TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY-NC-ND

Citation (APA)
Nielsen, J. A. H., Kicinski, M. J., Arge, T. N., Vijayadharan, K., Foldager, J., Borregaard, J., Meyer, J. J., Neergaard-Nielsen, J. S., Gehring, T., & Andersen, U. L. (2025). Variational quantum algorithm for enhanced continuous variable optical phase sensing. *NPJ Quantum Information*, *11*(1), Article 70. https://doi.org/10.1038/s41534-024-00947-1

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

npj | quantum information

Article

Published in partnership with The University of New South Wales

https://doi.org/10.1038/s41534-024-00947-1

Variational quantum algorithm for enhanced continuous variable optical phase sensing

Check for updates

Jens A. H. Nielsen[1], Mateusz J. Kicinski[1], Tummas N. Arge[1], Kannan Vijayadharan[2], Jonathan Foldager[3], Johannes Borregaard[4,5] ✉, Johannes Jakob Meyer[6], Jonas S. Neergaard-Nielsen[1], Tobias Gehring[1] & Ulrik L. Andersen[1] ✉

Variational quantum algorithms (VQAs) are hybrid quantum-classical approaches used for tackling a wide range of problems on noisy intermediate-scale quantum (NISQ) devices. Testing these algorithms on relevant hardware is crucial to investigate the effect of noise and imperfections and to assess their practical value. Here, we implement a variational algorithm designed for optimized parameter estimation on a continuous variable platform based on squeezed light, a key component for high-precision optical phase estimation. We investigate the ability of the algorithm to identify the optimal metrology process, including the optimization of the probe state and measurement strategy for small-angle optical phase sensing. Two different optimization strategies are employed, the first being a gradient descent optimizer using Gaussian parameter shift rules to estimate the gradient of the cost function directly from the measurements. The second strategy involves a gradient-free Bayesian optimizer, fine-tuning the system using the same cost function and trained on the data acquired through the gradient-dependent algorithm. We find that both algorithms can steer the experiment towards the optimal metrology process. However, they find minima not predicted by our theoretical model, demonstrating the strength of variational algorithms in modelling complex noise environments, a non-trivial task.

Hybrid quantum-classical algorithms[1,2] like the quantum approximate optimization algorithm (QAOA) and the variational quantum eigensolver (VQE) show promise for implementation on early quantum devices that are yet incapable of running quantum error correction. There have been several experimental demonstrations of these algorithms on quantum hardware such as trapped ions[3,4], neutral atoms[5], and superconducting qubits[6–8]. Although these algorithms are suitable for a wide range of optimization problems ranging from electronic structure calculations[9] to general combinatorial optimization problems[10], the demonstration of a practical quantum advantage compared to purely classical techniques remains an outstanding challenge.

Quantum metrology, which leverages inherently quantum effects such as entanglement to surpass classical limits on resolution and sensitivity[11], is a promising area for near-term applications of noisy quantum hardware. Numerous theoretical studies have shown that variational methods can be applied to identify optimal non-classical probe states and measurements in the presence of unknown noise processes and imperfections[12–16]. Recent experimental demonstrations with photonic platforms have also highlighted the feasibility of these techniques for multi-parameter sensing in the few photon limit[17] and supervised learning assisted by continuous variable entangled networks for multi-dimensional data classification[18].

We extend previous investigations to the practically relevant regime of optical phase estimation using a squeezed coherent state and homodyne measurements[19]. This combination has been successfully deployed to enhance gravitational wave-detection[20,21], magnetic field sensing[22], and biological imaging[23] beyond classical limits. We demonstrate that a variational algorithm can effectively steer the experiment towards the optimal probe state and measurement basis in the presence of various imperfections like phase fluctuations and loss. Our approach involves optimizing the classical Fisher information, a key determinant of metrological precision.

[1]Center for Macroscopic Quantum States bigQ, Department of Physics, Technical University of Denmark, Kgs. Lyngby, Denmark. [2]Department of Information Engineering, University of Padova, Padova, Italy. [3]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark. [4]Qutech, Delft University of Technology, Delft, Netherlands. [5]Department of Physics, Harvard University, Cambridge, USA. [6]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin, Germany. ✉e-mail: borregaard@fas.harvard.edu; ulrik.andersen@fysik.dtu.dk

We develop and implement parameter shift rules[24] to calculate and differentiate this quantity in continuous variable systems. This allows us to implement two approaches: one that estimates the gradient of the classical Fisher information through additional measurements for gradient descent optimization, and another that employs a gradient-free Bayesian optimizer to further fine-tune the optimization process.

Our results confirm the potential of variational techniques for quantum sensing tasks and highlight the remaining challenges for their practical application. We find that the additional time overhead incurred by calculating the gradient is balanced by the gradient-based optimizer's better handling of slow drifts in the sensing apparatus. Conversely, the gradient-free optimization method requires more fine-tuning of the exploration/exploitation trade-off, but enables faster convergence. Further investigation of variational methods in more complex setups with intricate parametrizations, and a detailed study of the potentials and limitations of different optimization techniques, will be crucial for the widespread adoption of variational metrology schemes.

Our task is to optimize a metrological protocol, which includes preparing a probe state and a suitable measurement, for estimation of a small phase shift imprinted on a mode of a continuous-variable quantum system. We choose to focus on a local estimation task considering small shifts around a fixed phase as this is of fundamental interest in applied quantum sensing. Without loss of generality, this fixed phase can be assumed to be zero since the phase of the probe can always be adjusted relative to the fixed phase. We have chosen the continuous-variable platform, as these types of systems are widely used and have proven their usefulness in practically relevant scenarios[20,21,23].

The conventional approach to developing metrological protocols typically starts with a theoretical description, followed by protocol development and implementation. This strategy has multiple downsides: A completely faithful model of a system is often hard to devise and when trying to push the system to its limits, every inaccuracy in the theoretical model can negatively impact performance. Furthermore, even with a faithful model, the system often depends on parameters that are not stable over time, causing parameter drift. Therefore, methods that adapt to the actual physical conditions during execution, without requiring a complete theoretical model, are highly desirable.

Variational quantum algorithms for quantum metrology[12–16] have been proposed for this purpose. In these approaches, a cost function is defined to represent the performance of a metrological protocol and is then optimized. This process does not require a full theoretical model, allowing the protocol to implicitly account unmodeled effects.

Our experiment broadly consists of three stages. First, a displaced squeezed state is prepared as a probe state, where the displacement phase angle relative to the squeezed quadrature, ϕ_α, is a free parameter. We fix the squeezing level r and displacement amplitude $|\alpha|$ to work with probes of a fixed photon number. Next, the state undergoes a phase shift that encodes the parameter to be sensed. Finally, we perform homodyne detection, where the homodyne detection angle ϕ_{HD}, relative to the squeezing angle, is another free parameter. Building on the approach of ref. 16, we use the inverse of the classical Fisher information[25] as a cost function $C(\phi_\alpha, \phi_{HD}) = 1/\mathcal{F}(\phi_\alpha, \phi_{HD})$, where the inverse of the Fisher information represents the lowest achievable variance with many repetitions of the experiment, serving as a good proxy for metrological precision. The "methods sections IIIB and IIIC") details the computation of this metric in continuous variable systems using Gaussian parameter-shift rules. As outlined in the introduction, we combine this setup with two different optimizers: a gradient-based one for ab initio optimization and a Bayesian one for fine-tuning. The optimization happens in real time, with the experiment continuously running while measurements are done, cost function is estimated and parameters are optimized and adjusted.

Results

Experimental principle

The principle behind the experiment is illustrated in Fig. 1. We prepare a displaced squeezed state by pumping a hemilithic optical parametric oscillator (OPO) (previously described in ref. 26) at 775 nm. This process generates squeezed vacuum at 1550 nm. The squeezed vacuum is then combined with a coherent state on a 99/1 beamsplitter, resulting in the displacement of the squeezed light at the 5 MHz sideband using a phase modulator in the coherent beam. After interacting with the lab environment, the squeezed light is measured by a homodyne detector. The relative phases between the squeezed light, the local oscillator and the displacement beam are locked using the coherent locking technique[27]. This technique, employing a 40 MHz phase-locked sideband mode transmitted alongside the squeezed light, serves as a phase reference and allows full access to the phase space for both the squeezed light and displacement.

A detailed description of the experimental setup can be found in the methods section fig. 4. For the measurements in this paper, the OPO's pump power is 2.7 mW resulting in the measurement of around 5 dB squeezing and 12 dB anti-squeezing. The displacement added to the squeezing is approximately $\alpha = 5.2$, a regime of interest since the contributions of the squeezed and the coherent photons to the classical Fisher Information are comparable.

The homodyne detector's output is sampled using a data-acquisition card after which it is downmixed to the 5 MHz sideband and subsequently lowpass filtered with bandwidth 1 MHz. The processed data are then used to characterize the measurement statistics. The cost function is estimated by varying the phases of the local oscillator according to the parameter-shift rules (see "methods section IIIB"). Depending on the algorithm employed, the local oscillator and displacement phases are either further shifted during measurements to generate gradients for the gradient descent algorithm, or the cost function is directly fed to the Bayesian optimizer for the gradient-free algorithm. Based on either the gradient of the cost function or the decision function of the Bayesian optimizer, a new set of phase parameters is found. The process is repeated until optimization is achieved.

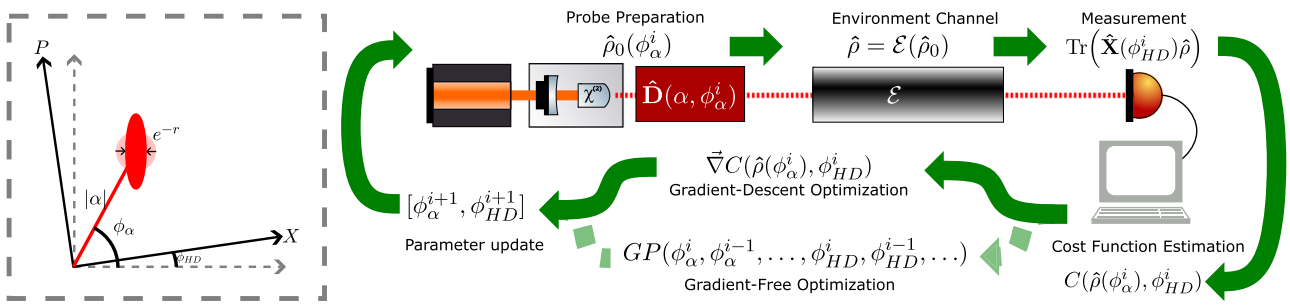


Fig. 1 | Variational experiment principle. A probe state is prepared as a squeezed, displaced state. The free parameters of our system are the measurement basis angle, ϕ_{HD}, and the displacement angle, ϕ_α, both relative to the squeezing angle ϕ_r. After interacting with the environment, the state is detected in a homodyne detector. The cost function is estimated by varying the measurement basis of the detector. Subsequently, the experiment either estimates the gradient of the cost function to determine the next set of initial parameters or employs a Gaussian Process (GP) in a Bayesian gradient-free optimization algorithm for parameter selection.

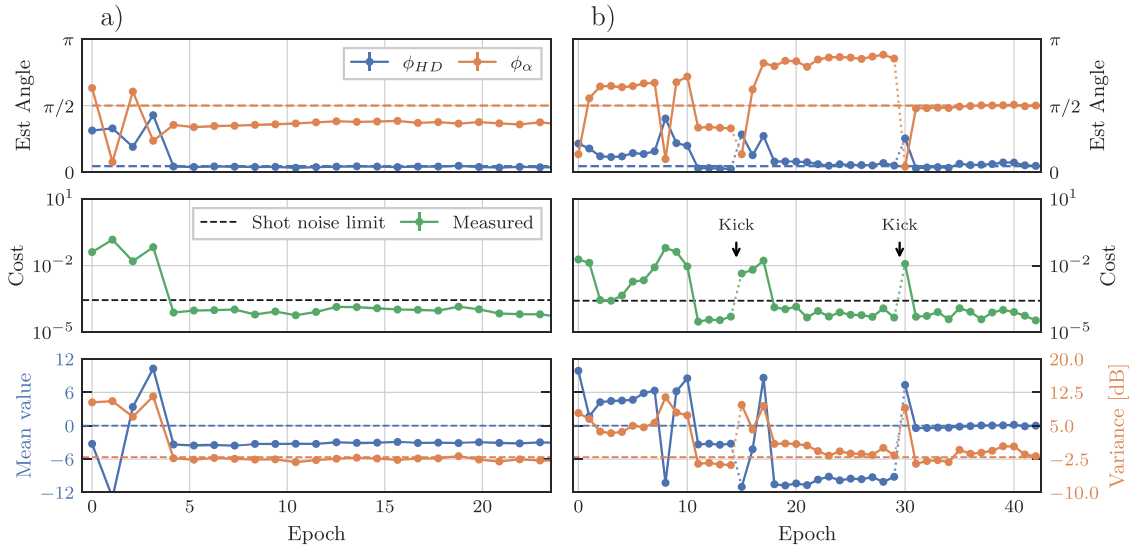


Fig. 2 | Results of the gradient descent experiment. a Demonstration of the gradient descent algorithm across 24 optimization epochs. (Top) The optical phase angles ϕ_{HD} and ϕ_α estimated from the measurements. (Middle) The measured cost function $C = 1/\mathcal{F}$. The dotted line represents the shot noise limit accounting for the number of photons in the measurement and the number of samples used to estimate the cost function. (Bottom) The measured quadrature mean values and variances. The dotted lines in the top and bottom plots indicate the optimal values as predicted by theory (See "section III"). Note that in this particular measurement, the displacement appears to be slightly larger than $\alpha = 5.2$, likely due to improved spatial overlap between the displacement beam and local oscillator, as the added modulation was consistent across all measurements. **b** Kick-test of the gradient descent optimization over 45 epochs. The arrows labeled "kick" indicate the points immediately following the application of a kick.

In the data discussed below, each experimental sequence consists of repeated measurements of the relevant quantities tracked in real time. This allows us to directly observe the fluctuations of the experimental parameters such as the cost function and their stabilization to an optimal setting as a result of the classical feedback cycle.

Gradient descent optimization

In Fig. 2a, we present a run of the gradient descent-based optimization over 24 epochs. There is a clear trend of the cost function starting at a high value and then converging to a low one. The black dotted line of the middle plot of Fig. 2a represents the theoretical shot noise limit, which is based on the average number of photons in the probe state and the number of samples used to estimate each value of the cost function. Notably, the optimized cost function falls below this limit after convergence, indicating that the algorithm successfully identifies an optimum below the classical limit. Examining the resulting quadrature values, we observe the variance dropping below shot noise and approaching a value determined by the basis angle that maximizes the quadrature variance contribution to our analytical model of the classical Fisher Information. The theoretical model is elaborated upon in great detail in sections IIIB and IIIC. In general, it can be described as a squeezed, displaced probe state undergoing optical loss. This can be modelled analytically (see eq. 8).

We observe unexpected local minima during the optimizer runs, evident from the mean value of the measurement quadrature. Ideally, we would expect this mean value to stabilize around 0, but it occasionally stabilizes around an intermediate value (in this case ~4). This is not expected from the analytical model of our experiment (eq. 8). The value of the cost function is roughly the same for this intermediate value and the one predicted by the analytical model indicating that both settings are valid minima of the cost function.

We try to simulate the gradient descent experiment by numerically introducing Gaussian distributed phase noise to approximate the experimental conditions as closely as possible, and while we can observe a similar trend, the simulation does not find the exact same minimum, showing the difficulty in modelling experimental noise environments. The exact cause of this phenomenon is a subject for further investigation, but it does, however, showcase the advantage of variational approaches, as they can naturally uncover minima not predicted by either the analytical or the numerical model of our experiment.

In Fig. 2b, we display the results of a different run of the experiment. To test the algorithm's robustness against disturbances, we allowed the system to optimize for 14 rounds. Then, on the 15th round, we apply a kick to both control parameters, effectively displacing the system from its optimum. Subsequently, the system was permitted to optimize for another 14 epochs, before being subjected to another kick.

Once again, we observe that the algorithm successfully optimizes the system after each kick, consistently reaching below the shot noise limit. While the quadrature variance converges towards the expected value, the mean value tends to stabilize around an intermediate point. Interestingly, in this particular round, during the final optimization step, the mean value does settle around the expected optimum. However, we also note that the difference in cost function between the various minima is very small and essentially indistinguishable in our measurements. This observation suggests that the optimization of the variance significantly impacts the cost function more than the mean value. This hypothesis is further supported by a simulation presented in Fig. 10 of the methods section. This simulation, a 2D simulation of the cost-function landscape with the kick-measurements superimposed, reveals that the minima are quite broad and shallow as a function of the displacement angle, especially once the measurement angle has been optimized.

Post-hoc Bayesian optimization for fine-tuning

In our search for an even better solution than those obtained from the gradient descent approach, we employ a post-hoc gradient-free optimization using data from low-cost areas. We use Bayesian optimization (BO), a probabilistic algorithm often used in machine learning for tuning model hyperparameters[28–30]. This is suitable for the local estimation task considered here. In the case of ab initio phase estimation, adaptive Baysian techniques[31] could potentially be employed to pinpoint approximate values around which local estimation could be performed. Bayesian optimization is particularly suited to our experiment as it utilizes a probabilistic approximation of the underlying model. This approach is beneficial for incorporating uncertainty related to our necessarily inexact modelling of the system and the to random processes occurring in the experiment. Furthermore, the

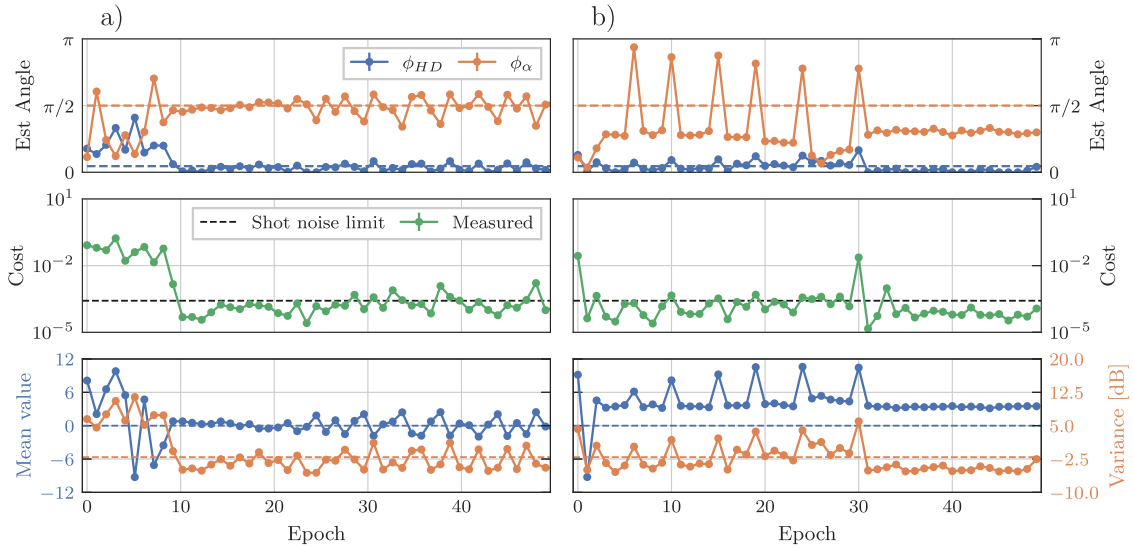


Fig. 3 | Results of the gradient free experiment. Demonstration of Bayesian optimization over 50 epochs, with (**a**) loose and (**b**) optimized hyper-parameters of the Gaussian Process (see "section IIIL"). (Top) The optical phase angles ϕ_{HD} and ϕ_α estimated from the measurements. (Middle) The measured cost function $C = 1/\mathcal{F}$. The dotted line represents the shot noise limit, accounting for the number of photons in the measurement and the number of samples used to estimate the cost function. (Bottom) The measured quadrature mean values and variances. The dotted lines in the top and bottom plots indicate the optimal values as predicted by theory ("section IIIB").

gradient-based algorithm requires five measurements to estimate the cost function and an additional eight measurements per control parameter to estimate the gradient, whereas the gradient-free algorithm only needs the five measurements for the cost function estimation. This results in a significant reduction in experimental resources per optimization step.

A Bayesian optimization routine comprises two parts. First, a surrogate function models the underlying cost function landscape $C(\phi_{HD}, \phi_\alpha)$ probabilistically, indicated by the predictive mean $\mu(C(\phi_{HD}, \phi_\alpha)) \approx C(\phi_{HD}, \phi_\alpha)$ and an uncertainty captured by the covariance matrix $\Sigma(C(\phi_{HD}, \phi_\alpha))$. Second, an acquisition function determines which data points to query during iterative optimization – essentially, it quantitatively identifies a "good" next point for the experiment.

We selected a Gaussian Process (GP)[32] for the surrogate function. For every point (ϕ_{HD}, ϕ_α), the predictive distribution of the surrogate model follows a normal distribution with mean $\mu(C(\phi_{HD}, \phi_\alpha))$ and covariance $\Sigma(C(\phi_{HD}, \phi_\alpha))$. The model's covariance is a key hyperparameter that balances the trade-off between exploration (favoring larger steps in parameter space) and exploitation (focusing on the vicinity of the current optimum). For the acquisition function, we opted for Expected Improvement (EI), a commonly used technique in Bayesian optimization[33]. Intuitively, EI selects the next point that according to the current model is expected to achieve the lowest value of the cost function.

Initially, the Gaussian Process was fitted with a data set of 136 data points obtained from the gradient descent experiment along with randomly sampled points in our search space. This approach is hence referred to as *post-hoc* Bayesian optimization for fine-tuning. The next parameter set for the experiment is determined by maximizing the acquisition function, which relies solely on the surrogate model of the Bayesian Optimizer.

After selecting this new set of parameters, the physical experiment is conducted, and the actual cost value is obtained. This new data point is then incorporated into the dataset, and the updated dataset is used to fit a new Gaussian Process for the subsequent step of the algorithm. This process is repeated iteratively for 50 epochs, and report the lowest obtained cost function value. See the methods section for technical details and Sec. 2.6–2.8 in ref. 34 for a full derivation.

From Fig. 3, we observe once again that the algorithm reaches an optimum of the system close to the expected theoretical minimum. In part a of Fig. 3, we set the hyperparameters of the Gaussian Process more loosely, resulting in the algorithm exploring the parameter landscape around the minimum. This exploration comes at the cost of optimization process's stability, causing the cost function to fluctuate to relatively high values. Optimizing the balance between exploration and exploitation is achievable by tuning the hyper-parameters of the model. The results of this tuning are evident in part b of Fig. 3, where the model initially focuses on exploring new areas of cost function landscape $C(\phi_{HD}, \phi_\alpha)$ for the first 30 epochs. It then shifts to a more stable exploitation of the achieved minimum in the subsequent 20 epochs. Although not depicted in this trace, the gradient-free algorithm was also exhibited susceptibility to the same local minima as those identified using the gradient descent-based algorithm. This adds further evidence that our theoretical model does not fully capture the entire noise landscape of the experiment.

The gradient-free Bayesian optimization method employed here offers two advantages over the gradient descent optimization. First, it reduces the risk of getting trapped in a local minimum, a common problem with gradient descent. Second, as it does not require the evaluation of the gradient, it uses less experimental resources. However, employing Bayesian optimization is not without its drawbacks. As discussed in previous sections, the gradient-based method will be able to follow slow drifts of the system as these drifts alter the gradients. In contrast, the Bayesian Optimizer is likely to be slower in optimizing a non-stationary system due to the need for an exploration phase whenever the experimental setup undergoes changes. Additionally, the performance of the gradient-free algorithm heavily depends on the careful optimization of its hyperparameters.

Discussion

In summary, we have investigated the feasibility of hybrid quantum-classical optimization algorithms for optical phase estimation with squeezed coherent light and homodyne detection. Specifically, we have investigated the performance of both a gradient-based optimization and post-hoc gradient-free Bayesian optimization.

Our results confirm that both algorithms can successfully adjust the control parameters of the system to achieve optimal estimation performance. This includes preparing of the optimal probe state and setting the measurement parameters. Notably, the algorithms achieve this without prior knowledge about the noise processes in the hardware, as evidenced by the discovery of optima that were not anticipated by our theoretical model.

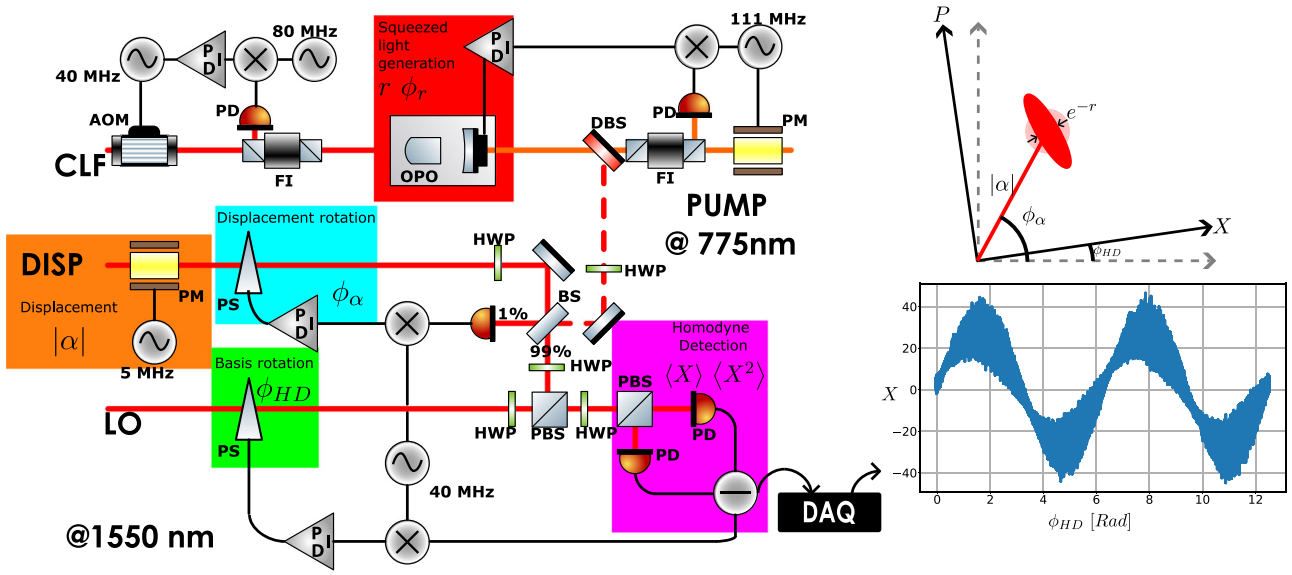


Fig. 4 | Sketch of the experimental setup. CLF Coherent-locking-field, DISP Displacement field, LO Local oscillator, DAQ Data acquisition, OPO Optical-parametric-oscillator, PD Photo-detector, FI Faraday Isolator, BS Beam-splitter, AOM Acousto-optic oscillator, PM Phase modulator, PS Phase shifter, PBS Polarizing beam-splitter, DBS Dichroic beam-splitter, HWP Half-wave plate.

Additionally, we have shown how the gradient-based algorithm can automatically adjust the system to the optimal setting when the phase to be estimated changes.

These findings underscore the potential of variational quantum algorithm-based quantum metrology, particularly for optical phase estimation with squeezed coherent light and homodyne detection.This motivates further investigation of such techniques in more complex quantum sensing systems with additional control parameters, such as the degree of squeezing and the level of coherent excitation. In addition, the application of these variational algorithms in multi-parameter sensing systems[35], which are relevant for quantum imaging[36], entangled sensor networks[37,38], and networked atomic clocks[39], is also a promising avenue. In particular, recent theoretical[12] and experimental work[18] on entanglement assisted supervised learning have shown that multi-mode entangled networks can outperform non-entangled methods for certain data classification tasks. These works considered an ad-hoc cost function and non-gradient based optimization methods. In comparison, we have considered a more general cost function based on the classical Fisher information, which provides a saturable lower bound on the performance of any unbiased estimator. While we have focused on phase estimation in this work, the application of similar techniques, including parameter shift rules for gradient estimation, to data-classification problems as studied in refs. 12,18 could further elucidate the promise of quantum enhanced sensing.

Our study also reveals that the choice of classical optimizers may vary depending on the specific estimation task. We observe that the gradient-based optimizer is effective for ab initio phase estimation but incurs a higher measurement overhead compared to the gradient-free Bayesian optimization. Conversely, our implementation of the Bayesian optimization was adept at fine-tuning control parameters, though it required a training set from the gradient descent optimization for optimal performance. We also noted that careful optimization of the hyper-parameters of the algorithms such as the learning rate and ratio of exploration/exploitation is crucial for good performance. Further investigations in this area, including the potential benefit of switching between different classical optimizers for practical phase estimation tasks, will be vital to further validate the practical applicability of variational techniques in quantum-enhanced metrology.

Methods

Experimental system

The experimental system is shown in figure Fig. 4. A hemilithic, double resonant, optical-parametrtic-oscillator (OPO) is pumped with a light field at 775 nm, and the squeezed light is generated at 1550 nm. The squeezed light source, which is a modification of the source presented in ref. 40, has a FWHM bandwidth of 66 MHz and a threshold power of around 6 mW. The source is pumped with 2.7 mW, and the system has around 72% efficiency and 30 mrad phase noise RMS (between squeezed light and local oscillator), measuring around 5 dB of squeezed light and 11.8 dB of anti-squeezed light at a sideband frequency of 5 MHz. From these a squeezing strength $r \sim 1.52$ can be estimated.

The OPO is stabilized via the Pound-Drever-Hall (PDH) technique[41], and a 40 MHz frequency-shifted beam is injected into the OPO to act as a phase reference in a coherent locking scheme[27]. This 40 MHz reference tone (CLF) is used to stabilize the phase between the squeezed light, the displacement beam (DISP) and the local oscillator (LO). Tuning the electrical down-mixing phase of the two phase locks allows full control of the displacement angle ϕ_α and the homodyne basis angle ϕ_{HD} relative to the squeezed quadrature angle ϕ_r, which is arbitrarily set by the phase of the pump light.

Classical Fisher information of a Gaussian state and the cost function

The cost function used in the experiment is the inverse of the classical Fisher Information

$$C = \frac{1}{\mathcal{F}} \tag{1}$$

We assume our system to be a CW Gaussian state described by quadrature operators $\hat{\mathbf{X}} \equiv (\hat{\mathbf{a}}^\dagger + \hat{\mathbf{a}})$ and $\hat{\mathbf{P}} \equiv i(\hat{\mathbf{a}}^\dagger - \hat{\mathbf{a}})$ and fully characterized by the first two statistical moments $\mu_x = \langle \hat{\mathbf{X}} \rangle$ and $V_x = \langle \hat{\mathbf{X}}^2 \rangle - \langle \hat{\mathbf{X}} \rangle^2$ (with similar moments for $\hat{\mathbf{P}}$) and with a photon number operator given by

$$\hat{\mathbf{n}} = \frac{1}{4}(\hat{\mathbf{X}}^2 + \hat{\mathbf{P}}^2 - 2\mathbb{I}), \tag{2}$$

using the commutator $[\hat{\mathbf{X}}, \hat{\mathbf{P}}] = 2i$.

The classical Fisher Information for the estimation of phase shifts ϕ, with a measurement probability distribution of $\hat{\mathbf{X}}$ quadrature values $P(x)$ parameterized by parameters $\{\Theta\}$ is given by

$$\mathcal{F} = \int \mathrm{d}x P(x|\phi,\{\Theta\}) \left(\frac{\partial \log(P(x|\phi,\{\Theta\}))}{\partial \phi} \right)^2. \tag{3}$$

Equation (3) can be evaluated, since we have a Gaussian state, as

$$\mathcal{F} = \frac{1}{V_x} \left(\frac{\partial \mu_x}{\partial \phi} \right)^2 + \frac{1}{2V_x^2} \left(\frac{\partial V_x}{\partial \phi} \right)^2, \tag{4}$$

where μ_x and V_x will be functions of the parameters $\{\Theta\}$.

The gradient of the cost function with respect to the parameters Θ is given by

$$\nabla_\Theta C = \frac{-\nabla_\Theta \mathcal{F}}{\mathcal{F}^2}. \tag{5}$$

The gradient of the classical Fisher Information is finally given by

$$\begin{aligned} \nabla_\Theta \mathcal{F} = & \frac{1}{V_x^3} \left(2V_x^2 \frac{\partial \mu_x}{\partial \phi} \nabla_\Theta \frac{\partial \mu_x}{\partial \phi} - V_x \left(\frac{\partial \mu_x}{\partial \phi} \right)^2 \nabla_\Theta V_x \right) \\ & + \frac{1}{V_x^3} \left(V_x \frac{\partial V_x}{\partial \phi} \nabla_\Theta \frac{\partial V_x}{\partial \phi} - \left(\frac{\partial V_x}{\partial \phi} \right)^2 \nabla_\Theta V_x \right). \end{aligned} \tag{6}$$

The derivative of the variance is in general given by

$$\frac{\partial V_x}{\partial \phi} = \frac{\partial}{\partial \phi} \langle \hat{\mathbf{X}}^2 \rangle - 2 \langle \hat{\mathbf{X}} \rangle \frac{\partial}{\partial \phi} \langle \hat{\mathbf{X}} \rangle. \tag{7}$$

The two main contributions to the loss of information are the loss of squeezed photons and phase noise between the three interacting fields (displacement- and squeezing fields and local oscillator). The loss of photons comes from the limited escape efficiency of the squeezer η_{esc}, the limited efficiency of optical components between the squeezer and the detector η_{opt}, the imperfect visibility between signal and local oscillator $\mathcal{V}^2$ and finally the imperfect quantum efficiency of the photodiodes η_{QE}. Phase noise comes mainly from the inability of the the phase-stabilization loops to remove all classical phase fluctuations due to having limited bandwidths and shot-noise of the light fields. In the case of pure loss, ignoring phase noise, the Classical Fisher Information can be evaluated analytically by starting from eq. (4)

$$\begin{aligned} \mathcal{F}(\phi) &= \frac{1}{V_x} \left(\frac{\partial \mu_x}{\partial \phi} \right)^2 + \frac{1}{2V_x^2} \left(\frac{\partial V_x}{\partial \phi} \right)^2 \\ &= \frac{1}{\eta\left(e^{-2r}\cos^2(\phi) + e^{2r}\sin^2(\phi)\right) + 1 - \eta} \left(\frac{\partial}{\partial \phi} 2\sqrt{\eta}|\alpha| \cos(\phi_\alpha - \phi) \right)^2 \\ &+ \frac{1}{\left(\eta\left(e^{-2r}\cos^2(\phi) + e^{2r}\sin^2(\phi)\right) + 1 - \eta\right)^2} \left(\frac{\partial}{\partial \phi} \left[\eta\left(e^{-2r}\cos^2(\phi) + e^{2r}\sin^2(\phi)\right) + 1 - \eta \right] \right)^2 \\ &= \frac{4\eta|\alpha|^2 \sin^2(\phi_\alpha - \phi)}{\eta\left(e^{-2r}\cos^2(\phi) + e^{2r}\sin^2(\phi)\right) + 1 - \eta} + \frac{2\eta^2 \sinh^2(2r)\sin^2(2\phi)}{\left(\eta\left(e^{-2r}\cos^2(\phi) + e^{2r}\sin^2(\phi)\right) + 1 - \eta\right)^2}. \end{aligned} \tag{8}$$

In the zero loss scenario, the variance term will always scale faster with increasing photon numbers, meaning that concentrating the photons in the squeezed state is the more effective strategy. In this case the optimal measurement phase is given by $\phi_{opt} = \arccos(\tanh(2r))/2$.

In the case of relatively low loss, there exists an optimal ratio between squeezed and coherent photons with an optimal relative angle being $\phi_\alpha - \phi = \pi/2$ and an optimal measurement angle given by

$$\phi_{opt}^{\mathcal{L}} = \frac{1}{2} \arccos \left(\frac{\frac{\eta e^{2r} + (1-\eta)}{\eta e^{-2r} + (1-\eta)} - 1}{\frac{\eta e^{2r} + (1-\eta)}{\eta e^{-2r} + (1-\eta)} + 1} \right). \tag{9}$$

In the high loss scenario, the mean value term will dominate meaning most photons should be put into the coherent state, with the optimal relative angle being $\phi_\alpha - \phi = \pi/2$ and the optimal measurement angle being $\phi = 0$.

Parameter-shift rules for quadratic operators

In general, an arbitrary operator can be expressed as

$$\hat{\mathbf{A}} = (a, b, c, d, e \ldots) \cdot \begin{pmatrix} \mathbb{I} \\ \hat{\mathbf{X}} \\ \hat{\mathbf{P}} \\ \hat{\mathbf{X}}^2 \\ \hat{\mathbf{P}}^2 \\ \vdots \end{pmatrix} \tag{10}$$

When a gate acts upon this arbitrary operator $\hat{\mathbf{A}}$, we in principle need to know the infinite-dimensional gate matrix describing the transformation of all entries of the vector $\hat{\mathbf{A}}_G = G[\hat{\mathbf{A}}] = M_G^T \hat{\mathbf{A}}$, where M_G is a matrix describing the action of the gate upon the operator. Please note the transposition of the matrix - this is a very subtle and important detail. The derivatives of operators linear in quadratures (e.g. $\nabla_\Theta \langle \hat{\mathbf{X}} \rangle$) can easily be evaluated using the parameter shift rules from[24] to find the higher-order entries of the gate matrix or truncate the vector space to only include the linear operators. It is also possible to derive parameter shift rules for operators quadratic in the quadratures, by finding and differentiating the higher order entries of the gate matrices using

$$G[\hat{\mathbf{A}}\hat{\mathbf{B}}] = G[\hat{\mathbf{A}}] G[\hat{\mathbf{B}}] \tag{11}$$

$$\frac{\partial}{\partial \Theta_i} G(\Theta_i) \left[\hat{\mathbf{A}}^2 \right] = \frac{\partial}{\partial \Theta_i} G(\Theta_i) [\hat{\mathbf{A}}] G(\Theta_i) [\hat{\mathbf{A}}] + G(\Theta_i) [\hat{\mathbf{A}}] \frac{\partial}{\partial \Theta_i} G(\Theta_i) [\hat{\mathbf{A}}], \tag{12}$$

where $G(\Theta_i)$ is a gate parameterized by the parameter Θ_i, and where eq. (12) has been truncated to only include linear quadrature operators $\hat{\mathbf{A}} \subseteq [\mathbb{I}, \hat{\mathbf{X}}, \hat{\mathbf{P}}]$.

In the following, we will reproduce the results of[24], finding the linear parameter shift rules for the relevant Gaussian gates. We will then extend this to also include quadratic operators. Throughout this we will adopt the definition of the Gaussian gates used in[24].

Squeezing gate for linear operators

The squeezing gate is parameterized by r the squeezing strength and the gate matrix for linear operators is given by

$$M_S(r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-r} & 0 \\ 0 & 0 & e^{r} \end{pmatrix}, \tag{13}$$

$$\frac{\partial}{\partial r} M_S(r) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -e^{-r} & 0 \\ 0 & 0 & e^{r} \end{pmatrix}. \tag{14}$$

We now want to express the derivate of the matrix as a linear superposition of the gate matrix itself

$$\begin{aligned}\frac{\partial}{\partial r}M_S(r) &= \begin{pmatrix}0 & 0 & 0\\ 0 & -e^{-r}\frac{e^s-e^{-s}}{2\sinh(s)} & 0\\ 0 & 0 & e^{r}\frac{e^s-e^{-s}}{2\sinh(s)}\end{pmatrix}\\ &= \frac{1}{2\sinh(s)}\begin{pmatrix}0 & 0 & 0\\ 0 & e^{-(r+s)}-e^{-(r-s)} & 0\\ 0 & 0 & e^{(r+s)}-e^{(r-s)}\end{pmatrix}\\ &= \frac{1}{2\sinh(s)}\big(M_S(r+s)-M_S(r-s)\big),\end{aligned} \tag{15}$$

which is the parameter shift rule for the squeezing gate, with s being an arbitrary shift in the squeezing strength.

Displacement gate for linear operators

The displacement gate is parameterized by α the displacement amplitude and ϕ_α the displacement angle and given by the matrix

$$M_D(\alpha,\phi_\alpha) = \begin{pmatrix}1 & 0 & 0\\ 2\alpha\cos(\phi_\alpha) & 1 & 0\\ 2\alpha\sin(\phi_\alpha) & 0 & 1\end{pmatrix}, \tag{16}$$

$$\frac{\partial}{\partial\alpha}M_D(\alpha,\phi_\alpha) = \begin{pmatrix}0 & 0 & 0\\ 2\cos(\phi_\alpha) & 0 & 0\\ 2\sin(\phi_\alpha) & 0 & 0\end{pmatrix}, \tag{17}$$

$$\frac{\partial}{\partial\phi_\alpha}M_D(\alpha,\phi_\alpha) = \begin{pmatrix}0 & 0 & 0\\ -2\alpha\sin(\phi_\alpha) & 0 & 0\\ 2\alpha\cos(\phi_\alpha) & 0 & 0\end{pmatrix}. \tag{18}$$

We can once again easily calculate the parameter shift rule for the linear operators

$$\begin{aligned}\tfrac{\partial}{\partial\alpha}M_D(\alpha,\phi_\alpha) &= \begin{pmatrix}0 & 0 & 0\\ 2\cos(\phi_\alpha)\frac{\alpha+s-(\alpha-s)}{2s} & 0 & 0\\ 2\sin(\phi_\alpha)\frac{\alpha+s-(\alpha-s)}{2s} & 0 & 0\end{pmatrix}\\ &= \frac{1}{2s}\big(M_D(\alpha+s,\phi_\alpha)-M_D(\alpha-s,\phi_\alpha)\big),\end{aligned} \tag{19}$$

$$\begin{aligned}\tfrac{\partial}{\partial\phi_\alpha}M_D(\alpha,\phi_\alpha) &= \begin{pmatrix}0 & 0 & 0\\ 2\alpha\frac{1}{2}\big(\cos(\phi_\alpha+\pi/2)-\cos(\phi_\alpha-\pi/2)\big) & 0 & 0\\ 2\alpha\frac{1}{2}\big(\sin(\phi_\alpha+\pi/2)-\sin(\phi_\alpha-\pi/2)\big) & 0 & 0\end{pmatrix}\\ &= \frac{1}{2}\big(M_D(\alpha,\phi_\alpha+\pi/2)-M_D(\alpha,\phi_\alpha-\pi/2)\big).\end{aligned} \tag{20}$$

Rotation gate for linear operators

The final gate of this analysis is the rotation gate $R(\phi)$ given by the matrix

$$M_R(\phi) = \begin{pmatrix}1 & 0 & 0\\ 0 & \cos(\phi) & -\sin(\phi)\\ 0 & \sin(\phi) & \cos(\phi)\end{pmatrix}, \tag{21}$$

$$\frac{\partial}{\partial\phi}M_R(\phi) = \begin{pmatrix}0 & 0 & 0\\ 0 & -\sin(\phi) & -\cos(\phi)\\ 0 & \cos(\phi) & -\sin(\phi)\end{pmatrix}, \tag{22}$$

From the derivative matrix we can (similar to the displacement gate) find the basic parameter shift rules by recognizing the linear difference of cosines shifted by $\pi/2$ gives sine and vice versa, and we arrive at the same parameter shift rules as with the displacement angle

$$\frac{\partial}{\partial\phi}M_R(\phi) = \frac{1}{2}\big(M_R(\phi+\pi/2)-M_R(\phi-\pi/2)\big). \tag{23}$$

Squeezing gate for quadratic operators

We begin by applying eq. (11) to the squeezing gate to find the higher-order entries of the gate matrix

$$S[\hat{\mathbf{X}}\hat{\mathbf{X}}] = S[\hat{\mathbf{X}}]S[\hat{\mathbf{X}}] = e^{-2r}\hat{\mathbf{X}}^2, \tag{24}$$

$$S[\hat{\mathbf{P}}\hat{\mathbf{P}}] = S[\hat{\mathbf{P}}]S[\hat{\mathbf{P}}] = e^{2r}\hat{\mathbf{P}}^2, \tag{25}$$

$$S[\hat{\mathbf{X}}\hat{\mathbf{P}}] = S[\hat{\mathbf{X}}]S[\hat{\mathbf{P}}] = \hat{\mathbf{X}}\hat{\mathbf{P}}, \tag{26}$$

$$S[\hat{\mathbf{P}}\hat{\mathbf{X}}] = S[\hat{\mathbf{P}}]S[\hat{\mathbf{X}}] = \hat{\mathbf{P}}\hat{\mathbf{X}}. \tag{27}$$

The gate matrix is then expanded to include

$$M_S(r) = \begin{pmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & e^{-r} & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & e^{r} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & e^{-2r} & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & e^{2r} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}. \tag{28}$$

If we truncate the matrix to only look at the quadratic terms, then we can derive parameters shift rules that apply to the quadratic operators

$$\begin{aligned}\tfrac{\partial}{\partial r}M_{S,quad}(r) &= \begin{pmatrix}-2re^{-2r} & 0 & 0 & 0\\ 0 & 2re^{2r} & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\\ &= \tfrac{1}{\sinh(2s)}\Big(M_{S,quad}(r+s)-M_{S,quad}(r-s)\Big)\end{aligned} \tag{29}$$

Displacement gate for quadratic operators

Once again we can repeat the calculation from before, finding the entries of the gate matrix for the quadratic operators

$$\begin{aligned}D(\alpha,\phi_\alpha)[\hat{\mathbf{X}}\hat{\mathbf{X}}] &= (2\alpha\cos(\phi_\alpha)+\hat{\mathbf{X}})(2\alpha\cos(\phi_\alpha)+\hat{\mathbf{X}}) = 4\alpha^2\cos^2(\phi_\alpha)\\ &\quad+\hat{\mathbf{X}}^2+4\alpha\cos(\phi_\alpha)\hat{\mathbf{X}},\end{aligned} \tag{30}$$

$$\begin{aligned}D(\alpha,\phi_\alpha)[\hat{\mathbf{P}}\hat{\mathbf{P}}] &= (2\alpha\sin(\phi_\alpha)+\hat{\mathbf{P}})(2\alpha\sin(\phi_\alpha)+\hat{\mathbf{P}}) = 4\alpha^2\sin^2(\phi_\alpha)\\ &\quad+\hat{\mathbf{P}}^2+4\alpha\sin(\phi_\alpha)\hat{\mathbf{P}},\end{aligned} \tag{31}$$

$$\begin{aligned}D(\alpha,\phi_\alpha)[\hat{\mathbf{X}}\hat{\mathbf{P}}] &= (2\alpha\cos(\phi_\alpha)+\hat{\mathbf{X}})(2\alpha\sin(\phi_\alpha)+\hat{\mathbf{P}}) = 4\alpha^2\sin(2\phi_\alpha)\\ &\quad+\hat{\mathbf{X}}\hat{\mathbf{P}}+2\alpha\big(\cos(\phi_\alpha)\hat{\mathbf{P}}+\sin(\phi_\alpha)\hat{\mathbf{X}}\big),\end{aligned} \tag{32}$$

$$D(\alpha,\phi_\alpha)[\hat{\mathbf{P}}\hat{\mathbf{X}}] = (2\alpha\sin(\phi_\alpha)+\hat{\mathbf{P}})(2\alpha\cos(\phi_\alpha)+\hat{\mathbf{X}}) = 4\alpha^2\sin(2\phi_\alpha) + \hat{\mathbf{P}}\hat{\mathbf{X}} + 2\alpha\big(\cos(\phi_\alpha)\hat{\mathbf{P}} + \sin(\phi_\alpha)\hat{\mathbf{X}}\big). \tag{33}$$

The resulting gate matrix including quadratic operators is then

$$M_D(\alpha,\phi_\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\alpha\cos(\phi_\alpha) & 1 & 0 & 0 & 0 & 0 & 0 \\ 2\alpha\sin(\phi_\alpha) & 0 & 1 & 0 & 0 & 0 & 0 \\ 4\alpha^2\cos^2(\phi_\alpha) & 4\alpha\cos(\phi_\alpha) & 0 & 1 & 0 & 0 & 0 \\ 4\alpha^2\sin^2(\phi_\alpha) & 0 & 4\alpha\sin(\phi_\alpha) & 0 & 1 & 0 & 0 \\ 4\alpha^2\sin(2\phi_\alpha) & 2\alpha\sin(\phi_\alpha) & 2\alpha\cos(\phi_\alpha) & 0 & 0 & 1 & 0 \\ 4\alpha^2\sin(2\phi_\alpha) & 2\alpha\sin(\phi_\alpha) & 2\alpha\cos(\phi_\alpha) & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{34}$$

Rotation Gate for quadratic operators

We begin again by finding the quadratic entries of the rotation matrix

$$R(\phi)[\hat{\mathbf{X}}\hat{\mathbf{X}}] = \cos^2(\phi)\hat{\mathbf{X}}^2 + \sin^2(\phi)\hat{\mathbf{P}}^2 - \cos(\phi)\sin(\phi)\big(\hat{\mathbf{X}}\hat{\mathbf{P}} + \hat{\mathbf{P}}\hat{\mathbf{X}}\big), \tag{38}$$

$$R(\phi)[\hat{\mathbf{P}}\hat{\mathbf{P}}] = \cos^2(\phi)\hat{\mathbf{P}}^2 + \sin^2(\phi)\hat{\mathbf{X}}^2 + \cos(\phi)\sin(\phi)\big(\hat{\mathbf{X}}\hat{\mathbf{P}} + \hat{\mathbf{P}}\hat{\mathbf{X}}\big), \tag{39}$$

$$R(\phi)[\hat{\mathbf{X}}\hat{\mathbf{P}}] = \cos^2(\phi)\hat{\mathbf{X}}\hat{\mathbf{P}} - \sin^2(\phi)\hat{\mathbf{P}}\hat{\mathbf{X}} + \cos(\phi)\sin(\phi)\Big(\hat{\mathbf{X}}^2 - \hat{\mathbf{P}}^2\Big), \tag{40}$$

$$R(\phi)[\hat{\mathbf{P}}\hat{\mathbf{X}}] = \cos^2(\phi)\hat{\mathbf{P}}\hat{\mathbf{X}} - \sin^2(\phi)\hat{\mathbf{P}}\hat{\mathbf{X}} + \cos(\phi)\sin(\phi)\Big(\hat{\mathbf{X}}^2 - \hat{\mathbf{P}}^2\Big). \tag{41}$$

The gate matrix for rotations including quadratic operators is then given by

$$M_{R,quad}(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) & 0 & 0 & 0 & 0 \\ 0 & -\sin(\phi) & \cos(\phi) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos^2(\phi) & \sin^2(\phi) & -\cos(\phi)\sin(\phi) & -\cos(\phi)\sin(\phi) \\ 0 & 0 & 0 & \sin^2(\phi) & \cos^2(\phi) & \cos(\phi)\sin(\phi) & \cos(\phi)\sin(\phi) \\ 0 & 0 & 0 & \cos(\phi)\sin(\phi) & -\cos(\phi)\sin(\phi) & \cos^2(\phi) & \sin^2(\phi) \\ 0 & 0 & 0 & \cos(\phi)\sin(\phi) & -\cos(\phi)\sin(\phi) & \sin^2(\phi) & \cos^2(\phi) \end{pmatrix}. \tag{42}$$

Once again, we limit ourselves to only quadratic operators and find the corresponding parameter shift rules. This is a bit more involved than with the squeezing gate, but if we consider the first column with terms proportional to α^2. For the displacement amplitude α, if we assume naively that the form of the parameter shift is the same as the linear one but with a different normalization, then differentiating gives us the following equation

$$\begin{aligned} \tfrac{\partial 4\alpha^2\cos^2(\phi_\alpha)}{\partial\alpha} &= 8\alpha\cos^2(\phi_\alpha) = \frac{4}{k}\big((\alpha+s)^2-(\alpha-s)^2\big)\cos^2(\phi_\alpha) \Rightarrow \\ 2\alpha &= \frac{1}{k}\big((\alpha+s)^2-(\alpha-s)^2\big) \Rightarrow \\ k &= 2s, \end{aligned} \tag{35}$$

which leads to the same parameter shift rule as with the linear operators

$$\frac{\partial}{\partial\alpha}M_{D,quad}(\alpha,\phi_\alpha) = \frac{1}{2s}\Big(M_{D,quad}(\alpha+s,\phi_\alpha) - M_{D,quad}(\alpha-s,\phi_\alpha)\Big). \tag{36}$$

The calculation for the displacement angle ϕ_α can be calculated by considering transformations of $\cos^2(\phi)\to-\sin(2\phi)$, $\sin^2(\phi)\to\sin(2\phi)$ and $\cos(\phi)\sin(\phi)\to\cos(2\phi)$. These results can be expressed by linear differences of the original functions shifted up and down by $\pi/4$ similar to the basic rotation gate parameter shift. The resulting parameter shift is then

$$\begin{aligned} \frac{\partial}{\partial\phi_\alpha}M_{D,quad}(\alpha,\phi_\alpha) &= \Big(M_{D,quad}(\alpha,\phi_\alpha+\pi/4) - M_{D,quad}(\alpha,\phi_\alpha-\pi/4)\Big) \\ &+ \Big(1-\tfrac{\sqrt{2}}{2}\Big)\Big(M_{D,quad}(\alpha,\phi_\alpha-\pi/2) - M_{D,quad}(\alpha,\phi_\alpha-\pi/2)\Big). \end{aligned} \tag{37}$$

The above expressions can be verified by looking at the derivatives of the number operator expectation value $\partial_\alpha\langle\hat{\mathbf{n}}\rangle = 2\alpha$ and $\partial_{\phi\alpha}\langle\hat{\mathbf{n}}\rangle = 0$, as we would expect.

Once again differentiating $M_{R,quad}(\phi)$ is calculated using the same considerations as with ϕ_α, and we find it to be

$$\frac{\partial}{\partial\phi}M_{R,quad}(\phi) = \Big(M_{R,quad}(\phi+\pi/4) - M_{R,quad}(\phi-\pi/4)\Big). \tag{43}$$

Calibration of displacement and homodyne angles

Since the parameter shift rules assume a certain shift of experimental parameters, it is necessary to calibrate the experimental apparatus to make sure the correct operations are implemented. This section will deal with the calibrations of the displacement angle and the homodyne angle.

In the coherent locking scheme that stabilizes the phase between the squeezed light and the displacement (and the squeezed light and the local oscillator), the changing of the phase of the 40 MHz electrical local oscillator that downmixes the error signal of the feedback loop rotates the angle ϕ_α (ϕ_{HD}). The correspondence between the set angle of the ELO and the actual quadrature angle is not linear, and thus needs to be calibrated. In order to do this, the phase of the displacement (homodyne) ELO function generator is swept through 2π over 2 seconds, while the homodyne output is recorded at 50 MSa s^{-1} and downmixed to the 5 MHz sideband using a 1 MHz lowpass filter. The data is normalized to the shot noise standard deviation. For the calibration of the displacement angle, a 5 MHz displacement is added and the homodyne angle is manually set to the squeezed quadrature. For the homodyne angle calibration, the displacement is removed leaving only vacuum squeezing. The raw data of these measurements are shown in Fig. 5.

Using the data we can estimate the quadrature phase. For the measurement of the homodyne phase using squeezed vacuum, the method used in[40] can be directly applied, however for the displacement phase

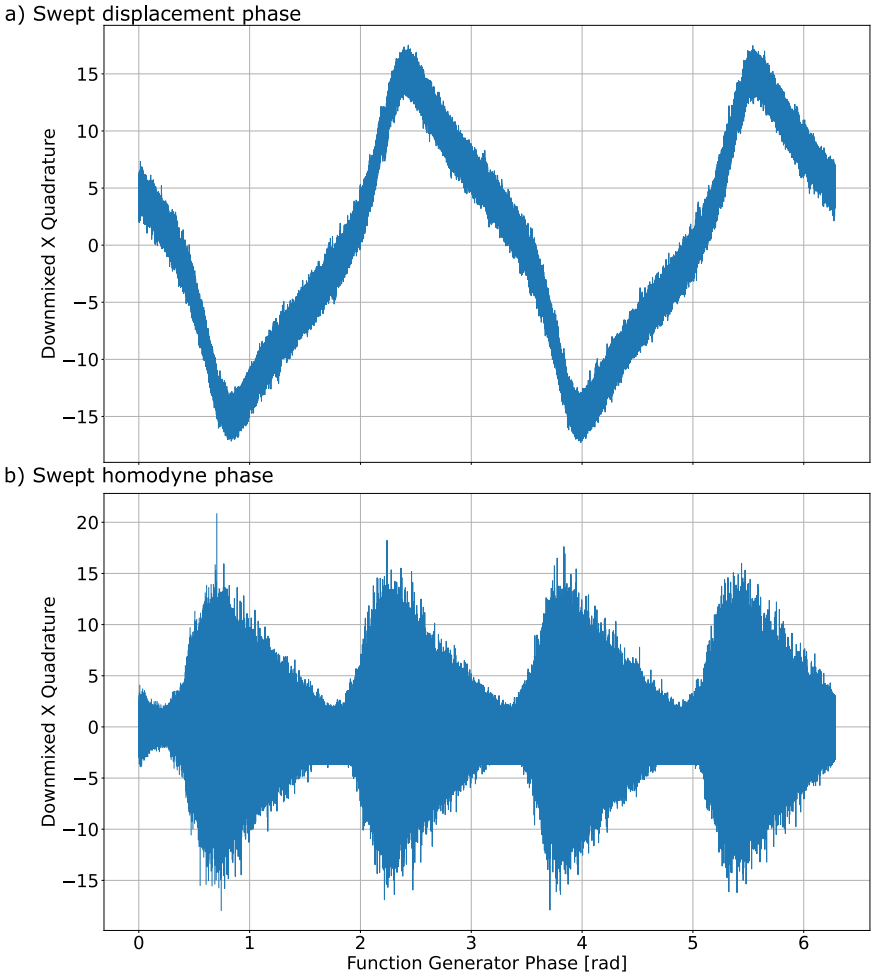


Fig. 5 | Swept homodyne outputs, no calibration. X quadrature measurements normalized to shot noise as a function of the swept ELO phase for (**a**) the displacement phase lock with the homodyne angle locked to squeezing and (**b**) the homodyne phase lock without displacement.

measurement, the marginal probability distribution has to be modified as

$$P(X|\phi_\alpha) = \frac{1}{\sqrt{2\pi V_-}} e^{-\frac{\left(X - 2|\alpha|\cos(\phi_\alpha)\right)^2}{2V_-}}. \tag{44}$$

The estimated phases are unwrapped, and the function generator phases of the displacement (homodyne) ELOs are interpolated as a function of the estimated phases using a 3-order B-spline with a smoothing parameter of 1.3. The estimated phases and corresponding spline-representations are shown in Fig. 6.

Using these B-spline interpolations, we can calibrate the system angles. Although the oscillatory behaviour of these calibrations generally result in the set angles to correspond to the desired angles, since the squeezing angle by itself is ill-defined, this can lead to the failure of these calibrations. This can result in a slight error between the set angle and the desired angle.

These calibrations can be verified by plotting the data in Fig. 5 as a function of the unwrapped estimated phases instead of the function generator phase as shown in Fig. 7.

Simulation model of the experiment

The experimental setup was modelled analytically using Mathematica and the numerical simulation of the gradient descent-based variational quantum algorithm was done using the PennyLane[42] Python library. Considering it's a continuous-variable system, a Gaussian quantum simulator was utilized as the backend.

A single-mode displaced squeezed state is prepared, characterized by a fixed degree of squeezing and displacement magnitude. The displacement angle, ϕ_α, and the homodyne detection angle, ϕ_{HD}, are taken as the free parameters. The outputs from the quantum system are the mean and variance of the probed quadrature.

The initial probe state is thus,

$$\hat{\rho_0}(\phi_\alpha) = \hat{D}(\alpha, \phi_\alpha)\hat{S}(r, \phi_r)|0\rangle \tag{45}$$

where $\hat{D}$ and $\hat{S}$ represent the displacement and quadrature squeezing operators respectively. r is the degree of squeezing and α the magnitude of displacement, corresponding to the experiment.

To simulate the environmental interaction on the probe state, the photon loss and thermal noise in the system are modelled by coupling the probe state mode with a thermal noise mode $|\bar{n}\rangle$ in a fictitious beamsplitter with a transmittivity of η (Fig. 8).

$$\hat{\rho} = \sqrt{\eta}\rho_0 + \sqrt{1-\eta}\,|\bar{n}\rangle \tag{46}$$

The phase noise in the measurement is modelled by encoding a random phase ϕ_p, sampled from a Gaussian distribution centred around the root mean square value corresponding to the experimental scheme. The variational algorithm is then simulated by using a gradient descent optimization scheme to minimize the cost function as shown in Fig. 9. It is to be noted that choosing a suitable learning rate and initial parameters is necessary for the optimizer to converge to the minima without requiring a large number of epochs.

In Fig. 10 the cost function landscape is visualized by setting the simulation model with the experimentally determined parameters and

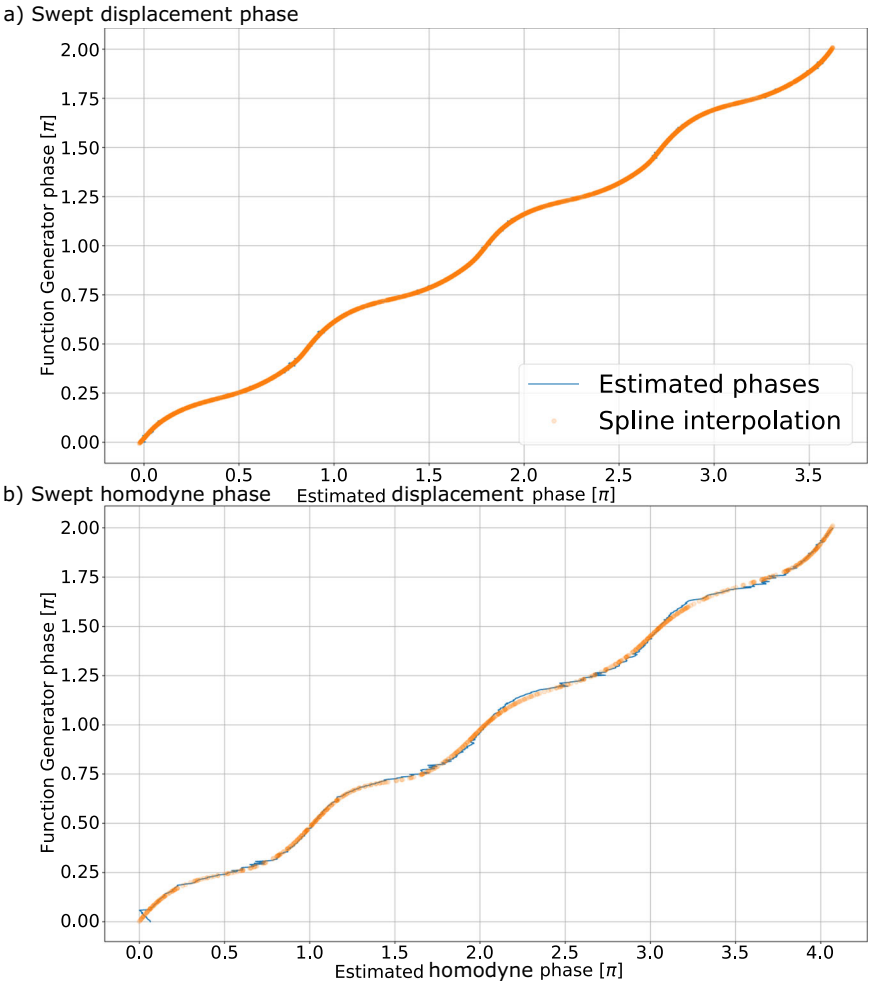


Fig. 6 | Phase calibration. Plot of the function generator phase for (**a**) the displacement phase lock and (**b**) homodyne phase lock as a function of the estimated phases. The orange dots are the corresponding B-spline representations.

probing the entire parameter space. The experimental data points from the kick-test described in the main text are superimposed on the landscape. We can observe that the optimizer effectively converges close to the theoretical minimum.

Bayesian optimization

Bayesian optimization is an iterative and gradient-free way of estimating the global minimum $\mathbf{x}_*$ of some function $f(\mathbf{x})$,

$$\mathbf{x}_* = \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x}) \tag{47}$$

In our experiment throughout the main paper, $\mathbf{x} = [\phi_\alpha, \phi_{HD}]$ and $f(\mathbf{x})$ is the corresponding cost function value measured for those parameters. However, in many real-world experiments, $f(\mathbf{x})$ might be corrupted with additive noise, such that the only measurable quantity $y(\mathbf{x})$ is given by

$$y(\mathbf{x}) = f(\mathbf{x}) + \epsilon \tag{48}$$

where it is often assumed that $\epsilon \sim \mathcal{N}(0, \sigma^2_{noise})$. For low-dimensional $\mathbf{x}$ (in our case only 2 dimensions) and expensive (e.g. time or monetary) queries to $y(\mathbf{x})$, Bayesian optimizer is a very relevant optimizer candidate[43]. This is indeed the case for our experiment outlined in the main text. In order to estimate $\mathbf{x}_*$, the Bayesian optimizer framework needs two quantities: 1) a surrogate function and 2) an acquisition function. The surrogate function "mimics" the observed datapoints $y(\mathbf{x})$, and is thus referred to as a *surrogate* function. Using this function, the acquisition function chooses which points to choose next by means of taking the maximum argument to the acquisition function. In our experiments, we choose a Gaussian Process as surrogate, and Expected Improvement as acquisition. In the next sections, we introduce these two quantities.

Gaussian Process. A Gaussian Process is a non-parametric model, which probabilistically models a variable $p(y_*|\mathbf{x}_*)$ with a normal distribution. It does so by conditioning on the corresponding input pair $\mathbf{x}_*$ as well as previously seen datapoints: a collection of N input/output datapoints $(\mathbf{x}_n, y_n)_{n=1}^N$ where $\mathbf{x}_n$ is the n'th input and y_n is the corresponding output. The entire collection of input datapoints, can be collected in an input matrix $\mathbf{X}$ and the corresponding outputs in a vector $\mathbf{y}$. Specifically, the distribution over any output y_*, which together with the corresponding input $\mathbf{x}_*$ we refer to as a testpoint, is given by the normal distribution

$$p(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}}, \boldsymbol{\Sigma}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}}), \tag{49}$$

$$\boldsymbol{\mu}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}} = K(\mathbf{x}_*, \mathbf{X})^\top [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} \mathbf{y}, \tag{50}$$

$$\boldsymbol{\Sigma}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}} = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X})^\top [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} K(\mathbf{X}, \mathbf{x}_*). \tag{51}$$

and where $K(\mathbf{X}, \mathbf{X})$ is called the kernel matrix. The kernel matrix is an $N \times N$ positive semi-definite matrix that contains pairwise similarity measures between the training points (vectors). Similarly $K(\mathbf{x}_*, \mathbf{X})$ is a N dimensional vector with a similarity measure between the test point and all the training points. It is easily verified from eq. (51) that both $\boldsymbol{\mu}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}}$ and $\boldsymbol{\Sigma}_{y_*|\mathbf{x}_*,\mathbf{X},\mathbf{y}}$ are scalars. For full derivation, we refer to[34].

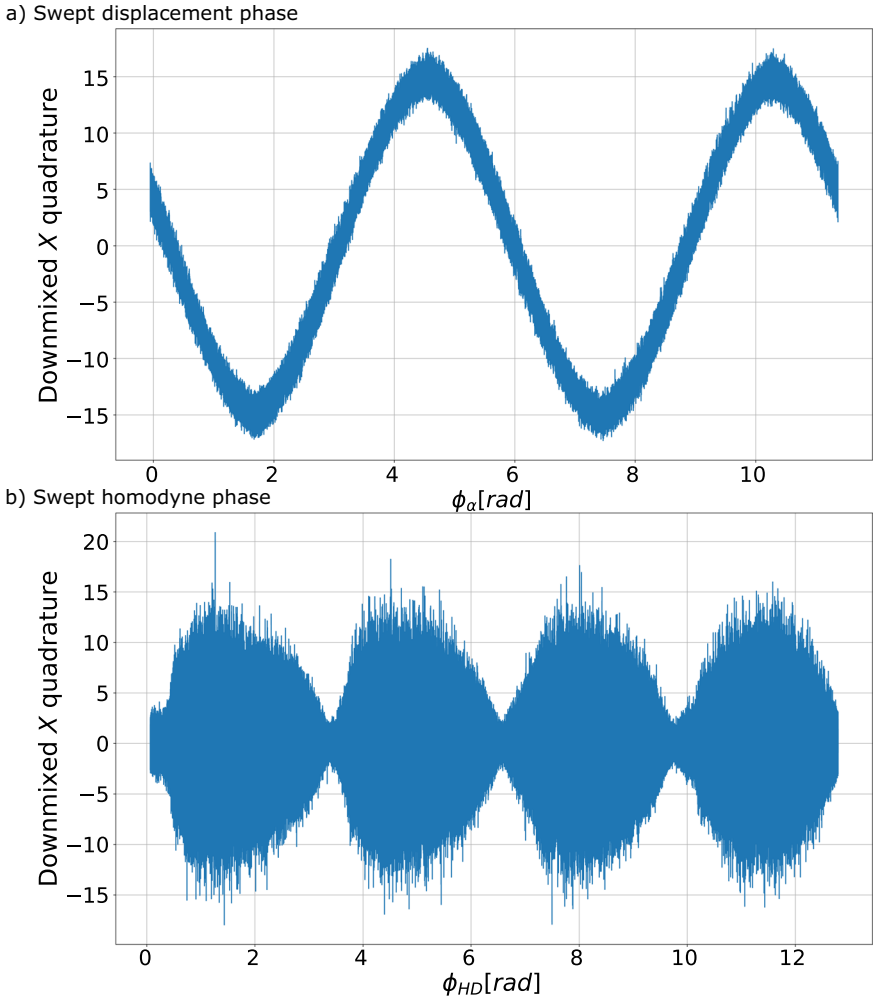


Fig. 7 | Swept homodyne outputs, with calibration. X quadrature measurements normalized to shot noise as a function of the estimated quadrature phase for (**a**) the displacement phase lock and (**b**) the homodyne phase lock.

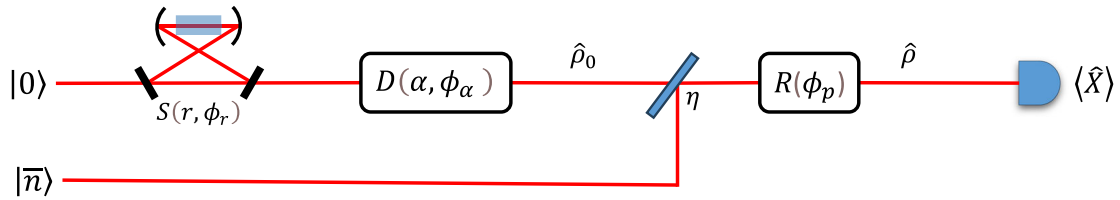


Fig. 8 | Schematic outline of the simulated system. Corresponding to the efficiency of the experimental system, the beamsplitter was simulated with a transmittivity of $\eta = 0.72$ and no thermal photons ($\bar{n} = 0$).

A popular choice of similarity measure between two datapoints (represented as vectors) is the Radial Basis Function (RBF) Kernel (also called a Gaussian Kernel) given by

$$K_{RBF}(\mathbf{x},\mathbf{x}') = exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{2\sigma^2}\right) \tag{52}$$

which is a function where similarity exponentially decays as the Euclidean distance between two points increases. The hyperparameter σ is called the length scale, as it defines the scale of how quickly the similarity should decrease. In our experiments, we use a Logarithmic Normal distribution prior and estimate it by maximizing the log-likelihood of the Gaussian Process. We also include an output scale hyperparameter σ_{scale} such that the final kernel function is given by $\sigma_{scale} \cdot K_{RBF}(\mathbf{x},\mathbf{x}')$.

Acquisition function. We use 136 datapoints and refer to this as our initial training set $\mathcal{D}_0 := [\mathbf{X},\mathbf{y}]$ and using the above equations we get the predictive distribution of the Gaussian Process. We refer to this Gaussian Process as our surrogate model, since it models the underlying loss landscape.

We now iteratively query points by finding the input point $\mathbf{x}_*$ that maximizes an acquisition function. A popular choice for acquisition function to go together with the Gaussian Process surrogate model is the expected improvement given by

$$\mathrm{EI}(\mathbf{x}_*) = (\mu(\mathbf{x}_*) - f(\mathbf{x}_*^+))\Phi(Z(\mathbf{x}_*)) + \sigma(\mathbf{x}_*)\phi(Z(\mathbf{x}_*)), \tag{53}$$

where $f(\mathbf{x}_*^+)$ is the current best guess of the global minimum and $\mathbf{x}_*^+$ is the parameter setting, Φ is the cumulative distribution function of a standard normal distribution, ϕ is the probability density function of a standard

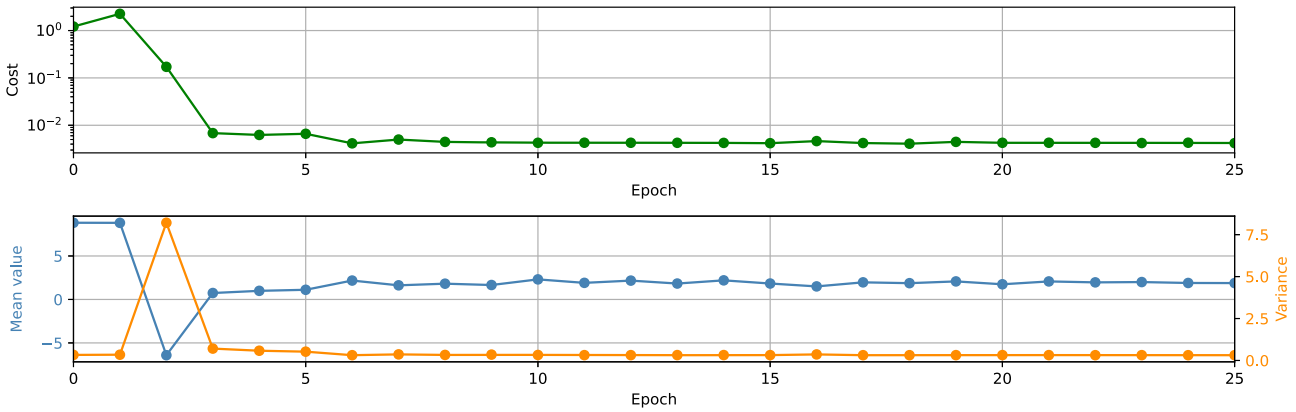
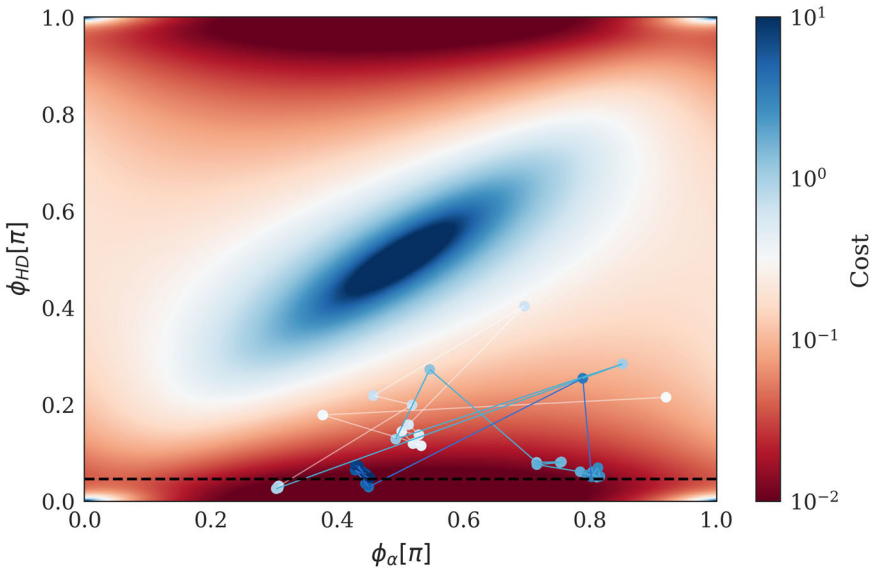


Fig. 9 | Simulation of the gradient descent-based experiment. The experimental parameters are taken to be $r = 1.52$, $|\alpha| = 5.2$, and phase noise RMS = 0.03 rad. The cost function represents the single-shot Fisher Information and does not take into account the number of samples used, hence the higher value compared to the experiment.

Fig. 10 | Simulated cost function landscape. The landscape has the kick-measurements superimposed. The points represent the cost for each epoch, with the colour intensity increasing progressively (the first epoch is depicted in white and the last epoch is the darkest blue). The black dotted line represents the theoretical optimum for the measurement angle ϕ_{HD}.



normal distribution and $\mu(\mathbf{x}_*)$ and $\sigma(\mathbf{x}_*)$ comes from the surrogate predictive distribution. EI is based on calculating expected marginal gain utility in the Gaussian Process after performing observation for candidate parameters[44]. The next queried input is thus given by

$$\mathbf{x}_{next} = \underset{\mathbf{x}}{\text{argmax}}\ \text{EI}(\mathbf{x}) \tag{54}$$

Note that $\mathbf{x}_{next}$ is a parameter combination $[\phi_{HD}, \phi_\alpha]$ which we have not used before. This parameter set is now used in the experiment to get the corresponding loss value $y(\mathbf{x}_{next})$. The dataset is now updated with this value to obtain $\mathcal{D}_1$, and so on. If the reader wants a pedagogical illustration of this process, have a look at Fig. 2.6 in ref. 34.

Hyper-parameters of the Gaussian process. Hyperparameters of the Gaussian Process were chosen very carefully. Their wrong definition might lead to the model's worse performance or in a critical case, the model not being able to perform optimization at all. Defining hyperparameters involves setting values of the mean and standard deviation of the aforementioned length scale and output scale hyperparameters. They can be either well-defined values (as for presented results in the main section of the paper), too strict (too low) or too loose (too high). On Fig. 11a we can see results for too loose definition of hyper-parameters and on Fig. 11b results for too strict values defining hyper-parameters are presented.

In figure Fig. 11b we see that too strict values defining hyperparameters lead to the optimizer performing very small changes of angle parameters resulting in it being stuck in areas that not necessarily are optimum. This argument is supported by relatively high values of a cost function which rarely goes below the limit of shot noise. In figure Fig. 11a we also see that too loose values of mean and variance of hyperparameters result in significant changes of angles in consecutive epochs. It is followed by considerable changes in the mean value of X quadrature, its variance relative to the shot noise and in values of a cost function as well. This behaviour is an example of a completely wrong choice of hyperparameters for which the model can't relate its definition with observations gathered during measurements. It results in the model "randomly walking" over the domain of angle parameters and not being able to perform optimization.

Both of these cases show the importance of defining the model correctly and prove that it should be done with meticulous attention.

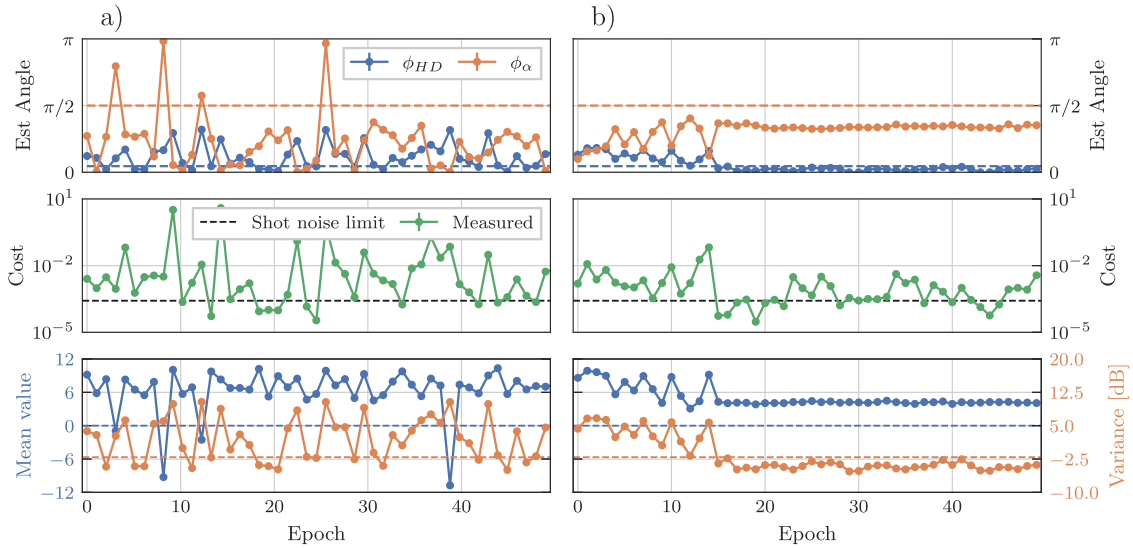


Fig. 11 | Bayesian optimization with bad parameters. Demonstration of the Bayesian optimization over 50 epochs for (**a**) very high and (**b**) very low mean and variance of the hyper-parameters of the Gaussian Process. (Top) The phase angles set by the algorithm. (Middle) The measured cost function $C = 1/F$. The dotted line is the shot noise limit taking into account the number of photons in the measurement and the number of samples used to estimate the cost function. (Bottom) The measured quadrature mean values and variances. The dotted lines are the optimal values predicted by the theory ("section IIIB").

Data availability

The data behind the plots is available upon reasonable request.

Received: 23 February 2024; Accepted: 12 December 2024;
Published online: 02 May 2025

References

1. Cerezo, M. et al. Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
2. Bharti, K. et al. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **94**, 015004 (2022).
3. Meth, M. et al. Probing phases of quantum matter with an ion-trap tensor-network quantum eigensolver. *Phys. Rev. X* **12**, 041035 (2022).
4. Hempel, C. et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**, 031022 (2018).
5. Graham, T. M. et al. Multi-qubit entanglement and algorithms on a neutral-atom quantum computer. *Nature* **604**, 457–462 (2022).
6. O'Malley, P. J. J. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
7. Self, C. N. et al. Variational quantum algorithm with information sharing. *npj Quantum Inf.* **7**, 116 (2021).
8. Zhang, H., Pokharel, B., Levenson-Falk, E. & Lidar, D. Predicting non-markovian superconducting-qubit dynamics from tomographic reconstruction. *Phys. Rev. Appl.* **17**, 054018 (2022).
9. Cao, Y. et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019).
10. Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. https://doi.org/10.48550/arXiv.1411.4028 (2014).
11. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).
12. Zhuang, Q. & Zhang, Z. Physical-layer supervised learning assisted by an entangled sensor network. *Phys. Rev. X* **9**, 041023 (2019).
13. Kaubruegger, R. et al. Variational spin-squeezing algorithms on programmable quantum sensors. *Phys. Rev. Lett.* **123**, 260505 (2019).
14. Koczor, B., Endo, S., Jones, T., Matsuzaki, Y. & Benjamin, S. C. Variational-state quantum metrology. *N. J. Phys.* **22**, 083038 (2020).
15. Yang, X. et al. Probe optimization for quantum metrology via closed-loop learning control. *NPJ Quantum Inf.* **6**, http://www.nature.com/articles/s41534-020-00292-z (2020).
16. Meyer, J. J., Borregaard, J. & Eisert, J. A variational toolbox for quantum multi-parameter estimation. *NPJ Quantum Inf.* **7**, 1–5 (2021).
17. Cimini, V. et al. Variational quantum algorithm for experimental photonic multiparameter estimation. arXiv e-prints arXiv:2308.02643 (2023).
18. Xia, Y., Li, W., Zhuang, Q. & Zhang, Z. Quantum-enhanced data classification with a variational entangled sensor network. *Phys. Rev. X* **11**, 021047 (2021).
19. Lawrie, B. J., Lett, P. D., Marino, A. M. & Pooser, R. C. Quantum sensing with squeezed light. *ACS Photonics* **6**, 1307–1318 (2019).
20. The LIGO Scientific Collaboration. Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nat. Photonics* **7**, 613 (2013).
21. Grote, H. et al. First long-term application of squeezed states of light in a gravitational-wave observatory. *Phys. Rev. Lett.* **110**, 181101 (2013).
22. Li, B.-B. et al. Quantum enhanced optomechanical magnetometry. *Optica* **5**, 850–856 (2018).
23. Taylor, M. A. et al. Biological measurement beyond the quantum limit. *Nat. Photonics* **7**, 229–233 (2013).
24. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **99**, 032331 (2019).
25. Meyer, J. J. Fisher information in noisy intermediate-scale quantum applications. *Quantum* **5**, 539 (2021).
26. Arnbak, J. et al. Compact, low-threshold squeezed light source. *Opt. Express* **27**, 37877–37885 (2019).
27. Vahlbruch, H. et al. Coherent control of vacuum squeezing in the gravitational-wave detection band. *Phys. Rev. Lett.* **97**, 011101 (2006).
28. Snoek, J., Larochelle, H. & Adams, R. P. Practical bayesian optimization of machine learning algorithms. *Adv. Neural Info. Process. Syst.* **25**, 2951–2959 (2012).
29. Görtler, J., Kehlbeck, R. & Deussen, O. A visual exploration of gaussian processes. *Distill* **4**, e17 (2019).
30. Agnihotri, A. & Batra, N. Exploring bayesian optimization. *Distill* **5**, e26 (2020).

31. Shi, H., Zhang, Z. & Zhuang, Q. Practical route to entanglement-assisted communication over noisy bosonic channels. *Phys. Rev. Appl.* **13**, 034029 (2020).
32. Rasmussen, C. E. Gaussian processes in machine learning. In *Summer school on machine learning*, 63–71 (Springer, 2003).
33. Frazier, P. I. Bayesian optimization. In *Recent advances in optimization and modeling of contemporary problems*, 255–278 (Informs, 2018).
34. Foldager, J. *Quantum Machine Learning in a World of Uncertainty: Towards Practical Applications with Quantum Neural Networks*. Ph.D. thesis, 2023 (Technical University of Denmark).
35. Guo, X. et al. Distributed quantum sensing in a continuous-variable entangled network. *Nat. Phys.* **16**, 281–284 (2020).
36. Kolobov, M. I. e.*Quantum Imaging* (Springer, New York, NY, 2007).
37. Brady, A. J. et al. Entangled sensor-networks for dark-matter searches. *PRX Quantum* **3**, 030333 (2022).
38. Zhang, Z. & Zhuang, Q. Distributed quantum sensing. *Quantum Sci. Technol.* **6**, 043001 (2021).
39. Kómár, P. et al. A quantum network of clocks. *Nat. Phys.* **10**, 582–587 (2014).
40. Nielsen, J. A. H., Neergaard-Nielsen, J. S., Gehring, T. & Andersen, U. L. Deterministic quantum phase estimation beyond N00N states. *Phys. Rev. Lett.* **130**, 123603 (2023).
41. Drever, R. et al. Laser phase and frequency stabilization using an optical resonator. *Appl. Phys. B* **31**, 97–105 (1983).
42. Bergholm, V. et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. https://doi.org/10.48550/arXiv.1811.04968 (2022).
43. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. Taking the human out of the loop: A review of bayesian optimization. *Proc. IEEE* **104**, 148–175 (2015).
44. Garnett, R. *Bayesian Optimization* (Cambridge University Press, 2023).

Acknowledgements

J.N., M.K., T.A., J.N.-N., T.G., and U.A. acknowledge support from the Danish National Research Foundation, Center for Macroscopic Quantum States (bigQ, DNRF142) and NNF project CBQS, EU project CLUSTEC (grant agreement no. 101080173), EU ERC project ClusterQ (grant agreement no. 101055224). K.V. acknowledges the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 956071 (AppQInfo). J.B. acknowledges funding and support from the NWO Gravitation Program Quantum Software Consortium (Project QSC No. 024.003.037) and The AWS Quantum Discovery Fund at the Harvard Quantum Initiative.

Author contributions

J.N. built the experiment and took measurements together with M.K. T.A., J.F., and M.K. wrote the experiment algorithms. K.V. provided theoretical simulations. J.N., J.B., and J.J.M. derived theoretical expressions. J.B., J.J.M., T.G., J. N.-N. and U.A. envisioned and planned the experiment.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Johannes Borregaard or Ulrik L. Andersen.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025