Quantum Science and Technology

Parameter estimation from quantum-jump data using neural networks

Enrico Rinaldi[1,2,3,4] ⓘ, Manuel González Lastre[5,6] ⓘ, Sergio García Herreros[5,6] ⓘ, Shahnawaz Ahmed[7] ⓘ, Maryam Khanahmadi[7] ⓘ, Franco Nori[3,4,8] ⓘ and Carlos Sánchez Muñoz[5,6,9,*] ⓘ

1 Quantinuum K.K., Otemachi Financial City Grand Cube 3F, 1-9-2 Otemachi, Chiyoda-ku, Tokyo, Japan
2 Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS) Program, RIKEN, Wakoshi, Saitama 351-0198, Japan
3 Center for Quantum Computing, RIKEN, Wakoshi, Saitama 351-0198, Japan
4 Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan
5 Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, 28049 Madrid, Spain
6 Instituto Nicolás Cabrera, Universidad Autónoma de Madrid, 28049 Madrid, Spain
7 Department of Microtechnology and Nanoscience, Chalmers University of Technology, 412 96 Gothenburg, Sweden
8 Physics Department, University of Michigan, Ann Arbor, MI 48109-1040, United States of America
9 Institute of Fundamental Physics IFF-CSIC, Calle Serrano 113b, 28006 Madrid, Spain
* Author to whom any correspondence should be addressed.

E-mail: carlossmwolff@gmail.com

Abstract

We present an inference method utilizing artificial neural networks for parameter estimation of a quantum probe monitored through a single continuous measurement. Unlike existing approaches focusing on the diffusive signals generated by continuous weak measurements, our method harnesses quantum correlations in discrete photon-counting data characterized by quantum jumps. We benchmark the precision of this method against Bayesian inference, which is optimal in the sense of information retrieval. By using numerical experiments on a two-level quantum system, we demonstrate that our approach can achieve a similar optimal performance as Bayesian inference, while drastically reducing computational costs. Additionally, the method exhibits robustness against the presence of imperfections in both measurement and training data. This approach offers a promising and computationally efficient tool for quantum parameter estimation with photon-counting data, relevant for applications such as quantum sensing or quantum imaging, as well as robust calibration tasks in laboratory-based settings.

1. Introduction

Quantum parameter estimation refers to the fundamental problem of inferring the value of unknown physical quantities within a quantum system [1]. It has important practical applications in device characterization [2–4], quantum feedback and control [5–10] and quantum communications [11, 12], and most importantly, it is at the core of the field of quantum metrology [13–21], whereby quantum correlations are exploited to estimate physical quantities with high precision.

The most conventional approach to quantum parameter estimation is based on the following steps: (i) prepare a quantum system in a given state; (ii) allow it to evolve through a unitary transformation that encodes the unknown parameters; (iii) perform a measurement on the system; (iv) estimate the parameter from the measurement results; (v) repeat the process to gather statistical information and increase the precision of the estimation [15, 17]. Beyond this standard prepare-evolve-measure approach, there are relevant situations where, instead, one performs a single-shot continuous quantum measurement on a quantum probe [1]. In this case, one is interested in obtaining an estimation that will be continuously updated as data is gathered over the time evolution of a single quantum trajectory. Protocols of quantum parameter estimation from continuous measurements have been developed and studied over the years

[22–35]; such parameter estimation techniques are particularly relevant for applications such as atomic magnetometry [36–38], quantum metrology with open quantum systems [28, 39–41] or estimation of classical time-dependent signals [42]. Previous works have established computational approaches to calculate the probability density $P(\theta|D)$ [23, 29–32] of unknown n-dimensional parameters $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$ conditioned on the continuously recorded data, D, which could correspond, for instance, to photon-counting records [30, 31, 43], continuous weak measurements [44–46], or homodyne measurements [32]. These calculations rely on the application of Bayes' theorem on the likelihood functions $P(D|\theta)$, obtained as the trace of un-normalized conditional density matrices with the evolution conditioned on the measured data record [47].

This Bayesian approach provides, in principle, the most accurate description of the knowledge that can be gained about θ given the measurement data D. However, the application of this technique faces several difficulties. Firstly, it requires exact modelling of the open quantum system and any source of error, which can become challenging for noisy, complex quantum systems [48]. Secondly, the method can be computationally demanding. For instance, when tackling the problem of multi-parameter estimation, where one wishes to estimate several parameters simultaneously, Bayesian inference calls for the use of stochastic methods to sample the posterior probability distribution [29]. Generally, the Bayesian parameter estimation method will be computationally time-consuming, even in simple systems. This hampers the prospects for real-time estimation and for the integration of the inference process in the actual measurement device taking the data, which would be desirable for reduced latency and energy consumption [49, 50].
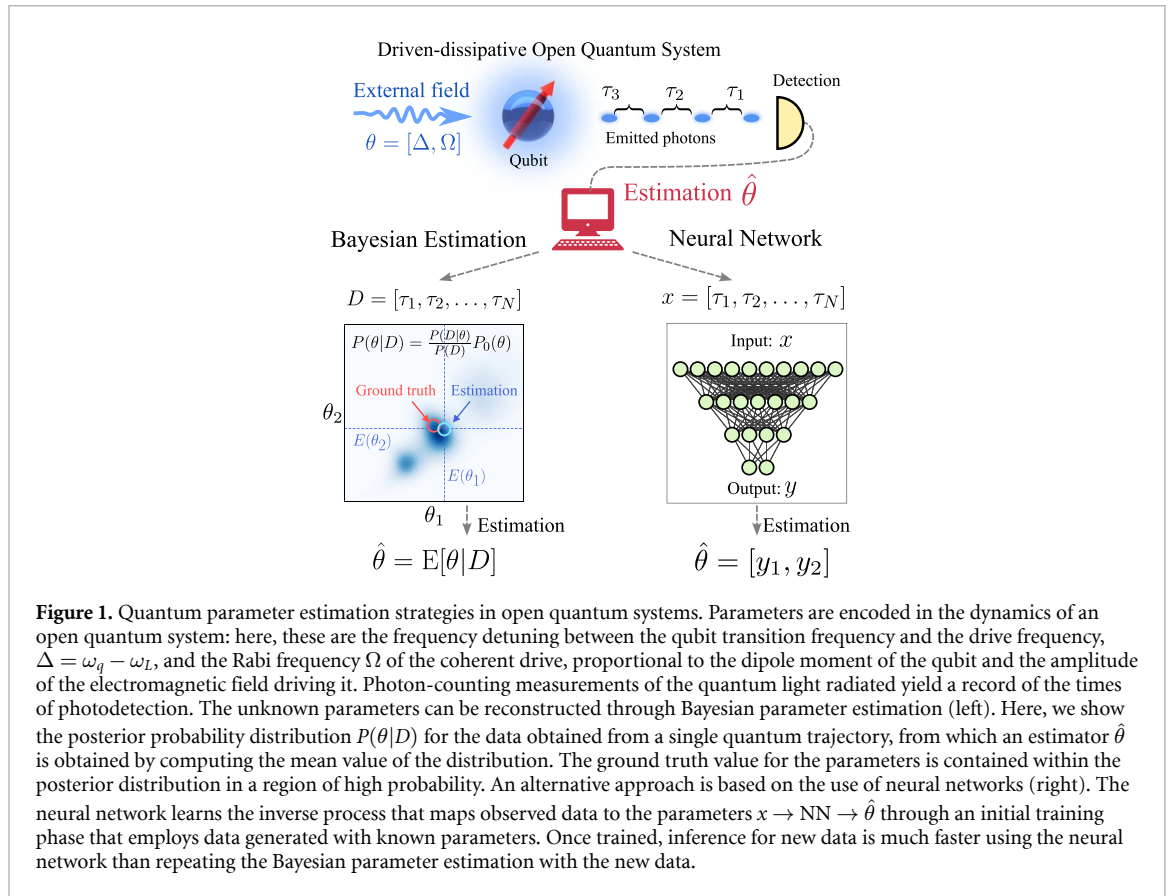
A novel approach to this problem that has gained attention in recent years is the application of machine-learning methods in which a neural network (NN) is trained to aid in the inference process. Crucially, this method allows to circumvent the requirement for detailed knowledge of the underlying physical system, as long as a reasonable amount of training data is available. The use of NNs have been demonstrated for a variety of tasks in quantum metrology, including parameter estimation in the presence of unavoidable noise [51–55], calibration [56, 57], phase estimation in interferometry [56–58], magnetometry [55, 59], control [54], and tomography [60, 61]. In the context of continuously monitored systems, NNs have been employed for the reconstruction of quantum dynamics [2, 62] and parameter estimation under continuous weak dispersive measurements [51, 52, 59, 63], which are characterized by a continuous, diffusive evolution of the quantum system [1, 64].

Here, we assess the potential of deep learning to tackle the parameter estimation task from continuous measurements of the photon-counting type. These measurements are described by point processes that, rather than yielding a diffusive evolution of the system, are characterized by sudden quantum jumps upon the acquisition of a discrete signal, such as the detection of a single photon [1]. This type of measurements, ubiquitous in the field of quantum optics and cavity-QED [65–72], provide insights into the particle-like nature of quantum systems and can produce strongly correlated signals that directly violate classical inequalities, as in the case of the renowned effect of photon antibunching [72]. Our work addresses the question of whether NNs can effectively process and exploit the quantum correlations present in this type of discrete signals.

Our proposed NN architectures approach this task as a regression problem, taking the time delays between individual photodetection events as an input and providing an estimate $\hat{\theta}$ for the desired parameters of interest as the output, as sketched in figure 1. Our main finding is that the NNs can take advantage of the quantum correlations within the signal, providing higher precision than using an equivalent signal that lacks these correlations, e.g. the integrated intensity value. By computing the root-mean-squared error (RMSE) of the estimations, we find that, after training, the NN can reach the limit of precision established by Bayesian inference, while requiring dramatically fewer computational resources. We also demonstrate that parameter estimation with NNs could be more resilient to noise than a simple Bayesian approach with no detailed knowledge of the underlying noise sources incorporated into the model, since it is possible for NNs to indirectly learn the noise model during the training phase. In contrast to a full Bayesian posterior distribution, the architectures studied in this work provide a single point estimate of the parameters, without including a confidence interval. Nevertheless, we evaluate their efficiency as estimators by conducting statistical analyses of their performance across numerous data samples.

2. System and data generation

The data we use to train the deep NNs and perform parameter inference consists of simulated measurements of single photo-detection times from the light emitted by a driven two-level system (TLS). This system represents a paradigmatic example of the types of quantum parameter estimation problems studied extensively in the literature [22, 23, 28]. It is a prototypical model that can describe various quantum emitters, such as quantum dots [73], molecules [74], or color centres [75]. We consider a TLS with states

Figure 1. Quantum parameter estimation strategies in open quantum systems. Parameters are encoded in the dynamics of an open quantum system: here, these are the frequency detuning between the qubit transition frequency and the drive frequency, $\Delta = \omega_q - \omega_L$, and the Rabi frequency Ω of the coherent drive, proportional to the dipole moment of the qubit and the amplitude of the electromagnetic field driving it. Photon-counting measurements of the quantum light radiated yield a record of the times of photodetection. The unknown parameters can be reconstructed through Bayesian parameter estimation (left). Here, we show the posterior probability distribution $P(\theta|D)$ for the data obtained from a single quantum trajectory, from which an estimator $\hat{\theta}$ is obtained by computing the mean value of the distribution. The ground truth value for the parameters is contained within the posterior distribution in a region of high probability. An alternative approach is based on the use of neural networks (right). The neural network learns the inverse process that maps observed data to the parameters $x \to \mathrm{NN} \to \hat{\theta}$ through an initial training phase that employs data generated with known parameters. Once trained, inference for new data is much faster using the neural network than repeating the Bayesian parameter estimation with the new data.

$\{|0\rangle, |1\rangle\}$, lowering operator $\hat{\sigma} = |0\rangle\langle 1|$ and transition frequency ω_q, continuously excited by a coherent drive with frequency ω_L and a Rabi frequency Ω. This quantum probe is coupled to the environment, giving rise to spontaneous emission with a rate γ. In the frame rotating at the drive frequency, the dynamics of such a system is described by the master equation (we set $\hbar = 1$ herein):

$$\partial_t \hat{\rho} = -i \left[\Delta \hat{\sigma}^\dagger \hat{\sigma} + \Omega \left(\hat{\sigma} + \hat{\sigma}^\dagger \right), \hat{\rho} \right] + \frac{\gamma}{2} D(\hat{\sigma}) [\hat{\rho}], \tag{1}$$
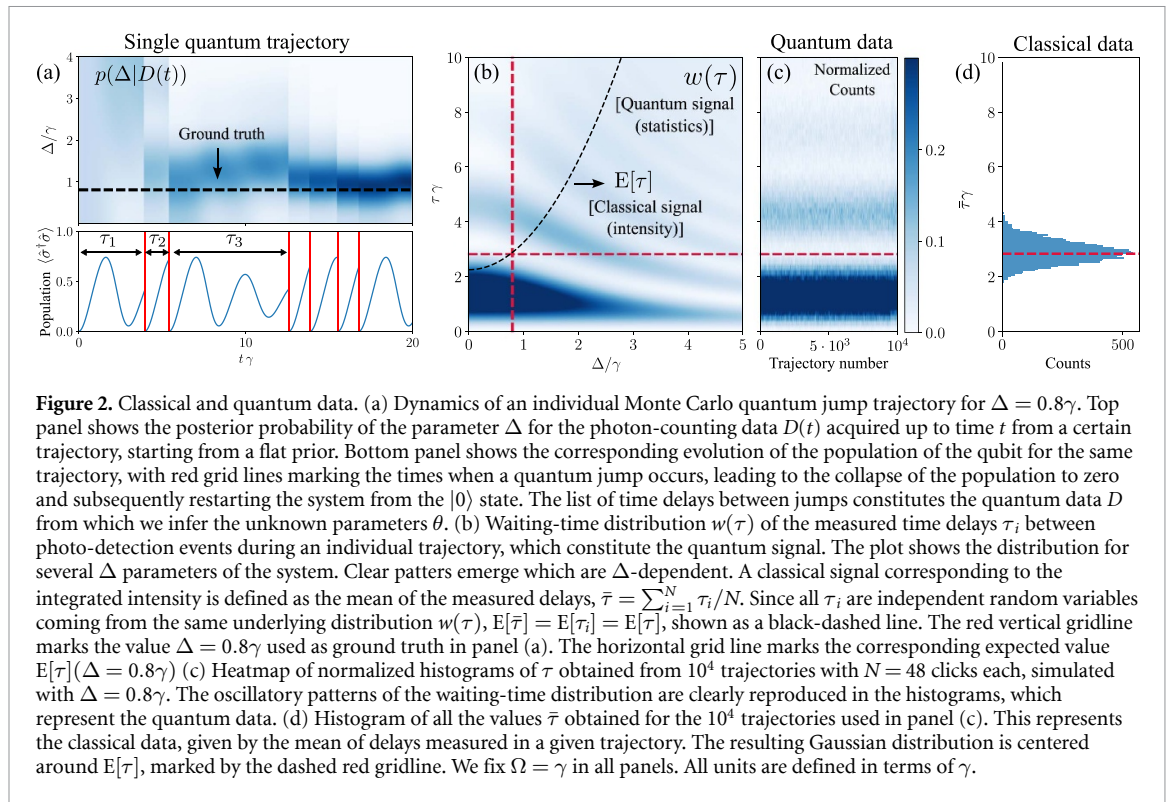
where $\Delta \equiv \omega_q - \omega_L$ and $D(\hat{x})[\hat{\rho}] \equiv 2\hat{x}\hat{\rho}\hat{x}^\dagger - \{\hat{x}^\dagger \hat{x}, \hat{\rho}\}$ follows the definition of the Lindblad superoperator [76]. The interplay between continuous driving and losses eventually brings the system into a steady state.

The estimation is performed using the data D generated by the continuous photon-counting measurement of the radiation emitted by the system during a single run of the experiment, considering an initial state $|\psi(0)\rangle = |0\rangle$ and a total evolution time, T, much longer than the relaxation time towards the steady state. We simulate this process using the Monte Carlo method of quantum trajectories [1, 47, 77] implemented in the QuTiP library [78, 79]. Under continuous photon-counting measurements, the system evolves stochastically under the following rules:

- at every differential time step, dt, the system can experience a quantum jump $|\psi(t + dt)\rangle \propto \hat{\sigma}|\psi(t)\rangle$ with a probability $p(t) = dt\gamma\langle\hat{\sigma}^\dagger\hat{\sigma}\rangle(t)$, which corresponds to the detection of a photon emitted by the system;
- otherwise, the system follows a non-Hermitian evolution $|\psi(t + dt)\rangle \propto [\mathbf{I} - i(\hat{H} - i\hat{\sigma}^\dagger\hat{\sigma}/2)dt]|\psi(t)\rangle$. The wavefunction is normalized in each time step.

Each stochastic simulation of a trajectory corresponds to an individual realization of the experiment, generating a data vector by storing the specific set of times $(t_1, \ldots, t_N)$ when a photodetection leading to a quantum jump occurs during the course of a total evolution time T, with N being the total number of jumps recorded. For convenience, we define our data as the set of time delays between these jumps, $D = [\tau_1, \ldots, \tau_N]$, where $\tau_i = t_i - t_{i-1}$, $i = 1, \ldots N$ and $t_0 = 0$ (see appendix A). An important feature of this particular system is that each time interval τ_i is an independent random variable, given that the system collapses to the ground state $|0\rangle$ after each emission [30].

Each trajectory is simulated for an evolution time T chosen to generate a data record D of precisely $N = 48$ jumps. This number is chosen to meet a compromise; it is sufficiently large to ensure that the system

Figure 2. Classical and quantum data. (a) Dynamics of an individual Monte Carlo quantum jump trajectory for $\Delta = 0.8\gamma$. Top panel shows the posterior probability of the parameter Δ for the photon-counting data $D(t)$ acquired up to time t from a certain trajectory, starting from a flat prior. Bottom panel shows the corresponding evolution of the population of the qubit for the same trajectory, with red grid lines marking the times when a quantum jump occurs, leading to the collapse of the population to zero and subsequently restarting the system from the $|0\rangle$ state. The list of time delays between jumps constitutes the quantum data D from which we infer the unknown parameters θ. (b) Waiting-time distribution $w(\tau)$ of the measured time delays τ_i between photo-detection events during an individual trajectory, which constitute the quantum signal. The plot shows the distribution for several Δ parameters of the system. Clear patters emerge which are Δ-dependent. A classical signal corresponding to the integrated intensity is defined as the mean of the measured delays, $\bar{\tau} = \sum_{i=1}^{N} \tau_i / N$. Since all τ_i are independent random variables coming from the same underlying distribution $w(\tau)$, $E[\bar{\tau}] = E[\tau_i] = E[\tau]$, shown as a black-dashed line. The red vertical gridline marks the value $\Delta = 0.8\gamma$ used as ground truth in panel (a). The horizontal grid line marks the corresponding expected value $E[\tau](\Delta = 0.8\gamma)$ (c) Heatmap of normalized histograms of τ obtained from 10^4 trajectories with $N = 48$ clicks each, simulated with $\Delta = 0.8\gamma$. The oscillatory patterns of the waiting-time distribution are clearly reproduced in the histograms, which represent the quantum data. (d) Histogram of all the values $\bar{\tau}$ obtained for the 10^4 trajectories used in panel (c). This represents the classical data, given by the mean of delays measured in a given trajectory. The resulting Gaussian distribution is centered around $E[\tau]$, marked by the dashed red gridline. We fix $\Omega = \gamma$ in all panels. All units are defined in terms of γ.

is in a steady state, while also allowing for inference with a limited number of photons in a relatively short timescale spanning just a few tens of emitter lifetimes, which is typically within the range of nanoseconds for solid-state quantum emitters [66]. It is important to note that fixing N requires that each trajectory will have a different evolution time T [80, 81]. While the choice of fixing N allows us to work straightforwardly with inputs of uniform dimension, there may be experimental scenarios where T is a more relevant metrological resource. In such cases, it may be necessary to maintain a fixed T while allowing N to vary. Even in this scenario, the analysis outlined below remains applicable, and the design and training of the NNs could be easily adapted, as we discuss in appendix D. More detailed information about the unravelling of the dynamics in quantum trajectories can be found in the supplemental material [82].

After obtaining the data D from a single trajectory or experiment, our goal is to estimate the unknown parameters θ that govern the dynamics of the system. In this study, we specifically focus on the parameters Δ and/or Ω, considering that γ is known. While this might be an idealized setting, it is instructive enough to benchmark the power of quantum correlations and NNs for the task of parameter estimation. We are interested in benchmarking the NNs against the ultimate precision limit set by the Bayesian inference process. This process yields a posterior probability distribution for the unknown parameters conditioned on the measured data D: following Bayes' rule we have $P(\theta|D) = P(D|\theta)/\int d\theta P(D|\theta)$. A comprehensive explanation of Bayesian inference from continuous measurements is provided in [28] and, within the context of this work, in appendix B and [82]. An illustrative individual trajectory, generated through our simulation approach, is depicted in figure 2(a), which shows the evolution of the posterior probability distribution for a single parameter, namely $\theta = [\Delta]$, in the top panel, and the TLS population in the bottom panel. Red grid lines mark the times when a quantum jump takes place. One can see how the regions of high probability in the posterior distribution concentrate more around the ground truth as time increases and more data is collected.

Once the posterior distribution is obtained, we build an estimator by taking the mean, $\hat{\theta} = E[\theta|D] = \int d\theta P(\theta|D)\theta$, which is known to be the estimator that minimizes the RMSE [83]. While reducing the full posterior to a single estimation $\hat{\theta}$ implies a loss of information and involves a subtle choice of the best estimator, this step allows us to straightforwardly compare the result with the estimation provided by the NNs. Here we also note that we use a flat (uniform) prior on the parameters of interest, while in general the Bayesian framework allows us to introduce our knowledge of the system using arbitrary priors.

A central question of our study is whether NNs can leverage the quantum correlations in the data. The emission from a TLS is strongly correlated; the fact that the emitter can only store one excitation at a time gives rise to the well-known non-classical phenomenon of antibunching [72], i.e. the suppressed probability of detecting two photons simultaneously. These quantum correlations can be evidenced through several

observables, such as the waiting-time distribution $w(\tau)$ (the probability of two consecutive emissions to be separated by a time delay τ), which presents a vanishing value at zero delay, $w(\tau = 0) = 0$, and an oscillatory behaviour that depends strongly on the system parameters. These features, shown in figure 2(b), are translated directly into the histograms of time delays observed in each individual trajectory, as we show in figure 2(c). We will refer to a signal containing these time delays as a quantum signal.

To assess the information contained in these correlations, we compare it to an equivalent classical signal where the only relevant value is the integrated intensity, i.e. the total number of counts detected per unit time, $I = N/T$. We can interpret such a classical signal as measurements lacking the necessary time resolution to detect individual photons, providing an integrated intensity without any access to the internal temporal correlations between photon emissions. In fact, we choose the inverse of the intensity as the classical signal, since it conveniently represents that we have reduced the list of measured delays, that constitute the quantum signal, by taking their average, $\bar{\tau} \equiv \sum_{i=1}^{N} \tau_i/N = I^{-1}$. Since $\bar{\tau}$ is a sum or independent random variables τ_i that all follow the probability distribution $w(\tau_i)$, the central limit theorem ensures that, when N is large, $\bar{\tau}$ is itself a random variable that follows instead a normal distribution, with $\mathrm{E}[\bar{\tau}] = \mathrm{E}[\tau_i]$. This normal distribution stands in contrast to the more complicated probability distribution $w(\tau_i)$ followed by the individual random variables τ_i. In the steady state, this distribution is centered around $\mathrm{E}[\tau_i] = \mathrm{E}[\tau] \equiv \int \mathrm{d}\tau w(\tau)\tau = (\langle \hat{\sigma}^\dagger \hat{\sigma} \rangle_{ss} \gamma)^{-1}$ (see figure 2(d)).

Access to $\mathrm{E}[\tau]$ instead of $w(\tau)$ represents an obvious loss of information, where any contributions from quantum correlations is completely discarded. Thanks to this, the RMSE of the estimation made by using the classical signal serves as a reference for comparison when doing estimations with the full quantum signal: whenever the obtained RMSE is lower than that of the classical signal, we understand that the estimation method is taking some advantage from quantum correlations. The process of parameter estimation from this classical signal, described in more detail in appendix C, also follows a Bayesian approach in which the chosen estimator $\hat{\theta}$ is the mean of the posterior $P(\theta|\bar{\tau})$.
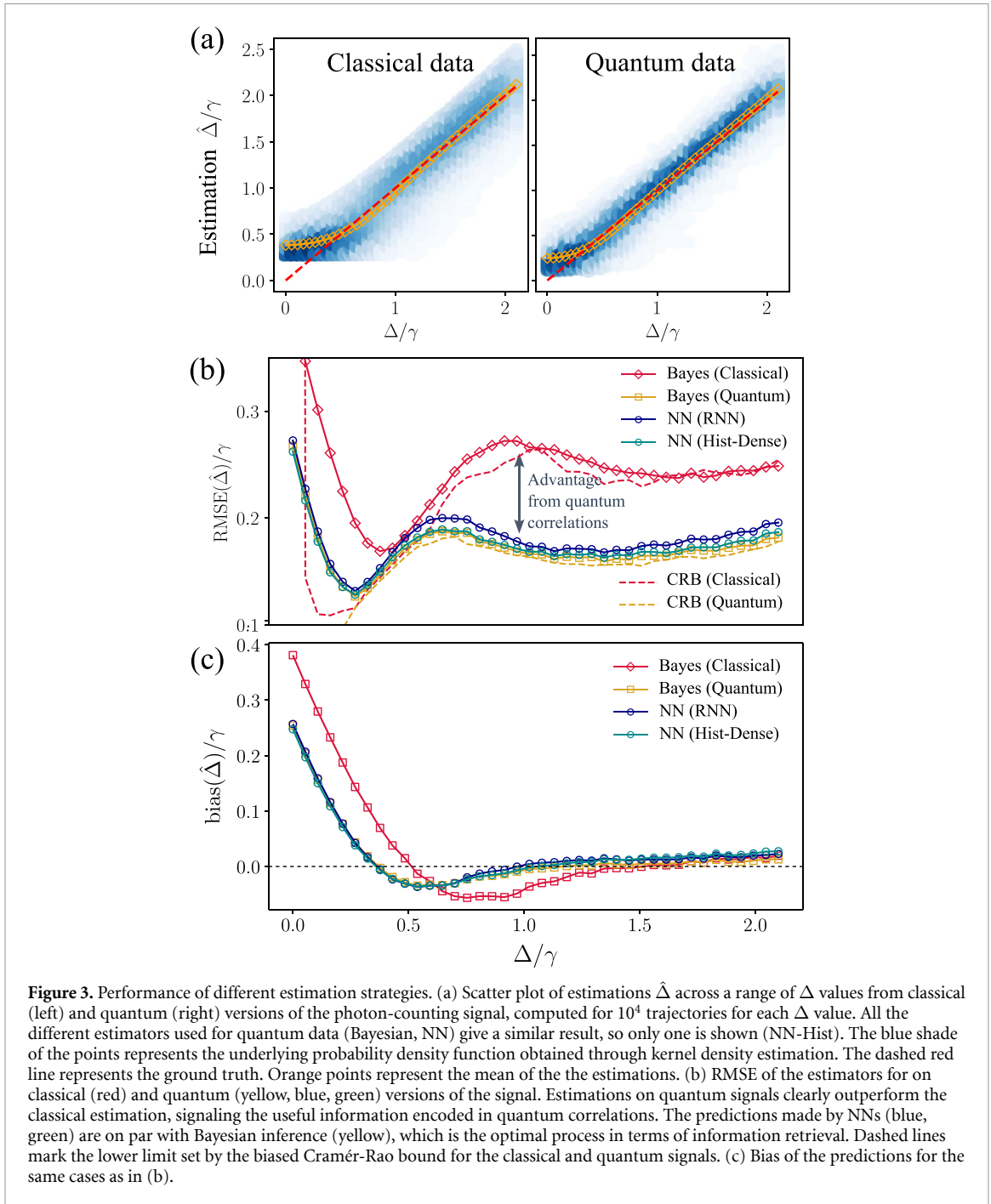
3. Training

We considered different neural-network architectures with the common structure of inputs and outputs that we illustrate in figure 1. The input consists of a vector $x = D = [\tau_1, \tau_2, \ldots, \tau_N]$ of time delays between photodetection events, while the output is a vector of size n corresponding to the estimations of the n-dimensional estimation problem, $y = [\hat{\theta}_1, \hat{\theta}_1, \ldots, \hat{\theta}_n]$. Here, we treat both one-dimensional estimation problems, with $\theta = [\Delta]$ in the 1D case, and $\theta = [\Delta, \Omega]$ in the 2D case. We fixed the size of the input vector to $N = 48$, indicating that our networks perform estimations based on a single experiment run once 48 photons have been detected.

Estimation via NNs requires a training stage akin to a calibration process (learning phase). In this work, the input data used for training, x_{train}, consists of a list of time delays obtained from trajectories simulated in QuTiP, with the corresponding target data y_{train} being the ground truth parameters used to run the simulations. Our best results are achieved using two types of architectures:

1. *RNN*: recurrent neural network (RNN) with two long short-term memory (LSTM) layers. These networks can learn correlations in sequential data, rendering them particularly promising for exploiting temporal quantum correlations. It is worth noting that there are no such correlations between time delays in our present system, since the τ_i are independent due to the collapse following a quantum jump. Nonetheless, these networks still deliver a very good performance. More information on these networks and their application in parameter estimation from continuous signals can be found in [59].

2. *Hist-Dense*: Fully connected NN with a first custom layer designed to return a histogram of the input data. The geometry of the bins of the histogram is defined by the number of bins and their range. We set these as non-trainable hyperparameters fixed at 700 total bins and a range given by $\tau_{\min} = 0$ and $\tau_{\max} = 100/\gamma$. This layer produces an efficient, physically-informed representation of the data, which leverages the absence of correlations between time delays in this particular system.
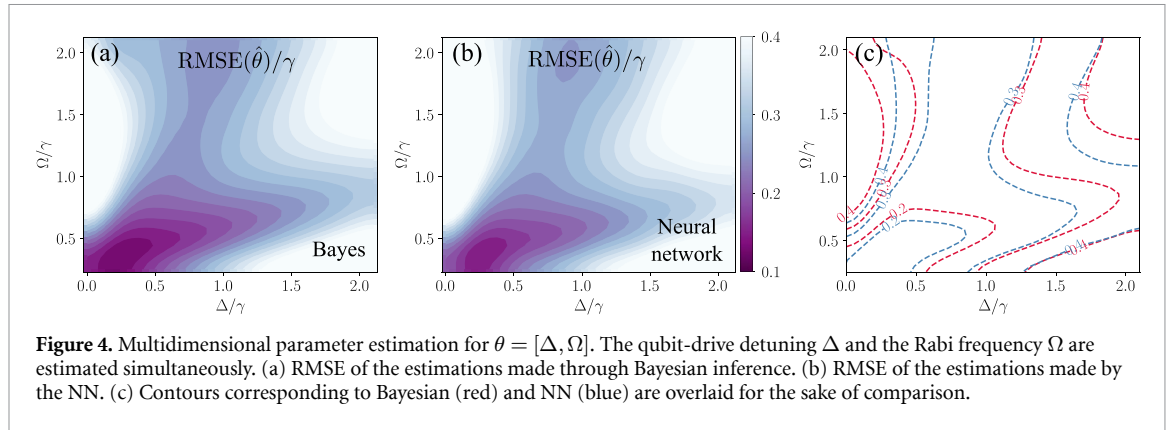
Further details on the networks and the training process, can be found in appendix D. There, we also discuss the minimal variations that would be required to handle the situation in which T is fixed and, consequently, the inputs have variable data size N. In all the numerical calculations, the unit of time is set by $1/\gamma$ (which we assume to be a known parameter) by fixing $\gamma = 1$. After training, we benchmark the estimations made by the NNs against the estimation made through Bayesian inference and, in the single-parameter estimation case, also against the estimation done on the classical version of the signal. Details of this validation process are in appendix E.

Figure 3. Performance of different estimation strategies. (a) Scatter plot of estimations $\hat{\Delta}$ across a range of Δ values from classical (left) and quantum (right) versions of the photon-counting signal, computed for 10^4 trajectories for each Δ value. All the different estimators used for quantum data (Bayesian, NN) give a similar result, so only one is shown (NN-Hist). The blue shade of the points represents the underlying probability density function obtained through kernel density estimation. The dashed red line represents the ground truth. Orange points represent the mean of the the estimations. (b) RMSE of the estimators for on classical (red) and quantum (yellow, blue, green) versions of the signal. Estimations on quantum signals clearly outperform the classical estimation, signaling the useful information encoded in quantum correlations. The predictions made by NNs (blue, green) are on par with Bayesian inference (yellow), which is the optimal process in terms of information retrieval. Dashed lines mark the lower limit set by the biased Cramér-Rao bound for the classical and quantum signals. (c) Bias of the predictions for the same cases as in (b).

4. Single-parameter estimation (1D)

First, we focus on the simplest situation in which a single parameter is estimated: the detuning $\theta = [\Delta]$. Figure 3(a) depicts a scatter plot of the predictions made by each method for different values of Δ. These plots show that the predictions made using the quantum signal—regardless of whether Bayesian inference or NNs are used—are more densely concentrated around the ground truth than the predictions obtained using the classical version of the signal. This already suggests that quantum correlations in the data can provide an advantage.

This advantage is clearly quantified in figure 3(b), which compares the RMSE of the predictions obtained by each of the methods computed on a different set of validation trajectories (see appendix E for details on the validation data). The estimation from classical data (red) always has a higher RMSE than estimation methods that have access to the quantum correlations in the data. Figure 3(b) also shows that the NN

Figure 4. Multidimensional parameter estimation for $\theta = [\Delta, \Omega]$. The qubit-drive detuning Δ and the Rabi frequency Ω are estimated simultaneously. (a) RMSE of the estimations made through Bayesian inference. (b) RMSE of the estimations made by the NN. (c) Contours corresponding to Bayesian (red) and NN (blue) are overlaid for the sake of comparison.

architectures used in this work learn to extract information encoded in these correlations, with the resulting RMSE reaching values very close to the ultimate limit set by Bayesian inference. The fact that NNs perform as well as the Bayesian approach by leveraging on quantum correlations in a completely model-agnostic manner is the main result of our work.

We note, however, that none of the approaches studied here provided an estimator that is unbiased for all Δ, as we show in figure 3(c). There, we plot the bias of several estimators, defined as $\mathrm{bias}(\hat{\theta}) \equiv \mathrm{E}(\hat{\theta} - \theta)$. This bias is not problematic *per se*, since it is known that biased estimators can be preferable to unbiased ones in some cases [83]. However, this bias needs to be taken into account at the time of comparing the obtained RMSE with the lower bound set by the Fisher information and the Cramér–Rao bound (CRB) [84, 85]. As we discuss in more detail in [82], one needs to consider the *biased* CRBs [86]. We calculated the Fisher information and the biased CRBs for the photon-counting measurements considered here [28–30], incorporating the calculated biases shown in figure 3(c). The resulting lower limits for the RMSE are shown as dashed lines in figure 3(b). We show that, for the quantum signal, both the Bayesian and NN estimators saturate the CRB for a wide range of values of Δ.

5. Multi-parameter estimation (2D)

We also benchmarked our results for the 2D parameter-estimation case, where both Δ and Ω are unknown and estimated simultaneously, $\theta = [\Delta, \Omega]$.

Our results for the 2D estimation case are shown in figure 4. In panel (a), we depict the RMSE of Bayesian estimation from the quantum signal in a region of the plane $[\Delta, \Omega]$. The resulting map shows a nontrivial structure that reflects the underlying physical complexity of the system and that is consistent with previous reports in the literature [29, 30]. The same physical complexity is captured by the NNs, as shown in figure 4(b), which depicts the RMSE obtained by the estimation of a *Hist-Dense* architecture and shows that the RMSE reflects the same underlying structure as the one obtained with Bayesian inference, yielding comparably small values overall. Both results can be more clearly compared in figure 4(c), depicting overlaying contour lines of the two plots in panels (a) and (b).

Our results clearly establish that NNs are able to exploit the information hidden in quantum correlations in order to perform parameter estimation. Given their comparable precision to Bayesian inference, a careful evaluation of the benefits of opting for NNs as an alternative becomes important.

6. Computation efficiency

A first aspect in which NNs can bring an important advantage is the efficiency of the computation, which is particularly relevant for the case in which one wishes to achieve real-time estimation. For the 2D case of simultaneous estimation of two parameters, the posterior for a single trajectory for Bayesian estimation can be computed using the UltraNest package [87] (see appendix B) in a typical wall time of 10 seconds on a MacBook Pro 2023 with an Apple M2 Max Chip. On the other hand, an estimation using the NNs built in this work, while being less informative (unlike the Bayesian posterior they provide a single estimation without a confidence interval) can be computed in a wall time of less than 1 ms, on the same device. The efficiency of the required calculation is such that trained NNs can be converted into compact and efficient TensorFlow Lite models, with a size of approximately 90 kB. These models are suitable for deployment and

execution on microcontrollers, which would minimize resource demands and enable data collection and estimation on the same device, leading to a significant reduction in latency [49, 50]. We can expect that, during data collection of photon-counting measurements, different chunks of measured time-delays are fed into the trained NN models for inference in real time. Such a situation would not only allow for precise and robust real time inference, but could also inform about time-drifting parameters by looking at the temporal sequence of predictions.

7. The challenge of noise

Another important challenge faced by the Bayesian inference approach is the requirement of a precise modelling of the underlying physical system. This become particularly problematic in the presence of noise, which not only requires a perfect characterization and modelling of the different sources of imperfections, but also makes the subsequent calculation of posteriors much more involved. In contrast, the model-agnostic approach of using NNs allows for a robust prediction without the need of any modelling of the sources of noise, only by training the network directly on noisy data. In the particular case of photon-counting with single-photon detectors, numerous sources of noise can be attributed just to the detection process, such as the efficiency, bandwidth, dead time, dark counts or jitter noise [71]. All these sources of noise have non-trivial consequences on the statistics of the photodetection events [88] and on the quantum trajectories associated to these imperfect measurements [89]. Hence, the necessity of a perfect modelling and characterization of these processes seriously hampers the utilization of Bayesian inference.

7.1. Noise in x: time jitter
We illustrate the potential advantage brought in by the estimation with NNs by considering the particular case of time jitter [88], which we describe by adding a noise term to the values of time delays τ used both in the training and estimation of the NNs for the 1D estimation case $\theta = [\Delta]$. We consider noise following a normal distribution with standard deviation σ_τ, so that $x \to x + x_{noise}(\sigma_\tau)$, with $x_{noise}(\sigma_\tau) \sim \mathcal{N}(0, \sigma_\tau)$. In order to assess the resilience of the NN predictors to this type of noise, we defined a list of values of σ_τ in the range $\sigma_\tau \in [0, \gamma^{-1}]$. For each value in this list, we sampled a population of x_{noise} and trained a model with the corresponding noisy data.

The resulting RMSE of the predictions returned by each of these models is computed using the same validation trajectories used in figure 3, but containing an added Gaussian noise with the same σ_τ as the data used to train the model. The resulting RMSEs are shown in figure 5(a), compared to the RMSE of the prediction via an imperfect Bayesian inference process in which the noise is not explicitly modeled. We observe that, in the case of NNs, the results are more insensitive to the jitter noise and the RMSE of the NN predictor is always lower than that of the Bayesian predictor. This result remains true for all values of Δ, as shown in figure 5(b). We emphasize that this did not require any modelling of the imperfections, as would have been required for a correct process of Bayesian inference. The latter would presumably recover the values reported by the NN, but requiring a perfect characterization of the sources of noise, which may be challenging in some scenarios.

7.2. Noise in y_{train}: imperfect calibration
Another relevant question is the effect of noise in the training target data y_{train}. In particular, we consider the case in which the target data used for training differs from the underlying ground truth used to generate the quantum trajectories by a Gaussian noise term with standard deviation σ_y, so that $y_{train} \to y_{train} + y_{noise}(\sigma_y)$, with $y_{noise} \sim \mathcal{N}(0, \sigma_y)$. This could reflect, for instance, the finite precision of a calibration device employed to experimentally measure the set of target values used to train the network.

The question we address is whether a NN will be able to perform parameter estimation with a lower RMSE than the standard deviation σ_y of the noisy target data used for training. A positive answer to this question would indicate that the precision of the NN estimator can outperform that of the calibration method or device used to generate the training data. Figures 5(c) and (d) illustrate that this scenario is indeed possible. Similarly to the process described for figure 5(a), we defined a series of values of σ_y and, for each value, we sampled a population of y_{noise} and trained a model with noisy target data. The validation of the trained models is done using the same set of trajectories as in figure 3. Figure 5(c) shows that, once the standard deviation of the noisy training target data σ_y (dot-dashed line) becomes greater than the RMSE achievable in the noiseless case (horizontal grid line), the predictions made by the NNs trained with this noisy data remain robust and approximately equal to the ideal RMSE. As σ_y increases, the RMSE of the NN predictor eventually reaches a threshold where it starts increasing as well, but remaining much smaller than

Figure 5. Parameter estimation with noisy data. (a) RMSE of NN predictor trained and evaluated with data containing jitter noise given by a Gaussian distribution with standard deviation σ_τ (green circles). This is compared to the RMSE of Bayesian inference lacking a description of the noise (yellow crosses). Results shown for a fixed $\Delta = 0.27\gamma$. (b) Same as in panel (a), shown versus Δ and with a fixed jitter noise $\sigma_\tau = 0.76\gamma$. RMSE of Bayesian predictor in the absence of noise shown for comparison (gray crosses). (c) RMSE of NN predictor trained with noisy target training data y_{train}, with noise following a Gaussian distribution with standard deviation σ_y. We observe that the RMSE of the NN predictor (green circles) can remain smaller than the standard deviation of the training data σ_y (gray dot-dashed line). The horizontal grid line marks the RMSE of Bayesian inference with noiseless data for comparison . Results shown for a fixed $\Delta = 0.27\gamma$. (d) Same RMSE as in panel (c), shown as a function of Δ. Different colors represent different values of σ_y, which are also represented as horizontal grid lines, confirming that the RMSE of NN predictor remains smaller than σ_y for all values of Δ.

σ_y. Such threshold is established by the size of the training dataset. This observation holds for all values of Δ, as shown in figure 5(d). This result highlights the potential of the NNs to improve sub-optimal estimations used to feed the training process.

8. Conclusions and outlook

We have established the potential of NNs to perform parameter estimation from the data generated by continuous measurements of the photon-counting type. We have carefully benchmarked their precision against the ultimate limit set by Bayesian inference, showing that NN approaches can reach comparable performance in a much more computationally efficient way, while also being more robust against noise.

These results pave the way for future research that can explore more complex quantum optical systems and measurement schemes. For example, future studies could investigate the radiation from cavity-QED systems with non-trivial quantum statistics [65, 70, 90, 91] or several quantum emitters, as well as multiple simultaneous detection channels, such as those provided by arrays of single-photon avalanche diodes for generalized Hanbury-Brown and Twiss measurements [66, 68], streak cameras [67], or single-photon fibre bundle cameras [69, 92]. We anticipate that such setups could yield more complex temporal correlations than those considered in this study, where the various time delays are independently sampled from the waiting time distribution. More complex temporal correlations have the potential to lead to non-trivial scaling of the Fisher information with time T or the number of detected photons N [39]. Investigating whether the saturation of the CRB by the NNs that we report in this work extends to these complex scenarios represents a promising avenue for further exploration.

One aspect we leave for future work is the exploration of NN methods that incorporate mechanisms to do uncertainty quantification (UQ) on the predictions (compute prediction intervals) [93]. For example, Bayesian Neural Networks [94] train a set of weights distributions and the predictions from those networks are obtained by sampling the posterior of the weights after training. Other approaches to UQ include bootstrapping and ensemble methods which are easily implemented and only require extra training time, maintaining all the benefits of NN estimators that we have shown in our work.

Another aspect deserving further exploration is leveraging domain knowledge and incorporating physical models into the learning process [2, 4]. We note that the *Hist-Dense* architecture used in this study could already be considered a physics-informed approach, taking advantage of a known feature about the model: the absence of information contained in the particular ordering of delays in the input data x, given that the source is a TLS that resets its state after every emission [30].

Moreover, it is worth noting that protocols of parameter estimation from photon-counting data, as the one described here, may find applications in fields such as fluorescence quantum microscopy [69, 92, 95–97] or quantum imaging [98, 99]. In these areas, experimental quantum signals beyond the capabilities of classical simulations—due to the number of emitters and detectors involved—are already within reach, calling for efficient methods of analysis for the optimal image reconstruction [100].

Data availability statement

All the data necessary to reproduce the results in this work, including simulated quantum trajectories for training of the NNs, validation trajectories, trained models and Bayesian inference calculations in 2D via nested sampling are openly available on Zenodo [101].

The codes used to generate the results in this work, including the implementation of the proposed NN architectures as TensorFlow models, are openly available on Zenodo and GitHub [102].

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/records/2009;509.

Acknowledgments

Appendix A. Photodetection time data

In practice, rather than working with the absolute times of detection, t_i, we find it more convenient to consider D to be equivalently composed of a list of time delays $\tau_i = t_i - t_{i-1}$, with $i = 1, \ldots N$ and $t_0 = 0$. For simplicity of calculations, we consider trajectories with different values of total evolution time T but a fixed number of total detections N, and assume that the measurement is immediately terminated after the last detection, i.e. $t = t_N$. This means that $t - t_N$ will always be zero, becoming an irrelevant time interval that we do not need to include in our data, which will then only consist of N values, $D = [\tau_1, \ldots \tau_N]$. This format provides a set of data records with a fixed number of photodetection events or 'clicks' that can be easily used as inputs for the NNs that we use in this work.

Appendix B. Bayesian Inference

Bayes' Rule.— We consider that, at the beginning of our estimation protocol, our knowledge of the unknown set of parameters is characterized by a constant prior probability distribution $P_0(\theta)$ over an arbitrarily large but finite support (we define an initial support that is big enough not to affect the final estimation, and alternative prior distributions can also be used). Our goal is to update this knowledge upon the new evidence provided by some measurement data D, i.e. to compute the posterior probability $P(\theta|D)$. According to Bayes' rule, this is given by

$$P(\theta|D) = \frac{P(D|\theta)}{P(D)} P_0(\theta) = \frac{P(D|\theta)}{\int d\theta P(D|\theta) P_0(\theta)} P_0(\theta), \tag{B1}$$

where the integral is defined over the domain of possible values of θ, i.e. the support of $P_0(\theta)$. If we have no prior knowledge of θ, as we assume here, $P_0(\theta)$ is constant over its support and can be taken out of the integral in the denominator in equation (B1), leaving us with a simple relation between $P(\theta|D)$ and $P(D|\theta)$:

$$P(\theta|D) = \frac{P(D|\theta)}{\int d\theta P(D|\theta)}. \tag{B2}$$

Thus, the required step to compute the posterior $P(\theta|D)$ is the calculation of the likelihood $P(D|\theta)$, which in general can be computed as the norm of a conditional density matrix evolving according to the list of quantum-jump times in data D, as explained in detail in [82].

Likelihood computation.— The calculation of the likelihood is particularly simple for the dynamics of a TLS, since each time interval is independent of the others, given that the system is completely reset to the ground state $|0\rangle$ after each emission [30]. Consequently, the probability for the set of data $D = [\tau_1, \ldots, \tau_N]$ is simply given by the product of the probabilities of each interval

$$P(D|\theta) = \prod_{i=1}^{N} w(\tau_i; \theta), \tag{B3}$$

where the waiting-time distribution $w(\tau; \theta)$ gives the probability of two successive clicks being separated by a time interval τ and we made explicit its dependence on the system parameters θ. In our system, governed by the master equation (1), the waiting-time distribution has the following analytical form,

$$
\begin{aligned}
w(\tau; \theta) = \frac{8\gamma\Omega^2}{R} \exp\left[-\gamma\tau/2\right] \\
\times \sum_{\eta=-1,1} \eta \cosh\left(\frac{\tau\sqrt{\gamma^2 - 4\left(\Delta^2 + 4\Omega^2\right) + \eta R}}{2\sqrt{2}}\right)
\end{aligned}
\tag{B4}
$$

with $R = \sqrt{\left[\gamma^2 + 4\left(\Delta^2 + 4\Omega^2\right)\right]^2 - 64\gamma^2\Omega^2}$. As explained in more detail in [30], we obtain this expression directly from

$$w(\tau) = \gamma\tilde{\rho}_{ee}(\tau), \tag{B5}$$

where $\tilde{\rho}_{ee}(\tau) \equiv \langle e|\tilde{\rho}(\tau)|e\rangle$ is the excited-state population of the conditional un-normalized density matrix solution to the master equation in the absence of any quantum jumps, given by

$$\partial_\tau \tilde{\rho} = -i\left[\hat{H}, \tilde{\rho}\right] - \frac{\gamma}{2}\left\{\hat{\sigma}^\dagger\hat{\sigma}, \tilde{\rho}\right\}, \tag{B6}$$

with $\tilde{\rho}(0) = \hat{\rho}(0)$.

Posterior computation.— Once the method to compute the likelihood $P(D|\theta)$ is established, the only obstacle to using Bayes rule in equation (B1) is the calculation of the normalization factor in the denominator. This factor is called the evidence or the marginal likelihood, as it marginalizes the likelihood over all possible values of the parameters. While this integral can be computed analytically for simple likelihood distributions, or numerically efficiently for a low-dimensional parameter space, many sampling techniques are typically used in Bayesian inference to bypass its direct computation. Markov Chain Monte Carlo [103] is a stochastic sampling method that allows us to compute samples from the posterior $P(\theta|D)$ as draws from multiple Markov chains evolving in Monte Carlo time according to a stationary distribution proportional to the posterior. Another method is Nested Sampling (NS) [104, 105], which is able to provide an estimate for the marginal likelihood itself by converting the original multi-dimensional integral over the parameter space to a one-dimensional integral over the inverse volume of the prior support. NS methods have been shown to be effective also in the case of multi-modal posteriors (related to parameter degeneracy) and when the posterior distribution has heavy or light tails, and they are routinely used in fields such as astrophysics and cosmology when Bayesian model selection is crucial.

For the multi-dimensional case $n = 2$, we compute the posterior probability distributions $P(\theta|D)$ using the UltraNest package [87] in Python. UltraNest contains several advanced algorithms to improve the efficiency and correctness of nested sampling, including bootstrapping uncertainties on the marginal likelihood and handling vectorized likelihood functions over parameter sets (together with additional MPI support).

Appendix C. Parameter estimation from classical signals

We consider that the classical version of the measured quantum signal is given by the sample mean of the N observed delays $\bar{\tau} = \sum_{i=1}^{N} \tau_i/N$. Being a sum of many independent random variables τ_i, $\bar{\tau}$ is itself a random variable that, in the limit of large N, will follow a Gaussian distribution $\bar{\tau} \sim \mathcal{N}(\mu, \sigma)$, where $\mu = \mathrm{E}[\tau_i]$ and

$\sigma^2 = \mathrm{Var}[\tau_i]/N$. Since the independent random variables follow the waiting-time distribution given by equation (B4), $\tau_i \sim w(\tau; \theta)$, μ and σ can be directly obtained:

$$\mu(\theta) = \left[\gamma\langle\hat{\sigma}^\dagger\hat{\sigma}\rangle_{\mathrm{ss}}(\theta)\right]^{-1} = \frac{\gamma^2 + 4\Delta^2 + 8\Omega^2}{4\gamma\Omega^2}, \tag{C1a}$$

$$\sigma(\theta)^2 = \frac{\left(\gamma^2 + 4\Delta^2\right)^2 - 8\left(\gamma^2 - 12\Delta^2\right)\Omega^2 + 64\Omega^4}{N16\gamma^2\Omega^4}, \tag{C1b}$$

were we assume that the emission is recorded when the driven emitter has reached a steady state and we made explicit the dependence of these quantities on the unknown parameters θ. Within a Bayesian framework, this means that the likelihood of the classical data is

$$P(\bar{\tau}|\theta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\bar{\tau} - \mu}{\sigma}\right)^2\right], \tag{C2}$$

and the posterior is given by equation (B2), assuming a constant prior. From here, we obtain a single estimation by taking the mean of the posterior, i.e. $\hat{\theta} = \mathrm{E}[\theta|D] = \int d\theta\,\theta P(\theta|D)$. We make this choice in order to be consistent with the estimator used for Bayesian inference with quantum data, and because it shows overall lower RMSE and bias than the maximum-likelihood estimator, see [82].

Appendix D. Training and design of the neural networks

For the 1D estimation case, the networks are trained with 4×10^6 trajectories in which the first 48 photodetection times are stored. From these, 80% are taken as training data, and 20% as validation data during the training process. Each trajectory is simulated by fixing $\Omega = \gamma$, and choosing randomly a value of the detuning within the domain $\Delta \in [0, 5\gamma]$, which should be large enough to cover the expected range of possible values of Δ. We choose this domain to be the same as the support of the prior in the Bayesian inference, so that both methods rely on the same underlying assumptions about the probability distribution of Δ. All the networks are trained using the Adam optimizer. A summary of these architectures—including number of trainable parameters and details about the batch size and number of epochs used in training—is provided in table 1. All the networks are defined and trained using the Tensorflow library [106].

For the 2D parameter estimation, we train the network with a new set of 4×10^6 trajectories generated for random values of Δ and Ω in the range $\Delta \in [0., 3\gamma]$ and $\Omega \in [0.25\gamma, 5\gamma]$ (lower values of Ω are not used in order to avoid longer simulations times associated with the lower pumping rate).

The first layers of both architectures (Histogram and LSTM) accept an input of any size. This would allow, in principle, to adapt our training and inference to the situation in which T is considered a relevant metrological resource to keep fixed, and N is let to change among trajectories. However, efficient batch processing during training typically requires a uniform input size. For the situation of non-uniform data obtained by fixing the evolution time T for all trajectories (resulting in a non-constant size N of each data record), this could easily be achieved through preprocessing, setting a consistent data size among trajectories by padding the quantum-jump data with neutral values that would be subsequently disregarded by the input layer.

Table 1. Summary of the NN architectures used for parameter estimation. All networks are trained with the Adam optimizer, with a learning rate 0.001, using a batch size of 12 800 and with a mean squared logarithmic error (MSLE) loss function.

Recurrent Neural Network (RNN)

Layer	Output shape	Activation	# Parameters
LSTM	17	ReLU	1292
LSTM	17	ReLU	2380
Dense	1	Linear	18
Trainable params.	3690		
Epochs	1200		

Hist-Dense

Layer	Output shape	Activation	# Parameters
Histogram	700	—	0
Dense	100	ReLU	70 100
Dense	50	ReLU	5050
Dense	30	ReLU	1530
Dense	1	Linear	31
Trainable params.	76 711		
Epochs	1200		

Hist-Dense 2D

Layer	Output shape	Activation	# Parameters
Histogram	700	—	0
Dense	100	ReLU	70 100
Dense	50	ReLU	5050
Dense	30	ReLU	1530
Dense	20	ReLU	620
Dense	10	ReLU	210
Dense	2	Linear	22
Trainable params.	77 532		
Epochs	1200		

Appendix E. Validation process

E.1. 1D case

In order to validate the performance of the NNs, we generate a set of validation trajectories for a uniform grid of 40 values of Δ (with $\Omega = \gamma$) in the range $\Delta \in [0, 2.1\gamma]$. For each value of Δ, we generated 10^4 trajectories containing 48 photodetection events each. For each trajectory, we obtained an estimation of the parameter for each of the methods, taking the mean of the posterior as the estimation in the Bayesian case. We then computed the MSE from the 10^4 different predictions with each of the methods for each value of Δ.

E.2. 2D case

Validation is done for a grid of (Δ, Ω) pairs built from a grid of 40 values of Δ in the range $\Delta \in [0., 2.1\gamma]$ and a grid of 40 values of Ω in the range $\Omega \in [0.25\gamma, 2.1\gamma]$, spanning a square grid of 1600 points. For each of these points, we simulate 10^4 validation trajectories with 48 clicks each.

ORCID iDs

Enrico Rinaldi ⓘ https://orcid.org/0000-0003-4134-809X
Manuel González Lastre ⓘ https://orcid.org/0009-0000-2299-6687
Sergio García Herreros ⓘ https://orcid.org/0009-0006-1420-4604
Shahnawaz Ahmed ⓘ https://orcid.org/0000-0003-1145-7279
Maryam Khanahmadi ⓘ https://orcid.org/0000-0002-3523-6449
Franco Nori ⓘ https://orcid.org/0000-0003-3682-7432
Carlos Sánchez Muñoz ⓘ https://orcid.org/0000-0001-8775-0135

References

[1] Wiseman H M and Milburn G J 2010 *Quantum Measurement and Control* (Cambridge University Press)

[2] Genois E, Gross J A, Di Paolo A, Stevenson N J, Koolstra G, Hashim A, Siddiqi I and Blais A 2021 Quantum-tailored machine-learning characterization of a superconducting qubit *PRX Quantum* **2** 040355

[3] Carrasco J, Elben A, Kokail C, Kraus B and Zoller P 2021 Theoretical and experimental perspectives of quantum verification *PRX Quantum* **2** 010102

[4] Youssry A, Yang Y, Chapman R J, Haylock B, Lenzini F, Lobino M and Peruzzo A 2024 Experimental graybox quantum system identification and control *npj Quantum Inf.* **10** 1

[5] Zhang J, Liu Y-x, Wu R-B, Jacobs K and Nori F 2017 Quantum feedback: theory, experiments and applications *Phys. Rep.* **679** 1

[6] Wiseman H M and Milburn G J 1993 Interpretation of quantum jump and diffusion processes illustrated on the Bloch sphere *Phys. Rev.* A **47** 1652

[7] Ashhab S and Nori F 2010 Control-free control: manipulating a quantum system using only a limited set of measurements *Phys. Rev.* A **82** 062103

[8] Sayrin C *et al* 2011 Real-time quantum feedback prepares and stabilizes photon number states *Nature* **477** 73

[9] Yonezawa H *et al* 2012 Quantum-enhanced optical-phase tracking *Science* **337** 1514

[10] Cui J-M, Sun F-W, Chen X-D, Gong Z-J and Guo G-C 2013 Quantum statistical imaging of particles without restriction of the diffraction limit *Phys. Rev. Lett.* **110** 153901

[11] Leverrier A 2015 Composable security proof for continuous-variable quantum key distribution with coherent states *Phys. Rev. Lett.* **114** 070501

[12] Liang K, Chai G, Cao Z, Yuan Y, Chen X, Lu Y and Peng J 2022 Bayesian parameter estimation for continuous-variable quantum key distribution *Phys. Rev. Appl.* **18** 054077

[13] Giovannetti V, Lloyd S and Maccone L 2006 Quantum metrology *Phys. Rev. Lett.* **96** 010401

[14] Giovannetti V, Lloyd S and Maccone L 2011 Advances in quantum metrology *Nat. Photon.* **5** 222

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

[16] Pirandola S, Bardhan B R, Gehring T, Weedbrook C and Lloyd S 2018 Advances in photonic quantum sensing *Nat. Photon.* **12** 724

[17] Pezzè L, Smerzi A, Oberthaler M K, Schmied R and Treutlein P 2018 Quantum metrology with non classical states of atomic ensembles *Rev. Mod. Phys.* **90** 035005

[18] Ma J, Wang X, Sun C P and Nori F 2011 Quantum spin squeezing *Phys. Rep.* **509** 89

[19] Liu Z-P, Zhang J, Özdemir Ş K, Peng B, Jing H, Lü X-Y, Li C-W, Yang L, Nori F and Liu Y-X 2016 Metrology with PT-symmetric cavities: enhanced sensitivity near the PT-phase transition *Phys. Rev. Lett.* **117** 110802

[20] Xu K *et al* 2022 Metrological characterization of non-gaussian entangled states of superconducting qubits *Phys. Rev. Lett.* **128** 150501

[21] Lee K-Y, Lin J-D, Miranowicz A, Nori F, Ku H-Y and Chen Y-N 2023 Steering-enhanced quantum metrology using superpositions of noisy phase shifts *Phys. Rev. Res.* **5** 013103

[22] Mabuchi H 1996 Dynamical identification of open quantum systems *Quantum Semiclass. Opt.* **8** 1103

[23] Gambetta J and Wiseman H M 2001 State and dynamical parameter estimation for open quantum systems *Phys. Rev.* A **64** 14

[24] Verstraete F, Doherty A C and Mabuchi H 2001 Sensitivity optimization in quantum parameter estimation *Phys. Rev.* A **64** 032111

[25] Chase B A and Geremia J M 2009 Single-shot parameter estimation via continuous quantum measurement *Phys. Rev.* A **79** 022314

[26] Guţă M 2011 Fisher information and asymptotic normality in system identification for quantum Markov chains *Phys. Rev.* A **83** 062324

[27] Ralph J F, Jacobs K and Hill C D 2011 Frequency tracking and parameter estimation for robust quantum state estimation *Phys. Rev.* A **84** 052119

[28] Gammelmark S and Mølmer K 2014 Fisher information and the quantum Cramér-Rao sensitivity limit of continuous measurements *Phys. Rev. Lett.* **112** 170401

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

[30] Kiilerich A H and Mølmer K 2014 Estimation of atomic interaction parameters by photon counting *Phys. Rev.* A **89** 052110

[31] Kiilerich A H and Mølmer K 2015 Parameter estimation by multichannel photon counting *Phys. Rev.* A **91** 12119

[32] Kiilerich A H and Mølmer K 2016 Bayesian parameter estimation by continuous homodyne detection *Phys. Rev.* A **94** 032103

[33] Cortez L, Chantasri A, García-Pintos L P, Dressel J and Jordan A N 2017 Rapid estimation of drifting parameters in continuously measured quantum systems *Phys. Rev.* A **95** 012314

[34] Ralph J F, Maskell S and Jacobs K 2017 Multiparameter estimation along quantum trajectories with sequential Monte Carlo methods *Phys. Rev.* A **96** 052306

[35] Albarelli F, Rossi M A C, Tamascelli D and Genoni M G 2018 Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment *Quantum* **2** 110

[36] Kominis I K, Kornack T W, Allred J C and Romalis M V 2003 A subfemtotesla multichannel atomic magnetometer *Nature* **422** 596

[37] Geremia J M, Stockton J K, Doherty A C and Mabuchi H 2003 Quantum Kalman filtering and the Heisenberg limit in atomic magnetometry *Phys. Rev. Lett.* **91** 250801

[38] Amorós-Binefa J and Kołodyński J 2021 Noisy atomic magnetometry in real time *New J. Phys.* **23** 123030

[39] Macieszczak K, Guţă M, Lesanovsky I and Garrahan J P 2016 Dynamical phase transitions as a resource for quantum enhanced metrology *Phys. Rev.* A **93** 022103

[40] Ilias T, Yang D, Huelga S F and Plenio M B 2022 Criticality-enhanced quantum sensing via continuous measurement *PRX Quantum* **3** 010354

[41] Yang D, Huelga S F and Plenio M B 2023 Efficient information retrieval for sensing via continuous measurement *Phys. Rev.* X **13** 031012

[42] Tsang M, Wiseman H M and Caves C M 2011 Fundamental quantum limit to waveform estimation *Phys. Rev. Lett.* **106** 090401

[43] Clark L A, Markowicz B and Kołodyński J 2022 Exploiting non-linear effects in optomechanical sensors with continuous photon-counting *Quantum* **6** 812

[44] Smith G A, Silberfarb A, Deutsch I H and Jessen P S 2006 Efficient quantum-state estimation by continuous weak measurement and dynamical control *Phys. Rev. Lett.* **97** 180403

[45] Ashhab S, You J Q and Nori F 2009 The information about the state of a qubit gained by a weakly coupled detector *New J. Phys.* **11** 083017

[46] Gong B, Yang Y and Cui W 2018 Quantum parameter estimation via dispersive measurement in circuit QED *Quantum Inf. Process.* **17** 301

[47] Wiseman H M and Milburn G J 1993 Quantum theory of field-quadrature measurements *Phys. Rev.* A **47** 642

[48] Chen Y, Ban Y, He R, Cui J-M, Huang Y-F, Li C-F, Guo G-C and Casanova J 2022 A neural network assisted ^{171}Yb$^+$ quantum magnetometer *npj Quantum Inf.* **8** 1

[49] Verhelst M and Murmann B 2020 Machine learning at the edge *Nano-Chips 2030: On-Chip AI for an Efficient Data-Driven World* (*The Frontiers Collection*) ed B Murmann and B Hoefflinger (Springer International Publishing) pp 293–322

[50] Homrighausen J, Horsthemke L, Pogorzelski J, Trinschek S, Glösekötter P and Gregor M 2023 Edge-machine-learning-assisted robust magnetometer based on randomly oriented NV-Ensembles in diamond *Sensors* **23** 1119

[51] Liu W, Huang J, Li Y, Li H, Fang C, Yu Y and Zeng G 2019 Parameter estimation via weak measurement with machine learning *J. Phys. B: At. Mol. Opt. Phys.* **52** 045504

[52] Ban Y, Echanobe J, Ding Y, Puebla R and Casanova J 2021 Neural-network-based parameter estimation for quantum detection *Quantum Sci. Technol.* **6** 045012

[53] Nolan S P, Pezzè L and Smerzi A 2021 Frequentist parameter estimation with supervised learning *AVS Quantum Sci.* **3** 034401

[54] Xiao T, Fan J and Zeng G 2022 Parameter estimation in quantum sensing based on deep reinforcement learning *npj Quantum Inf.* **8** 1

[55] Ban Y, Casanova J and Puebla R 2022 Neural networks forBayesian quantum many-body magnetometry (arXiv:2212.12058 [quant-ph])

[56] Cimini V, Gianani I, Spagnolo N, Leccese F, Sciarrino F and Barbieri M 2019 Calibration of quantum sensors by neural networks *Phys. Rev. Lett.* **123** 230502

[57] Cimini V, Polino E, Valeri M, Gianani I, Spagnolo N, Corrielli G, Crespi A, Osellame R, Barbieri M and Sciarrino F 2021 Calibration of multiparameter sensors via machine learning at the single-photon level *Phys. Rev. Appl.* **15** 044003

[58] Cimini V, Valeri M, Polino E, Piacentini S, Ceccarelli F, Corrielli G, Spagnolo N, Osellame R and Sciarrino F 2023 Deep reinforcement learning for quantum multiparameter estimation *Adv. Photon.* **5** 016005

[59] Khanahmadi M and Mølmer K 2021 Time-dependent atomic magnetometry with a recurrent neural network *Phys. Rev.* A **103** 032406

[60] Ahmed S, Sánchez Muñoz C, Nori F and Kockum A F 2021 Quantum state tomography with conditional generative adversarial networks *Phys. Rev. Lett.* **127** 140502

[61] Ahmed S, Sánchez Muñoz C, Nori F and Kockum A F 2021 Classification and reconstruction of optical quantum states with deep neural networks *Phys. Rev. Res.* **3** 033278

[62] Flurin E, Martin L S, Hacohen-Gourgy S and Siddiqi I 2020 Using a recurrent neural network toreconstruct quantum dynamics of a superconducting qubit from physical observations *Phys. Rev.* X **10** 011006

[63] Greplova E, Andersen C K and Mølmer K 2017 Quantum parameter estimation with a neural network (arXiv:1711.05238)

[64] Bartolo N, Minganti F, Lolli J and Ciuti C 2017 Homodyne versus photon-counting quantum trajectories for dissipative Kerr resonators with two-photon driving *Eur. Phys. J. Spec. Top.* **226** 2705

[65] Zubizarreta Casalengua E, López Carreño J C, Laussy F P and del Valle E 2020 Conventional and unconventional photon statistics *Laser Photon. Rev.* **14** 1900279

[66] De Santis L *et al* 2017 A solid-state single-photon filter *Nat. Nanotechnol.* **12** 663

[67] Silva B *et al* 2016 The colored Hanbury Brown-Twiss effect *Sci. Rep.* **6** 37980

[68] Rundquist A, Bajcsy M, Majumdar A, Sarmiento T, Fischer K, Lagoudakis K G, Buckley S, Piggott A Y and Vučković J 2014 Nonclassical higher-order photon correlations with a quantum dot strongly coupled to a photonic-crystal nanocavity *Phys. Rev.* A **90** 023846

[69] Israel Y, Tenne R, Oron D and Silberberg Y 2017 Quantum correlation enhanced super-resolution localization microscopy enabled by a fibre bundle camera *Nat. Commun.* **8** 1

[70] Lambert N, Chen Y-N and Nori F 2010 Unified single-photon and single-electron counting statistics: from cavity QED to electron transport *Phys. Rev.* A **82** 063840

[71] Hadfield R H 2009 Single-photon detectors for optical quantum information applications *Nat. Photon.* **3** 696

[72] Kimble H J, Dagenais M and Mandel L 1977 Photon antibunching in resonance fluorescence *Phys. Rev. Lett.* **39** 691

[73] Lodahl P, Mahmoodian S and Stobbe S 2015 Interfacing single photons and single quantum dots with photonic nanostructures *Rev. Mod. Phys.* **87** 347

[74] Toninelli C *et al* 2021 Single organic molecules for photonic quantum technologies *Nat. Mater.* **20** 1615

[75] Sipahigil A *et al* 2016 An integrated diamond nanophotonics platform for quantum-optical networks *Science* **354** 847

[76] Breuer H-P and Petruccione F 2002 *The Theory of Open Quantum Systems* (Oxford University Press)

[77] Plenio M B and Knight P L 1998 The quantum-jump approach to dissipative dynamics in quantum optics *Rev. Mod. Phys.* **70** 101

[78] Johansson J R, Nation P D and Nori F 2012 QuTiP: an open-source Python framework for the dynamics of open quantum systems *Comput. Phys. Commun.* **183** 1760

[79] Johansson J R, Nation P D and Nori F 2013 QuTiP 2: a Python framework for the dynamics of open quantum systems *Comput. Phys. Commun.* **184** 1234

[80] Gneiting C, Rozhkov A V and Nori F 2021 Jump-time unraveling of Markovian open quantum systems *Phys. Rev.* A **104** 062212

[81] Gneiting C, Koottandavida A, Rozhkov A V and Nori F 2022 Unraveling the topology of dissipative quantum systems *Phys. Rev. Res.* **4** 023036

[82] See Supplemental Material for further details on the theory of quantum jump trajectories, the choice of different estimators from posterior distributions, the Fisher information and the Cramér-Rao bound, and the deployment of models in TensorFlow Lite.

[83] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge University Press)

[84] Paris M G 2009 Quantum estimation for quantum technology *Int. J. Quantum Inf.* **7** 125

[85] Cramér H 1999 *Mathematical Methods of Statistics* (Princeton University Press)

[86] Matson C L and Haji A 2006 BiasedCramér-Rao lower bound calculations for inequality-constrained estimators *J. Opt. Soc. Am.* A **23** 2702

[87] Buchner J 2021 UltraNest - a robust, general purpose Bayesian inference engine *J. Open Sour. Softw.* **6** 3001

[88] López Carreño J C, Zubizarreta Casalengua E, Silva B, Del Valle E and Laussy F P 2022 Loss of antibunching *Phys. Rev.* A **105** 023724

[89] Warszawski P, Wiseman H M and Mabuchi H 2002 Quantum trajectories for realistic detection *Phys. Rev.* A **65** 023802

[90] Sánchez Muñoz C, del Valle E, Tudela A G, Müller K, Lichtmannecker S, Kaniber M, Tejedor C, Finley J J and Laussy F P 2014 Emitters of N-photon bundles *Nat. Photon.* **8** 550

[91] Hamsen C, Tolazzi K N, Wilk T and Rempe G 2017 Two-photon blockade in an atom-driven cavity QED system *Phys. Rev. Lett.* **118** 133604

[92] Tenne R, Rossman U, Rephael B, Israel Y, Krupinski-Ptaszek A, Lapkiewicz R, Silberberg Y and Oron D 2019 Super-resolution enhancement by quantum image scanning microscopy *Nat. Photon.* **13** 116

[93] Khosravi A R, Nahavandi S, Creighton D and Atiya A F 2011 Comprehensive review of neural network-based prediction intervals and new advances *IEEE Trans. Neural Netw.* **22** 1341

[94] Wang H and Yeung D 2020 A survey on Bayesian deep learning *ACM Comput. Surv.* **53** 1

[95] Delaubert V, Treps N, Fabre C, Bachor H A and Réfrégier P 2008 Quantum limits in image processing *Europhys. Lett.* **81** 44001

[96] Speirits F C, Sonnleitner M and Barnett S M 2017 From retrodiction to Bayesian quantum imaging *J. Opt.* **19** 044001

[97] He Z, Zhang Y, Tong X, Li L and Wang L V 2023 Quantum microscopy of cells at the Heisenberg limit *Nat. Commun.* **14** 2441

[98] Albarelli F, Barbieri M, Genoni M G and Gianani I 2020 A perspective on multiparameter quantum metrology: from theoretical tools to applications in quantum imaging *Phys. Lett.* A **384** 126311

[99] Wolley O, Gregory T, Beer S, Higuchi T, Padgett M and Photonics H K 2022 Quantum imaging with a photon counting camera *Sci. Rep.* **12** 1

[100] Altmann Y, McLaughlin S, Padgett M J, Goyal V K, Hero A O and Faccio D 2018 Quantum-inspired computational imaging *Science* **361** eaat2298

[101] Sánchez Muñoz C and Rinaldi E 2023 Dataset: parameter estimation by learning quantum correlations in continuous photon-counting data using neural networks (available at: https://zenodo.org/records/8305509)

[102] Sánchez Muñoz C and Rinaldi E 2023 Code: parameter estimation by learning quantum correlations in continuous photon-counting data using neural networks (available at: https://zenodo.org/records/8392691)

[103] Speagle J S 2020 A conceptual introduction to Markov Chain Monte Carlo methods (arXiv:1909.12313 [astro-ph, physics:physics, stat])

[104] Skilling J 2004 Nested sampling *AIP Conf. Proc.* **735** 395–405

[105] Buchner J 2023 Nested sampling methods *Stat. Surv.* **17** 169

[106] Abadi M *et al* TensorFlow: a system for large-scale machine learning (arXiv:1605.08695 [cs])