

# Machine Learning for Precise Quantum Measurement

Alexander Hentschel and Barry C. Sanders

*Institute for Quantum Information Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

Adaptive feedback schemes are promising for quantum-enhanced measurements yet are complicated to design. Machine learning can autonomously generate algorithms in a classical setting. Here we adapt machine learning for quantum information and use our framework to generate autonomous adaptive feedback schemes for quantum measurement. In particular our approach replaces guesswork in quantum measurement by a logical, fully-automatic, programmable routine. We show that our method yields schemes that outperform the best known adaptive scheme for interferometric phase estimation.

In classical physics, it is assumed that detectors and controls can be arbitrarily accurate, restricted only by technical limitations. However, this paradigm is valid only on a scale where quantum effects can be ignored. The ‘standard quantum limit’ (SQL) [1] restricts achievable precision, beyond which measurement must be treated on a quantum level. Heisenberg’s uncertainty principle provides a much lower but insurmountable bound for the accuracy of measurement and feedback. Approaching the Heisenberg limit is an important goal of quantum measurement.

The problem of quantum measurement can be stated as follows. A quantity such as spatial displacement, energy fluctuation, phase shift, or combination thereof, must be measured precisely within a specific duration of time. A typical device has an input and output, and the relation between the input and output yields information from which the quantity of interest can be inferred.

Important examples of practical quantum measurement problems within limited time include atomic clocks [2] and gravitational-wave detection [3]. Extensive efforts are underway to detect gravitational waves with laser-interferometers. The precision of these detectors is ultimately limited by the number of photons available to the interferometer within the duration of the gravitational-wave pulse [4]. The SQL to measurement is a concern for opening up a new field of gravitational-wave astronomy [5].

For the typical two-channel interferometer, shown in Fig. 1, the goal is to estimate the relative phase shift  $\varphi$  between the two channels. The interferometer has two input ports and two output ports, and we consider each input and output field as being a single mode.

Each input photon to the interferometer provides a single quantum bit, or ‘qubit’, as the photon is a superposition of  $|0\rangle$ , which represents the photon proceeding down one channel, or  $|1\rangle$ , corresponding to traversing the other channel. Each photon is either detected as leaving one of the two output ports or is lost. Thus, quantum measurement can extract no more than one bit of information about  $\varphi$  per qubit in the input state [6, 7].

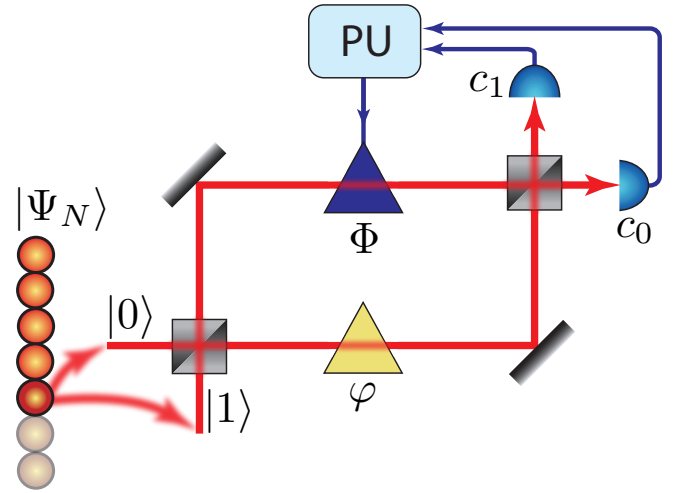
The fundamental precision bound is given by the ‘Heisenberg limit’: the standard deviation  $\Delta\varphi$  of the phase

estimate scales as  $1/N$  for  $N$  the number of input qubits used for the measurement.  $\Delta\varphi$  is determined by the error probability distribution  $P(\varsigma)$  for estimating  $\varphi$  with error  $\varsigma$ . As  $\varsigma$  is cyclic over  $2\pi$ ,  $\Delta\varphi$  is related to the Holevo variance  $V$  by [8]

$$V = (\Delta\varphi)^2 = S^{-2} - 1, \quad S = \left| \int_{-\pi}^{\pi} P(\varsigma) e^{i\varsigma} d\varsigma \right|. \quad (1)$$

$S$  is the ‘sharpness’ [9] of  $P(\varsigma)$ . In contrast, classical measurements only manage to achieve the SQL scaling  $\Delta\varphi \sim 1/\sqrt{N}$  due to partition noise for photons passing through the beam splitter. Quantum alternatives such as injecting squeezed light into one port of the interferometer can partially evade partition noise [10].

Since for any time-limited interferometric measure-



**Figure 1. Adaptive feedback scheme for interferometric phase estimation:** Mach-Zehnder Interferometer with an unknown phase difference  $\varphi$  between the two arms and an additional controllable phase shifter  $\Phi$ . The input state  $|\Psi_N\rangle$  is stored in a quantum memory and one qubit at a time is transformed into a photonic qubit and sent through the interferometer. The processing unit (PU) sets the value of the phase shifter  $\Phi$  depending on the measurement outcome of the single photon detectors  $c_0$  and  $c_1$ . Adaptive feedback at step  $m = 2$  is depicted. That is, two of the  $N$  input photons (the lowest two circles) have been sent through the interferometer and measured in previous steps.

ment, the number of input-qubits  $N$  determines the achievable precision, we define  $N$  as the relevant cost for the measurement. However, it is important to discriminate resources required to operate a measurement device from the ones used to develop it. Accordingly we distinguish between operational and developmental cost. The strategic question concerns the design of a device with a certain operational cost, so that its precision surpasses the SQL and scales as close to the Heisenberg limit as possible.

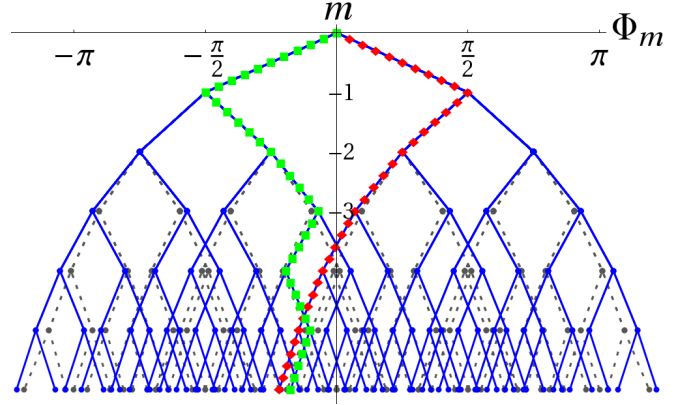
Quantum measurement schemes employing adaptive feedback are most effective, since accumulated information from measurements is exploited to maximize the information gain in subsequent measurements. Such adaptive measurements have been experimentally shown to be a powerful technique to achieve precision beyond the standard quantum limit [11, 12]. However, devising ‘policies’, which determine feedback actions, is generally challenging and typically involves guesswork. Our aim is to deliver a method for an automated design of policies based on machine learning [13]. To show the power of our framework, we apply it to adaptive phase estimation. As we will show, the policies generated by our method outperform the best known solutions for this problem.

Fig. 1 shows how a two-channel quantum interferometric measurement with feedback operates. We inject a  $N$ -photon input state  $|\Psi_N\rangle$  into a Mach-Zehnder interferometer with an unknown phase shift  $\varphi$  in one arm and a controllable phase shift  $\Phi$  in the other arm. Detectors at the two output ports measure which way the photon left, and this information is transmitted to a processing unit (PU), which determines how  $\Phi$  should be adjusted for the next input-qubit. We show that, after all  $N$  input qubits have been sent through the interferometer,  $\varphi$  can be inferred with a precision that scales closely to the Heisenberg limit.

We use the input  $|\Psi_N\rangle = \sum_{n,k=0}^N c_{n,k} |n, N-n\rangle$  from [14], with  $c_{n,k} = (\frac{N}{2} + 1)^{-\frac{1}{2}} \sin(\frac{k+1}{N+2} \pi) e^{\frac{i}{2}\pi(k-n)} d_{n,k}^{N/2}(\frac{\pi}{2})$  and  $d_{\nu,\mu}^j(\beta)$  Wigner’s (small)  $d$ -matrix [15].  $|n, N-n\rangle$  denotes a symmetrized state of  $N$  suitable delayed photons with  $n$  photons in channel  $|0\rangle$  and  $N-n$  in  $|1\rangle$  [16].

The challenge is to find a feedback policy, i.e. algorithm to run in the PU, that adjusts  $\Phi$  optimally. Fortunately, the area of machine learning suggests a promising approach. However, standard machine learning assumes classical bits as input and output. We inject a sequence of entangled qubits and obtain output bits. Due to the entanglement, the state of the remaining input qubits is progressively updated by the measurement. Consequently, the input to the system (except the first qubit) depends on the unknown system parameters. As a result, the space of quantum measurement policies is generically non-convex, which makes policies hard to optimize.

Particle swarm optimization (PSO) algorithms [17] are



**Figure 2. Decision tree representations of two adaptive feedback policies for  $N=6$  photons.** The PSO-generated policy is graphed in blue solid lines, the BWB-policy in gray dotted lines. All  $2^N$  possible experimental runs are represented by paths in the tree. The path corresponding to an experiment with detections  $u_1 u_2 \dots u_6 = 100000$  is marked by  $\blacklozenge$ , the path corresponding to the detections  $u_1 u_2 \dots u_6 = 011010$  is highlighted by  $\blacksquare$ . For each path in the tree, the inner nodes represent the applied feedback phases  $\Phi_m$  and the leaf shows the final phase estimate  $\tilde{\varphi}$ .

remarkably successful for solving non-convex problems. PSO is a ‘collective intelligence’ strategy from the field of machine learning that learns via trial and error and performs as well as or better than simulated annealing and genetic algorithms [18–20]. Here we show that PSO algorithms also deliver automated approaches to devising successful quantum measurement policies for implementation in the PU.

Our method is effective even if the quantum system is a black box, i.e. complete ignorance about the system itself. The only prerequisite is a comparison criterion during the training phase by which the success of candidate policies can be evaluated.

To explain how we use machine learning for the quantum measurement problem, consider the decision tree required by the PU to update the feedback  $\Phi$ . The measurement of the  $i^{\text{th}}$  qubit yields one bit  $u_i$  of information about which way the photon exited. (If the photon is lost, there is no detection at all and hence no bit. Therefore, a policy must be robust against loss.) After  $m$  photons have been processed, the PU stores the  $m$ -bit string  $n_m = (u_m u_{m-1} \dots u_1)$  and computes the feedback phase  $\Phi_m$ . In the most general case of a uniform prior distribution for  $\varphi \in [0, 2\pi)$ , there is no optimal setting for the initial feedback so we set  $\Phi_0 = 0$ , without loss of generality. All subsequent  $\Phi_m$  are chosen according to a prescribed decision tree.

In order to show that our method not only works, but is superior to existing feedback-based quantum measurements, we choose the Berry-Wiseman-Breslin (BWB) policy [14] as a benchmark. The BWB-policy is the most precise policy known to date for interferometric phase estimation with direct measurement of the interferometer

output. Furthermore, its practicality has been demonstrated in a recent experiment [11]. The BWB-policy achieves its best performance with the input state  $|\Psi_N\rangle$ . We use the same input state to provide fair premises. However, any more practical input state can be used and the PSO will autonomously learn good feedback policies.

In Fig. 2 we depict the decision trees of the BWB-policy and of our six-photon policy. At depth  $m$ , a measurement  $u_{m+1} = 0$  directs the path to the left and  $u_{m+1} = 1$  to the right. The final destination of the path yields an estimate  $\tilde{\varphi}$  of  $\varphi$ , which is solely determined by the measurement record  $n_N$ . Each experimental course corresponds to a path in the decision tree, where a path is a string of applied feedback phases  $\Phi_0, \Phi_1(n_1), \dots, \Phi_{N-1}(n_{N-1})$  plus a final phase estimate  $\tilde{\varphi}(n_N)$ .

A policy is entirely characterized by all the actions it can possibly take, thus by the  $2^{(N+1)} - 1$  phase values  $\Phi_0, \Phi_1(0), \Phi_1(1), \Phi_2(00), \Phi_2(01), \dots \in [0, 2\pi)$ . Therefore, a policy can be parametrized as a vector  $\rho$  in the policy space  $[0, 2\pi)^{2^{(N+1)} - 1}$ , and any such vector  $\rho$  forms a valid policy.

For addition and scalar multiplication modulo  $2\pi$ , the policy space forms a vector space. However, the dimension  $2^{(N+1)} - 1$  of this space grows exponentially with  $N$  making numeric optimization computationally intractable. Hence, we have to decrease the dimension of the search space exponentially by excluding policies.

In the case of logarithmic search, the adjustments of the feedback phase,  $\Delta\Phi_m := |\Phi_m - \Phi_{m-1}|$ , follow the recursive relation  $\Delta\Phi_m = \frac{1}{2}\Delta\Phi_{m-1}$ . Here, we generalize this search approach and treat  $\Delta\Phi_1, \dots, \Delta\Phi_N$  as independent variables. In the emerging trees, the adjustment  $\Delta\Phi_m$  depends only on the depth  $m$ , i.e. the number of measurements performed, but not on the full measurement history  $n_m$ :

$$\Phi_m = \Phi_{m-1} - (-1)^{u_m} \Delta\Phi_m. \quad (2)$$

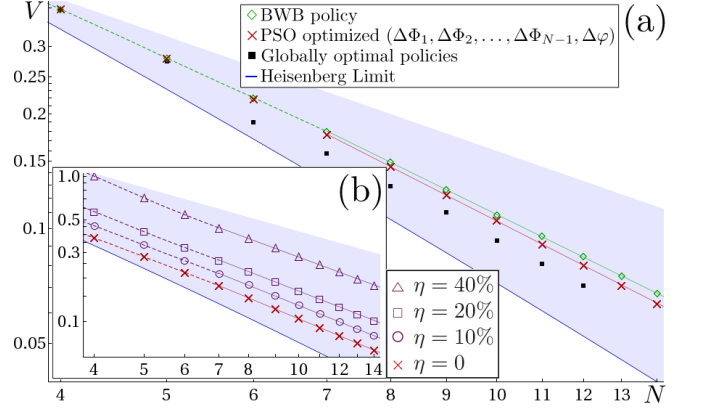
Equivalently, the final phase estimate is determined via

$$\tilde{\varphi} = \Phi_{N-1} - (-1)^{u_N} \Delta\varphi. \quad (3)$$

By this parametrization,  $(\Delta\Phi_1, \dots, \Delta\Phi_{N-1}, \Delta\varphi)$  fully define a decision tree, because the initial feedback phase is set to  $\Phi_0 = 0$ . The dimension of the resulting policy space  $\mathcal{P} = [0, 2\pi)^N$  is linear in  $N$ .

Furthermore our dimensional reduction is promising because  $\mathcal{P}$  includes a good approximation of the BWB-policy. Therefore, the best policies of this class will presumably outperform the BWB-policy.

Now that the policy space is appropriately small, we employ a PSO algorithm. This population-based stochastic optimization algorithm is inspired by social behavior of birds flocking or fish schooling to locate feeding sites [21]. Instead of birds and flocks, we employ the standard terms particles and swarms.



**Figure 3. (a) Holevo phase variance  $V_\varphi$  of the PSO-optimized policies in comparison to the BWB-policy and globally optimal policies for varying operational cost  $N$ . The blue shaded area shows the domain of quantum enhanced measurements. (b) Performance of the PSO policies with probability of photon loss  $\eta$ . All curves follow a power law for  $N \geq 7$  indicated by solid lines.**

To search for the optimal phase estimation policy, the PSO algorithm models a swarm of particles moving in the search space  $\mathcal{P}$ . The position  $\rho^{(i)} = (\Delta\Phi_1, \dots, \Delta\Phi_{N-1}, \Delta\varphi) \in \mathcal{P}$  of particle  $i$  represents a candidate policy for estimating  $\varphi$ , which is initially chosen randomly. Given the policy  $\rho^{(i)}$ , the sharpness  $S(\rho^{(i)})$  is analytically computed and disclosed to the particle.

The PSO algorithm updates the candidate policies of all particles, i.e. the positions in the policy space  $\mathcal{P}$ , in sequential rounds. At every time step, each particle displays the sharpest policy  $g^{(i)} \in \mathcal{P}$  it has found so far to the rest of the swarm. Then all particles try other policies by moving in the policy space  $\mathcal{P}$ . The moving direction for each particle is based on its own experience and also on what other particles in its neighborhood have discovered is the best overall policy.

The computation of the sharpness  $S(\rho)$  has exponential time complexity in  $N$ . Consequently, policies can be optimized only for small  $N$ . In practice the values of  $N$  achieved in experiments are quite small, much less than 14. So small  $N$  simulations are of practical value.

We have trained the quantum learning algorithm for phase estimation up to a total photon number of  $N = 14$ . In each case, the PSO algorithm tries to find the sharpest policy  $(\Delta\Phi_1, \dots, \Delta\Phi_{N-1}, \Delta\varphi)$ . However, as the algorithm involves stochastic optimization, it is not guaranteed to learn the optimal policy every time. So it must be run several times independently for each  $N$ . Rerunning the PSO-algorithm increases the developmental cost for the policies but does not affect their operational cost.

Fig. 3(a) depicts the performance of our quantum learning algorithm and compares it to the BWB-policy. Within the limits of the available computational resources, the PSO policies outperform the BWB-policy. To provide a quantitative estimate of the performance

difference, we calculated the scaling of the Holevo phase variance  $V_\varphi$  for  $N \geq 7$ , where both curves follow a clear power law (solid lines). Our policy yields  $V_\varphi \propto N^{-\alpha}$  with a scaling of  $\alpha_{\text{PSO}} = -1.472 \pm 0.005$ , compared to BWB's  $\alpha_{\text{BWB}} = -1.408 \pm 0.005$ .

Any practical policy has to be robust to photon loss. In Fig. 3(b), we have graphed the performance of our policies for loss rates  $\eta$  up to 40% and calculated the scaling  $\alpha_\eta$  for  $N \geq 7$ . We found  $\alpha_{0.1} = 1.421 \pm 0.006$ ,  $\alpha_{0.2} = 1.377 \pm 0.008$ , and  $\alpha_{0.4} = 1.307 \pm 0.009$ . This shows that our PSO-generated policies, which are optimized for a loss-less interferometer, are robust against moderate loss (which is also true for the BWB-policy). Moreover, one could train the PSO algorithm for a fixed loss rate  $\eta$ , which increases the computation time for the sharpness evaluation by a factor of  $N$ .

The dimensional reduction of the search space comes at the price of possibly excluding superior policies. In addition to proposing the BWB-policy, the authors performed in [14] a brute force search for ‘globally optimal policies’ in the exponential space. This was done by approximating  $[0, 2\pi)$  with a mesh and evaluating every possible combination of feedback phases. The performance of the optimal policies of this search is shown in Fig. 3(a). We

found that the phase variance of the globally optimal policies is better than the performance of policies from our reduced space  $\mathcal{P}$  only by a constant factor  $0.89 \pm 0.01$ .

In summary, we have developed a framework which utilizes machine learning to autonomously generate adaptive feedback measurement policies for single parameter estimation problems. Within the limits of the available computational resources, our PSO generated policies achieve an optimal scaling of precision for singleshoot interferometric phase estimation with direct measurement of the interferometer output. Our method can be extended to allow training using a real experimental setup by adapting a noise tolerant PSO algorithm [22]. This algorithm does not require prior knowledge about the physical processes involved. Specifically, it can learn to account for all systematic experimental imperfections, thereby making time-consuming error modeling and extensive calibration dispensable.

*Acknowledgments:* We are grateful to D. Schuurmans, C. Jacob, A. S. Shastry and N. Khemka for intellectual contributions. We thank B. Bunk and the Humboldt-Universität zu Berlin for computational resources and iCORE for financial support. BCS is a CIFAR Associate.

- 
- [1] C. M. Caves *et al.*, “On The Measurement Of A Weak Classical Force Coupled To A Quantum Mechanical Oscillator,” *Rev. Mod. Phys.*, vol. 52, pp. 341–392, 1980.
  - [2] J. J. Bollinger *et al.*, “Laser-cooled-atomic frequency standard,” *Phys. Rev. Lett.*, vol. 54, pp. 1000–1003, 1985.
  - [3] A. Abramovici *et al.*, “LIGO: The Laser Interferometer Gravitational-Wave Observatory,” *Science*, vol. 256, no. 5055, pp. 325–333, 1992.
  - [4] C. M. Caves, “Quantum-mechanical noise in an interferometer,” *Phys. Rev. D*, vol. 23, no. 8, pp. 1693–1708, 1981.
  - [5] K. Goda *et al.*, “A quantum-enhanced prototype gravitational-wave detector,” *Nature Physics*, vol. 4, pp. 472–476, 2008, 0802.4118.
  - [6] A. S. Holevo, “Some estimates of information transmitted through quantum communication channel,” *Prob. Pere-dachi Inf.*, vol. 9, no. 3, pp. 3–11, 1973.
  - [7] A. Cabello, “Quantum key distribution in the holevo limit,” *Phys. Rev. Lett.*, vol. 85, p. 5635, 2000.
  - [8] A. S. Holevo, “Covariant measurements and imprimitivity systems,” in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes*, pp. 153–172, Springer, Berlin, 1984.
  - [9] J.-M. Lévy-Leblond, “Who is Afraid of Nonhermitian Operators? a Quantum Description of Angle and Phase,” *Annals of Physics*, vol. 101, no. 1, pp. 319 – 341, 1976.
  - [10] M. Xiao, L.-A. Wu, and H. J. Kimble, “Precision measurement beyond the shot-noise limit,” *Phys. Rev. Lett.*, vol. 59, no. 3, pp. 278–281, 1987.
  - [11] B. L. Higgins *et al.*, “Entanglement-free heisenberg-limited phase estimation,” *Nature*, vol. 450, no. 7168, pp. 393–396, 2007.
  - [12] M. A. Armen *et al.*, “Adaptive homodyne measurement of optical phase,” *Phys. Rev. Lett.*, vol. 89, no. 13, p. 133602, 2002.
  - [13] Per definition, a machine learning algorithm is one that has the ability to improve its performance based on past experience.
  - [14] D. W. Berry and H. M. Wiseman, “Optimal states and almost optimal adaptive measurements for quantum interferometry,” *Phys. Rev. Lett.*, vol. 85, no. 24, pp. 5098–5101, 2000.  
D. W. Berry, H. M. Wiseman, and J. K. Breslin, “Optimal input states and feedback for interferometric phase estimation,” *Phys. Rev. A*, vol. 63, no. 5, p. 053804, 2001.
  - [15] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*. Academic Press, New York, 1971.
  - [16] A. Hentschel and B. C. Sanders, “Ordered Measurements of Permutation-Invariant Qubit-Strings,” *in preparation*.
  - [17] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 1995*, pp. 39–43, IEEE, New York, 1995.
  - [18] S. Ethni *et al.*, “Comparison of particle swarm and simulated annealing algorithms for induction motor fault identification,” in *Proceedings of the 7th IEEE International Conference on Industrial Informatics, Cardiff, England, 2009*, IEEE, New York, 2009.
  - [19] J. Kennedy and W. M. Spears, “Matching algorithms to problems: An experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator,” in *Proceedings of the IEEE Congress on Evolutionary Computation, Anchorage, USA, 1998*, pp. 78–83, IEEE, New York, 1998.
  - [20] P. Fourie and A. Groenwold, “The particle swarm optimization algorithm in size and shape optimization,” *Structural Multidisciplinary Optimization*, vol. 23, pp. 259–267, 2002.



- [21] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*. Morgan Kaufmann, San Francisco, 2001.
- [22] J. Pugh, Y. Zhang, and A. Martinoli, “Particle swarm optimization for unsupervised robotic learning,” in *Proceedings of the Swarm Intelligence Symposium, Pasadena, USA, 2005*, pp. 92–99, IEEE, New York, 2005.
- [23] B. Yurke, S. L. McCall, and J. R. Klauder, “SU(2) and SU(1,1) interferometers,” *Phys. Rev. A*, vol. 33, no. 6,

pp. 4033–4054, 1986.

- [24] H. M. Wiseman and R. B. Killip, “Adaptive single-shot phase measurements: A semiclassical approach,” *Phys. Rev. A*, vol. 56, no. 1, pp. 944–957, 1997.
- [25] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

## APPENDIX

### A. Interferometer Description

We use the convention

$$|0\rangle = \hat{a}^\dagger |\text{vac}\rangle, \quad |1\rangle = \hat{b}^\dagger |\text{vac}\rangle, \quad (4)$$

where  $\hat{a}^\dagger$  and  $\hat{b}^\dagger$  are the creation operators for the field modes  $a$  and  $b$ , and  $|\text{vac}\rangle$  denotes the vacuum. We consider a Mach-Zehnder interferometer, where the first 50:50 beam splitter combining the two inputs has a scattering matrix

$$B_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (5)$$

The second beam splitter  $B_2$  is chosen such that it recovers the input if the phases  $\Phi$  and  $\varphi$  of both arms are equal, i.e.  $B_2 = B_1^{-1}$ . The operator of the Mach-Zehnder interferometer is given by [23]

$$\mathcal{I}(\theta) = \exp \left[ -\theta(\hat{a}^\dagger \hat{b} - \hat{a} \hat{b}^\dagger) \right], \quad (6)$$

with  $\theta = \frac{1}{2}(\varphi - \Phi)$ .

### B. Input States

For single-shot interferometric phase estimation, so called ‘minimum uncertainty states’ have been proposed to reduce the Holevo variance of the estimates of  $\varphi$  [14, 24]. These states are symmetric with respect to permutations of qubits and therefore the relevant quantities are the number  $n_a$  and  $n_b$  of photons in mode  $a$  and  $b$ . In this case, the product of two Fock states for the two modes  $a$  and  $b$ , denoted  $|n_a, n_b\rangle$  with  $N = n_a + n_b$  is a convenient basis.

The minimum uncertainty state  $|\Psi_N\rangle$  with  $N$  qubits is given by

$$|\Psi_N\rangle = \left(\frac{N}{2} + 1\right)^{-\frac{1}{2}} \times \sum_{n,k=0}^N \sin\left(\frac{k+1}{N+2}\pi\right) e^{\frac{i}{2}\pi(k-n)} d_{n,k}^j\left(\frac{\pi}{2}\right) |n, N-n\rangle \quad (7)$$

where  $d_{\nu,\mu}^j(\beta)$  is Wigner’s (small)  $d$ -matrix [15].

The minimum uncertainty state has been found to be the optimal input state for single-shot adaptive interferometric phase estimation [14], but, due to its entanglement, it is naturally hard to prepare. The BWB-policy achieves its best performance under the use of the input state (7), we use the same input state to provide fair premises. As for the BWB-policy, any other, more practical, input state can be used and the PSO will autonomously learn a good adaptive strategy.

### C. Feedback Technique

The value of  $\theta$  changes with the progress of the experiment due to the varying feedback  $\Phi$ . In our notation,  $\Phi_m$  is the feedback phase applied *after* the  $m^{\text{th}}$  detection. Hence, at the time when the  $m^{\text{th}}$  particle of the input state passes the interferometer, the phase difference between the two arms is parametrized by

$$\theta_m := \frac{1}{2}(\varphi - \Phi_{m-1}). \quad (8)$$

The remaining input state  $|\psi(n_m, \varphi)\rangle$  after  $m$  photo-detections is given by

$$|\psi(n_m, \varphi)\rangle = \hat{c}_{u_m}(\theta_m) \cdots \hat{c}_{u_2}(\theta_2) \hat{c}_{u_1}(\theta_1) |\Psi_N\rangle, \quad (9)$$

where

$$\hat{c}_{u_k}(\theta_k) = \frac{\hat{a} \cos(\theta_k - u_k \frac{\pi}{2}) - \hat{b} \sin(\theta_k - u_k \frac{\pi}{2})}{\sqrt{N - k + 1}} \quad (10)$$

is the Kraus operator [25] representing a measurement of the  $k^{\text{th}}$  particle with outcome  $u_k$  [14]. The states (9) are not normalized. In fact, their norm represents the probability

$$P(n_m|\varphi) = \langle \psi(n_m, \varphi) | \psi(n_m, \varphi) \rangle. \quad (11)$$

for obtaining the measurement record  $n_m$  given  $\varphi$ .

### D. Performance Measure For Policies

In this section, we will show how the sharpness (1) can be analytically computed for a given policy  $\rho$ . Our derivation follows the procedure in [14]. The sharpness is

determined by the probability that  $\rho$  produces an estimate  $\tilde{\varphi}_\rho$  with error  $\varsigma = \tilde{\varphi}_\rho - \varphi$ ,

$$P_\rho(\varsigma|\varphi) = \sum_{\substack{n_N \in \\ \{0,1\}^N}} P_\rho(n_N|\varphi) \delta(\varsigma - (\tilde{\varphi}_\rho(n_N) - \varphi)) . \quad (12)$$

Here  $P_\rho(n_N|\varphi)$  is the probability that the experiment, with feedback actions determined by  $\rho$ , produces the measurement string  $n_N$  given the phase value  $\varphi$ . Here we use a flat prior for  $\varphi$ , i.e.  $P(\varphi) = \frac{1}{2\pi}$ .

$$\begin{aligned} P_\rho(\varsigma) &= \int_0^{2\pi} P(\varphi) P_\rho(\varsigma|\varphi) d\varphi \\ &= \frac{1}{2\pi} \sum_{\substack{n_N \in \\ \{0,1\}^N}} P_\rho(n_N|\tilde{\varphi}_\rho(n_N) - \varsigma) \end{aligned} \quad (13)$$

From this probability distribution, we determine the sharpness with equation (1)

$$\begin{aligned} S(\rho) &= \left| \frac{1}{2\pi} \sum_{\substack{n_N \in \\ \{0,1\}^N}} \int_0^{2\pi} P_\rho(n_N|\tilde{\varphi}_\rho(n_N) - \varsigma) e^{i\varsigma} d\varsigma \right| \\ &= \left| \frac{1}{2\pi} \sum_{\substack{n_N \in \\ \{0,1\}^N}} e^{i\tilde{\varphi}_\rho(n_N)} \int_0^{2\pi} P_\rho(n_N|\varsigma) e^{-i\varsigma} d\varsigma \right|. \end{aligned} \quad (14)$$

The probability  $P_\rho(n_N|\varsigma)$  is given by equation (11) and can be directly computed for a given policy  $\rho$ . (For more details see [14].) From equation (14) it is obvious that computing the sharpness of a policy  $\rho$  has complexity  $\mathcal{O}(2^N)$ . This is because the summand has to be evaluated for every bit-string  $n_N$  of length  $N$ .

### E. Optimization Problem

Given the policy space  $\mathcal{P}$ , the optimization problem is defined as finding a policy

$$\rho_{\max} \in \arg \max_{\rho \in \mathcal{P}} S(\rho), \quad (15)$$

i.e. find a  $\rho_{\max}$  such that  $S(\rho_{\max}) \geq S(\rho)$  for all  $\rho \in \mathcal{P}$ .

### F. Details of the employed PSO algorithm

In this section the details of the PSO algorithm we employed are presented. The swarm  $\mathcal{S} = \{p_1, p_2, \dots, p_\Xi\}$  is composed of a set of particles  $i = 1, 2, \dots, \Xi$ , where  $p_i$  is the set of properties of the  $i^{\text{th}}$  particle and  $\Xi \in \mathbb{N}$  is the population size. At any time step  $t$ ,  $p_i$  includes the position  $\rho^{(i)} \in \mathcal{P}$  of particle  $i$  and  $\hat{\rho}^{(i)}$ , which is the best position  $i$  has visited until time step  $t$ .

$N$	$\Xi$	$\Delta$	$\varphi_1$	$\varphi_2$	$\omega$	$\nu_{\max}$	$r$	$\lambda$
4	50	700	0.5	1	1	0.05	1	100%
5	50	700	0.5	1	1	0.05	1	100%
6	50	700	0.5	1	1	0.05	1	100%
7	50	500	0.5	1	0.8	0.2	4	100%
8	60	300	0.5	1	0.8	0.2	6	35%
9	60	500	0.5	1	0.8	0.2	6	33%
10	60	400	0.5	1	0.8	0.2	6	25%
11	60	400	0.5	1	0.8	0.2	6	66%
12	120	1000	0.5	1	0.8	0.2	12	20%
13	375	300	0.5	1	0.8	0.2	30	17%
14	441	100	0.5	1	0.8	0.2	35	20%

Table I. PSO settings for  $N$ -Photon input state: velocity damping  $\omega$ , swarm size  $\Xi$ , number of PSO-Steps  $\Delta$ , max step-size  $\nu_{\max}$ , exploitation weight  $\varphi_1$ , exploration weight  $\varphi_2$ , fraction  $\lambda$  of PSO runs produced policies with variance depicted in Fig. 3.

Particle  $i$  communicates with other particles in its neighborhood  $\mathcal{N}^{(i)} \subseteq \mathcal{S}$ . The neighborhood relations between particles are commonly represented as a graph, where each vertex corresponds to a particle in the swarm and each edge establishes a neighbor relationship between a pair of particles. This graph is commonly referred to as the swarm's population topology.

We have adapted the common approach to set the neighborhood  $\mathcal{N}^{(i)}$  of each particle in a pre-defined way regardless of the particles' position. For that purpose the particles are arranged in a ring topology. For particle  $i$ , all particles with a maximum distance of  $r$  on the ring are in  $\mathcal{N}^{(i)}$ .

The PSO algorithm updates the position of all particles in a round based manner as follows. At time step  $t$

1. Each particle  $i = 1, 2, \dots, \Xi$  assesses the sharpness  $S_\varsigma(\rho_t^{(i)})$  of its current position  $\rho_t^{(i)}$  in the policy space (and updates  $\hat{\rho}^{(i)}$  if necessary).
2. Each particle  $i$  communicates the sharpest policy  $\hat{\rho}^{(i)}$  it has found so far to all members of its neighborhood  $\mathcal{N}^{(i)}$ .
3. Each particle  $i$  determines the sharpest policy  $g^{(i)} = \max_{j \in \mathcal{N}^{(i)}} \hat{\rho}^{(j)}$  found so far by any one particle in its neighborhood  $\mathcal{N}^{(i)}$  (including itself).
4. Each particle  $i$  changes its position according to

$$\begin{aligned} \rho_{t+1}^{(i)} &= \rho_t^{(i)} + \Delta \rho_t^{(i)} \\ \Delta \rho_t^{(i)} &= \omega (\Delta \rho_{t-1}^{(i)} + \varphi_1 \cdot \text{rand}() \cdot (\hat{\rho}^{(i)} - \rho_t^{(i)}) \\ &\quad + \varphi_2 \cdot \text{rand}() \cdot (g^{(i)} - \rho_t^{(i)})) . \end{aligned} \quad (16)$$

The parameter  $\omega$  represents a damping factor that assists convergence, and  $\text{rand}()$  is a function returning uniformly distributed random numbers in  $[0, 1]$ . The 'exploitation weight'  $\varphi_1$  parametrizes the attraction of a

$N$	$\Delta\Phi_1, \dots, \Delta\Phi_{N-1}$	$\Delta\varphi$	$V_\varphi$
4	1.5701, 0.7862, 0.5043	0.3507	0.37621
5	1.5722, 0.7816, 0.5293, 0.3684	0.2739	0.27922
6	1.5708, 0.7830, 0.5669, 0.3881, 0.2889	0.2306	0.21835
7	1.5708, 0.7854, 0.6159, 0.4130, 0.3073, 0.2421	0.1988	0.17630
8	1.5708, 0.7854, 0.6663, 0.4399, 0.3264, 0.2551, 0.2080	0.1750	0.14561
9	1.5708, 0.7854, 0.7079, 0.4620, 0.3440, 0.2671, 0.2164, 0.1811	0.1554	0.12253
10	1.5708, 0.7854, 0.7392, 0.4788, 0.3599, 0.2780, 0.2240, 0.1867, 0.1597	0.1393	0.10482
11	1.5706, 0.7850, 0.7613, 0.4934, 0.3744, 0.2875, 0.2313, 0.1920, 0.1642, 0.1421	0.1260	0.09094
12	1.5708, 0.7854, 0.7800, 0.5023, 0.3890, 0.2983, 0.2384, 0.1973, 0.1677, 0.1456, 0.1285	0.1149	0.07985
13	1.5695, 0.7847, 0.7920, 0.5119, 0.4029, 0.3083, 0.2457, 0.2027, 0.1720, 0.1487, 0.1310, 0.1170	0.1054	0.07083
14	1.5703, 0.7860, 0.8018, 0.5195, 0.4171, 0.3179, 0.2529, 0.2077, 0.1756, 0.1517, 0.1335, 0.1190, 0.107326	0.0975	0.06337

Table II. Parameters for the best PSO-generated policies for  $N = 4, \dots, 14$ .

particle to its best previous position  $\hat{\rho}^{(i)}$ , and the ‘exploration weight’  $\varphi_2$  describes the attraction to the best position  $g^{(i)}$  in the neighborhood. To increase convergence, we bound each component of  $\Delta\rho_s^{(i)}$  by a maximum value of  $\nu_{\max}$ . In summary, the properties of the swarm, such as size and behavior, are defined by the following parameters.

$$\begin{aligned}
\omega &\in [0, 1] && \text{velocity damping factor} \\
\varphi_1 &\in [0, 1] && \text{exploitation weight} \\
\varphi_2 &\in [0, 1] && \text{exploration weight} \\
\Xi &&& \text{population size} \\
\nu_{\max} &&& \text{maximum step size} \\
&&& \text{particles are allowed to move} \\
r &&& \text{interaction range of particles}
\end{aligned} \tag{17}$$

Clearly, the success and the number of required PSO steps to find the maximum is highly dependent on the values of these parameters. For instance, with increasing  $N$ , a bigger population size is required to account for the raising dimensionality of the search space. The most successful settings we found are listed in Table I. For each  $N = 1, \dots, 14$ , the best policy  $(\Delta\Phi_1, \dots, \Delta\Phi_{N-1}, \Delta\varphi)$  the PSO algorithm learned is given Table II.

### G. Noise resistance

As with the BWB-policy, which works with an idealized noiseless model, the training of our PSO algorithm is performed based on the simulation of a noiseless Mach-Zehnder interferometer. However, Higgins et al. recently used the BWB-policy as a component for their experiment [11], which shows that the feedback policies we considered for optimization are robust against noise.

Therefore, the policies generated by our learning approach are applicable to moderately noisy experiments,

even though the PSO algorithm was trained on a simulated noise-free experiment.

Figure 3 shows the Holevo variance of the PSO-generated policies for different photon-loss rates. We have calculated the variance as follows. For a fixed loss-rate  $\eta$ , the probability of detecting  $k$  of the  $N$  input-photons is given by the binomial distribution  $B(k; N, \eta) = \binom{N}{k} \eta^k (1 - \eta)^{N-k}$ . Then the probability that the policy  $\rho$  produces an estimate  $\tilde{\varphi}_\rho$  with error  $\varsigma = \tilde{\varphi}_\rho - \varphi$  is

$$P_\rho(\varsigma|\varphi) = \sum_{k=0}^N B(k; N, \eta) P(\varsigma|\varphi, k). \tag{18}$$

An analogous calculation to the one in appendix D yields

$$S(\rho) = \left| \sum_{k=0}^N B(k; N, \eta) \mathcal{S}(k) \right| \tag{19}$$

with

$$\mathcal{S}(k) = \frac{1}{2\pi} \sum_{\substack{n_k \in \\ \{0,1\}^k}} e^{i\tilde{\varphi}_\rho(n_k)} \int_0^{2\pi} P_\rho(n_k|\varsigma) e^{-i\varsigma} d\varsigma. \tag{20}$$

The probability  $P_\rho(n_k|\varsigma)$  is given by equation (11).

Figure 3 shows that our PSO-generated policies with the state (7) as input are remarkably robust against photon loss. Even with a loss rate of  $\eta = 40\%$  the variance scales as  $N \propto N^{-1.307 \pm 0.009}$ , and the measurement lies in the domain of quantum enhanced measurements.

The strong robustness against photon loss is mainly due to the nature of the input state (11), which is highly entangled and symmetric with respect to qubit permutations. As a consequence, this state remains entangled even if a high percentage of photons are lost.