Deep Learning of Quantum Correlations for Quantum Parameter Estimation of Continuously Monitored Systems

Enrico Rinaldi

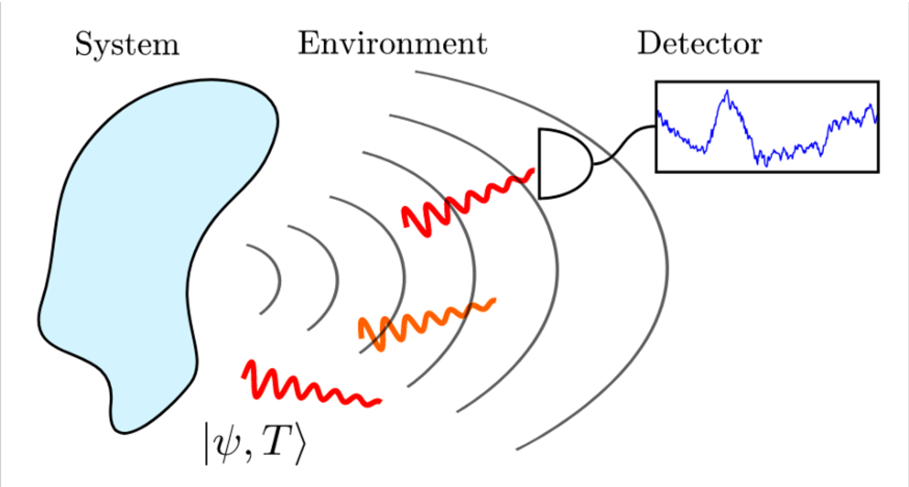
(Main) Quantinuum K.K. ⊗ (Visitor) RIKEN [iTHEMS + RQC]

2023 November 22, Quantum Techniques in Machine Learning

Gammelmark, S. & Mølmer, K. Fisher Information and the Quantum Cramér-Rao Sensitivity Limit of Continuous Measurements. Physical Review Letters 112, 170401 (2014)

Parameter estimation in open quantum systems

Estimate the unknown parameters that govern the dynamics of a sensor that is coupled to an environment and it is continuously monitored

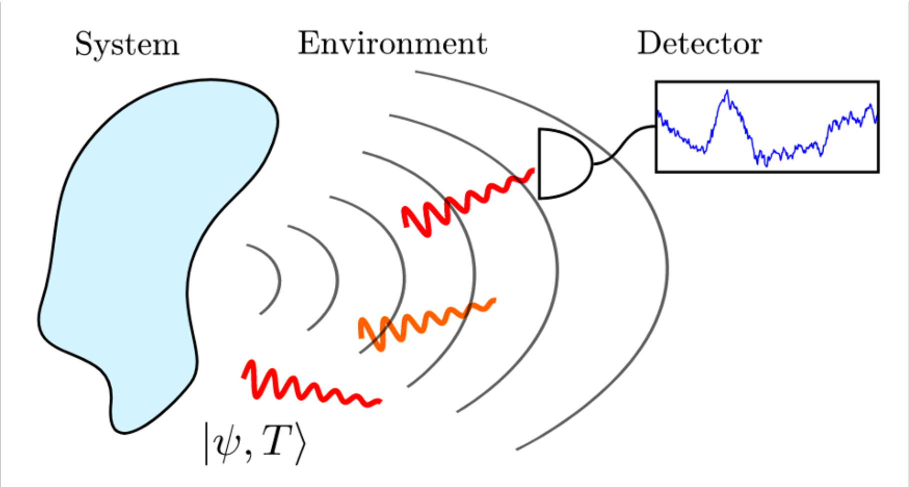


Gammelmark, S. & Mølmer, K. Fisher Information and the Quantum Cramér-Rao Sensitivity Limit of Continuous Measurements. Physical Review Letters 112, 170401 (2014)

Parameter estimation in open quantum systems

Estimate the unknown parameters that govern the dynamics of a sensor that is coupled to an environment and it is continuously monitored

- Metrology with non isolated systems
- Metrology of time-varying signals
- Magnetometry
- Spectroscopy
- Fluorescence microscopy
- Device characterization
- …

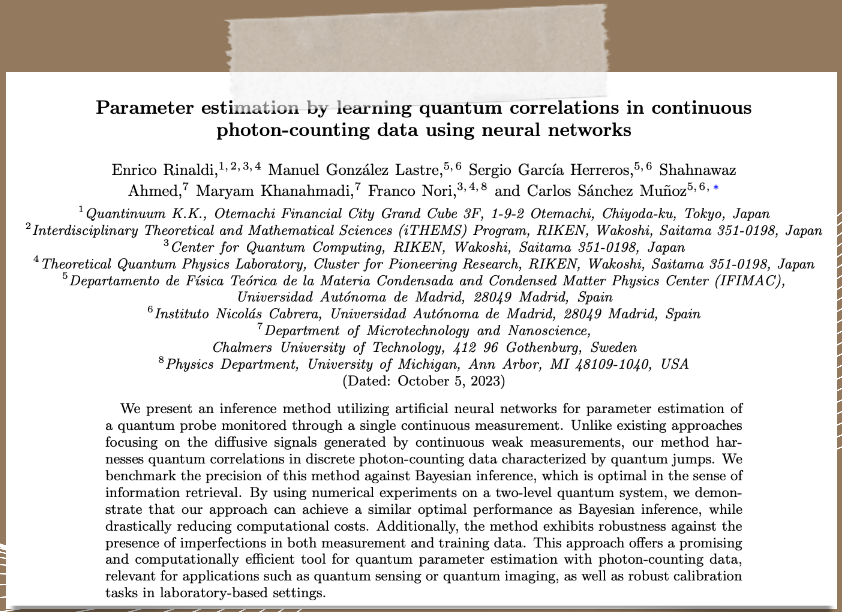


- Can we leverage modern deep learning techniques to **improve parameter estimation**?

- Can we leverage modern deep learning techniques to **improve parameter estimation**?

Yes! See our preprint https://arxiv.org/abs/2310.02309

Project lead by Carlos Sànchez Muñoz at UAM



Parameter estimation by learning quantum correlations in continuous photon-counting data using neural networks

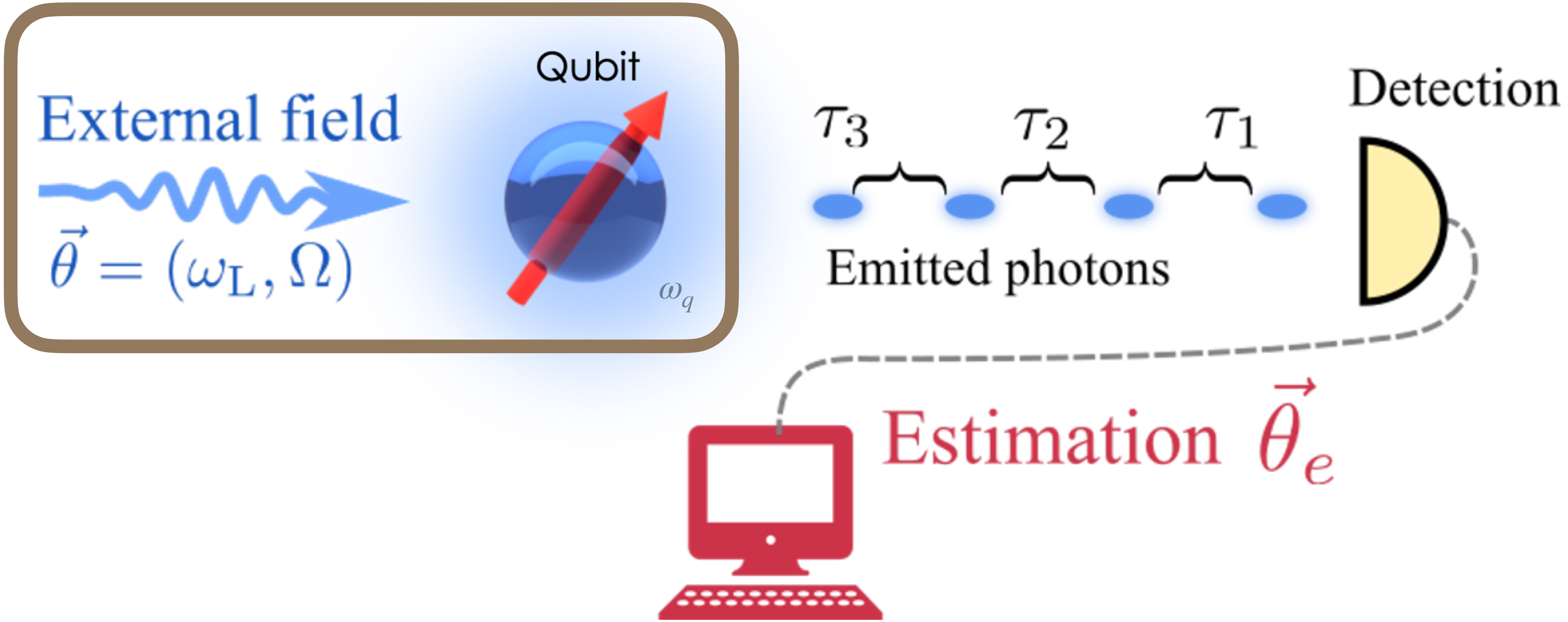
Enrico Rinaldi,[1,2,3,4] Manuel González Lastre,[5,6] Sergio García Herreros,[5,6] Shahnawaz Ahmed,[7] Maryam Khanahmadi,[7] Franco Nori,[3,4,8] and Carlos Sánchez Muñoz[5,6,*]

[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, USA*

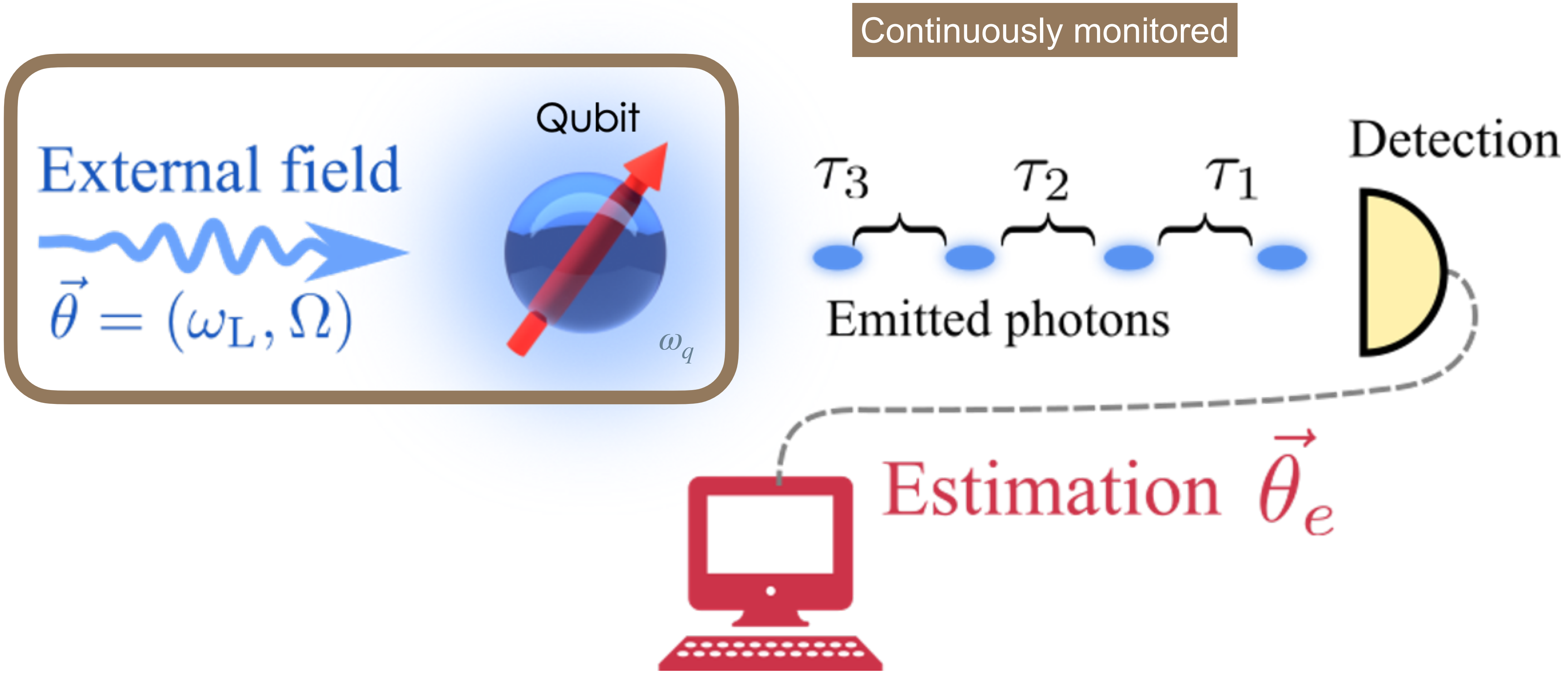
(Dated: October 5, 2023)

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.

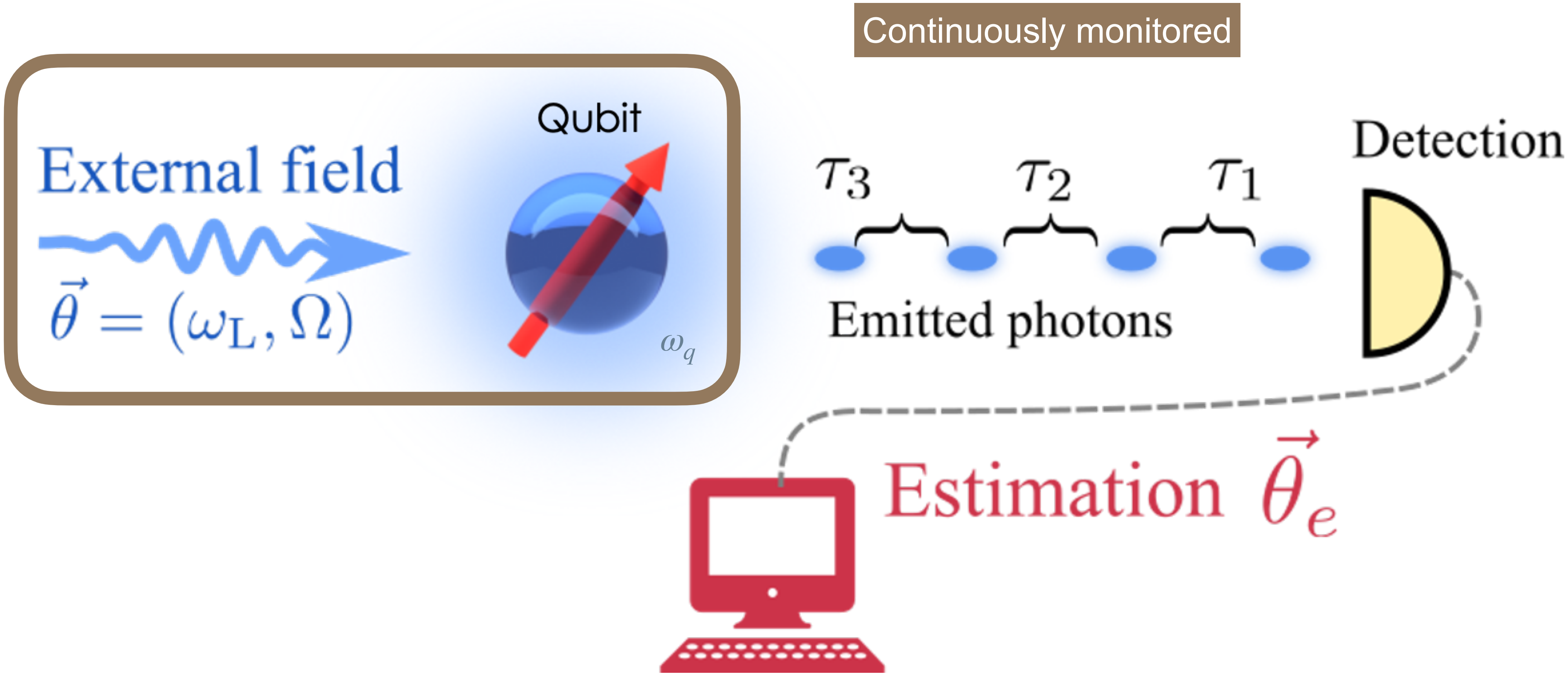
A simple model of an open quantum sensor



A simple model of an open quantum sensor

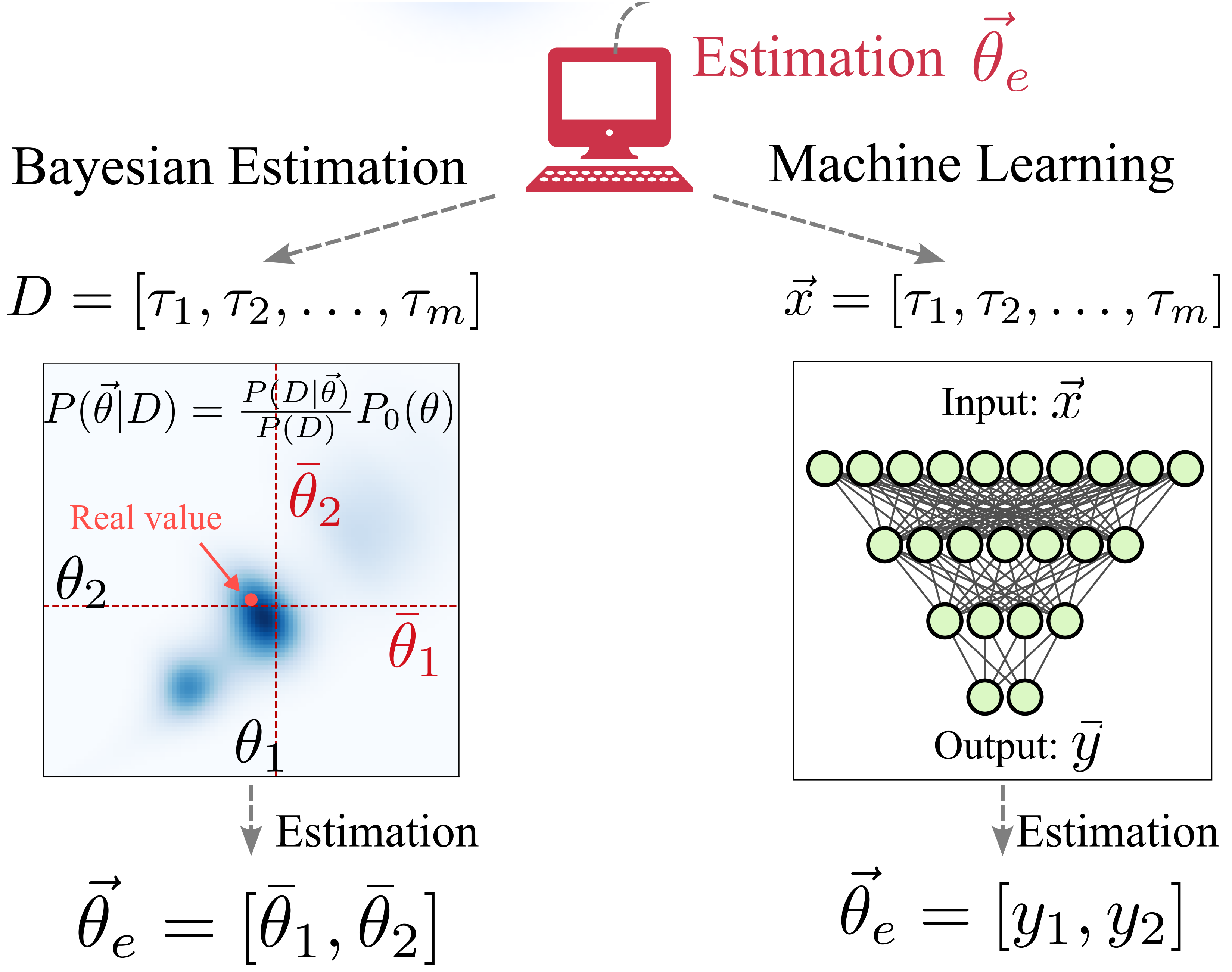


A simple model of an open quantum sensor



Can we precisely and robustly extract the value of the system's parameters?

$\vec{\theta}_e = \{\Delta = \omega_q - \omega_{\mathrm{L}}$ and $\Omega\}$



Estimation $\vec{\theta}_e$
Bayesian Estimation
Machine Learning
$D = [\tau_1, \tau_2, \ldots, \tau_m]$
$\vec{x} = [\tau_1, \tau_2, \ldots, \tau_m]$
$P(\vec{\theta}|D) = \frac{P(D|\vec{\theta})}{P(D)} P_0(\theta)$
$\bar{\theta}_2$
Real value
θ_2
$\bar{\theta}_1$
θ_1
Input: $\vec{x}$
Output: $\vec{y}$
Estimation
Estimation
$\vec{\theta}_e = [\bar{\theta}_1, \bar{\theta}_2]$
$\vec{\theta}_e = [y_1, y_2]$

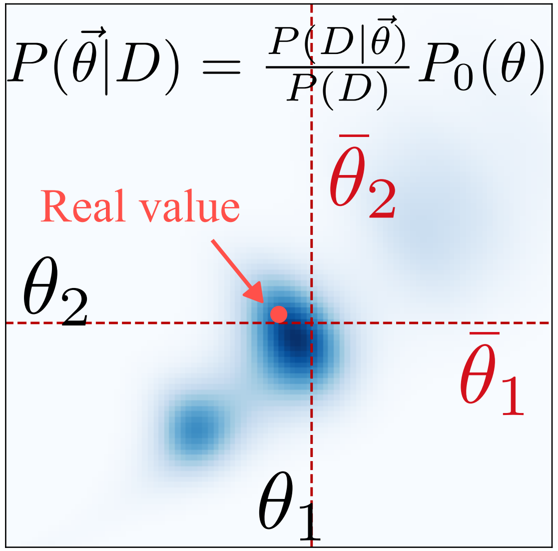
Bayesian parameter estimation

Use Bayes' Theorem to compute the *probability density function* of the parameters, conditioned on the observed data

$$P(\vec{\theta} \mid D) = \frac{P(D \mid \vec{\theta})}{P(D)} P_0(\vec{\theta}) = \frac{P(D \mid \vec{\theta})}{\int d\vec{\theta} P(D \mid \vec{\theta}) P_0(\vec{\theta})} P_0(\vec{\theta})$$

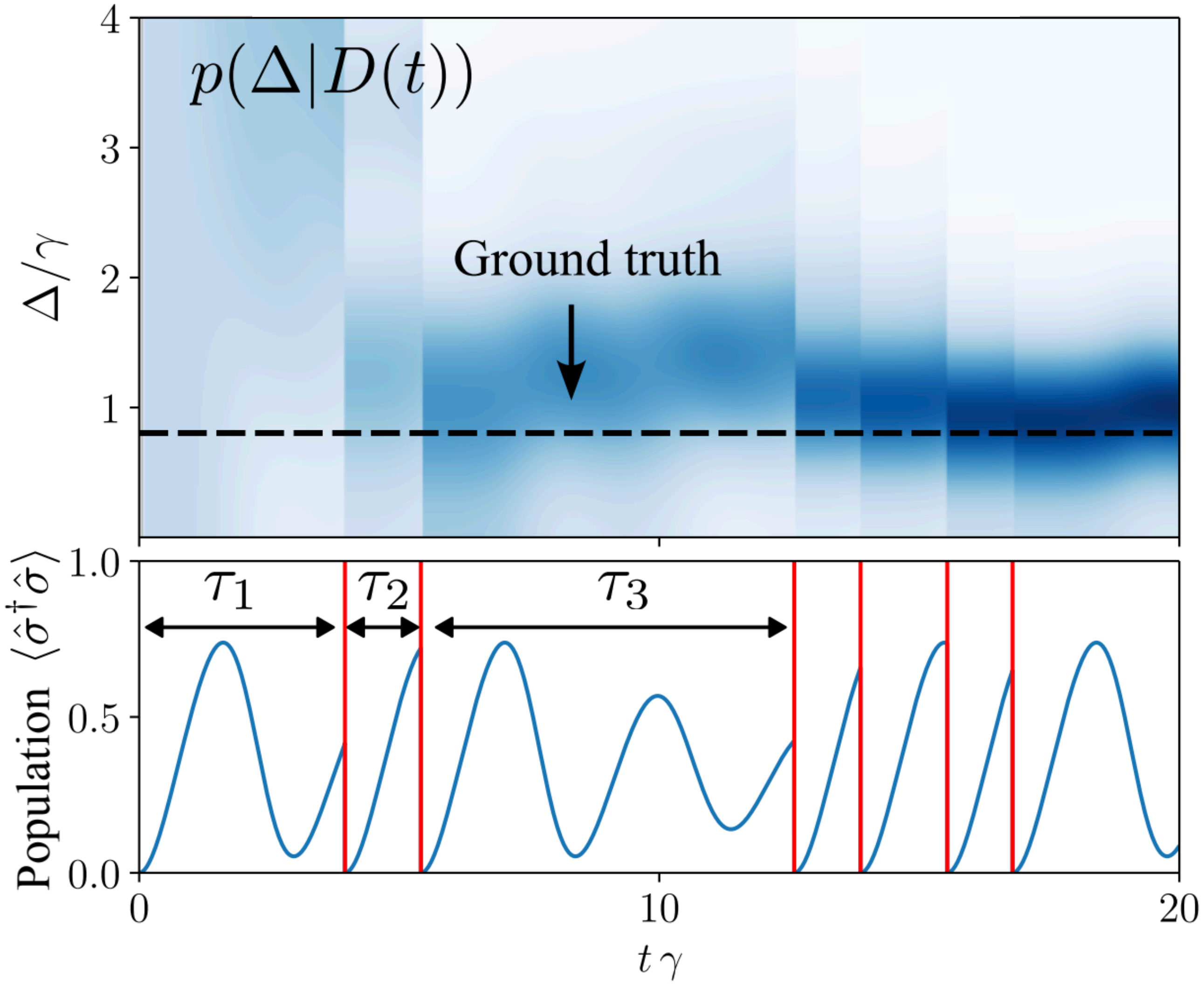
Bayesian Estimation

$$D = [\tau_1, \tau_2, \ldots, \tau_m]$$



Estimation

$$\vec{\theta}_e = [\bar{\theta}_1, \bar{\theta}_2]$$



Posterior

Data

Example for 1 parameter

- The posterior changes as different data D(t) is observed
- The longer time t we observe, the more precise our prediction

Challenges in Bayesian parameter estimation

Challenges in Bayesian parameter estimation

- One of the ingredients of Bayesian estimation is the Likelihood:
  - This requires a full model of the quantum system
  - A model of the quantum system, when available, might require solving complicated differential equations or finding eigenfunctions of large matrices.
  - In many cases these likelihoods can become extremely small (for unlikely events) and numerical precision becomes an issue

Challenges in Bayesian parameter estimation

- One of the ingredients of Bayesian estimation is the Likelihood:
  - This requires a full model of the quantum sy
  - A model of the quantum system, when avail … complicated differential equations or finding … ces.
  - In many cases these likelihoods can becom … events) and numerical precision becomes an issue

Not practical

Challenges in Bayesian parameter estimation

- One of the ingredients of Bayesian estimation is the Likelihood:
  - This requires a full model of the quantum sy
  - A model of the quantum system, when avail … complicated differential equations or finding … ces.
  - In many cases these likelihoods can becom … events) and numerical precision becomes an issue

Not practical

- Another ingredient is the Evidence:
  - The evidence (or marginal likelihood) requires an integral over all possible parameter values
  - When the parameter space is high dimensional, computing this integral is numerically expensive
  - Monte Carlo methods are routinely employed in this case

Challenges in Bayesian parameter estimation

- One of the ingredients of Bayesian estimation is the Likelihood:
  - This requires a full model of the quantum system
  - A model of the quantum system, when avail… complicated differential equations or finding …ces.
  - In many cases these likelihoods can becom… events) and numerical precision becomes an issue

Not practical

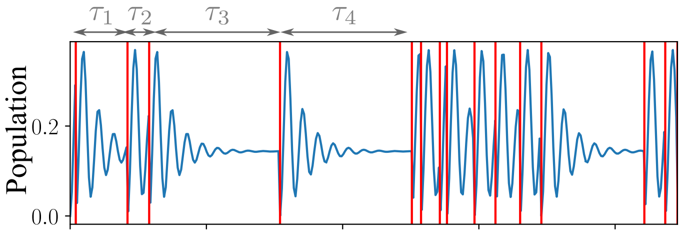
- Another ingredient is the Evidence:
  - The ev…lihood) requires an integral over all possible param…
  - When … high dimensional, computing this integral is numer…
  - Monte Carlo methods are routinely employed in this case

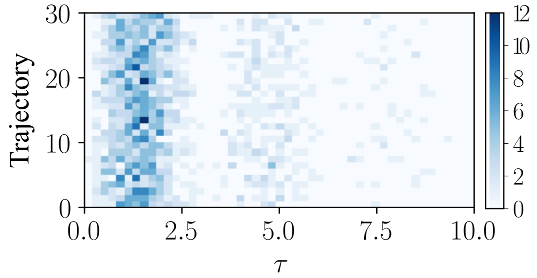
Expensive

Deep learning parameter estimation

Use trained neural networks to output a *point* estimate of the parameters, given an observation as input

Two types of data





Sequence

input = $[\tau_1, \ldots \tau_N]$

RNN:
LSTM layers

output = $[\theta_1, \ldots \theta_n]$

Histogram Counts

input = $[c_1, \ldots c_{400}]$

FCNN:
Dense layers

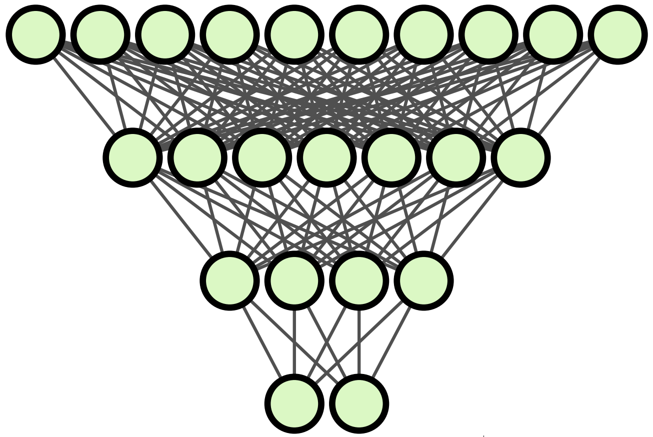
output = $[\theta_1, \ldots \theta_n]$

TensorFlow
Implementation
And
Deployment

Machine Learning

$$\vec{x} = [\tau_1, \tau_2, \ldots, \tau_m]$$

Input: $\vec{x}$



Output: $\vec{y}$

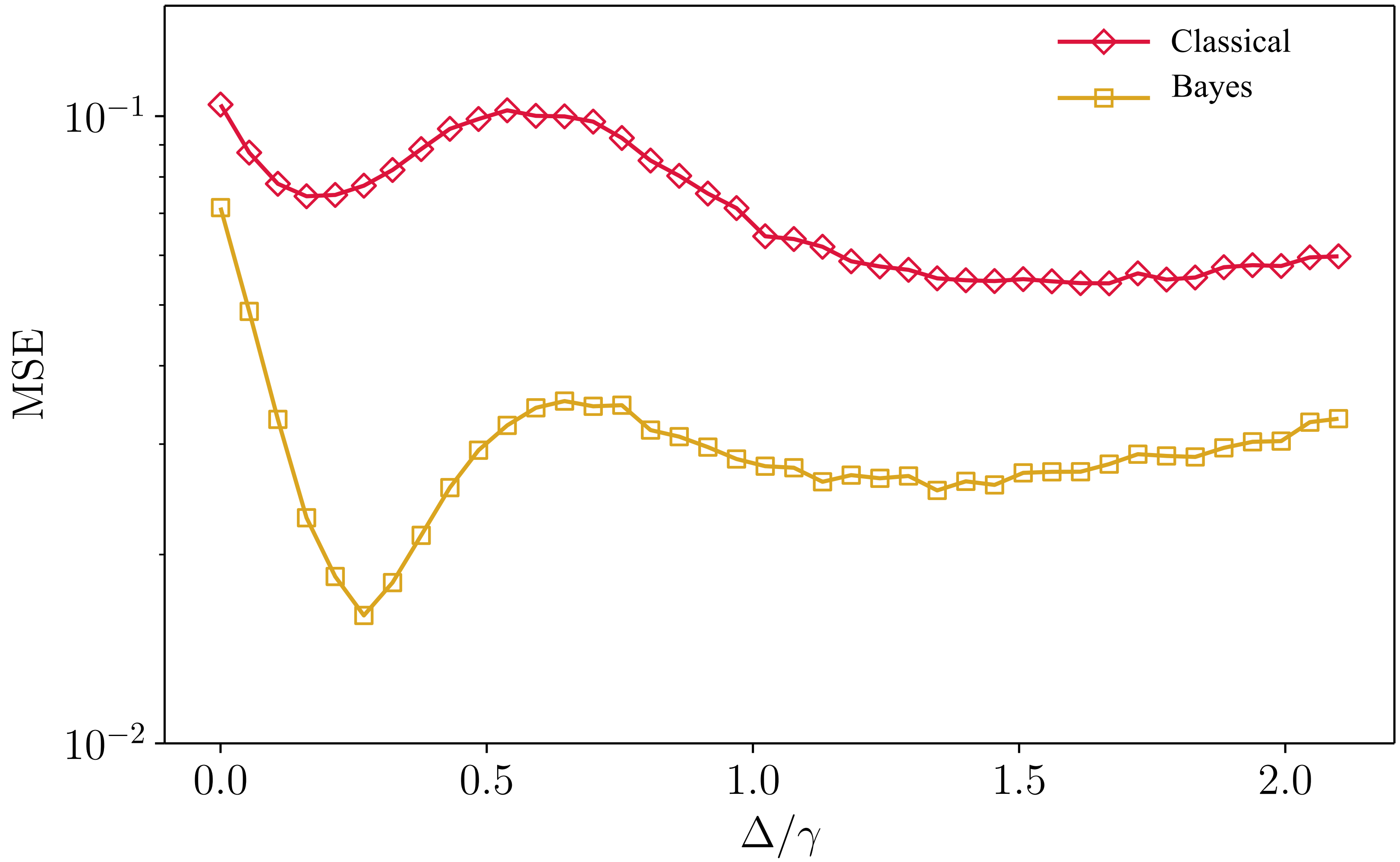
Estimation

$$\vec{\theta}_e = [y_1, y_2]$$

Test with 1 parameter

$$\text{MSE} = \frac{1}{N_{\text{traj}}} \sum^{N_{\text{traj}}} (\Delta_e - \Delta)^2$$

Measure how well the “prediction” compares to the “truth” (averaged over many samples)



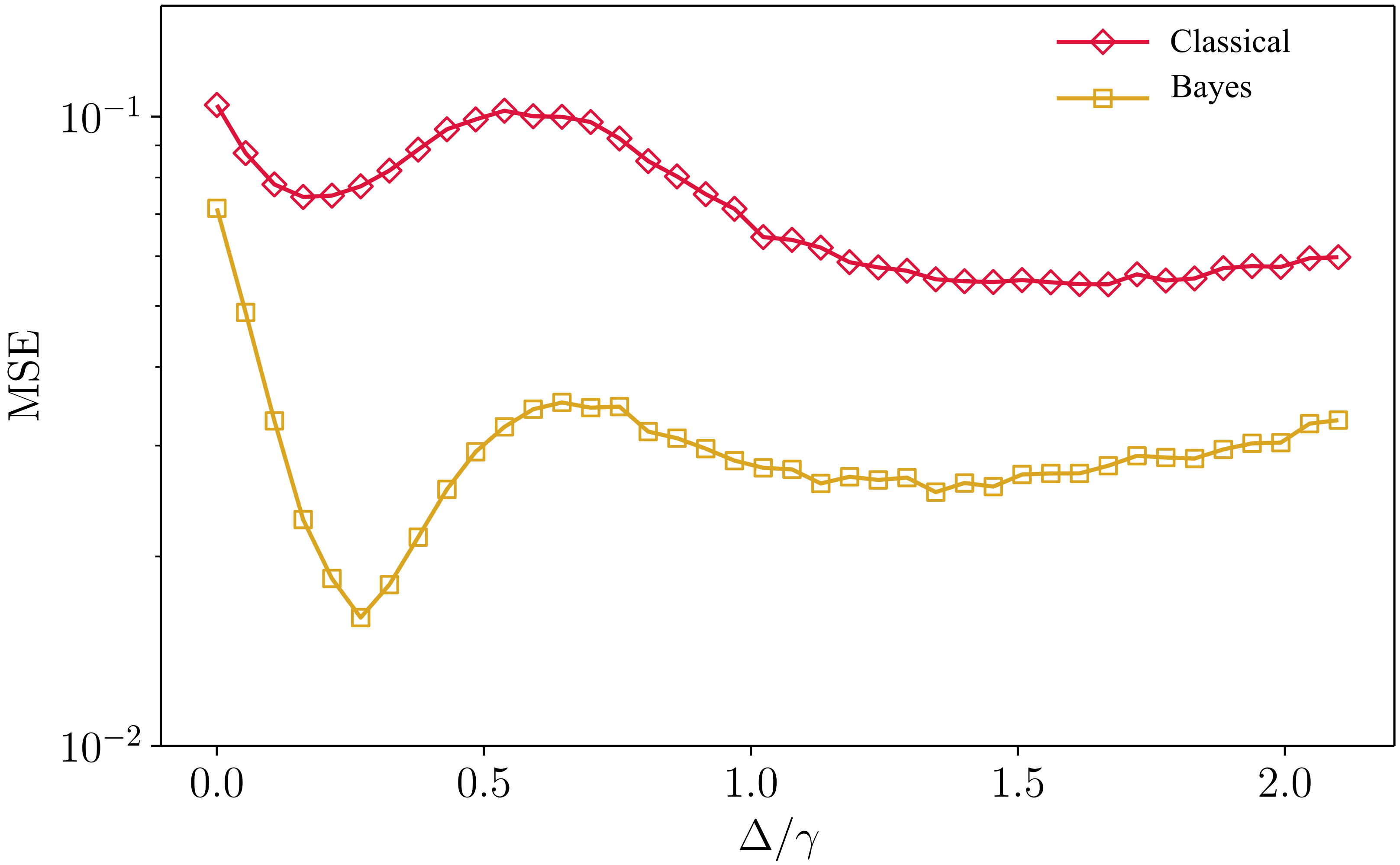
True value

Test with 1 parameter

$$\mathrm{MSE} = \frac{1}{N_{\mathrm{traj}}} \sum^{N_{\mathrm{traj}}} (\Delta_e - \Delta)^2$$

Measure how well the “prediction” compares to the “truth” (averaged over many samples)

- Using data without quantum correlations we obtain a larger deviation from the true parameter value



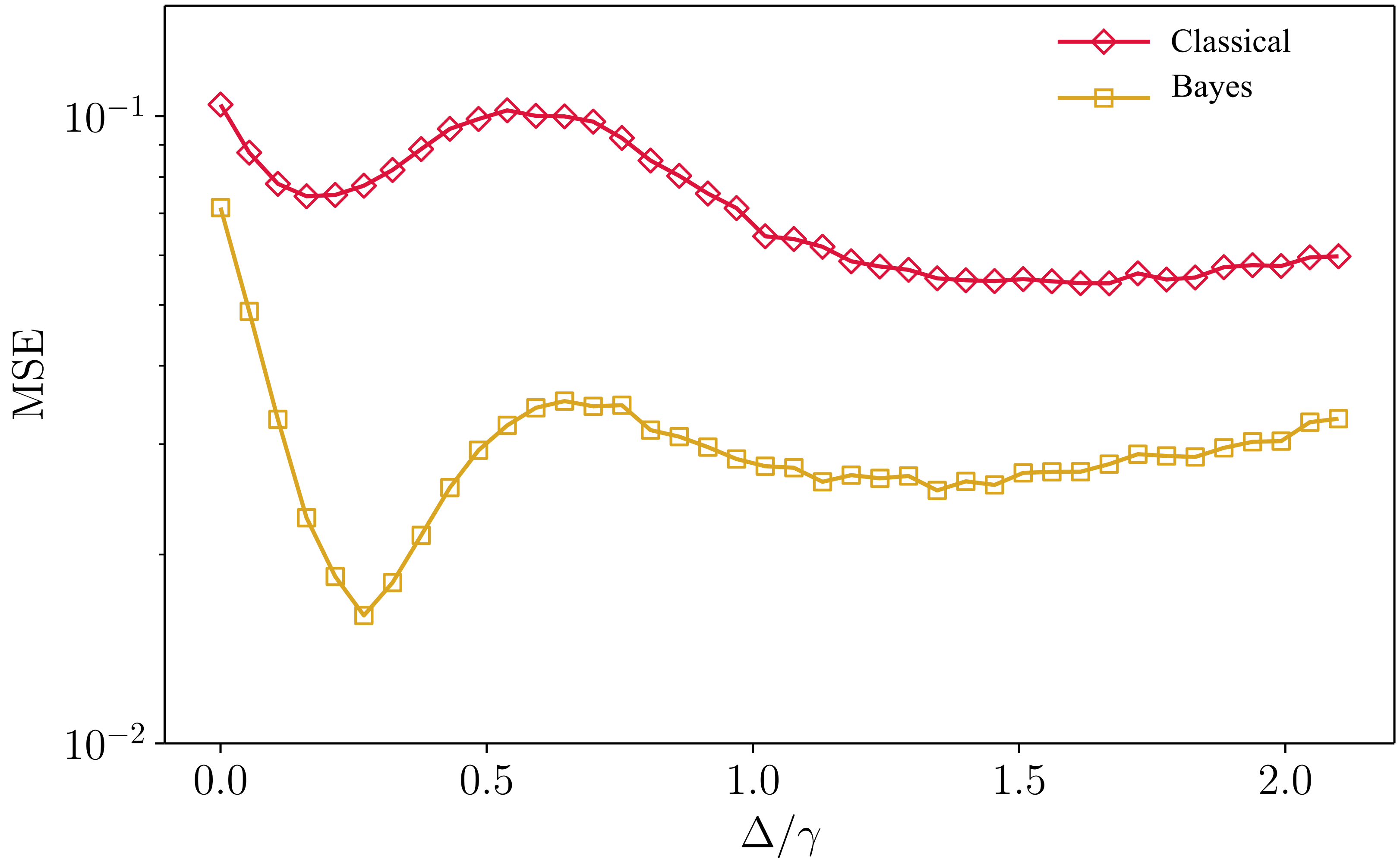
True value

Test with 1 parameter

$$\mathrm{MSE} = \frac{1}{N_{\mathrm{traj}}} \sum^{N_{\mathrm{traj}}} (\Delta_e - \Delta)^2$$

Measure how well the “prediction” compares to the “truth” (averaged over many samples)

- Using data without quantum correlations we obtain a larger deviation from the true parameter value
- Using Bayesian inference on data with quantum correlations we significantly lower the deviation



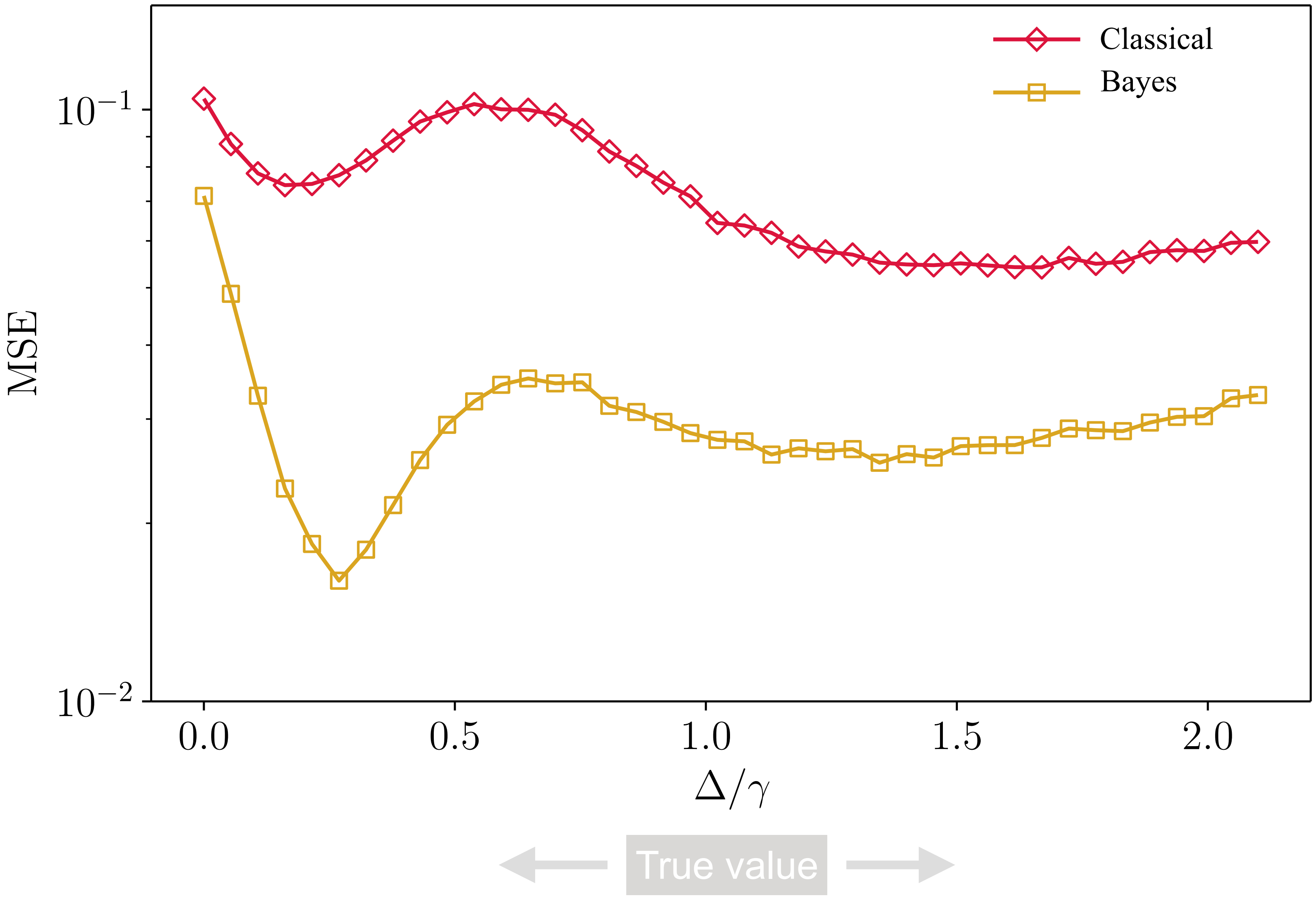
True value

Test with 1 parameter

$$\text{MSE} = \frac{1}{N_{\text{traj}}} \sum^{N_{\text{traj}}} (\Delta_e - \Delta)^2$$

Measure how well the "prediction" compares to the "truth" (averaged over many samples)

- Using data without quantum correlations we obtain a larger deviation from the true parameter value
- Using Bayesian inference on data with quantum correlations we significantly lower the deviation
- The same test trajectories are used in both cases

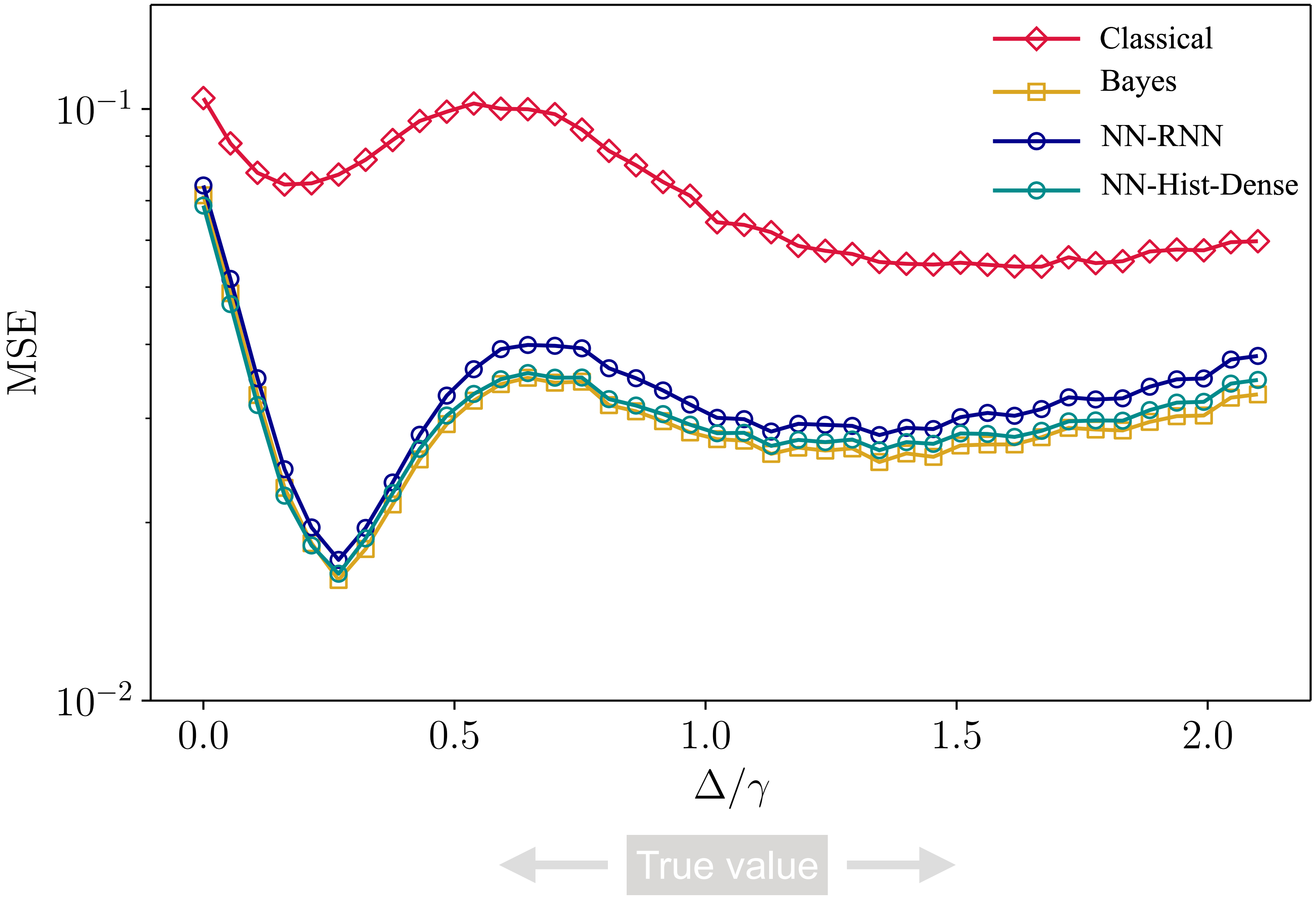


Test with 1 parameter

$$\mathrm{MSE} = \frac{1}{N_{\mathrm{traj}}} \sum^{N_{\mathrm{traj}}} (\Delta_e - \Delta)^2$$

Measure how well the "prediction" compares to the "truth" (averaged over many samples)

- Using data without quantum correlations we obtain a larger deviation from the true parameter value
- Using Bayesian inference on data with quantum correlations we significantly lower the deviation
- The same test trajectories are used in both cases
- Using neural networks we obtain MSE values similar to Bayesian inference

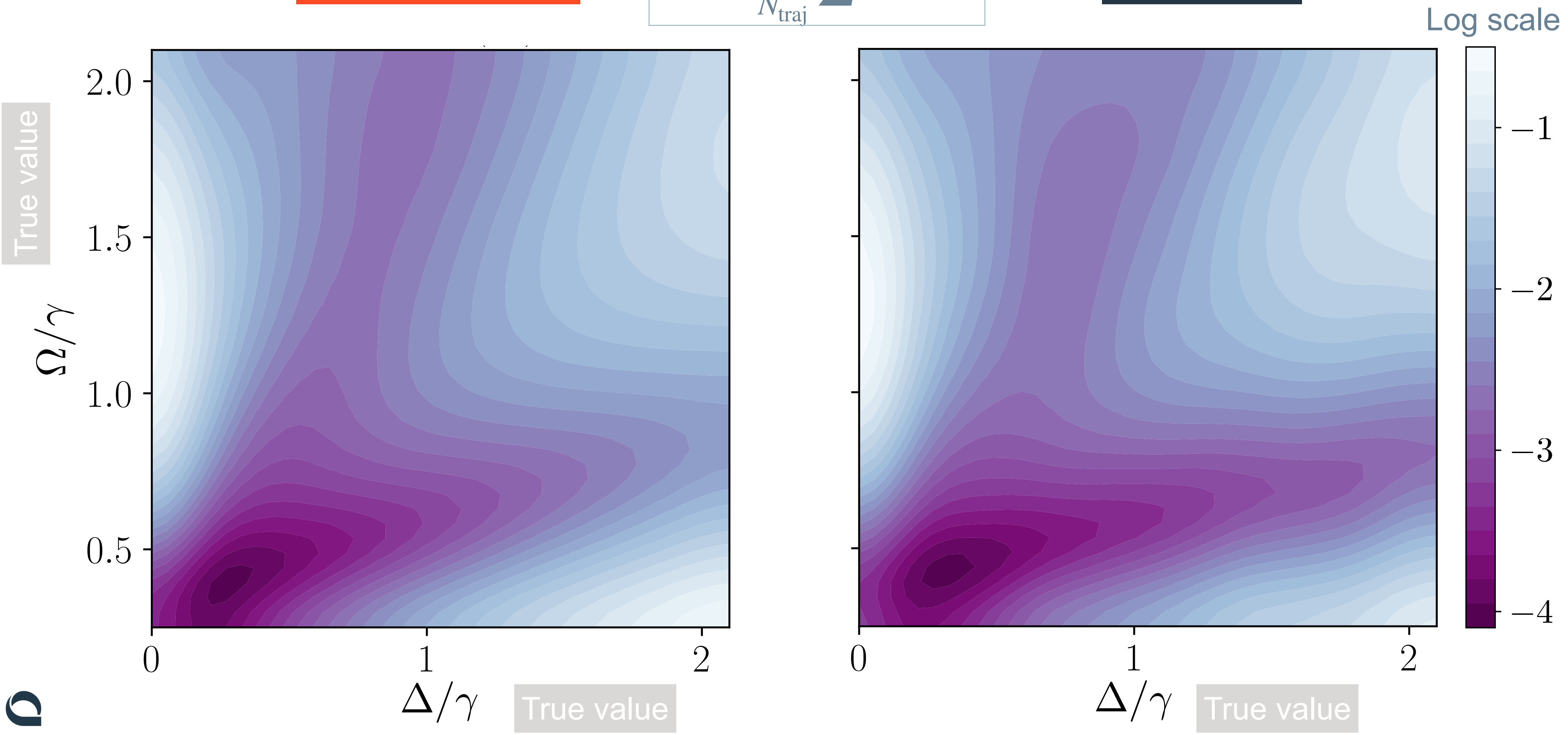


Test with 2 parameters

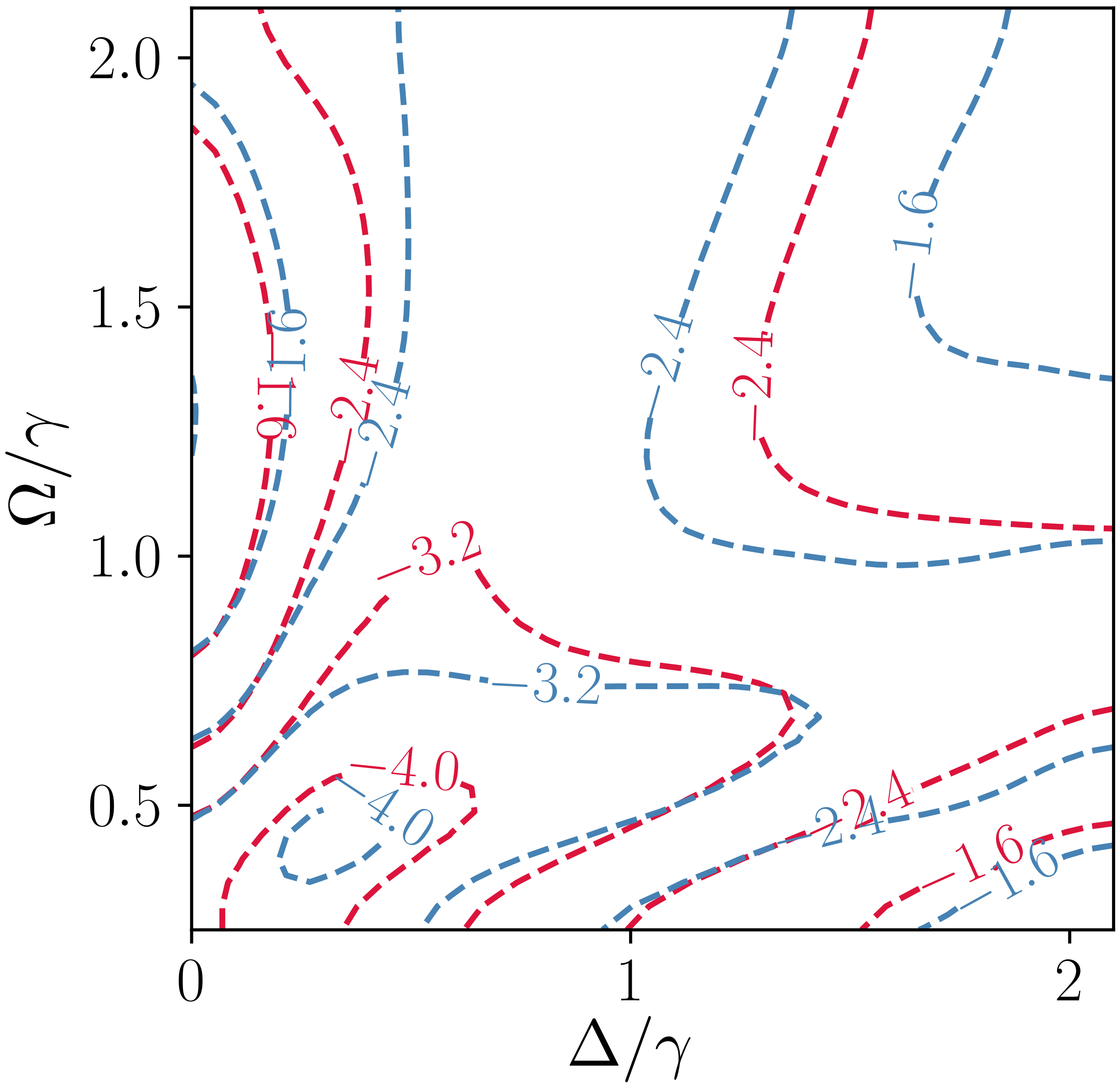
Bayesian Inference

$$\mathrm{MSE} = \frac{1}{N_{\mathrm{traj}}} \sum^{N_{\mathrm{traj}}} (\Delta_e - \Delta)^2$$

NN Inference



Test with 2 parameters

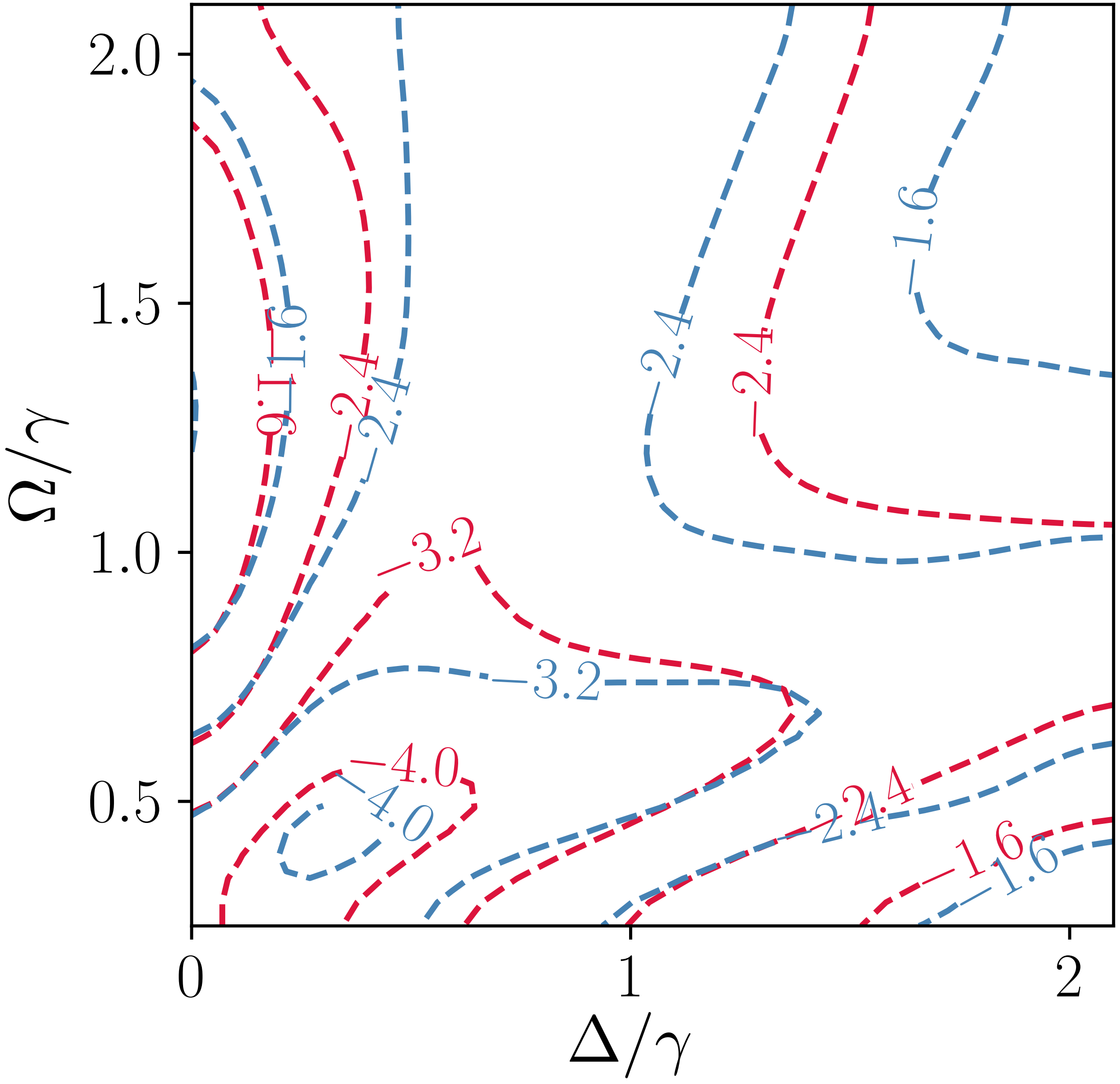


Bayesian Inference

vs

NN Inference

Test with 2 parameters



Bayesian Inference

vs

NN Inference

- Plot showing only the deviation on the parameter Δ
- Both parameter are estimated together from the data
- Using ~10K trajectories to gather statistics on the MSE
- <u>Using neural networks we obtain results x10000 times faster</u>

Results from noisy data

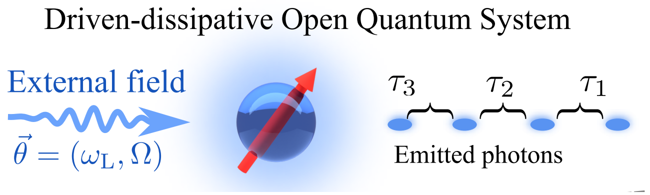
What happens when the training data is not “correct”?

Results from noisy data

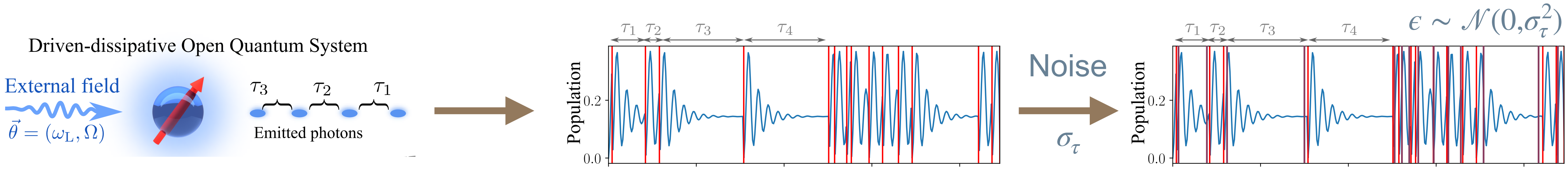
What happens when the training data is not “correct”?

- The time delays from photon counting have jitter noise
- The true parameters used in training are mis-calibrated

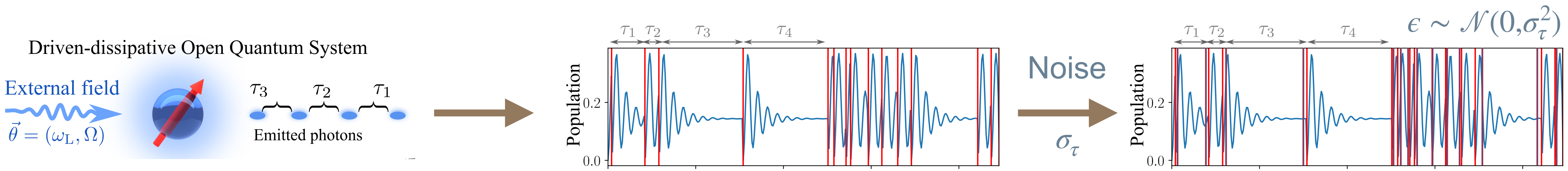
The time delays from photon counting have jitter noise



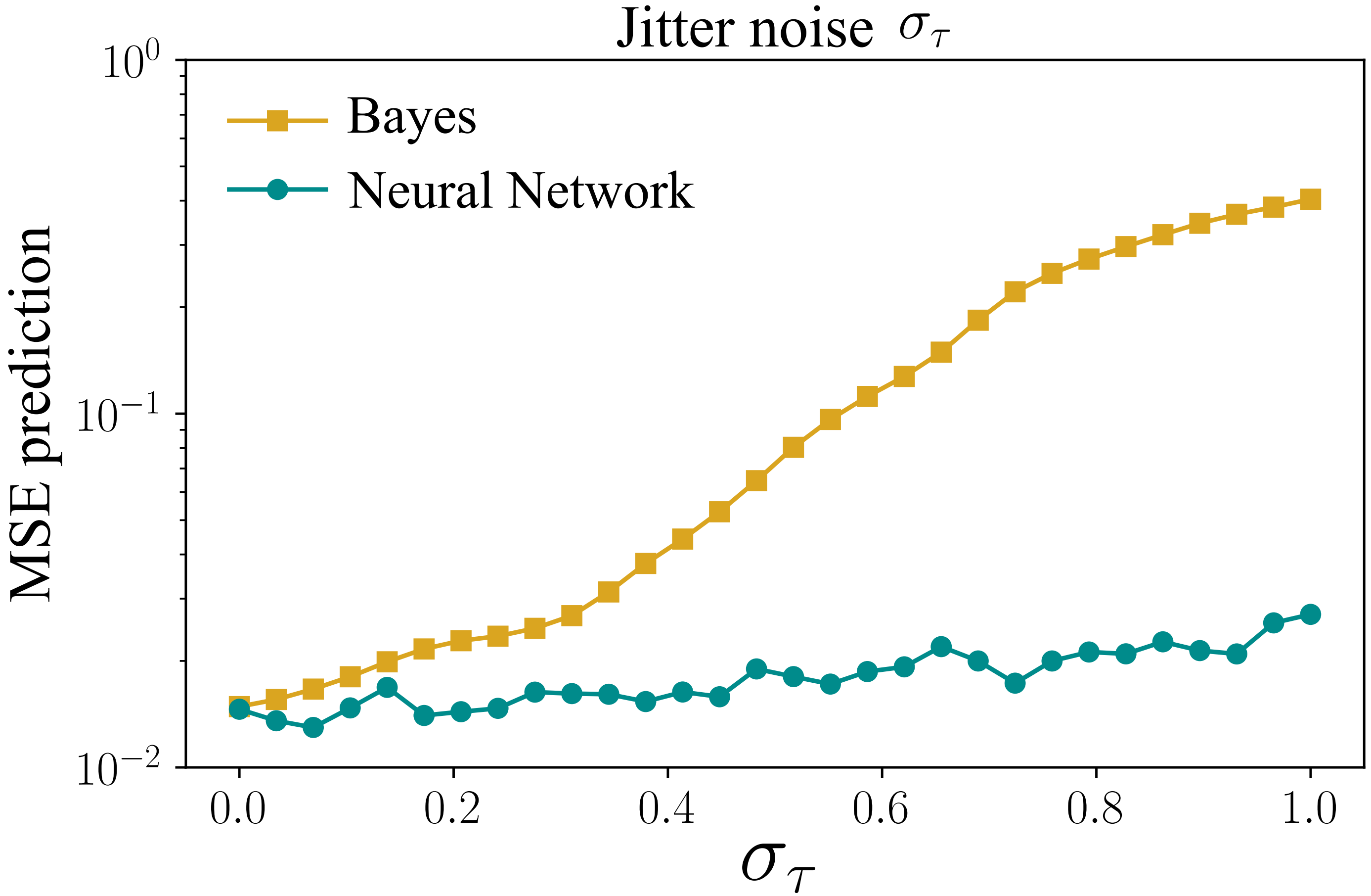
The time delays from photon counting have jitter noise



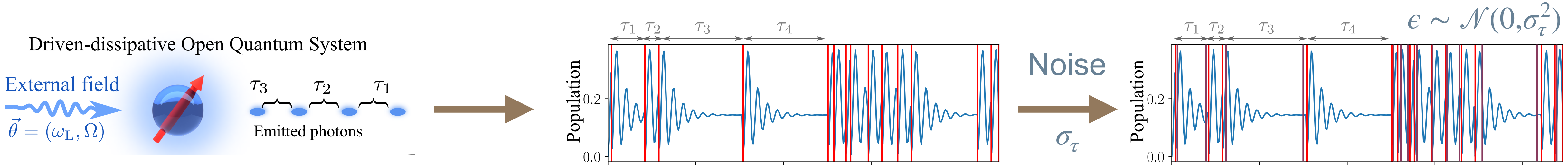
The time delays from photon counting have jitter noise



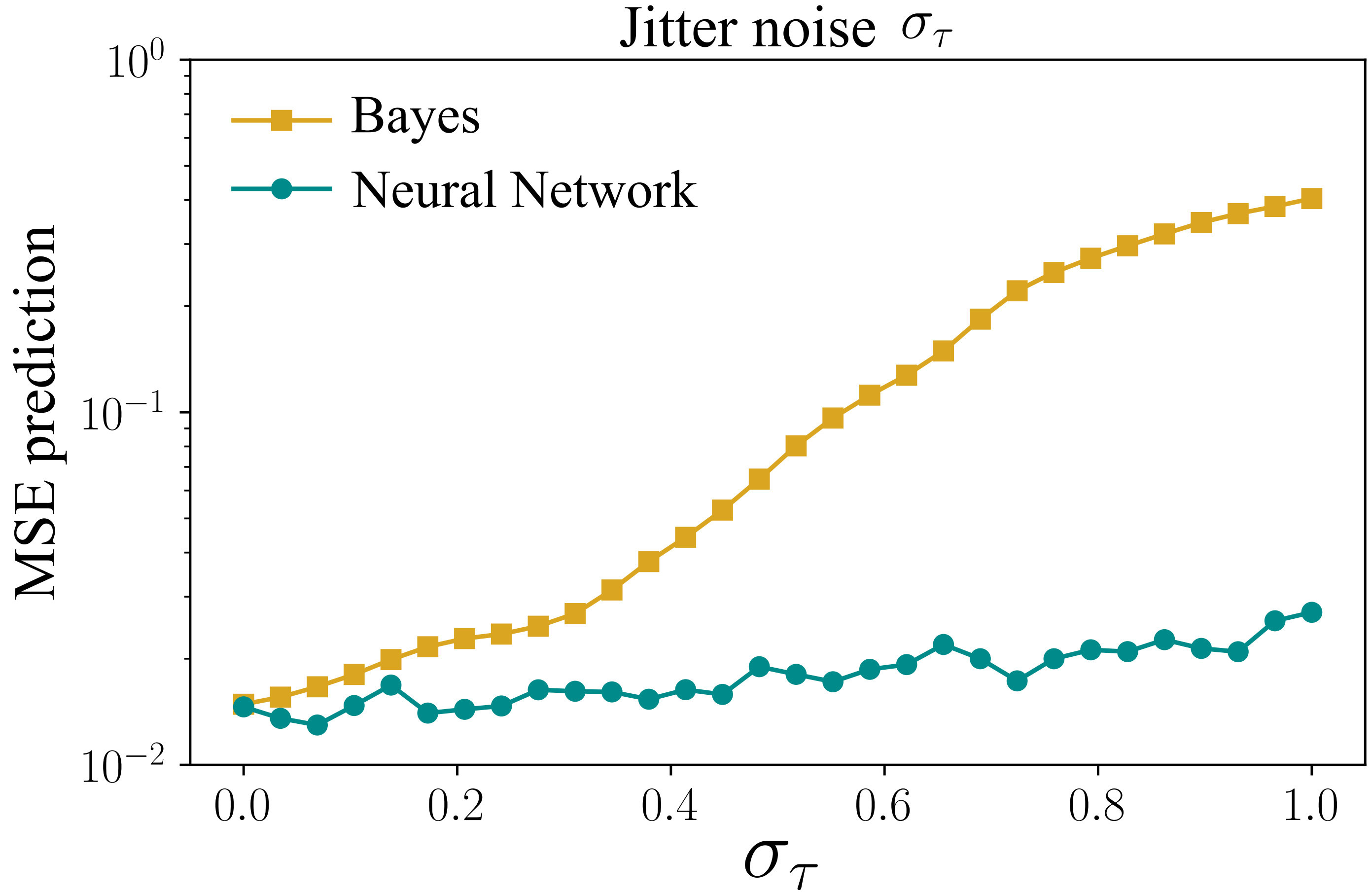
$$\Delta = 0.27\gamma$$



The time delays from photon counting have jitter noise

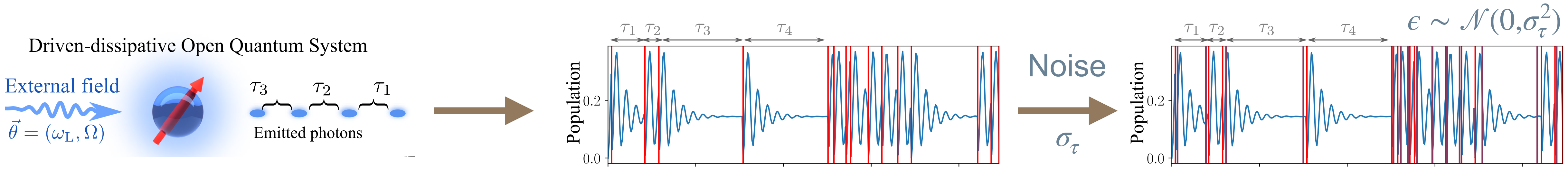


$\Delta = 0.27\gamma$

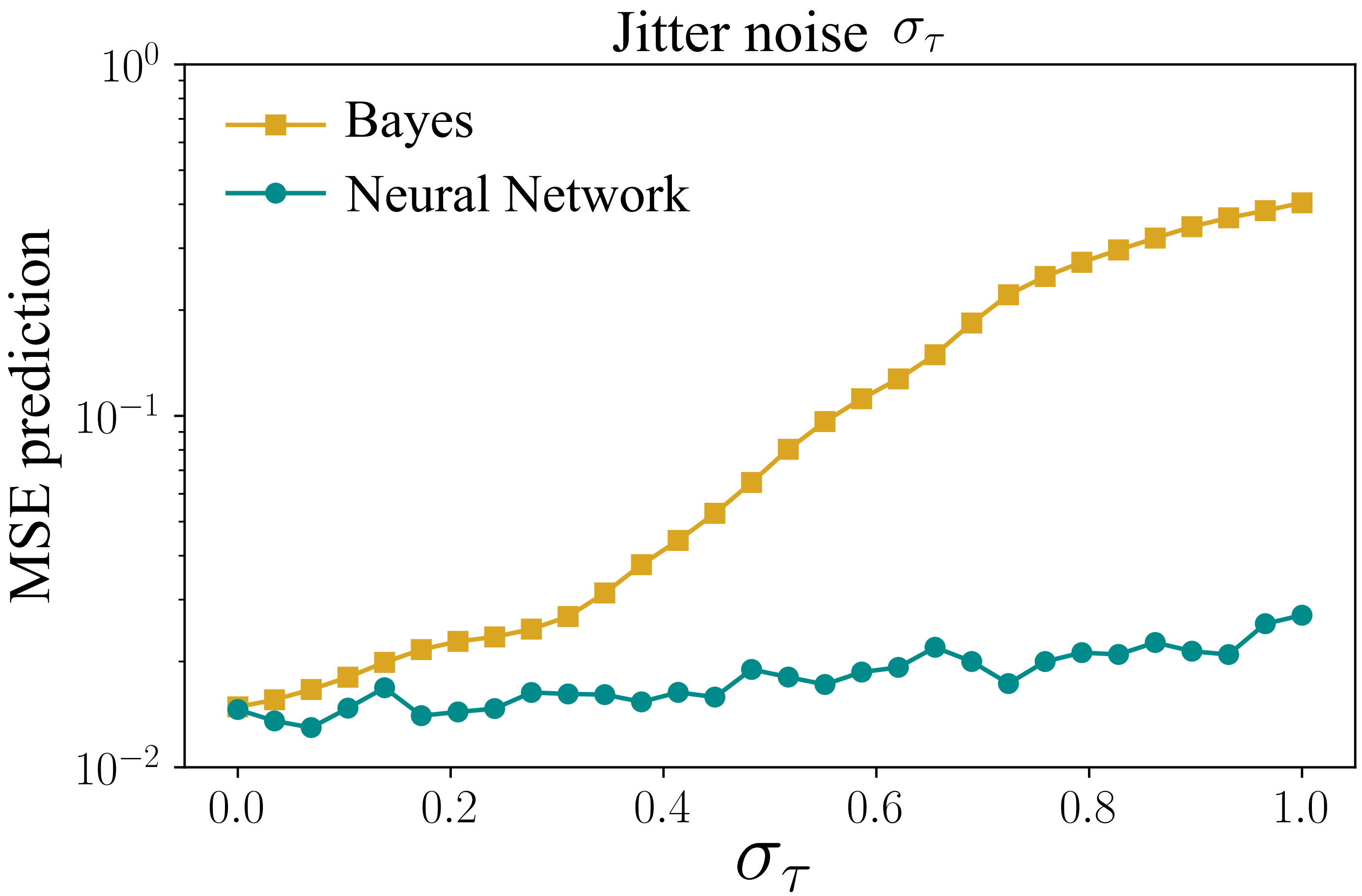


- Plot showing only the deviation on the parameter Δ which is fixed

The time delays from photon counting have jitter noise

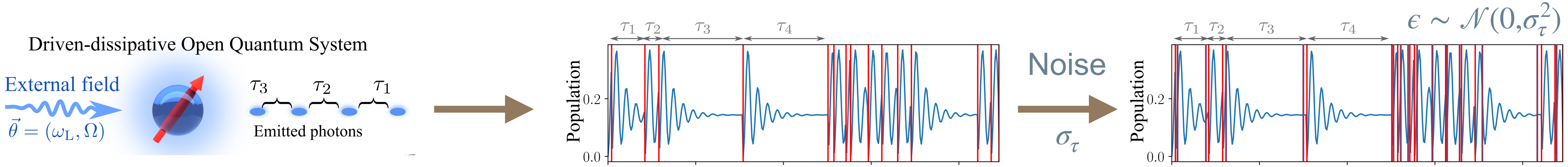


$\Delta = 0.27\gamma$

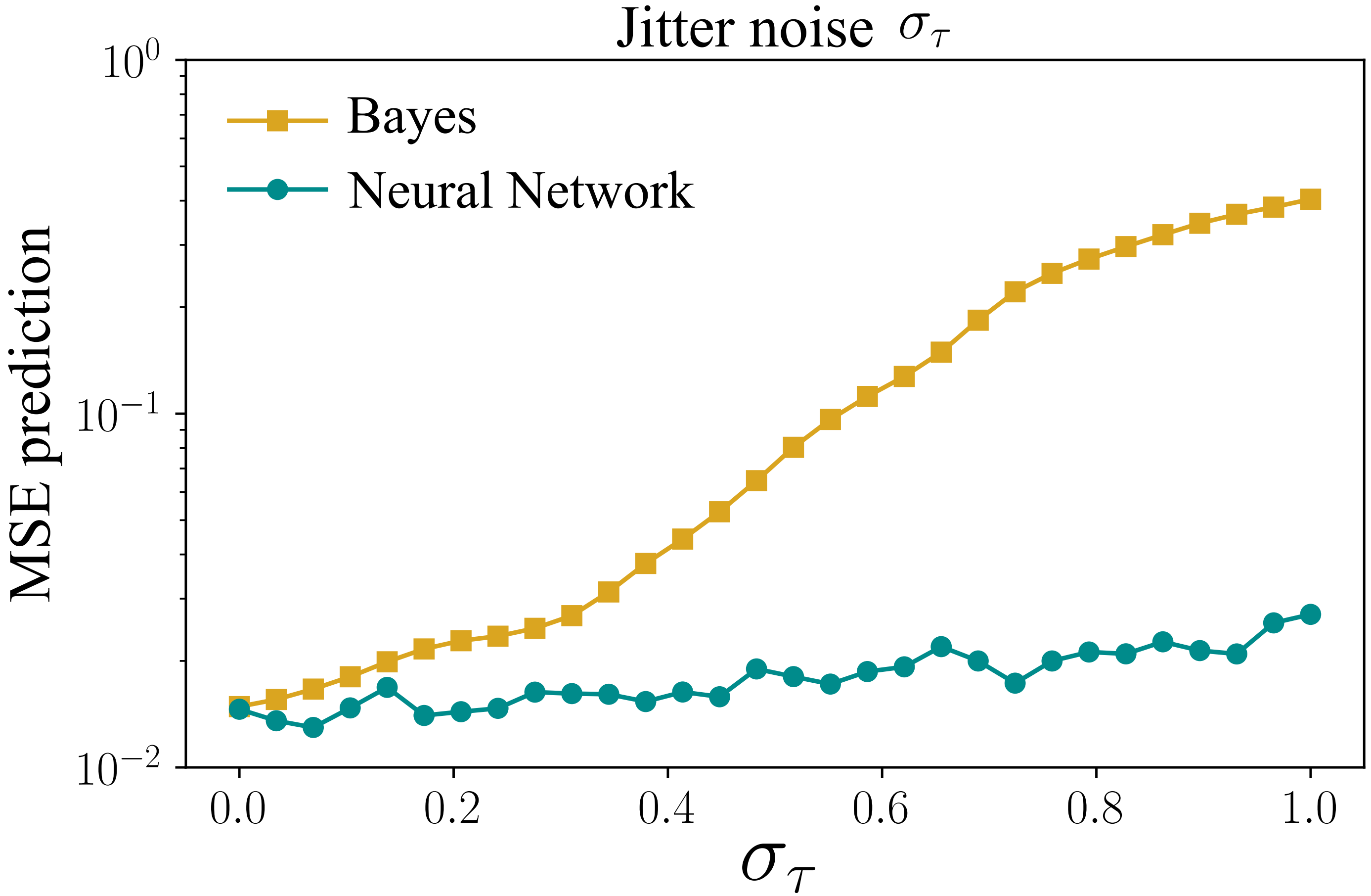


- Plot showing only the deviation on the parameter Δ which is fixed
- The Bayesian inference is using the “wrong” likelihood: no modeling of the noise!

The time delays from photon counting have jitter noise

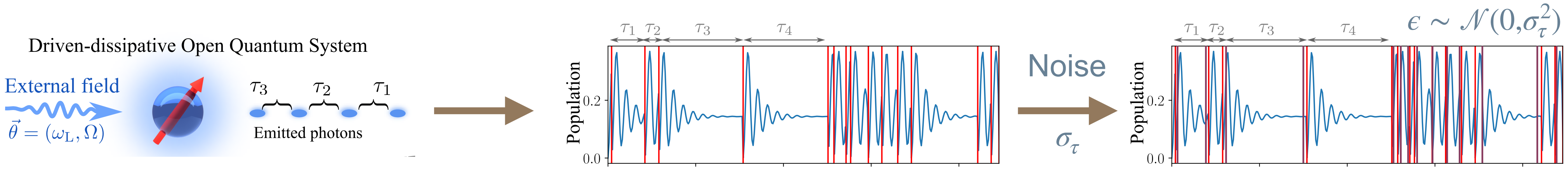


$$\Delta = 0.27\gamma$$

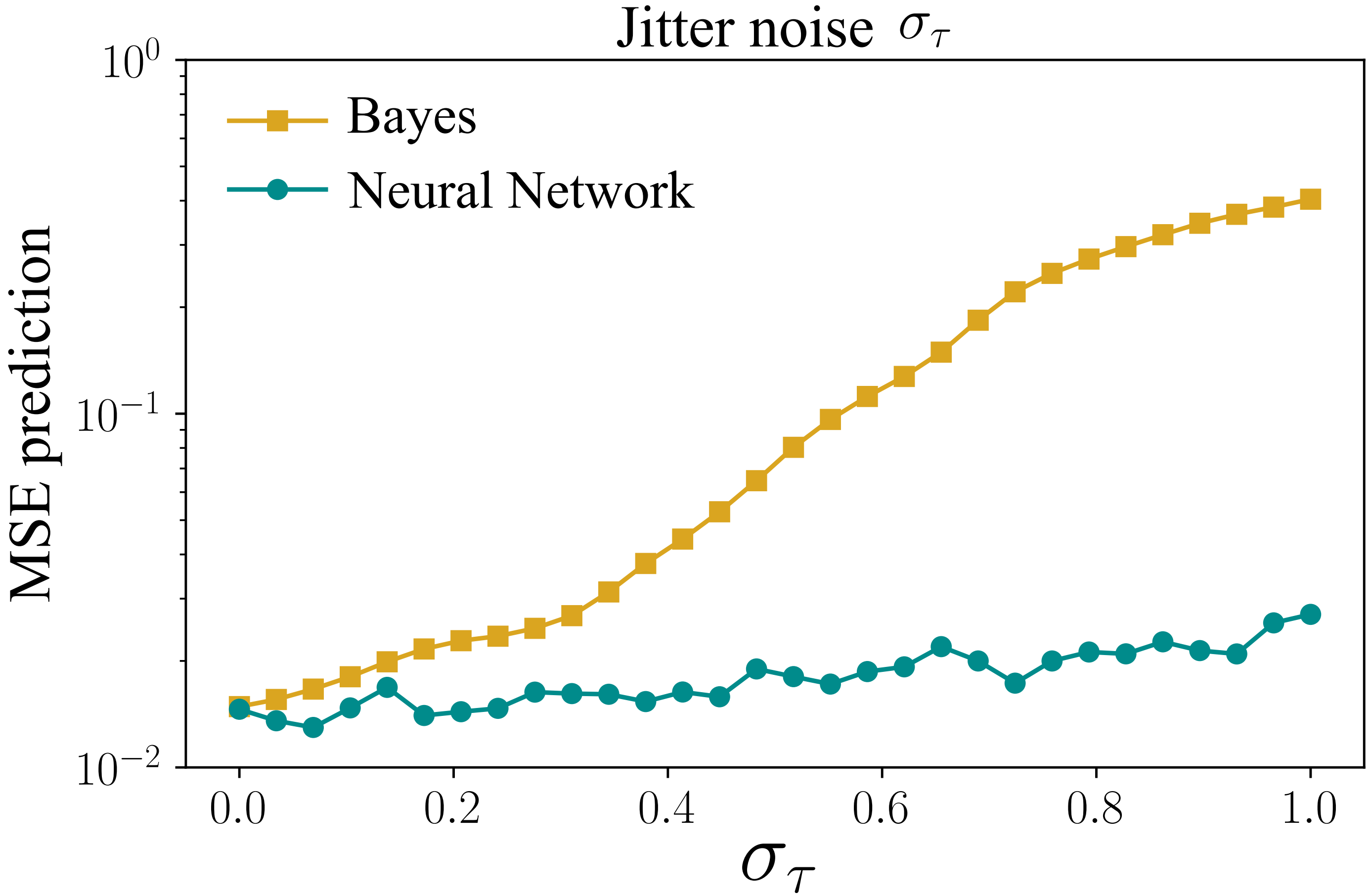


- Plot showing only the deviation on the parameter Δ which is fixed
- The Bayesian inference is using the “wrong” likelihood: no modeling of the noise!
- Larger noise means larger mis-modeling errors

The time delays from photon counting have jitter noise

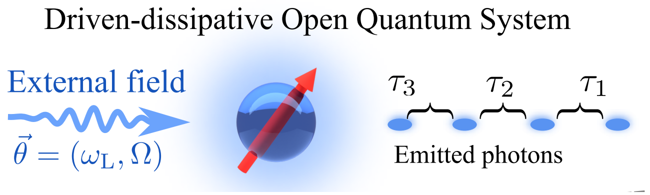


$$\Delta = 0.27\gamma$$

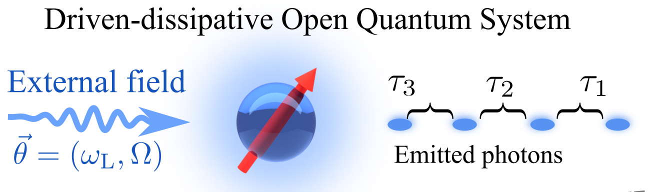


- Plot showing only the deviation on the parameter Δ which is fixed
- The Bayesian inference is using the "wrong" likelihood: no modeling of the noise!
- Larger noise means larger mis-modeling errors
- <u>Using neural networks we obtain results that are robust to jitter noise</u>

The true parameters used in training are mis-calibrated



The true parameters used in training are mis-calibrated



$x_i \rightarrow y_i$

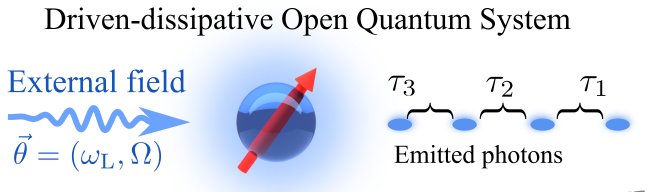
Training Noise

σ_y

$x_i \rightarrow y_i + \epsilon$

$\epsilon \sim \mathcal{N}(0, \sigma_y^2)$

The true parameters used in training are mis-calibrated

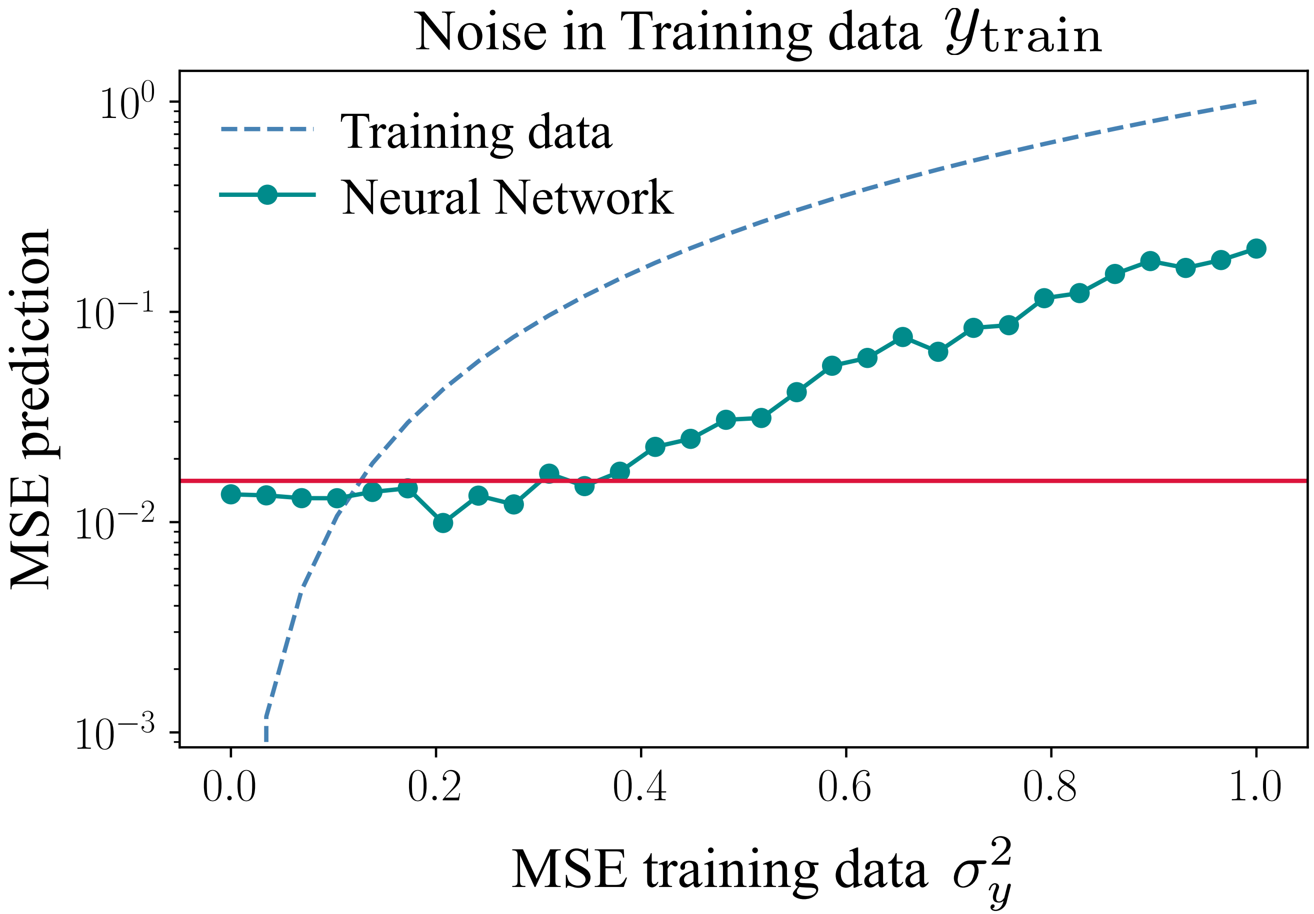


$x_i \to y_i$

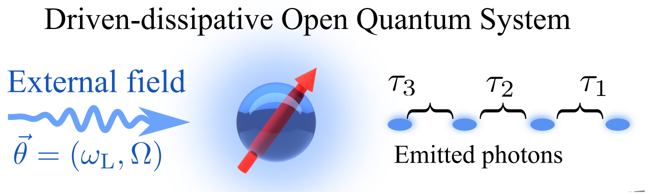
Training Noise σ_y

$x_i \to y_i + \epsilon$ $\quad \epsilon \sim \mathcal{N}(0,\sigma_y^2)$

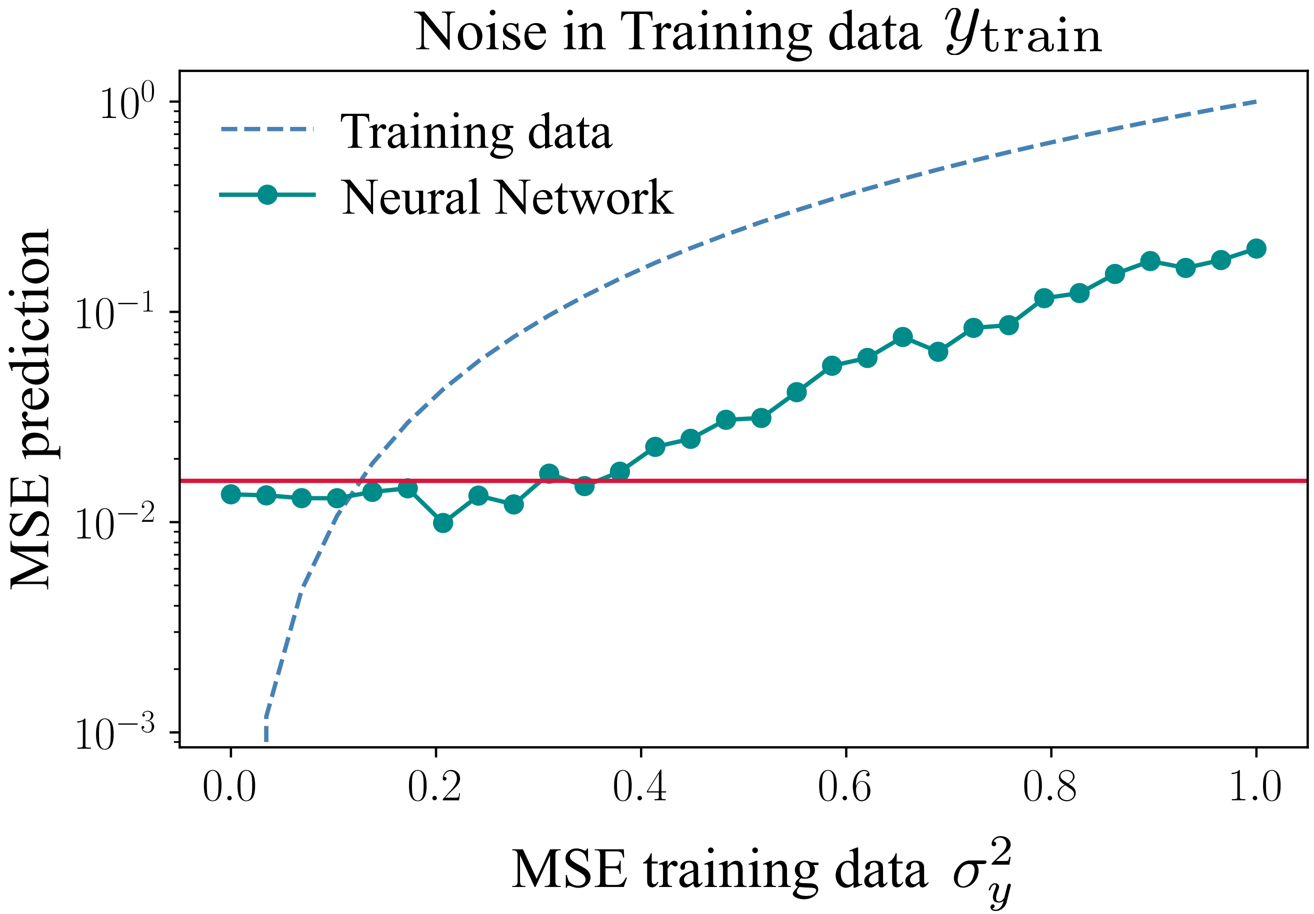
$\Delta = 0.27\gamma$



The true parameters used in training are mis-calibrated



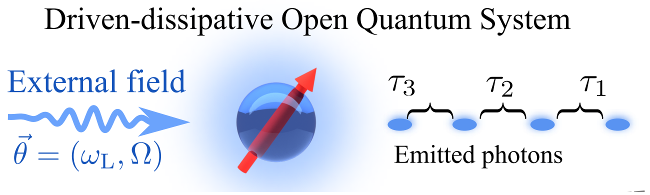
$x_i \rightarrow y_i$ — Training Noise σ_y → $x_i \rightarrow y_i + \epsilon$ $\quad \epsilon \sim \mathcal{N}(0, \sigma_y^2)$



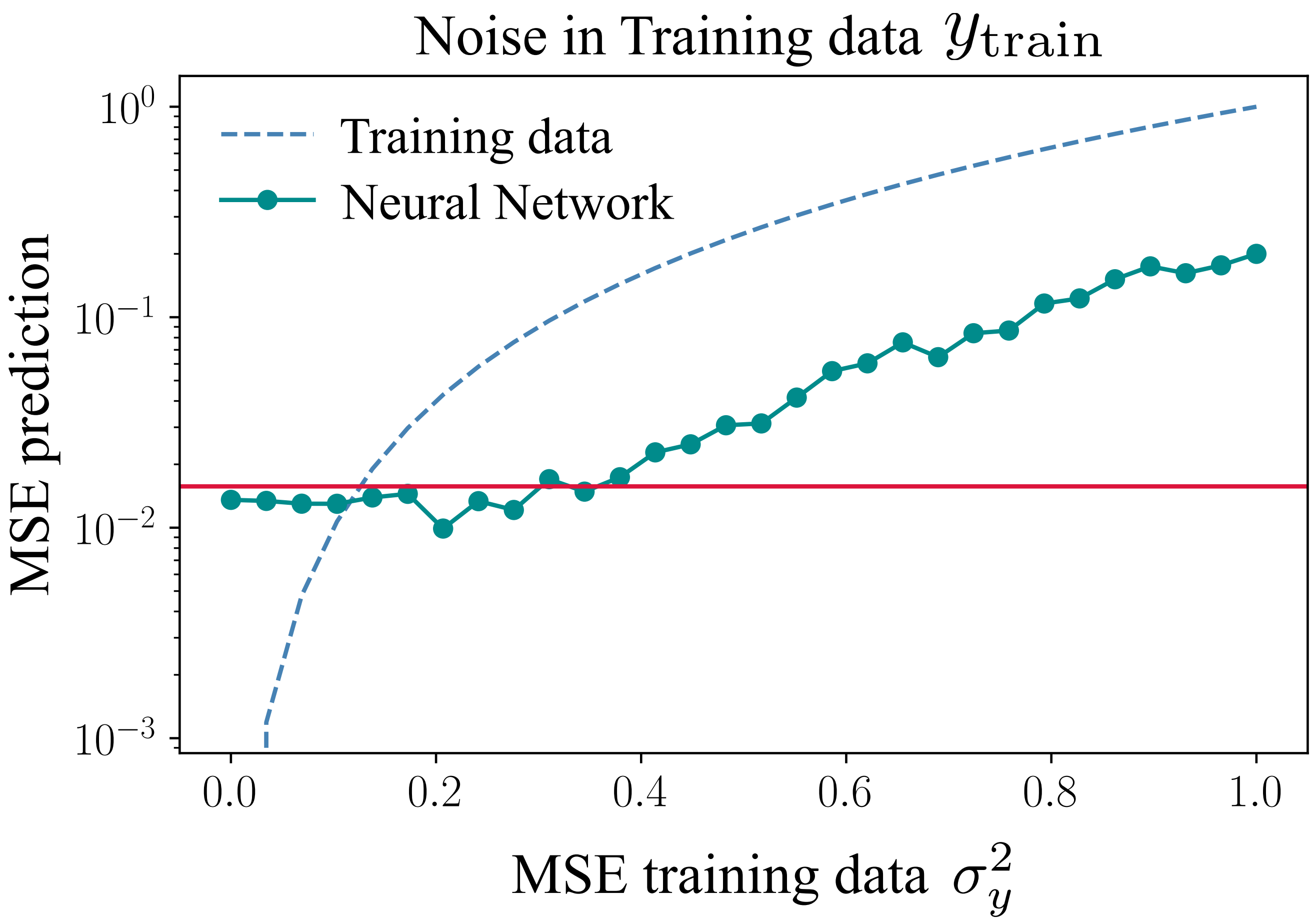
$$\Delta = 0.27\gamma$$

- Bayesian ideal result in red

The true parameters used in training are mis-calibrated



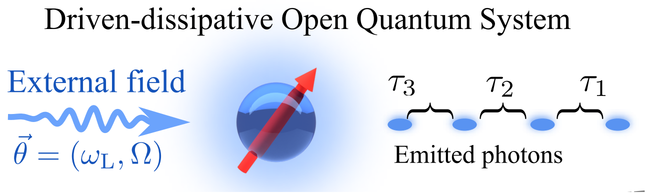
$x_i \rightarrow y_i$ —Training Noise, σ_y→ $x_i \rightarrow y_i + \epsilon$ $\quad \epsilon \sim \mathcal{N}(0, \sigma_y^2)$



$$\Delta = 0.27\gamma$$

- Bayesian ideal result in red
- Training data is “wrong”: intrinsic MSE=σ_y^2

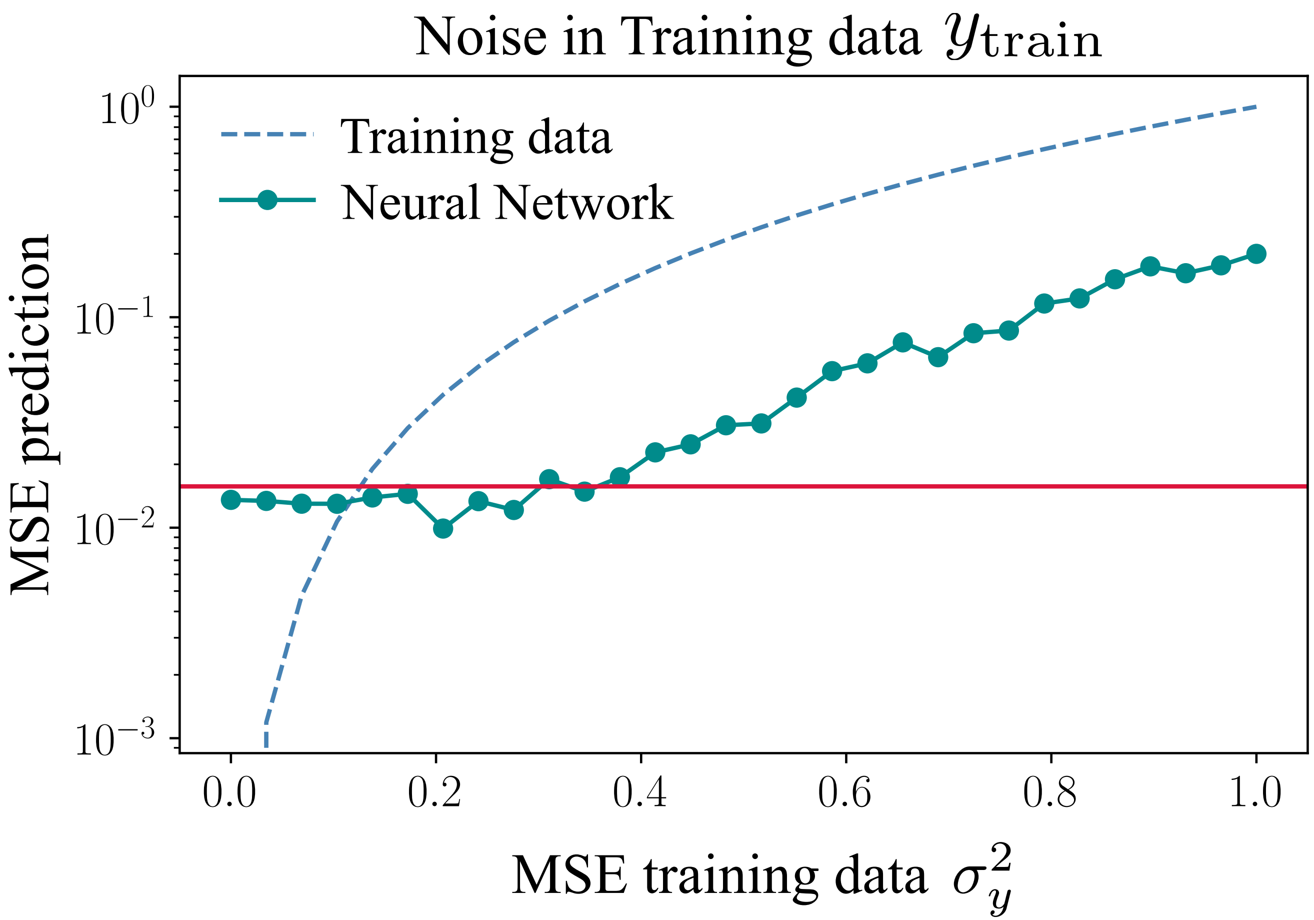
The true parameters used in training are mis-calibrated



$x_i \to y_i$

Training Noise σ_y

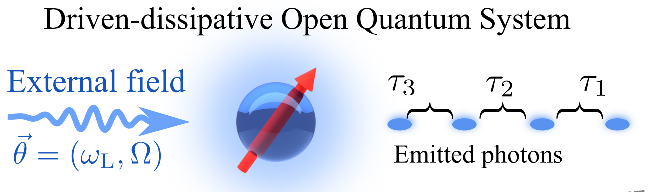
$x_i \to y_i + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \sigma_y^2)$



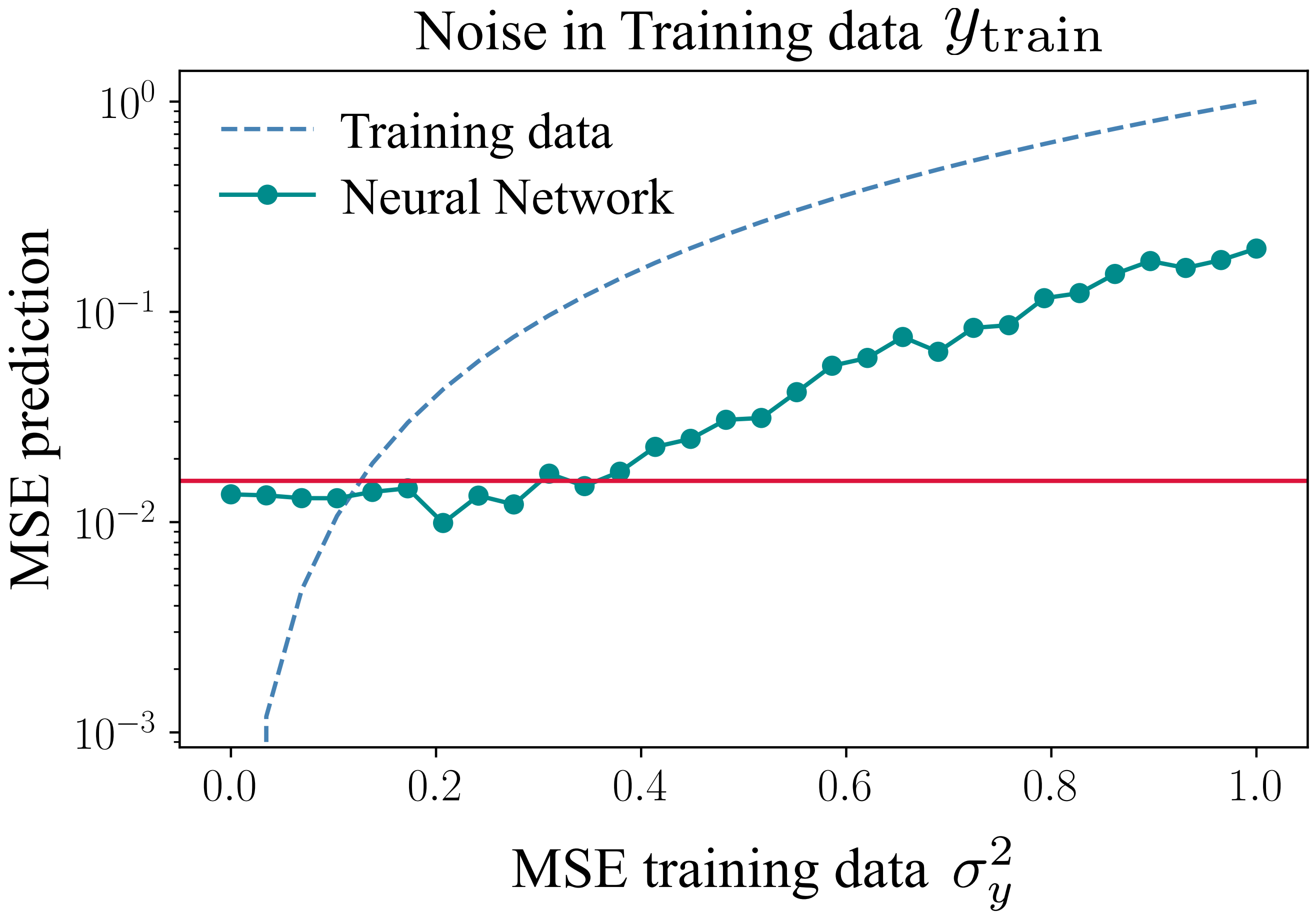
$\Delta = 0.27\gamma$

- Bayesian ideal result in red
- Training data is “wrong”: intrinsic MSE=σ_y^2
- Dashed line is MSE expected from training data

The true parameters used in training are mis-calibrated



$x_i \rightarrow y_i$ — Training Noise σ_y → $x_i \rightarrow y_i + \epsilon$ $\quad \epsilon \sim \mathcal{N}(0, \sigma_y^2)$



$$\Delta = 0.27\gamma$$

- Bayesian ideal result in red
- Training data is “wrong”: intrinsic MSE=σ_y^2
- Dashed line is MSE expected from training data
- Using neural networks we obtain results that are robust to training noise and consistent with ideal Bayes results

Robust and speedy inference with NN

- Does not need exact modeling of quantum sensors
- Does not need "big data"
- Reaches the limits of precision of Bayesian Inference
- Deployable to edge devices inside quantum labs
- Zenodo code and data release for reproducibility

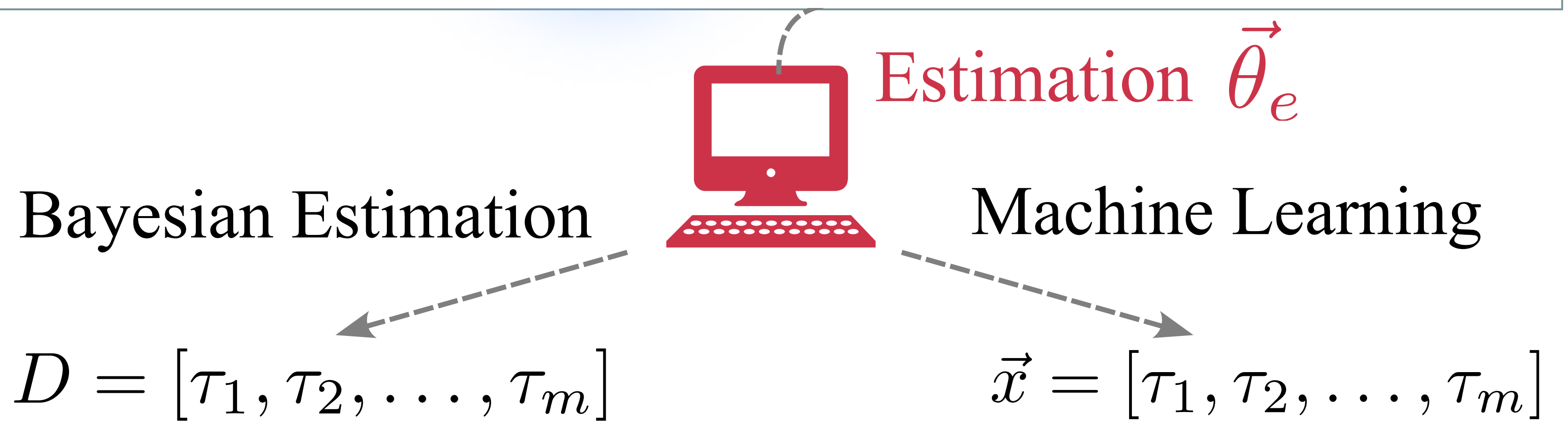
QUANTINUUM

Importance of data

Classical vs Quantum data

Importance of data

Classical vs Quantum data

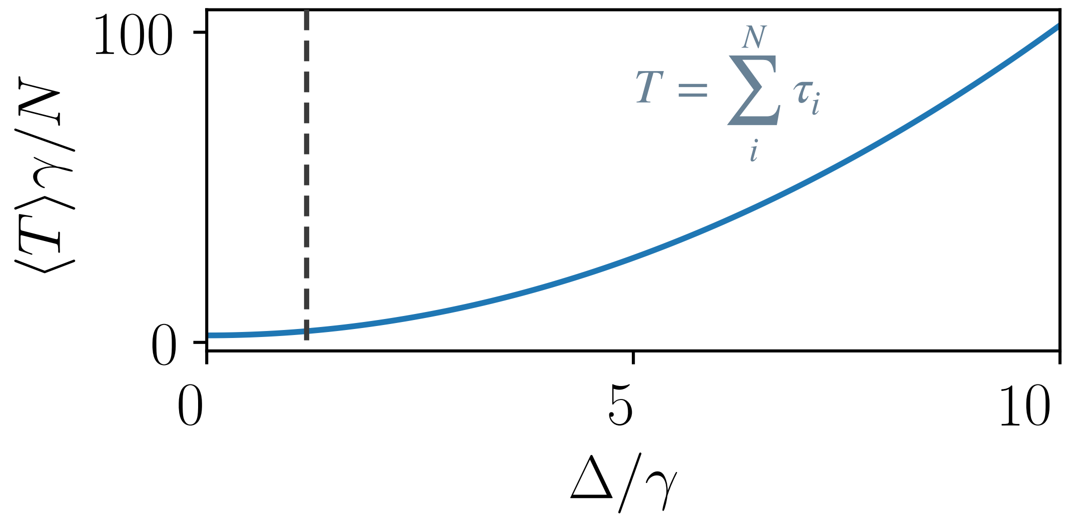


Classical

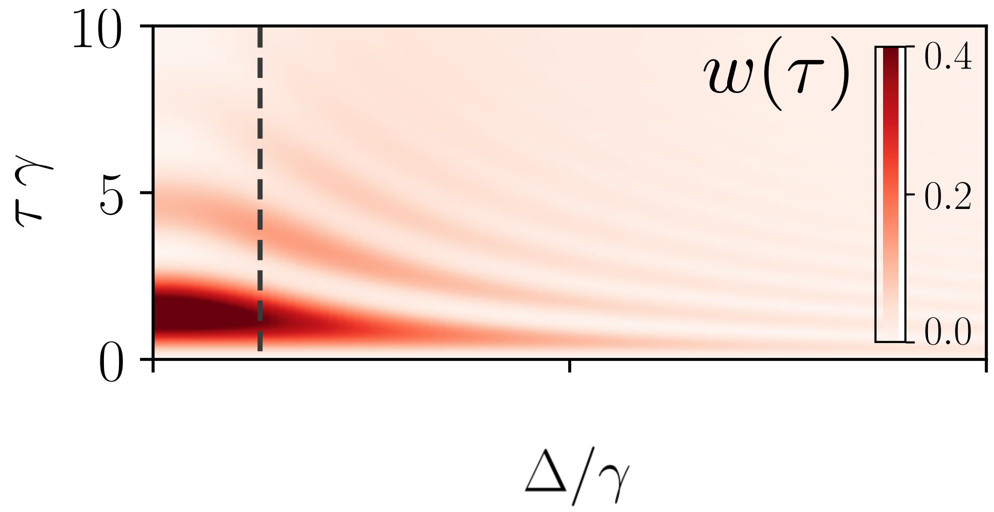
Quantum

Driven-dissipative Open Quantum System

No single-photon resolution



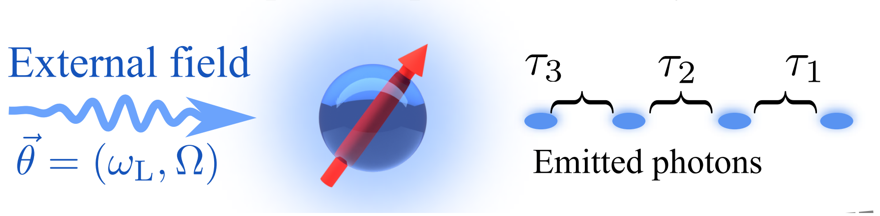
Single-photon resolution ("clicks")



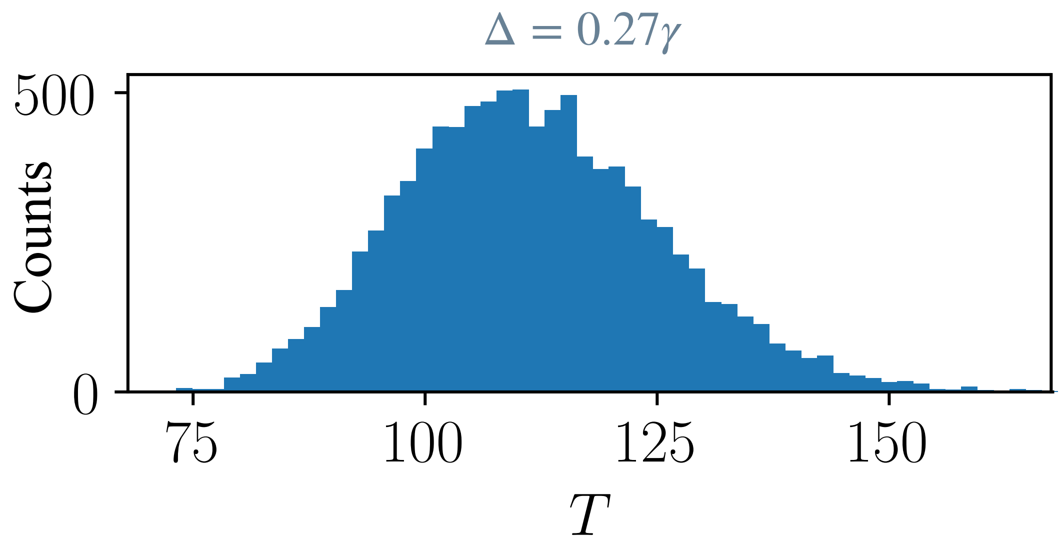
Classical

Quantum

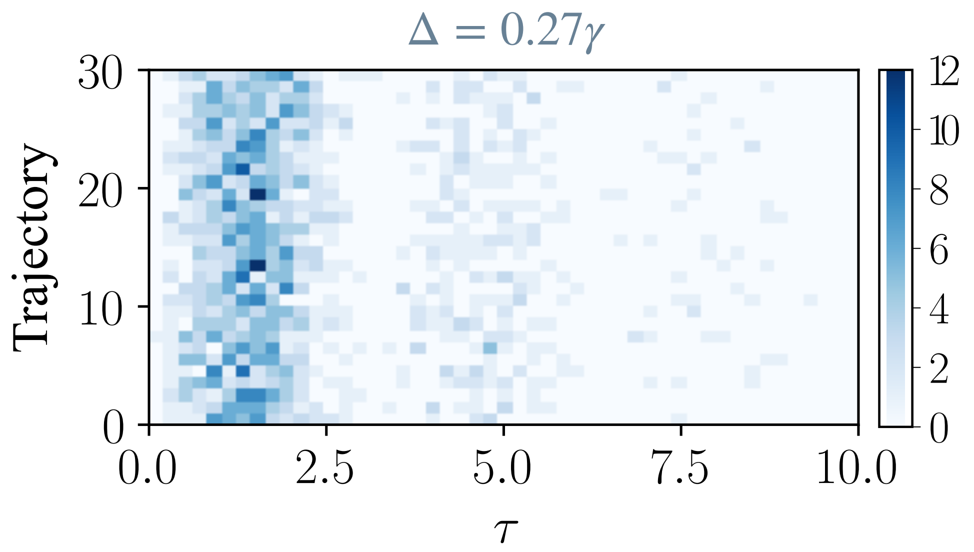
Driven-dissipative Open Quantum System



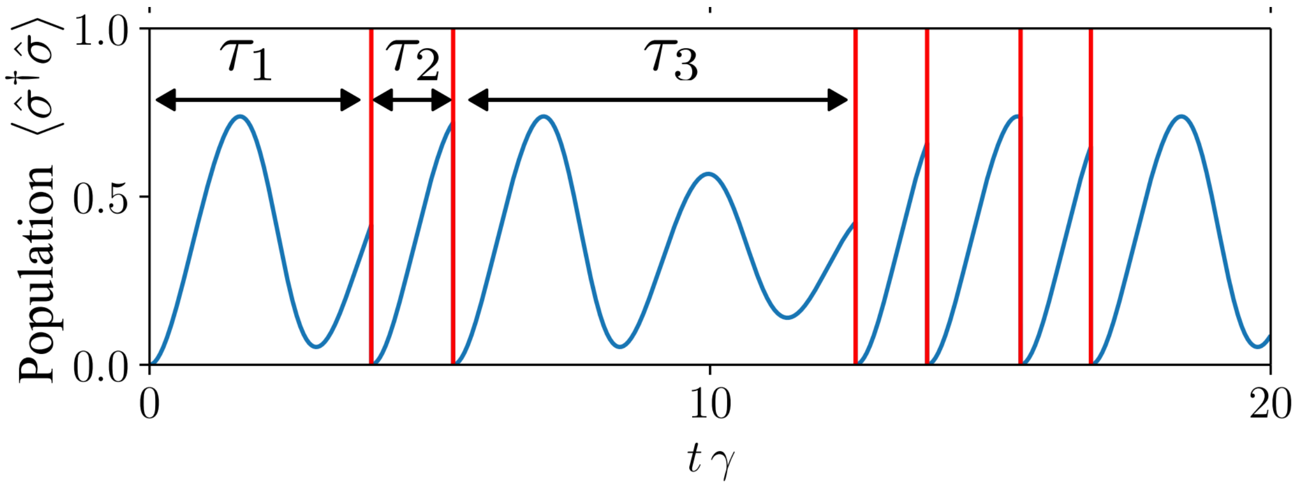
No single-photon resolution



Single-photon resolution ("clicks")

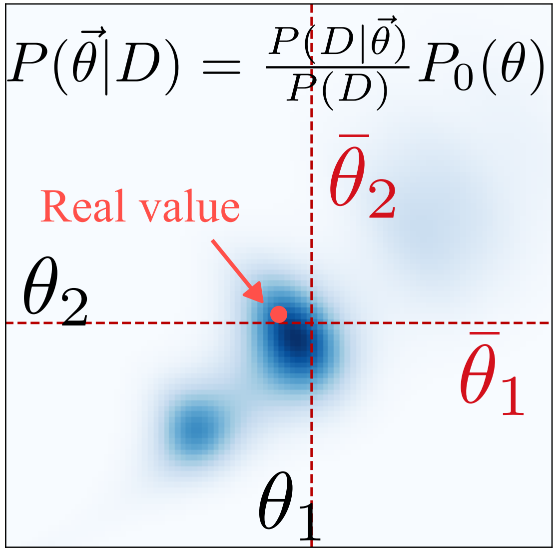


Bayesian Estimation



Data

$$D = [\tau_1, \tau_2, \ldots, \tau_m]$$



Estimation

$$\vec\theta_e = [\bar\theta_1, \bar\theta_2]$$

$$P(\vec\theta \mid D) = \frac{P(D \mid \theta)}{P(D)} P_0(\vec\theta) = \frac{P(D \mid \theta)}{\int d\vec\theta P(D \mid \vec\theta) P_0(\vec\theta)} P_0(\vec\theta)$$

Posterior

$$P(D \mid \vec\theta) = \prod_{i=1}^{N} w\left(\tau_i; \vec\theta\right)$$

This is the probability to observe D given a pair $\vec\theta$

Likelihood

$$P_0(\vec\theta)$$

This is the probability of a pair $\vec\theta$ before any observation

Prior

$$\int d\vec\theta P(D \mid \vec\theta) P_0(\vec\theta)$$

This is a normalization factor

Evidence

Gambetta, J. & Wiseman, H. M. *Phys. Rev. A* **64**, 14 (2001)

Neural Network workflow

Code on GitHub

Training

- Collect "simulated" data using Qutip
  - ~4M trajectories (80-20 split for train-validation)
  - One case for 1D estimation (fix Ω)
  - One case for 2D estimation
  - Fix trajectories with 48 jumps
- Fix the network architecture (no hyper-parameter optimization)
- Minimize the Mean Squared Logarithmic Error (MSLE) with the Adam optimizer

Test

- Collect "simulated" data using Qutip
  - ~10K trajectories
  - These are completely different from the training set
  - Fix trajectories with 48 jumps
- For each true value of the parameters, compute the Mean Squared Error (MSE) with the predicted value

Inference

- If the performance is satisfactory, save the model for inference
- The model can be stored in a very compact form and be ready for deployment
- Use the model on edge devices right next the experimental setup to do "real time" inference of the parameters

Neural Network details

Definition of the two different architectures used in this work:

1. Recurrent Neural Network (RNN)
2. Fully connected Neural Network with Histogram layer (Hist-Dense) for both 1D and 2D estimation

Example of number of parameters in the case of the estimation of Δ for 1D and (Δ, Ω) for 2D.

Training is done with the Adam optimizer with default learning rate for 1200 epoch and batch size of 12800 on TPUs

Recurrent Neural Network (RNN)

Layer	Output shape	Activation	# Parameters
LSTM	17	ReLU	1292
LSTM	17	ReLU	2380
Dense	1	Linear	18
Trainable params.	3,690		
Epochs	1200		

Hist-Dense

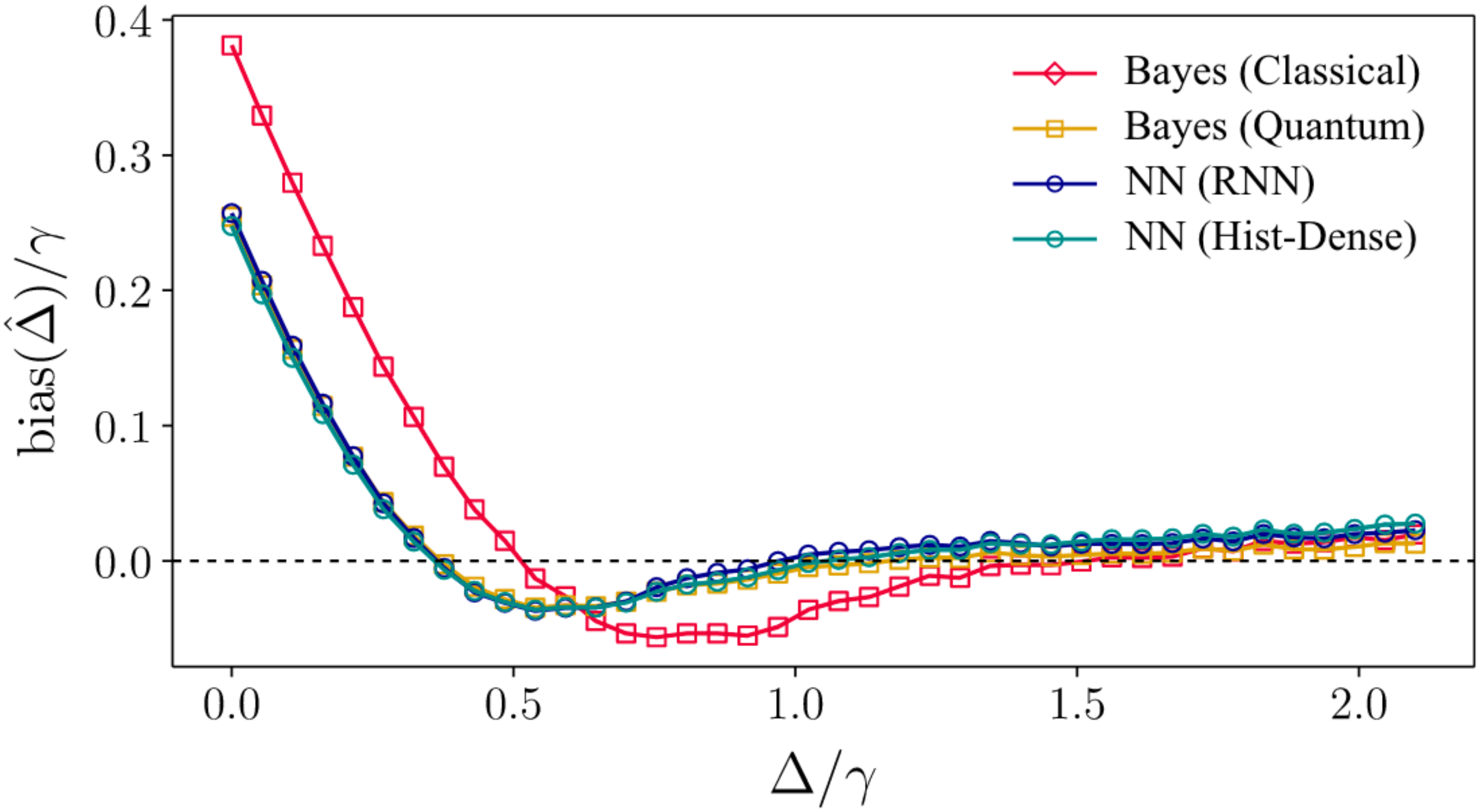
Layer	Output shape	Activation	# Parameters
Histogram	700	-	0
Dense	100	ReLU	70100
Dense	50	ReLU	5050
Dense	30	ReLU	1530
Dense	1	Linear	31
Trainable params.	76,711		
Epochs	1200		

Hist-Dense 2D

Layer	Output shape	Activation	# Parameters
Histogram	700	-	0
Dense	100	ReLU	70100
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		

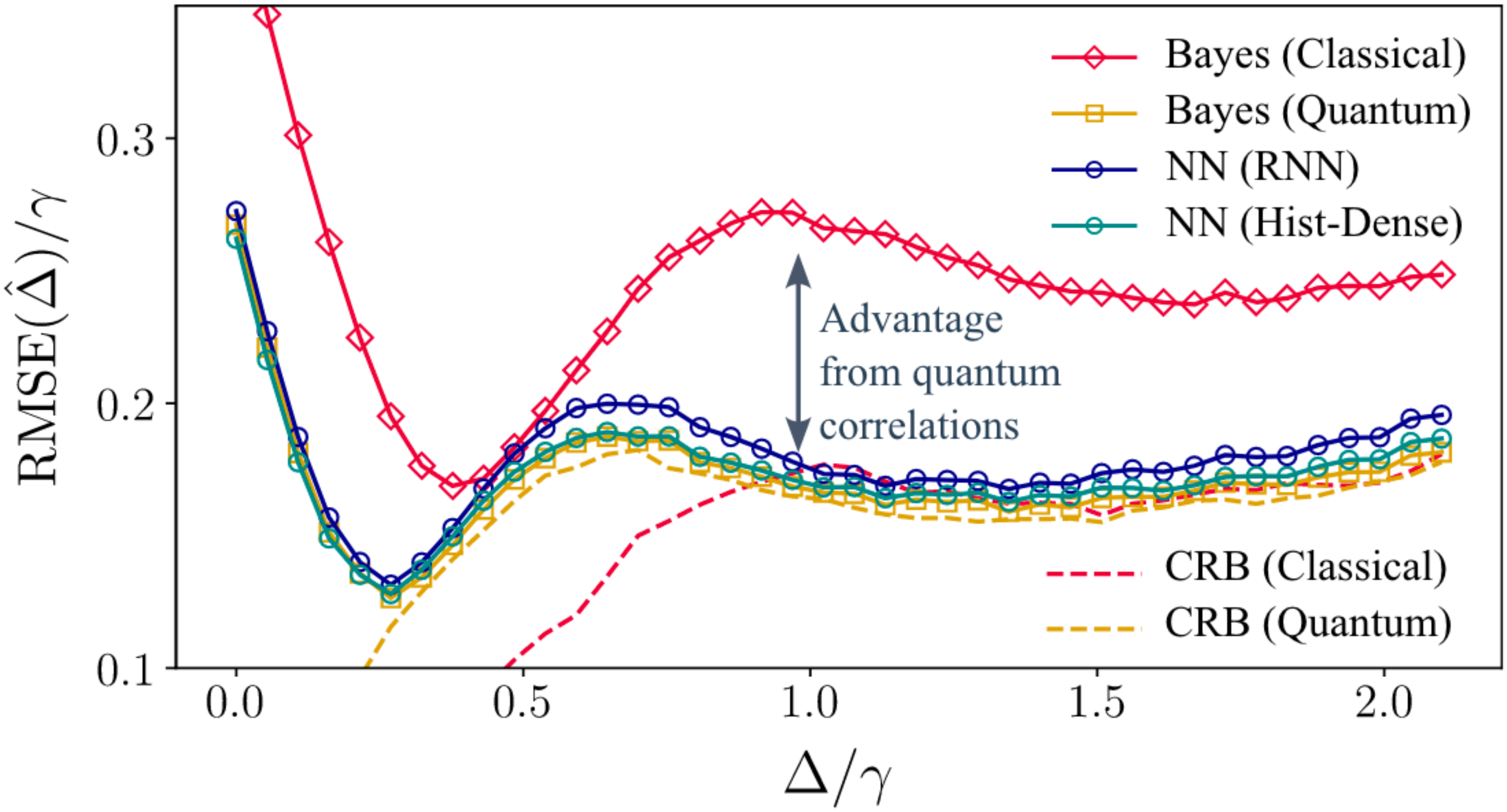
Code on [GitHub]

Bias of the estimators in 1D



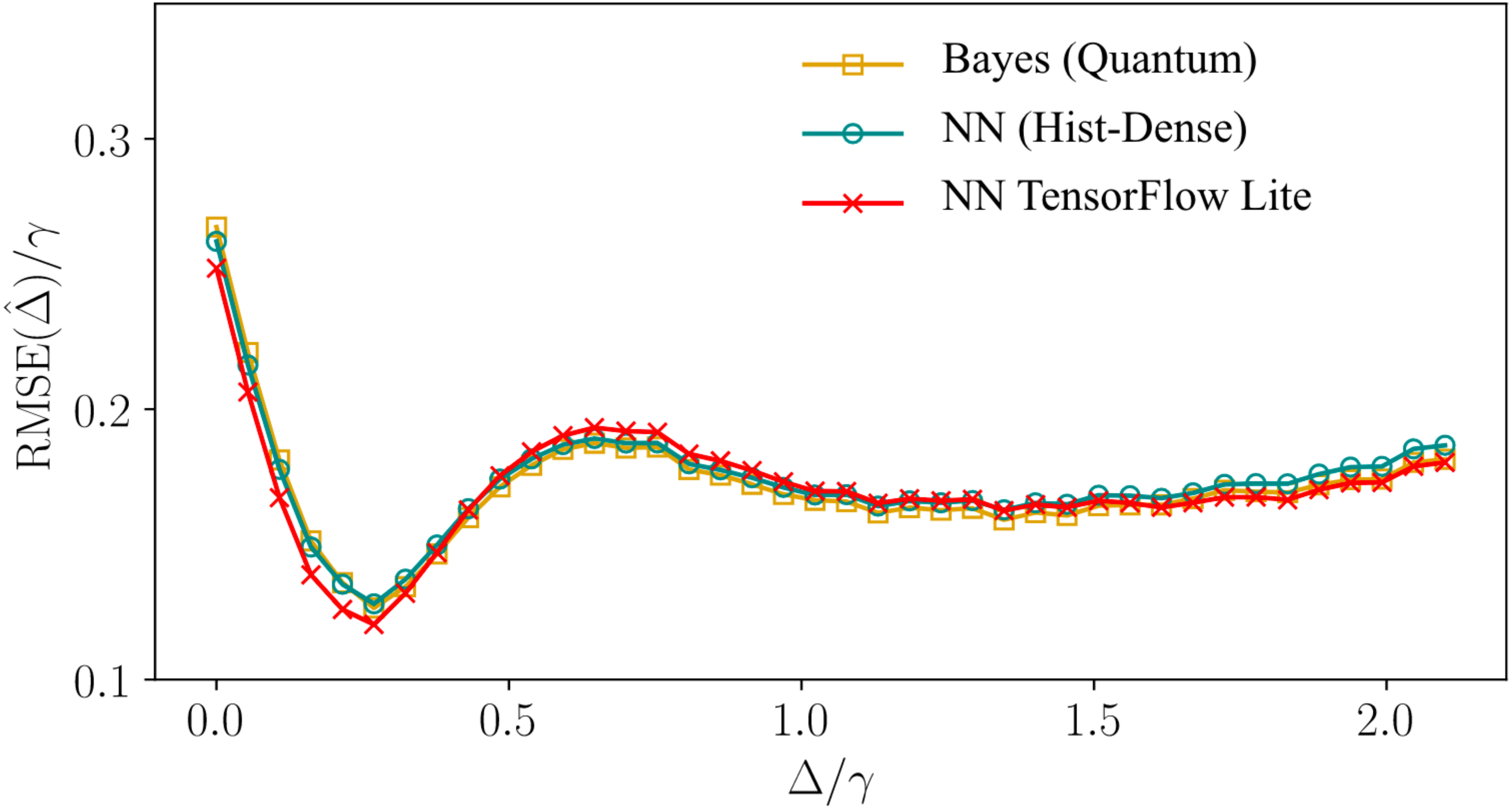
Code on
GitHub

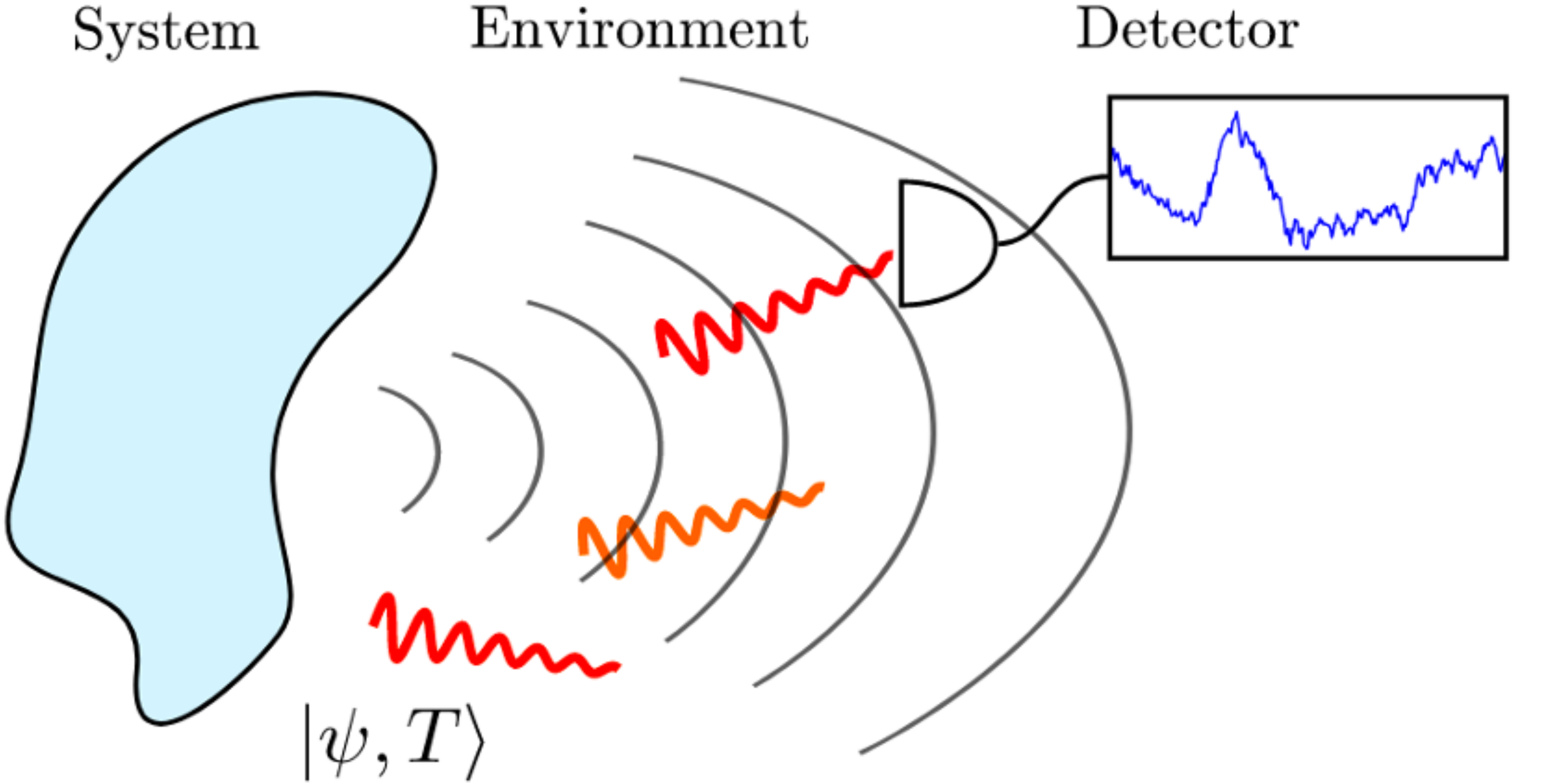
Generalized (biased) Cramer-Rao bounds



Code on [GitHub]

Deployment to edge devices



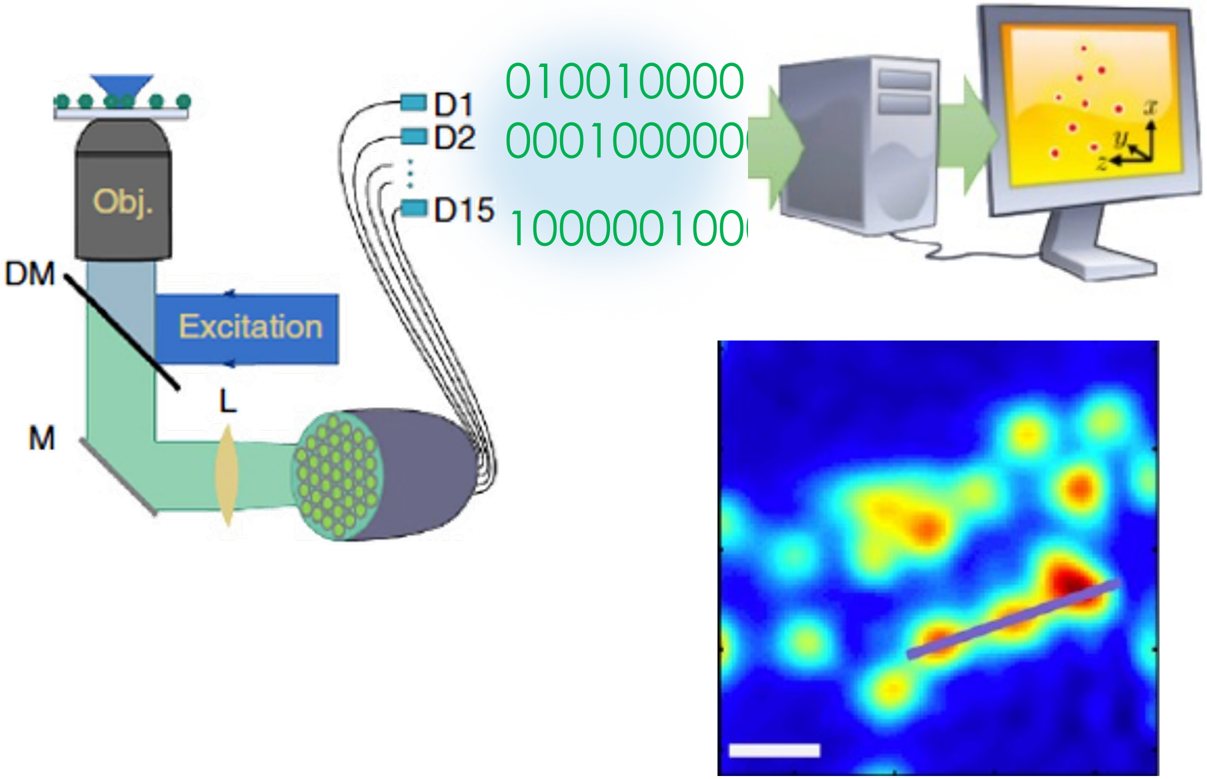


- Metrology with non isolated systems
- Metrology of time-varying signals
- Magnetometry
- Spectroscopy
- Fluorescence microscopy
- Device characterization
- …

Gammelmark, S. & Mølmer, K. Fisher Information and the Quantum Cramér-Rao Sensitivity Limit of Continuous Measurements. *Physical Review Letters* **112**, 170401 (2014)

Slide from Carlos Muñoz

Prospects



Fields such as quantum imaging are already capitalizing on the photonic correlations for quantum metrology.

A more powerful data analysis based on the combination of Bayesian inference and machine learning could boost the resolution in quantum microscopy and spectroscopy.

Schwartz, O. *et al.* Superresolution Microscopy with Quantum Emitters. *Nano Letters* **13**, 5832–5836 (2013).

Israel, Y. *et al.*, Quantum correlation enhanced super-resolution localization microscopy enabled by a fibre bundle camera. *Nature Communications* **8**, 1–5 (2017).

Tenne, R. *et al.* Super-resolution enhancement by quantum image scanning microscopy. *Nature Photonics* **13**, 116–122 (2019).

Lightweight training

Well trained with less than 10k data points (trajectories)