e^{-2x}} \end{aligned} \quad (\text{A44})$$

$$= \delta \beta^2 T T_2, \quad (\text{A45})$$

TABLE IV. Lower bounds for the precision of the characterization of a dephasing noise on an NV center.

	Time		Measurement	
$\beta$ nuis.	$\frac{1}{\delta T \mathbb{E}[T_2] \mathbb{E}[\beta^2] + I(T_2^{-1})}$	(A33)	$\frac{1}{\mu M \mathbb{E}[T_2^2] \mathbb{E}[\beta^2] + I(T_2^{-1})}$	(A34)
$\beta = 2$	$\frac{1}{4\epsilon T \mathbb{E}[T_2] + I(T_2^{-1})}$	(A35)	$\frac{1}{4\mu M \mathbb{E}[T_2^2] + I(T_2^{-1})}$	(A36)
Both	$\frac{1}{\psi T \mathbb{E}[T_2^{-1}] \mathbb{E}[\beta^{-2}] + I(\beta)} + \frac{1}{\delta T \mathbb{E}[T_2] \mathbb{E}[\beta^2] + I(T_2^{-1})}$	(A37)	$\frac{1}{\chi M \mathbb{E}[\beta^{-2}] + I(\beta)} + \frac{1}{\mu M \mathbb{E}[T_2^2] \mathbb{E}[\beta^2] + I(T_2^{-1})}$	(A38)

TABLE V. Lower bounds for the estimation precision of two precession frequencies, split by the hyperfine interaction of the NV center with a  $^{13}\text{C}$  nucleus.

	Time		Measurement	
$T_2 = \infty$	$\frac{2}{\frac{1}{4}T^2 + I(\omega)}$	(A47)	$\frac{2^{-M}}{24} \text{MHz}^2$	(A48)
$T_2 < \infty$	$\frac{2}{\frac{1}{8}T\mathbb{E}[T_2] + I(\omega)}$	(A49)	$\max \left\{ \frac{2^{-M}}{24} \text{MHz}^2, \frac{2}{\frac{1}{4}\mu MT_2^2 + I(\omega)} \right\}$	(A50)

where  $\delta$  is the supremum of the function in  $x \in \mathbb{R}_+$  and  $\beta \in (1.5, 4)$ , realized for  $\beta = 1.5$ . Inserting the expression above in Eq. (A8) and taking the expectation values of the parameters produces Eq. (A34).

(5) The  $\beta = 2$  bound can be obtained from Eq. (A45) by maximizing only on  $x$ . The coefficient  $\varepsilon$  in Eq. (A34) is defined as  $\varepsilon = \sup_x [(e^{-2x}x^{\frac{3}{2}})/(1 - e^{-2x})]$ . After adding the FI from the prior we get Eq. (A34).

(6) For the estimation of  $\beta$  we again maximize the normalized FI in  $x \in \mathbb{R}_+$ :

$$\begin{aligned} \frac{I(\beta|\tau, T_2)}{\tau} &= \frac{1}{T_2\beta^2} \frac{x^{2-\frac{1}{\beta}} e^{-2x} \log^2 x}{1 - e^{-2x}} \\ &\leq \frac{1}{T_2\beta^2} \sup_{\beta, \tau} \frac{x^{2-\frac{1}{\beta}} e^{-2x} \log^2 x}{1 - e^{-2x}} \\ &\leq \psi T_2^{-1} \beta^{-2}, \end{aligned} \quad (\text{A46})$$

which inserted in Eq. (A8) gives the second part of the bound in Eq. (A37), the first being identical to Eq. (A34).

## 6. Hyperfine-coupling estimation

In this section, we report the precision lower bounds for the estimation of the parallel hyperfine coupling of the NV center electron spin with a carbon nucleus, which are plotted in Fig. 6 of Sec. III F. These bounds are based on the one computed for DC magnetometry in Section A 3. The numerical coefficients in the table are  $\mu = 0.16190$  and  $I(\omega) = 18.18181 \mu\text{s}^2$ . The true values of both frequencies  $\omega_0$  and  $\omega_1$  are extracted uniformly in  $(0, 1)$  MHz in the simulations, however, since they are completely symmetrized in the model likelihood, the true prior is the uniform distribution over the triangle in the  $(\omega_0, \omega_1)$  plane identified by the points  $(0, 0)$  MHz,  $(1, 0)$  MHz, and  $(1, 1)$  MHz. Given  $\sigma^2$  the variance of such distribution we define  $I(\omega) := \frac{1}{2}\sigma^2$  the prior FI for a single frequency. In the following we give a derivation for the bounds in the Table V.

(1) Starting from Eq. (A11), based on the bit counting argument, we observe that  $2M$  measurements are needed for having  $N$  bits of information for each of the phases  $\omega_{0,1} \in (0, 1)$  MHz. The MSE on a single phase after  $N$  measurement is thus limited by  $2^{-M}/12$ . For two phases the bound becomes  $2^{-M}/6$ . Since it is not easy to implement the information on the triangular prior in this bound, we consider a square subset

of the triangular prior only, i.e., we fix  $\omega_0 \in (0.5, 1)$  MHz and  $\omega_1 \in (0, 0.5)$  MHz. This means dividing the bound further by a factor of four, which gives us Eq. (A48) of Table V.

(2) The FI for a single frequency is 1/4 that of the DC magnetometry example of Eq. (A10), i.e.,  $I(\omega_{0,1}) = T^2/4$  in the infinite coherence case. This comes from the factor 1/2 in the likelihood. For two parameters this gives the bound of Eq. (A47).

(3) The same argument can be made to go from Eqs. (A13)–(A50) and from Eqs. (A12)–(A49).

## APPENDIX B: LOWER BOUNDS FOR PHOTONIC CIRCUITS

In this section, we present the precision lower bound for the photonic based applications. These are the agnostic Dolinar receiver, the multiphase discrimination task on the four bosonic lines interferometer, the linear classifier of multiple states, and the QML classifier of three coherent states. All these examples are based on the identification of the correct hypothesis among a finite number of them. Accordingly, the figure of merit for their performances is the probability of a wrong classification. The bounds presented here for this family of tasks are based either on the Helstrom bound [92] and on its generalizations, or on the pretty good measurement (PGM), which gives good results for the one shot classification of states [97]. This approach does not really give a lower bound for the precision, in all cases but one, nevertheless we use it as a reference value, and given the fact that performing the PGM requires in general entangled measurements we consider its precision to be unachievable by means of photon counting and linear optics only.

### 1. Helstrom bound

Consider a pure state  $|\psi\rangle$  in the Hilbert space  $\mathcal{H}$ . We are assured that this state is either  $|\psi_0\rangle$  or  $|\psi_1\rangle$ , both of which also reside in  $\mathcal{H}$ . Our objective is to correctly identify which state it is. If we are free to perform whatever POVM measurement on  $|\psi\rangle$ , then the error probability is bounded as

$$P_e \geq P_e^{\text{H}} := \frac{1}{2} - \frac{1}{2}\sqrt{1 - |\langle\psi_0, \psi_1\rangle|^2}. \quad (\text{B1})$$

This is called the Helstrom bound [92] and for two orthogonal states is  $P_e^{\text{H}} = 0$ . In the Dolinar receiver, where the task is to

discriminate  $|\pm\alpha\rangle$ , the Helstrom bound reads

$$\begin{aligned} P_e^H &= \frac{1}{2} - \frac{1}{2}\sqrt{1 - |\langle -\alpha, +\alpha \rangle|^2} \\ &= \frac{1}{2} - \frac{1}{2}\sqrt{1 - e^{-4|\alpha|^2}}. \end{aligned} \quad (\text{B2})$$

This is represented by the shaded gray region on the precision plots of the Dolinar receiver in Fig. 8 of Sec. IV A. In this work we study however the agnostic Dolinar receiver, where the task is to discriminate the states  $|\alpha\rangle \otimes |\alpha\rangle^{\otimes n}$  and  $|\alpha\rangle \otimes |\alpha\rangle^{\otimes n}$ , without having any classical knowledge on the value of  $\alpha$ . It can be proved [31] that the error lower bound in this case is

$$P_e^H := \frac{1}{2} \left( 1 - \frac{1}{2} \sum_{k=0}^{\infty} \mathcal{P}(k; \sqrt{n+1}\alpha) \sqrt{1 - \left( \frac{n-1}{n+1} \right)^{2k}} \right), \quad (\text{B3})$$

where  $\mathcal{P}(k; \mu)$  is the probability distribution of a Poissonian variable, i.e.,

$$\mathcal{P}(k; \mu) := \frac{\mu^k e^{-\mu}}{k!}. \quad (\text{B4})$$

The Helstrom bound with limited  $n$  is plotted as the red lines of Fig. 8.

## 2. Pretty good measurement

Suppose we are given a state  $|\psi\rangle \in \mathcal{H}$  and we are promised it belongs to the set of states  $\{|\psi_j\rangle\}_{j=1}^m$ , which are classically known. The *a priori* probabilities for the states are  $p_j$  for  $j = 1, \dots, m$ . We can perform whatever measurement on  $|\psi\rangle$  and the task is to identify the state in the set. A good approach to this problem, which does not require solving a semidefinite program, is the PGM [97], which prescribes performing the POVM measurement  $\mathcal{M}_{\text{PGM}} := \{M_j\}_{j=1}^m$ , whose operators are

$$M_j := p_j S^{-\frac{1}{2}} |\psi_j\rangle \langle \psi_j| S^{-\frac{1}{2}}, \quad (\text{B5})$$

where  $S := \sum_{j=1}^m p_j |\psi_j\rangle \langle \psi_j|$ . If in the measurement the outcome corresponding to the  $j$ th operator is observed, then the guess for the state classification is  $|\psi_j\rangle$ . The error probability of the PGM is therefore

$$P_e^{\text{PGM}} = 1 - \sum_{j=1}^m p_j \text{Tr}(M_j |\psi_j\rangle \langle \psi_j|). \quad (\text{B6})$$

In our applications to photonic circuits such measurements are not optimal and in general their performance cannot be achieved with linear optics and photon measurement. They provide nevertheless a useful reference value for the estimation precision. Throughout the thesis the computation of  $P_e^{\text{PGM}}$  has been carried out numerically.

## 3. Multiphase discrimination

In this work we studied the problem of multiphase discrimination on a four arms interferometer, for four different input states. To give a reference value on the precision of the discrimination we apply the PGM. Let us break the interferometer and remove the control phases  $\{c_i\}_{i=1}^3$ , the closing quarter, and the photodetectors. The photon states after the encoding with the phases  $\{\varphi_i\}_{i=1}^3$  are, respectively,

$$(1) \quad \left| \frac{1}{2} e^{-i\varphi_0}, \frac{1}{2} i e^{-i\varphi_1}, \frac{1}{2} i e^{-i\varphi_2}, -\frac{1}{2} \right\rangle, \quad (\text{B7})$$

$$(2) \quad \left| \frac{1}{2} (1+i) e^{-i\varphi_0}, \frac{1}{2} (-1+i) e^{-i\varphi_1}, \frac{1}{2} (1+i) e^{-i\varphi_2}, \frac{1}{2} (-1+i) \right\rangle, \quad (\text{B8})$$

$$(3) \quad \left| \frac{1}{2} + i e^{-i\varphi_0}, \frac{1}{2} i e^{-i\varphi_1}, \frac{1}{2} i e^{-i\varphi_2}, -\frac{1}{2} + i \right\rangle, \quad (\text{B9})$$

$$(4) \quad |i e^{-i\varphi_0}, i e^{-i\varphi_1}, i e^{-i\varphi_2}, i\rangle \quad (\text{B10})$$

for the input states with one, two, three, and four photons on average. We are allowed to consume  $n$  input states in order to produce  $n$  copies of the encoded coherent states reported above. These are then sent to an adder, which perform the operation

$$U |\alpha\rangle^{\otimes n} = |\sqrt{n}\alpha\rangle \quad (\text{B11})$$

on the  $n$  coherent states for each of the four bosonic wires. On this resulting state of the interferometer the PGM is then executed. The error probability can be expressed as a function of the number of input states  $n$  or as a function of the total number of employed photons, and is the gray area represented in Fig. 14 of Sec. IV C. Although this is not rigorously a lower bound on the precision it serves as a reference value, which is most certainly impossible to saturate with the device at our disposal.

## 4. Quantum machine-learning classifier

In the QML task we have a device that identifies the class  $s$  of a signal  $|\alpha_s\rangle$ , given the quantum training set  $|\alpha_0\rangle^{\otimes n} \otimes |\alpha_1\rangle^{\otimes n} \otimes |\alpha_2\rangle^{\otimes n}$ . To perform the PGM we must know classically the states that we aim to distinguish and this is not possible with a finite quantum training set. We suppose that by consuming double the amount of reference states, i.e.,  $2n$  for each class, we can estimate the complex numbers  $\alpha_0, \alpha_1, \alpha_2$  exactly and then apply the optimal PGM. For many instances of the classification task, with the reference states  $|\alpha_0\rangle, |\alpha_1\rangle, |\alpha_2\rangle$  and the signal  $|\alpha_s\rangle$  extracted according to their prior, we compute the error probability of the PGM, i.e.,  $P_e^{\text{PGM}}$ , and associate to it the average number of consumed photons

$$N_{\text{ph}} = 2n(|\alpha_0|^2 + |\alpha_1|^2 + |\alpha_2|^2) + |\alpha_s|^2. \quad (\text{B12})$$

We then average the data points  $(N_{\text{ph}}, P_e^{\text{PGM}})$  in order to compute the expected error probability  $\langle P_e^{\text{PGM}} \rangle$  for a certain average number of photons. This curve defines the shaded gray area in Fig. 11 of Sec. IV B. Although this is not rigorously a lower bound on the precision it serves as a reference value, which is most certainly impossible to saturate with the device at our disposal.

## 5. Linear classifier of multiple states

The last example on the photonic circuit platform of Sec. IV D concerns the classification of a signal state  $|\alpha_s\rangle$ , that can take values in the set of known states  $\{|\alpha_j\rangle\}_{j=1}^m$ . For this discrimination we employ only a network of beam splitters,



which is statically optimized, and no adaptivity is taken into account. By virtue of the cyclic symmetry of the states to discriminate, the PGM is optimal [98–100] and its error rate defines the shaded gray area in the plots of Fig. 16. Again we are using an adder, with action reported in Eq. (B11), to build the state  $|\sqrt{M}\alpha_s\rangle$  out of  $M$  copies of the signal, which in the device are passed to the  $M$  layers of the discriminator. On the state  $|\sqrt{M}\alpha_s\rangle$  the PGM is executed and its precision is plot as a function of  $M$ .

### APPENDIX C: ANALYTICAL OPTIMIZATION FOR THE NITROGEN-VACANCY CENTER

In this Appendix, we optimize the precision of magnetic-field estimation on an NV center, with the tools developed in Ref. [59]. In the present paper we have so far allowed the continuous tuning of the parameters  $\tau$  and  $\varphi$ , but now we suppose that only  $\tau_j = 2^{j-1}$  can be measured, for  $j = 1, 2, \dots, K$ , and similarly we are limited to  $\varphi = 0, \frac{\pi}{2}$ . We parametrize a strategy under these constraints by defining  $v_j$  the number of measurements performed with time  $\tau_j$ . The total number of Ramsey measurements is then

$$M := 2 \sum_{j=1}^K v_j. \quad (\text{C1})$$

By applying the phase algorithm presented in Ref. [59] and using the corresponding analytical tool to upper bound its precision we can write

$$\Delta^2 \tilde{\omega} \leq \left(\frac{2\pi}{3}\right)^2 \left(\frac{1}{4^K} + 16 \sum_{j=1}^K \frac{A}{4^{j-1}} C^{-v_j}\right), \quad (\text{C2})$$

where  $C \simeq 1.66$  and  $A \simeq 0.60$  are numerical constants. We can find the optimal measurement distribution that optimizes this upper bound with the methods of Lagrangian multiplier:

$$\mathcal{L} := \left(\frac{2\pi}{3}\right)^2 \left(\frac{1}{4^K} + 16 \sum_{j=1}^K \frac{A}{4^{j-1}} C^{-v_j}\right) - \lambda \left(2 \sum_{j=1}^K v_j - M\right). \quad (\text{C3})$$

The optimization on  $v_j$  gives

$$v_j = v_K + \frac{2}{\log_2 C} (K - j). \quad (\text{C4})$$

The number of measurements  $v_j$  should be integers, but in the following we neglect the rounding as we are only interested in extracting the scaling of the precision of the phase estimation procedure. By substituting Eq. (C4) in Eq. (C2) we get

$$\Delta^2 \tilde{\omega} \leq \left(\frac{2\pi}{3}\right)^2 (1 + 64KAC^{-v_K}) \frac{1}{4^K}, \quad (\text{C5})$$

and the resource resummation reads

$$M = K \left(v_K - \frac{1}{\log_2 C}\right) + \frac{K^2}{\log_2 C}. \quad (\text{C6})$$

We now extract  $v_K$  as a function of  $M$  and  $K$  from Eq. (C6), and substitute it in Eq. (C5) to get

$$\Delta^2 \tilde{\omega} \leq \left(\frac{2\pi}{3}\right)^2 \left(\frac{1}{4^K} + 32KA \frac{2^{-\frac{M \log_2 C}{K}}}{2^K}\right). \quad (\text{C7})$$

Observe that  $2^{-(M \log_2 C)/K} < 1 \forall K \geq 0$  and that the right-hand side of Eq. (C7) tends to zero for  $K \rightarrow \infty$ , because the term  $K$  at numerator of

$$32KA \frac{2^{-\log_2 C \frac{M}{K}}}{2^K}$$

cannot compensate the exponential in the denominator. As  $K \rightarrow \infty$  we have  $v_K \rightarrow 0$ , but this is a nonphysical solution because it must be  $v_K \geq 1$ . Therefore the maximum allowed  $K$  ( $:= K^*$ ) corresponds to  $v_{K^*} = 1$ . In this case, the resource resummation reads

$$M = K^* \left(1 - \frac{1}{\log_b C}\right) + \frac{K^{*2}}{\log_b C}. \quad (\text{C8})$$

For  $K^* \gg 1$  the above equation gives

$$K^* = \sqrt{\log_2 C \sqrt{M}}. \quad (\text{C9})$$

This result led us to the choice of the prefactor  $h$  that corresponds the maximum  $\tau$ . According to this derivation, neglecting the factor  $\sqrt{\log_2 C}$ , which is of order one, we get the maximum free evolution time  $\tau_{K^*} \simeq 2^{\sqrt{M}}$ . Inserting  $K^*$  into Eq. (C7) and neglecting the subleading terms, we arrive at the scaling

$$\Delta^2 \tilde{\theta} \lesssim O\left(\frac{\sqrt{M}}{4\sqrt{\log_2 C \sqrt{M}}}\right). \quad (\text{C10})$$

This scaling for the precision is exponential in the square root of the number of measurements. If the free evolution time follows instead the power law  $\tau_j = b^{j-1}$ , then the corresponding Lagrangian for the optimization of the precision is

$$\begin{aligned} \mathcal{L} := & \left(\frac{\pi}{nb^{K-1}}\right)^2 + \sum_{j=1}^K \left(\frac{2\pi}{nb^{j-2}} \frac{b}{b-1}\right)^2 AC^{-v_j} \\ & + \lambda \left(2 \sum_{j=1}^K v_j - M\right), \end{aligned} \quad (\text{C11})$$

with  $n = b + 1$ . The optimization with respect to  $v_j$  gives the expression

$$v_j = v_K + \frac{2}{\log_b C} (K - j), \quad (\text{C12})$$

and the resource resummation is

$$M = K \left(v_K - \frac{1}{\log_b C}\right) + \frac{K^2}{\log_b C}. \quad (\text{C13})$$

The same analysis done before give the precision scaling

$$\Delta^2 \tilde{\theta} \lesssim O\left(\frac{\sqrt{M}}{b^2 \sqrt{\log_b C \sqrt{M}}}\right). \quad (\text{C14})$$

We wish to emphasize that this is an achievable scaling of the precision for a fixed number of measurements  $M$ , which

has to be compared with the unachievable lower bound on the precision, obtained through a bit-counting argument [27]:

$$\Delta^2 \hat{\theta} \geq \frac{2^{-2(M+1)}}{3}, \quad (\text{C15})$$

which has been used as a lower bound in Section A. This bound is computed assuming that each measurement gives exactly one bit of information about the phase  $\theta$ . In summary, while the precision lower bound for this counting of the resources scales as  $O(4^{-N})$ , we have been able to prove the achievability of the (much worse)  $O(\sqrt{N}4^{-\sqrt{N}})$  scaling.

- 
- [1] F. Flamini, A. Hamann, S. Jerbi, L. M. Trenkwalder, H. P. Nautrup, and H. J. Briegel, Photonic architecture for reinforcement learning, *New J. Phys.* **22**, 045002 (2020).
- [2] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. H. Yoo, S. V. Isakov, P. Massey, R. Halavati, M. Y. Niu, A. Zlokapa, E. Peters, O. Lockwood, A. Skolik, S. Jerbi, V. Dunjko, M. Leib, M. Streif, D. Von Dollen, H. Chen, S. Cao *et al.*, TensorFlow quantum: A software framework for quantum machine learning, [arXiv:2003.02989](https://arxiv.org/abs/2003.02989).
- [3] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi, J. M. Arrazola, U. Azad, S. Banning, C. Blank, T. R. Bromley, B. A. Cordier, J. Ceroni, A. Delgado, O. Di Matteo, A. Dusko *et al.*, PennyLane: Automatic differentiation of hybrid quantum-classical computations, [arXiv:1811.04968](https://arxiv.org/abs/1811.04968).
- [4] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Reinforcement learning in different phases of quantum control, *Phys. Rev. X* **8**, 031086 (2018).
- [5] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, When does reinforcement learning stand out in quantum control? A comparative study on state preparation, *npj Quantum Inf.* **5**, 85 (2019).
- [6] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, *npj Quantum Inf.* **5**, 33 (2019).
- [7] R. Porotti, A. Essig, B. Huard, and F. Marquardt, Deep reinforcement learning for quantum state preparation with weak nonlinear measurements, *Quantum* **6**, 747 (2022).
- [8] R. Porotti, V. Peano, and F. Marquardt, Gradient-ascent pulse engineering with feedback, *PRX Quantum* **4**, 030305 (2023).
- [9] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement learning with neural networks for quantum feedback, *Phys. Rev. X* **8**, 031084 (2018).
- [10] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, Calibration of quantum sensors by neural networks, *Phys. Rev. Lett.* **123**, 230502 (2019).
- [11] Y. Ban, J. Echanobe, Y. Ding, R. Puebla, and J. Casanova, Neural-network-based parameter estimation for quantum detection, *Quantum Sci. Technol.* **6**, 045012 (2021).
- [12] S. Nolan, A. Smerzi, and L. Pezzè, A machine learning approach to Bayesian parameter estimation, *npj Quantum Inf.* **7**, 169 (2021).
- [13] S. P. Nolan, L. Pezzè, and A. Smerzi, Frequentist parameter estimation with supervised learning, *AVS Quantum Sci.* **3**, 034401 (2021).
- [14] V. Nguyen, S. B. Orbell, D. T. Lennon, H. Moon, F. Vigneau, L. C. Camenzind, L. Yu, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, D. Sejdinovic, and N. Ares, Deep reinforcement learning for efficient measurement of quantum devices, *npj Quantum Inf.* **7**, 100 (2021).
- [15] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, *npj Quantum Inf.* **6**, 20 (2020).
- [16] Y. Quek, S. Fort, and H. K. Ng, Adaptive quantum state tomography with neural networks, *npj Quantum Inf.* **7**, 105 (2021).
- [17] H.-Y. Hsieh, J. Ning, Y.-R. Chen, H.-C. Wu, H. L. Chen, C.-M. Wu, and R.-K. Lee, Direct parameter estimations from machine learning-enhanced quantum state tomography, *Symmetry* **14**, 874 (2022).
- [18] F. Marquardt, Machine learning and quantum devices, *SciPost Phys. Lect. Notes*, **1**, 29 (2021).
- [19] F. Marquardt, Online course: Advanced machine learning for physics, science, and artificial scientific discovery (2021).
- [20] M. Krenn, J. Landgraf, T. Foesel, and F. Marquardt, Artificial intelligence and machine learning for quantum technologies, *Phys. Rev. A* **107**, 010101 (2023).
- [21] R. A. Fisher, *The Design of Experiments* (Oliver & Boyd, Oxford, 1935).
- [22] A. E. Foster, Variational, monte carlo and policy-based approaches to bayesian experimental design, Ph.D. thesis, University of Oxford, 2021.
- [23] A. G. Baydin, K. Cranmer, P. D. C. Manzano, C. Delaere, D. Derkach, J. Donini, T. Dorigo, A. Giammanco, J. Kieseler, L. Layer, G. Louppe, F. Ratnikov, G. C. Strong, M. Tosi, A. Ustyuzhanin, P. Vischia, and H. Yarar, Toward machine learning optimization of experimental design, *Nucl. Phys. News* **31**, 25 (2021).
- [24] Z. Ballard, C. Brown, A. M. Madni, and A. Ozcan, Machine learning and computation-enabled intelligent sensor design, *Nat. Mach. Intell.* **3**, 556 (2021).
- [25] D. R. Ivanova, A. Foster, S. Kleinegesse, M. U. Gutmann, and T. Rainforth, Implicit deep adaptive design: Policy-based experimental design without likelihoods, in *Advances in Neural Information Processing Systems* (Curran Associates, 2021), Vol. 34, pp. 25785–25798.
- [26] A. Foster, D. R. Ivanova, I. Malik, and T. Rainforth, Deep adaptive design: Amortizing sequential Bayesian experimental design, in *Proceedings of the 38th International Conference on Machine Learning*, edited by M. Meila and T. Zhang, Proceedings of Machine Learning Research Vol. 139 (PMLR, 2021), pp. 3384–3395.
- [27] L. J. Fiderer, J. Schuff, and D. Braun, Neural-network heuristics for adaptive Bayesian quantum estimation, *PRX Quantum* **2**, 020303 (2021).
- [28] A. Ścibior and F. Wood, Differentiable particle filtering without modifying the forward pass, [arXiv:2106.10314](https://arxiv.org/abs/2106.10314).
- [29] F. BelliarDO, F. Zoratti, and V. Giovannetti, qsensoropt: Quantum sensor optimisation (2022).
- [30] F. BelliarDO, F. Zoratti, and V. Giovannetti, qsensoropt documentation (2022).

- [31] F. Zoratti, N. D. Pozza, M. Fanizza, and V. Giovannetti, An agnostic-Dolinar receiver for coherent states classification, *Phys. Rev. A* **104**, 042606 (2021).
- [32] J. J. Meyer, J. Borregaard, and J. Eisert, A variational toolbox for quantum multi-parameter estimation, *npj Quantum Inf.* **7**, 89 (2021).
- [33] M. Zhang, H.-M. Yu, H. Yuan, X. Wang, R. Demkowicz-Dobrzański, and J. Liu, QuanEstimation: An open-source toolkit for quantum parameter estimation, *Phys. Rev. Res.* **4**, 043057 (2022).
- [34] C. Granade, C. Ferrie, I. Hincks, S. Casagrande, T. Alexander, J. Gross, M. Kononenko, and Y. Sanders, Qinfer: Statistical inference software for quantum applications, *Quantum* **1**, 5 (2017).
- [35] R. D. McMichael, S. M. Blakley, and S. Dushenko, Optbayesxpt: sequential Bayesian experiment design for adaptive measurements, *J. Res. NIST* **126**, 126002 (2021).
- [36] J. Bavaresco, P. Lipka-Bartosik, P. Sekatski, and M. Mehboudi, Designing optimal protocols in Bayesian quantum parameter estimation with higher-order operations, [arXiv:2311.01513](https://arxiv.org/abs/2311.01513).
- [37] J. Liu, and H. Yuan, Quantum parameter estimation with optimal control, *Phys. Rev. A* **96**, 012117 (2017).
- [38] H. Xu, J. Li, L. Liu, Y. Wang, H. Yuan, and X. Wang, Generalizable control for quantum parameter estimation through reinforcement learning, *npj Quantum Inf.* **5**, 82 (2019).
- [39] P. Rembold, N. Oshnik, M. M. Müller, S. Montangero, T. Calarco, and E. Neu, Introduction to quantum optimal control for quantum sensing with nitrogen-vacancy centers in diamond, *AVS Quantum Sci.* **2**, 024701 (2020).
- [40] J. Schuff, L. J. Fiderer, and D. Braun, Improving the dynamics of quantum sensors with reinforcement learning, *New J. Phys.* **22**, 035001 (2020).
- [41] H. Xu, L. Wang, H. Yuan, and X. Wang, Generalizable control for multiparameter quantum metrology, *Phys. Rev. A* **103**, 042615 (2021).
- [42] J. Liu, M. Zhang, H. Chen, L. Wang, and H. Yuan, Optimal scheme for quantum metrology, *Adv. Quantum Technol.* **5**, 2100080 (2022).
- [43] T. Xiao, J. Fan, and G. Zeng, Parameter estimation in quantum sensing based on deep reinforcement learning, *npj Quantum Inf.* **8**, 2 (2022).
- [44] Y. Qiu, M. Zhuang, J. Huang, and C. Lee, Efficient and robust entanglement generation with deep reinforcement learning for quantum metrology, *New J. Phys.* **24**, 083011 (2022).
- [45] T. Xiao, J. Huang, H. Li, J. Fan, and G. Zeng, Quantum generative adversarial imitation learning, *New J. Phys.* **25**, 033034 (2023).
- [46] S. S. Vedaie, A. Dalal, E. J. Pérez, and B. C. Sanders, Framework for learning and control in the classical and quantum domains, *Ann. Phys.* **458**, 169471 (2023).
- [47] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, *Nat. Rev. Phys.* **5**, 141 (2023).
- [48] Z. Ma, P. Gokhale, T.-X. Zheng, S. Zhou, X. Yu, L. Jiang, P. Maurer, and F. T. Chong, Adaptive circuit learning for quantum metrology, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2021), pp. 419–430.
- [49] R. Kaubruegger, D. V. Vasilyev, M. Schulte, K. Hammerer, and P. Zoller, Quantum variational optimization of Ramsey interferometry and atomic clocks, *Phys. Rev. X* **11**, 041045 (2021).
- [50] C. D. Marciniak, T. Feldker, I. Pogorelov, R. Kaubruegger, D. V. Vasilyev, R. van Bijnen, P. Schindler, P. Zoller, R. Blatt, and T. Monz, Optimal metrology with programmable quantum sensors, *Nature (London)* **603**, 604 (2022).
- [51] R. Kaubruegger, A. Shankar, D. V. Vasilyev, and P. Zoller, Optimal and variational multiparameter quantum metrology and vector field sensing, *PRX Quantum* **4**, 020333 (2023).
- [52] E. Köse and D. Braun, Superresolution imaging with multiparameter quantum metrology in passive remote sensing, *Phys. Rev. A* **107**, 032607 (2023).
- [53] A. M. de las Heras, C. Tabares, J. T. Schneider, L. Tagliacozzo, D. Porras, and A. González-Tudela, Photonic quantum metrology with variational quantum optical non-linearities, *Phys. Rev. Res.* **6**, 013299 (2024).
- [54] J. Yang, S. Pang, Z. Chen, A. N. Jordan, and A. del Campo, Variational principle for optimal quantum controls in quantum metrology, *Phys. Rev. Lett.* **128**, 160505 (2022).
- [55] F. Belliardo, F. Zoratti, F. Marquardt, and V. Giovannetti, Model-aware reinforcement learning for high-performance Bayesian experimental design in quantum metrology, [arXiv:2312.16985](https://arxiv.org/abs/2312.16985).
- [56] F. Belliardo, F. Zoratti, and V. Giovannetti, Application of machine learning to experimental design in quantum mechanics, *Int. J. Quantum Inf.* **0**, 2450002 (2024).
- [57] F. Belliardo, V. Cimini, E. Polino, F. Hoch, B. Piccirillo, N. Spagnolo, V. Giovannetti, and F. Sciarrino, Optimizing quantum-enhanced Bayesian multiparameter estimation in noisy apparatus, [arXiv:2211.04747](https://arxiv.org/abs/2211.04747).
- [58] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [59] F. Belliardo and V. Giovannetti, Achieving Heisenberg scaling with maximally entangled states: an analytic upper bound for the attainable root-mean-square error, *Phys. Rev. A* **102**, 042613 (2020).
- [60] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
- [61] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction* (De Gruyter, Berlin, Boston, 2019).
- [62] M. W. Doherty, C. R. Du, and G. D. Fuchs, Quantum science and technology based on color centers with accessible spin, *J. Appl. Phys.* **131**, 010401 (2022).
- [63] M. Chen, C. Meng, Q. Zhang, C. Duan, F. Shi, and J. Du, Quantum metrology with single spins in diamond under ambient conditions, *Natl. Sci. Rev.* **5**, 346 (2018).
- [64] J. Maze, Quantum manipulation of nitrogen-vacancy centers in diamond: From basic properties to applications, Ph.D. thesis, Harvard University, 2010.
- [65] 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).
- [66] 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).
- [67] T. Joas, S. Schmitt, R. Santagati, A. A. Gentile, C. Bonato, A. Laing, L. P. McGuinness, and F. Jelezko, Online adaptive quantum characterization of a nuclear spin, *npj Quantum Inf.* **7**, 56 (2021).
- [68] M. Valeri, Experimental adaptive Bayesian estimation of multiple phases with limited data, *npj Quantum Inf.* **6**, 92 (2020).
- [69] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, Experimental Bayesian quantum phase estimation on a silicon photonic chip, *Phys. Rev. Lett.* **118**, 100503 (2017).
- [70] E. Polino, M. Riva, M. Valeri, R. Silvestri, G. Corrielli, A. Crespi, N. Spagnolo, R. Osellame, and F. Sciarrino, Experimental multiphase estimation on a chip, *Optica* **6**, 288 (2019).
- [71] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino, Photonic quantum metrology, *AVS Quantum Sci.* **2**, 024703 (2020).
- [72] M. Barbieri, A. Datta, I. A. Walmsley, and P. C. Humphreys, Quantum enhanced multiple phase estimation, *Phys. Rev. Lett.* **111**, 070403 (2013).
- [73] S. Schmitt, T. Gefen, D. Louzon, C. Osterkamp, N. Staudenmaier, J. Lang, M. Markham, A. Retzker, L. P. McGuinness, and F. Jelezko, Optimal frequency measurements with quantum probes, *npj Quantum Inf.* **7**, 55 (2021).
- [74] C. Ferrie, C. E. Granade, and D. G. Cory, How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies, *Quantum Inf. Process.* **12**, 611 (2013).
- [75] 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).
- [76] R. D. McMichael, S. Dushenko, and S. M. Blakley, Sequential Bayesian experiment design for adaptive Ramsey sequence measurements, *J. Appl. Phys.* **130**, 144401 (2021).
- [77] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online Hamiltonian learning, *New J. Phys.* **14**, 103013 (2012).
- [78] N. Oshnik, P. Rembold, T. Calarco, S. Montangero, E. Neu, and M. M. Müller, Robust magnetometry with single nitrogen-vacancy centers via two-step optimization, *Phys. Rev. A* **106**, 013107 (2022).
- [79] K. Craigie, E. M. Gauger, Y. Altmann, and C. Bonato, Resource-efficient adaptive Bayesian tracking of magnetic fields with a quantum sensor, *J. Phys.: Condens. Matter* **33**, 195801 (2021).
- [80] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, Optimized quantum sensing with a single electron spin using real-time adaptive measurements, *Nat. Nanotechnol.* **11**, 247 (2016).
- [81] R. Santagati, A. A. Gentile, S. Knauer, S. Schmitt, S. Paesani, C. Granade, N. Wiebe, C. Osterkamp, L. P. McGuinness, J. Wang, M. G. Thompson, J. G. Rarity, F. Jelezko, and A. Laing, Magnetic-field learning using a single electronic spin in diamond with one-photon readout at room temperature, *Phys. Rev. X* **9**, 021019 (2019).
- [82] I. Zohar, B. Haylock, Y. Romach, M. J. Arshad, N. Halay, N. Drucker, R. Stöhr, A. Denisenko, Y. Cohen, C. Bonato, and A. Finkler, Real-time frequency estimation of a qubit without single-shot-readout, *Quantum Sci. Technol.* **8**, 035017 (2023).
- [83] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, High-dynamic-range magnetometry with a single electronic spin in diamond, *Nat. Nanotechnol.* **7**, 109 (2012).
- [84] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, Experimental quantum Hamiltonian learning, *Nat. Phys.* **13**, 551 (2017).
- [85] 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).
- [86] C. Bonato and D. W. Berry, Adaptive tracking of a time-varying field with a quantum sensor, *Phys. Rev. A* **95**, 052348 (2017).
- [87] C. Ferrie, C. E. Granade, and D. G. Cory, Adaptive Hamiltonian estimation using Bayesian experimental design, *AIP Conf. Proc.* **1443**, 165 (2012).
- [88] G. Liu, M. Chen, Y.-X. Liu, D. Layden, and P. Cappellaro, Repetitive readout enhanced by machine learning, *Mach. Learn.: Sci. Technol.* **1**, 015003 (2020).
- [89] M. Tsukamoto, S. Ito, K. Ogawa, Y. Ashida, K. Sasaki, and K. Kobayashi, Machine-learning-enhanced quantum sensors for accurate magnetic field imaging, *Sci. Rep.* **12**, 13942 (2022).
- [90] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Hamiltonian learning and certification using quantum resources, *Phys. Rev. Lett.* **112**, 190501 (2014).
- [91] V. Cimini, E. Polino, F. Belliardo, F. Hoch, B. Piccirillo, N. Spagnolo, V. Giovannetti, and F. Sciarrino, Experimental metrology beyond the standard quantum limit for a wide resources range, *npj Quantum Inf.* **9**, 20 (2023).
- [92] C. W. Helstrom, Quantum detection and estimation theory, *J. Stat. Phys.* **1**, 231 (1969).
- [93] A. S. Holevo, Statistical problems in quantum physics, in *Proceedings of the Second Japan-USSR Symposium on Probability Theory*, edited by G. Maruyama and Y. V. Prokhorov, Lecture Notes in Mathematics (Springer, Berlin, Heidelberg, 1973), pp. 104–119.
- [94] F. Belliardo and V. Giovannetti, Incompatibility in quantum parameter estimation, *New J. Phys.* **23**, 063055 (2021).
- [95] V. Cimini, M. Valeri, E. Polino, S. Piacentini, F. Ceccarelli, G. Corrielli, N. Spagnolo, R. Osellame, and F. Sciarrino, Deep reinforcement learning for quantum multiparameter estimation, *Adv. Photon.* **5**, 016005 (2023).
- [96] J. B. Brask, Gaussian states and operations – a quick reference, [arXiv:2102.05748](https://arxiv.org/abs/2102.05748).
- [97] P. Hausladen and W. K. Wootters, A ‘pretty good’ measurement for distinguishing quantum states, *J. Mod. Opt.* **41**, 2385 (1994).
- [98] Y. C. Eldar, A. Megretski, and G. C. Verghese, Designing optimal quantum detectors via semidefinite programming, *IEEE Trans. Inf. Theory* **49**, 1007 (2003).
- [99] Y. C. Eldar and G. D. Forney, On quantum detection and the square-root measurement, *IEEE Trans. Inf. Theory* **47**, 858 (2006).
- [100] S. Izumi, M. Takeoka, M. Fujiwara, N. D. Pozza, A. Assalini, K. Ema, and M. Sasaki, Displacement receiver for phase-shift-keyed coherent states, *Phys. Rev. A* **86**, 042328 (2012).