2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
Banff Center, Banff, Canada, October 5-8, 2017

Robustness of Learning-Assisted Adaptive Quantum-Enhanced Metrology in the Presence of Noise

Pantita Palittapongarnpim*, Peter Wittek†‡ and Barry C. Sanders*§¶‖
*Institute for Quantum Science and Technology, University of Calgary, Calgary, Alberta T2N 1N4 Canada
†ICFO-The Institute of Photonic Sciences, Castelldefels (Barcelona), 08860 Spain
‡University of Borås, Borås, 501 90 Sweden
§Program in Quantum Information Science, Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8 Canada
¶Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China
‖Shanghai Branch, CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, People's Republic of China

***Abstract*—Reinforcement learning algorithms have been shown to generate procedures for executing quantum control tasks to desired performances. Although reinforcement learning has been effective, its robustness in generating quantum control procedures under general noise condition needs to be tested. Here we consider adaptive quantum-enhanced interferometric phase estimation as a case study in quantum feedback control. The algorithm is determined to deliver a robust quantum-metrological procedure under various phase-noise models when the imprecision surpasses the standard quantum limit. The robustness against unknown environmental variations positions the reinforcement learning as a practical approach for devising quantum control procedures.**

I. Introduction

Reinforcement learning in quantum physics involves two complementary approaches. On the one side is the application of reinforcement learning in providing classical learning and control over quantum systems, for instance, in adapting to a stray magnetic field in measurement-based quantum computing [1], compensating for qubit decoherence [2], and estimating an unknown phase [3]. This approach contrasts with the alternative of introducing quantum physics into reinforcement learning, providing a transition point into quantum-enhanced machine learning, where classical or quantum data are processed by a quantum system. Quantum-enhanced reinforcement-learning procedures [4] for large-scale quantum information processing systems are likely to be a major application area of quantum machine learning [5] and have potential in quantum-enhanced protocols, for instance, in trapped-ion systems [6] and superconducting qubits [7].

One quantum technology beneficiary of reinforcement learning is adaptive quantum-enhanced metrology (AQEM) [8], [9], whose aim is to infer the estimate $\widetilde{\phi}$ of an unknown parameter ϕ such that the scaling of imprecision $\Delta\widetilde{\phi}$ exceeds the standard quantum limit (SQL) [10]. The imprecision scales with the number of particles N used in the measurement process as $\Delta\widetilde{\phi} \in \mathcal{O}\left(N^{-\wp}\right)$ [11]. The SQL is an asymptotic limit defined for $N \to \infty$ where $\wp = 1/2$. Surpassing the SQL is possible with N-entangled state inputs, enabling a power-law scaling that approaches $\wp = 1$ known as the Heisenberg limit (HL) [11]. If sequential single-particle measurements are performed, the HL can be approached by using an adaptive procedure based on the outcomes of preceding measurements [8]. The rules for applying the adaptive procedure is called a policy and is difficult to devise due to the exponentially increasing number of possible outcomes with N. A reinforcement-learning approach is introduced to make this process tractable.

Phase estimation is a widely-studied problem in metrology as it captures the essence of cutting-edge scientific investigations and technologies, such as gravitational-wave detection [12], atomic clocks [13], and magnetometers [14]. As a result, the SQL and the HL are well known for phase estimation and can be used to compare the performance of quantum-enhanced measurement procedures [15]. Reinforcement-learning-based AQEM can deliver imprecision exceeding SQL under Gaussian phase noise by directly searching for feasible feedback policies [16]. Whether this reinforcement-learning algorithm can generate successful policies under other phase-noise models has never been tested.

We call a reinforcement-learning algorithm robust if it delivers $\wp > 1/2$ for all phase-noise models. Of course, we can only test robustness against a selection of phase noise, so we call the algorithm robust if it is robust against these test cases of phase-noise models. The phase noise is simulated by making ϕ a random number according a probability distribution, and the functional form of the distribution is what we refer to here as a model. As Gaussian distribution is symmetric, we choose one symmetric and two asymmetric distributions for our robustness test, all of which are unimodal. Random telegraph noise exhibits as discrete shifts in the value of ϕ [17], which we modify to be an example of non-Gaussian symmetric distribution. The two asymmetric distributions are

978-1-5386-1644-4/17/$31.00 ©2017 IEEE

skew-normal [18] and log-normal distribution [19], which are chosen because their variance shift in opposite direction when their parameters for skewness increase and hence can be used to determine if the change in $\wp$ has any correlation with skewness.

We show that our reinforcement-learning-based adaptive phase estimation incorporating three types of phase noises is robust by numerical analysis of the imprecision and the scaling $\wp$, assuming a power-law relationship between the imprecision and N.

II. Background

In this section, we describe the procedure of adaptive interferometric phase estimation including phase noise. We then discuss employing reinforcement learning to generate a successful policy and outline the algorithm based on differential evolution (DE).

A. Noisy adaptive interferometric phase estimation

In this subsection, we explain the procedure of adaptive interferometric phase estimation (Fig. 1). An interferometer that has two input and two output ports accepts an input state of N particles injected into the interferometer one at a time. Inside the interferometer is a noisy unknown phase shifter ϕ and a controllable phase shifter Φ. The output path for the m^{th} particle is detected as $x_m \in \{0,1\}$. The outcome of the detection is used to update the controllable phase shifter $\Phi_m = \Phi_{m-1} - (-1)^{x_m}\Delta_m$, which is a form of generalized logarithmic search [20]. After all the particles are used, the estimate $\widetilde{\phi} \equiv \Phi_N$ is the output.

The quantum input state of the particles $|\psi_N\rangle$ is a normalized vector in a complex Hilbert space. For our adaptive phase estimation procedure, we consider the sine state [21]–[23], where particles in different time bins are entangled and is spanned on a permutationally symmetric basis $\{|n_{\text{a}}, N-n_{\text{a}}\rangle\}$, where n_{a} is the number of particles entering the interferometer through input port a. The Hilbert space for the sine state is of $N+1$ dimension.

The unitary operator representing the interferometer $\hat{U}_m(\phi - \Phi_{m-1})$ acts on the m^{th} particle. In the case of noisy phase estimation, ϕ takes a value according to $P(\phi)$, which we consider to be unimodal that peaks at $\phi_0 \in [0, 2\pi)$. The value of ϕ changes for each particle independently.

A measurement on a quantum state is a positive-operator valued measure, which transforms the quantum state contingent upon the measurement outcome x_m. The outcome also determines the update $\Phi_{m-1} \to \Phi_m$ according to a real-vector policy $\varrho_N = (\Delta_1, \Delta_2, \ldots, \Delta_N) \in [0, 2\pi)^N$. As such, the policy determines $\widetilde{\phi}$ and consequently $\Delta\widetilde{\phi}$ and $\wp$.

The imprecision of estimate $\widetilde{\phi}$, which is a variable with a period of 2π, is quantified by a modified Holevo variance [24]

$$V_{\text{H}} = S_K(\varrho_N)^{-2} - 1, \tag{1}$$

$$S_K(\varrho_N) = \left| \sum_{k=1}^{K} \frac{\mathrm{e}^{\mathrm{i}(\phi_0^{(k)} - \widetilde{\phi}_k(\varrho_N))}}{K} \right|. \tag{2}$$

Input: number of particles N, input quantum state $|\psi_N\rangle$, policy ϱ_N, unknown phase shift ϕ_0
Output: estimate $\widetilde{\phi}$
Initialization: controllable phase shift $\Phi_0 \leftarrow 0$
for $1 \le m \le N$ **do**
  $\phi \leftarrow$ RandomNumber(ϕ_0)
  $x_m, |\psi_{N-m}\rangle \leftarrow$ Measure($|\psi_{N-m+1}\rangle, \phi, \Phi_{m-1}$)
  if $x_m = 0$ **then**
    $\Phi_m \leftarrow \Phi_{m-1} - \Delta_m$
  else
    $\Phi_m \leftarrow \Phi_{m-1} + \Delta_m$
  end if
  $m \leftarrow m + 1$
end for
return Φ_N

Fig. 1. Pseudocode describing adaptive interferometric phase estimation. The phase ϕ_0 is the most likely value of ϕ. The phase shift ϕ is a random number generated by function RandomNumber() following a probability distribution. The process of injecting and measuring a particle is simulated by function Measure(), which outputs a measurement outcome x_m and updates the state $|\psi_{N-m+1}\rangle \mapsto |\psi_{N-m}\rangle$ accordingly [22]. The state $|\psi_{N-m}\rangle$ is used in the next step of the measurement until all particles are detected.

The sharpness S_K here is calculated from estimates of uniformly sampled $\left\{\phi_0^{(k)}\right\}$ from $[0, 2\pi)$. The sample size $K = 10N^2$ is used as it has been found to approximate the exact sharpness [16].

Because $\phi_0^{(k)} - \widetilde{\phi}_k$ is in the exponent, the bias of $\widetilde{\phi}$ is not quantified by the modified Holevo variance, and therefore we cannot ascertain the accuracy of $\widetilde{\phi}$ using this information. We leave the subject of bias for later work as the program needs to be modified to compute and output bias properly and independently of the modified Holevo variance. Here we decide whether the learning algorithm is robust based solely on whether it can generate policies that deliver the scaling of the modified Holevo variance exceeds the SQL for all phase noise models.

B. Reinforcement learning for adaptive phase estimation

In this subsection, we explicate how reinforcement learning is used to overcome the intractability in generating the feedback policy by expressing the adaptive measurement scheme as a partially observable Markov decision process (POMDP) [25]. From there, the reinforcement-learning algorithm is formulated to perform a direct search of the policy. We outline the algorithm and the heuristics employed to make the search tractable.

Generating a feedback policy for an adaptive measurement is an optimization process over all possible sequences of outcomes x_m and all possible value of $\phi_0 \in [0, 2\pi)$, which is computationally not tractable because of the infinitely many possible values of ϕ_0. A learning algorithm can be used to find a generalized policy optimized over a finite set of training data [26]. Reinforcement learning, in particular, is useful for quantum control in the absence of a trusted model as the

policy can be generated through interactions with the quantum system [27].

Reinforcement learning can be applied to adaptive phase estimation by recognizing that the procedure is a POMDP. The processing unit makes a decision to tweak the current estimate Φ_{m-1} by $\pm\Delta_m$ depending on x_m without the full knowledge of the environment state (the quantum state of particles $|\psi_{N-m+1}\rangle$, the unknown phase shift ϕ_0, and the noise model and parameters). As we are not concerned with the trajectory of the state evolution and only the imprecision of the estimate, we choose to search directly for successful policies instead of back-propagating for a value function [28].

To select an appropriate optimization algorithm, we first consider the dimension of the search space and the fitness landscape. The generalized logarithmic given in Subsec. II-A is Markovian and symmetric, so for each measurement step only one parameter Δ_m is associated with the update. Policies for the procedure are in $[0, 2\pi)^N$, whose dimension scales linearly with the number N of particles. We use the sharpness function (2) as the measure of a policy's performance due to its relation to the modified Holevo variance.

The fitness landscape of the sharpness function is non-convex [3], and, therefore, we select a global optimization algorithm to generate a policy. The algorithm is chosen based on two criteria: its ability to generate solutions for high-dimensional problems, and its independence of models. The latter criterion is imposed to make the reinforcement learning as close to model-free as possible. To this end, we select DE [29], which has been found to generate successful policies over $N = 90$ for an ideal interferometer as opposed to particle swarm optimization (PSO), which only delivers power-law scaling up to $N = 45$ [30]. With modification [31], DE is able to deliver successful policies for Gaussian phase noise [32]. Moreover, DE does not rely on a model of the quantum system or a model of the fitness landscape.

Further reduction of search-space size is implemented by initializing the DE population for N-particle measurement randomly according to Gaussian distributions around the policy for $N-1$ in the N-dimensional search space [3]. This initialization heuristic allows our algorithm to generate successful policies within an evaluation budget of 300 generations for $N = 4$ and 100 generations for $4 < N \leq 93$ using a population size $N_\mathrm{p} = 48$. For $N > 93$, we implement a different termination criterion based on whether the variance follows the power-law trend set by data from $N \leq 93$, allowing us to achieve a better than SQL scaling up to 100 particles [32].

Each Δ_m in policy ϱ_N is randomly initialized following a specified probability distribution and iteratively modified according to DE/rand/1/bin algorithm. The value are kept within the domain by setting the upper limit to 2π and lower limit to 0. The outline of the algorithm is in Fig. 2.

Input: number of particles N, policy ϱ_{N-1}, DE population size N_p, DE scaling factor $F = 0.1$, DE crossover rate $C_\mathrm{r} = 0.6$
Output: ϱ_N
Initialization: generation $t \leftarrow 1$, candidate solutions $\boldsymbol{V}(t)$, candidate offsprings $\boldsymbol{D}(t)$, fitness from one instance $S(t)$, average fitness $f(t)$
for $1 \leq p \leq N_\mathrm{p}$ **do**
  $\boldsymbol{V}^{(p)}(t) \leftarrow$ Initialize(ϱ_{N-1})
  $f^{(p)}(t) \leftarrow$ Sharpness($\boldsymbol{V}^{(p)}(t)$)
end for
while termination condition not met **do**
  for $1 \leq p \leq N_\mathrm{p}$ **do**
    $S^{(p)}(t) \leftarrow$ Sharpness($\boldsymbol{V}^{(p)}(t)$)
    $f^{(p)}(t) \leftarrow$ Average($f^{(p)}(t), S^{(p)}(t)$)
    Randomly select 3 candidates $\{\boldsymbol{V}^{(i\neq p)}(t)\}$
    Generate $\boldsymbol{D}^{(p)}(t)$ from selected candidates using F, C_r
    $S^{(p)}(t) \leftarrow$ Sharpness($\boldsymbol{D}^{(p)}(t)$)
    Select $\boldsymbol{D}^{(p)}(t)$ or $\boldsymbol{V}^{(p)}(t)$ to be $\boldsymbol{V}^{(p)}(t+1)$ based on whether $S^{(p)}(t)$ or $f^{(p)}(t)$ is higher
  end for
  $\varrho_N \leftarrow \boldsymbol{V}^{(p)}(t)$ with highest $f^{(p)}(t)$
  $t \leftarrow t + 1$
end while
return ϱ_N

Fig. 2. Pseudocode outlining the noise-resistant reinforcement-learning algorithm for adaptive interferometric phase estimation [32]. t runs up to 300 for $N = 4$ and 100 for $4 < N \leq 93$. For $N > 93$, the algorithm accepts a policy that follows a power-law trend line within 0.98 confidence interval. The function Initialize() creates a random starting policy following a specified distribution. The Sharpness() computes the sharpness function by calling $K = 10N^2$ instances of adaptive phase estimation using randomly sampled ϕ_0. An offspring is generated using DE/rand/1/bin and compete with its parent using the average sharpness.

III. Simulation of Phase Noise

In this section, we outline the methods for generating phase noise in the simulation of adaptive interferometric phase estimation, i.e., the function RandomNumber() in Fig. 1. We apply our reinforcement-learning algorithm to adaptive phase measurements that include three examples of symmetric and asymmetric noise models, namely, random telegraph noise, skew-normal noise and log-normal noise. The random number is divided by 2π and the remainder is used as the phase shift in simulation in order to keep $\phi \in [0, 2\pi)$.

A. Properties of phase noise distribution

In this subsection, we discuss variance and skewness of a phase-noise distribution and their effect on $\Delta\widetilde{\phi}$ and $\wp$. Whereas variance has been previously observed to increase $\Delta\widetilde{\phi}$, the effect of skewness has never been studied. Because changing the parameter that determines skewness also changes the variance, we choose two asymmetric distributions whose variances response to the change in skewness parameters in opposite directions. The tail probability of a skewed-normal distribution approaches zero faster than a Gaussian distribution of the same σ [33], meaning the variance decreases with increasing skewness. The opposite is true for a log-normal

distribution [19]. Contrasting the two distributions allows us to determine the effect of skewness on $\Delta\widetilde{\phi}$ and $\wp$.

The phase noise imparts additional imprecision to $\widetilde{\phi}$ by adding uncertainty to ϕ in each measurement step. As such, the effect of the phase noise should scale with N, although the mathematical relationship between the properties of the distribution and $\wp$ is not known. We found that the skewness does not correlate with $\wp$ whereas the variance does. Therefore, we only report the variance for the selected noise models and parameters in this paper.

B. Random telegraph noise

A random telegraph noise [17] follows a discrete probability distribution

$$p(\phi) = \begin{cases} 1 - P_{\mathrm{s}}, & \phi = \phi_0, \\ \frac{P_{\mathrm{s}}}{2}, & \phi = \phi_0 - \delta, \\ \frac{P_{\mathrm{s}}}{2}, & \phi = \phi_0 + \delta. \end{cases} \tag{3}$$

We indicate the probability switching to an erroneous value of ϕ by P_{s}. The parameter δ indicates the distance between ϕ_0 and two side peaks at $\phi_0 + \delta$ and $\phi_0 - \delta$. We keep $P_{\mathrm{s}} = 1/2$, which leads to the distribution being unimodal. The variance of a symmetric random telegraph noise is $V(\phi) = P_{\mathrm{s}}\delta^2$.

The random telegraph noise is implemented by conditioning the value of ϕ on a uniform random number. If the random number is less than P_{s}, ϕ is either $\phi_0 - \delta$ or $\phi_0 + \delta$ with 50/50 chance.

C. Skew normal noise

Skew normal noise follows a continuous distribution [18]

$$p(\phi) = \frac{\mathrm{e}^{-\frac{(\phi-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \left(1 + \mathrm{erf}\left(\frac{\alpha}{\sqrt{2}\sigma}(\phi - \mu)\right)\right), \tag{4}$$

defined for $\phi \in (-\infty, \infty)$, where erf() is the error function. A Gaussian distribution is recovered when $\alpha = 0$, with μ indicating the mean and σ the width of the distribution. As the ratio $|\alpha/\sigma|$ increases, the distribution narrows and becomes asymmetric. The distribution peaks close to μ, although the closed-form for the mode is not known. The variance for this distribution is $V(\phi) = \sigma^2\left(1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}\right)$.

The method for generating the random number starts with a sampling from a skew-normal distribution for $\mu = 0$, $\sigma = 1$, and a desired α [34]. We then rescale and shift the random number by $\phi \mapsto \phi\sigma + \phi_0$.

D. Log-normal noise

A log-normal noise follows a heavy-tailed distribution of the form [35]

$$p(\phi) = \frac{\mathrm{e}^{-\frac{(\log\phi-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma\phi}. \tag{5}$$

The variance of the distribution is $V(\phi) = \left(\mathrm{e}^{\sigma^2} - 1\right)\mathrm{e}^{2\mu+\sigma^2}$.

As the log-normal distribution is defined for $\phi \in (0, \infty)$, we first generate a random number within the supporting domain given μ and σ using the rejection sampling approach [36]. To shift the mode to ϕ_0, we compute the mode for the distribution from $\mathrm{e}^{\mu-\sigma^2}$ and the distance $\phi_0 - \mathrm{e}^{\mu-\sigma^2}$ between the mode and the true phase shift. The random number generated is then shifted by $\phi \mapsto \phi + (\phi_0 - \mathrm{e}^{\mu-\sigma^2})$.

TABLE I
THE POWER-LAW SCALING $\wp$ COMPUTED FROM THE LINEAR FIT OF $\log V_{\mathrm{H}}$ VS. $\log N$

Noise model	**Parameters**	ϕ **variance**	$2\wp$	R^2
SQL			1	
HL			2	
ideal			1.4397	0.9991
telegraph	$P_{\mathrm{s}} = 0.5, \delta = 0.2$	0.02	1.4147	0.9994
	$P_{\mathrm{s}} = 0.5, \delta = 0.5$	0.125	1.3495	0.9994
	$P_{\mathrm{s}} = 0.5, \delta = 1.0$	0.5	1.2703	0.9995
skew-normal	$\sigma = 0.5, \alpha = 0$	0.25	1.3411	0.9995
	$\sigma = 0.5, \alpha = 2.5$	0.112797	1.3815	0.9994
	$\sigma = 0.5, \alpha = 10$	0.0924208	1.3915	0.9993
log-normal	$\mu = 0.2, \sigma = 0.2$	0.063367	1.4785	0.9944
	$\mu = 0.2, \sigma = 0.5$	0.544062	1.3062	0.9985
	$\mu = 0.2, \sigma = 1.0$	6.96798	1.1363	0.9976

IV. METHOD

In this section, we explain how the data are generated and the scaling of the modified Holevo variance are computed. We run the reinforcement-learning-based adaptive phase estimation including phase noise on a computer cluster of 48 CPUs and collect the modified Holevo variance for accepted policies.

For random telegraph noise and skew-normal noise, we generate the policies for $N = \{4, 5, \ldots, 100\}$, giving the wall time of 48 hours. For the log-normal noise, we generate data up to $N = 30$ as the rejection-sampling method for generating random numbers are time-consuming, taking 5 hours to generate data for $N = \{4, 5, \ldots, 30\}$. As such, the power-law scaling $\wp$ of V_{H} from the log-normal noise should not be directly compared with the data from random telegraph noise and skew-normal noise but taken as suggestions for whether the learning algorithm is able to generate successful policies.

For each noise model, we collect data for three parameter sets. For asymmetric distributions, we vary only the parameters associated with skewness in order to determine how the asymmetry affects $\Delta\widetilde{\phi}$ and $\wp$. We also compute the variance of the distributions to show the correlation with the scaling $\wp$.

The power-law scaling $\wp$ is computed by linearly fitting $\log V_{\mathrm{H}}$ vs $\log N$. The data from the ideal interferometer is used to generate the SQL and HL from the intercept of the linear fit. The SQL and HL computed in this fashion are used to provide theoretical benchmarks for the adaptive scheme given an ideal interferometer.

V. RESULTS

In this section, we present the modified Holevo variance obtained from applying the reinforcement-learning algorithm to adaptive phase estimation in the presence of one of the three noise models (Fig. 3) and $\wp$ computed from these data (Table I).

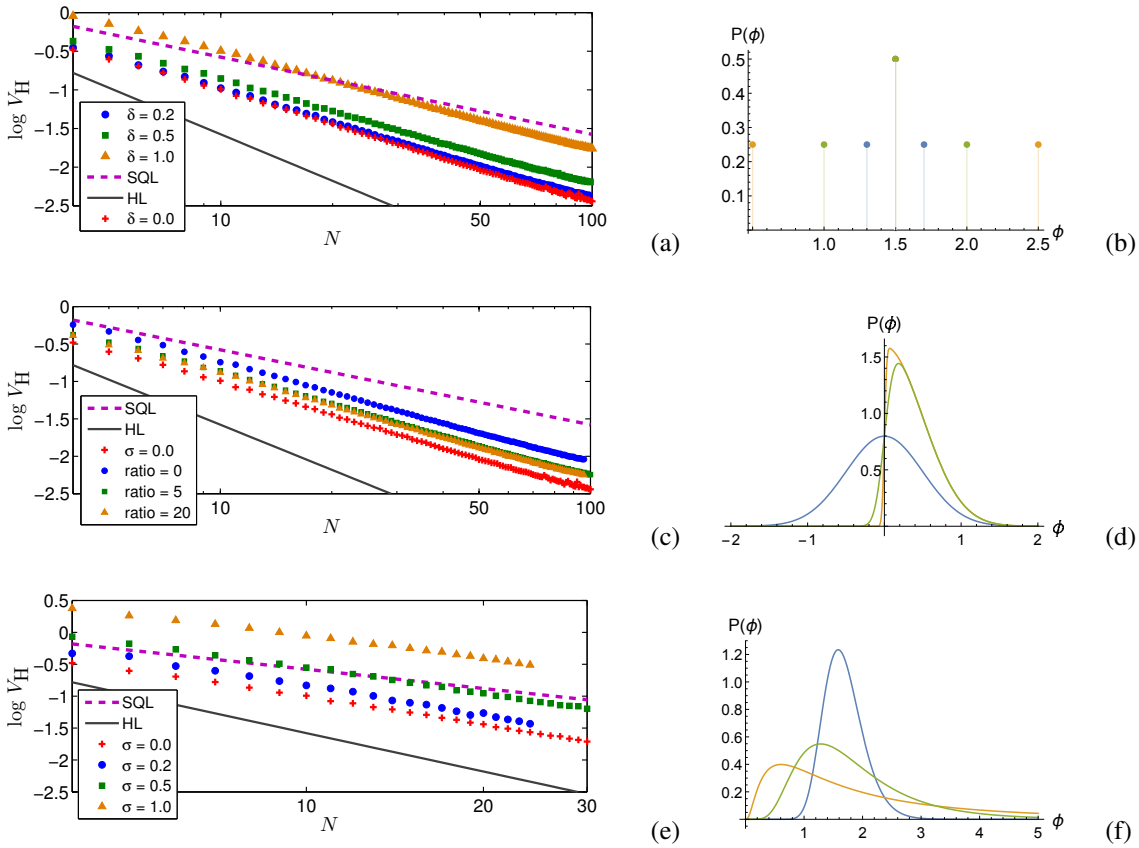

Fig. 3. Logarithmic plots of $V_{\rm H}$ versus number of particles in the presence of phase noise. The red crosses are the results from an ideal interferometer. The intercept from these data are used to calculate the SQL (purple dash line) and HL (solid black line). (a) $V_{\rm H}$ in the presence of random telegraph noise with $P_{\rm s} = 0.5$. Three values of δ are used in the simulation: 0.2 rad (blue circles), 0.5 (green squares), and 1.0 (orange triangle). (b) The random telegraph distributions for $P_{\rm s} = 0.5$ and three values of δ: 0.2 (blue), 0.5 (green) and 1.0 (orange). (c) $V_{\rm H}$ in the presence of skew-normal noise with $\sigma = 0.5$. Three values of α/σ are used in the simulation: 0 (blue circles), 5 (green squares), and 20 (orange triangle). (d) The skew-normal distributions for $\sigma = 0.5$ and three values of α/σ: 0 (blue), 5 (green) and 20 (orange). (e) $V_{\rm H}$ in the presence of log-normal noise with $\mu = 0.2$. Three values of σ are used in the simulation: 0.2 (blue circles), 0.5 (green squares), and 1.0 (orange triangle). (f) The log-normal distributions for $\mu = 0.2$ and three values of σ: 0.2 (blue), 0.5 (green) and 1.0 (orange).

Fig. 3b shows the distributions for random telegraph noise plotted for three parameter sets that are used to generate $V_{\rm H}$ in Fig. 3a. The scaling $\wp$ computed from these data exceed SQL (Table I) with the tendency towards the SQL as δ, and therefore the variance, increases.

A skew-normal distribution narrows as skewness, determined by α, increases (Fig. 3d). At $\alpha = 0$, the distribution takes the form of a Gaussian distribution with the width of σ. Without changing σ, we increases α and found that the values of $\wp$ increases, correlating to the decrease in variance rather than the increase in skewness (Fig. 3c).

The data from the log-normal distributions (Fig. 3f) shows the same trend between $\wp$ and the variance of the distribution (Fig. 3e). Fixing the parameter μ, the variance of log-normal distribution increases with skewness determined by σ. As a result, $\wp$ computed from the distribution with the largest σ in the three sets also is the closest to $\wp$ of the SQL.

VI. Discussion

In this section, we discuss the findings in Sec. V. The power-law scaling of the modified Holevo variance in the presence of the three noise models all exceed the SQL for the selected noise parameters (Table I), including the one for log-normal noise with variance of 6.96798, which is a large variance for a variable in $[0, 2\pi)$. Hence, our reinforcement-learning algorithm based on Gaussian phase noise is robust.

The data in Table I also suggest that $\wp$ decreases with the increase in noise variance, and the algorithm is speculated to not be robust for strong noise beyond the variance of 6.96798. However, the limit for when the learning algorithm fails is not yet known. Also, the inverse correlation between variance and $\wp$ is only shown for the phase noise of the same functional form. A conclusive relationship between variance and $\wp$ cannot yet be established due to limited data.

VII. Conclusion

In this paper, we test the robustness of our reinforcement-learning algorithm for adaptive interferometric phase estima-

tion under various phase-noise models. The algorithm is said to be robust if $\wp > 1/2$ for all models. We show that for random telegraph noise, skewed-normal noise, and log-normal noise, our reinforcement-learning algorithm is robust. The decrease in $\wp$ with the increase in noise variance suggests that the algorithm is not robust for strong noise, although the value of the variance in which $\wp$ approaches SQL is not known. Determining the noise parameters where the algorithm fails and determining the bias due to asymmetric phase noise are topics for future work.

ACKNOWLEDGMENT

We thank NSERC and AITF for financial support. P.W. acknowledges financial support from the ERC (Consolidator Grant QITBOX), MINECO (Severo Ochoa grant SEV-2015-0522 and FOQUS), Generalitat de Catalunya (SGR 875), and Fundació Privada Cellex. B.C.S. also acknowledges support from China's 1000 Talent Plan (Grant No. GG2340000241). The computational work is enabled by the support of WestGrid (www.westgrid.ca) through Compute Canada Calcul Canada (www.computecanada.ca).

REFERENCES

[1] M. Tiersch, E. J. Ganahl, and H. J. Briegel, "Adaptive quantum computation in changing environments using projective simulation," *Sci. Rep.*, vol. 5, 2015, doi:10.1038/srep12874.

[2] S. Mavadia, V. Frey, J. Sastrawan, S. Dona, and M. J. Biercuk, "Prediction and real-time compensation of qubit decoherence via machine-learning," *Nat. Commun*, vol. 8, Jan 2017, doi:10.1038/ncomms14106.

[3] A. Hentschel and B. C. Sanders, "Machine learning for precise quantum measurement," *Phys. Rev. Lett.*, vol. 104, no. 6, Feb 2010, doi:10.1103/PhysRevLett.104.063603.

[4] V. Dunjko, J. M. Taylor, and H. J. Briegel, "Quantum-enhanced machine learning," *Phys. Rev. Lett.*, vol. 117, no. 13, Sep 2016, doi:10.1103/physrevlett.117.130501.

[5] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *arXiv:1611.09347*, 2016. [Online]. Available: https://arxiv.org/abs/1611.09347

[6] V. Dunjko, N. Friis, and H. J. Briegel, "Quantum-enhanced deliberation of learning agents using trapped ions," *New J. Phys.*, vol. 17, no. 2, Jan 2015, doi:10.1088/1367-2630/17/2/023006.

[7] L. Lamata, "Basic protocols in quantum reinforcement learning with superconducting circuits," *Sci. Rep.*, vol. 7, no. 1, May 2017, doi:10.1038/s41598-017-01711-6.

[8] H. M. Wiseman, D. W. Berry, S. D. Bartlett, B. L. Higgins, and G. J. Pryde, "Adaptive measurements in the optical quantum information laboratory," *IEEE J. Sel. Top. Quantum Electron.*, vol. 15, no. 6, pp. 1661–1672, Nov 2009, doi:10.1109/JSTQE.2009.2020810.

[9] P. Cappellaro, "Spin-bath narrowing with adaptive parameter estimation," *Phys. Rev. A*, vol. 85, no. 3, Mar 2012, doi:10.1103/PhysRevA.85.030301.

[10] V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," *Nat. Photon*, vol. 5, no. 4, pp. 222–229, 2011, doi:10.1038/nphoton.2011.35.

[11] G. Tóth and I. Apellaniz, "Quantum metrology from a quantum information science perspective," *J. Phys. A: Math. Theor.*, vol. 47, no. 42, 2014, doi:10.1088/1751-8113/47/42/424006/.

[12] B. P. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett.*, vol. 116, no. 6, Feb 2016, doi:10.1103/PhysRevLett.116.061102.

[13] J. Borregaard and A. S. Sørensen, "Near-Heisenberg-limited atomic clocks in the presence of decoherence," *Phys. Rev. Lett.*, vol. 111, no. 9, Aug 2013, doi:10.1103/PhysRevLett.111.090801.

[14] J. A. Jones, S. D. Karlen, J. Fitzsimons, A. Ardavan, S. C. Benjamin, G. A. D. Briggs, and J. J. L. Morton, "Magnetic field sensing beyond the standard quantum limit using 10-spin NOON states," *Science*, vol. 324, no. 5931, pp. 1166–1168, 2009, doi:10.1126/science.1170730.

[15] M. J. W. Hall, D. W. Berry, M. Zwierz, and H. M. Wiseman, "Universality of the Heisenberg limit for estimates of random phase shifts," *Phys. Rev. A*, vol. 85, no. 4, Apr 2012, doi:10.1103/PhysRevA.85.041802.

[16] A. Hentschel and B. C. Sanders, "Efficient algorithm for optimizing adaptive quantum metrology processes," *Phys. Rev. Lett.*, vol. 107, no. 23, Nov 2011, doi:10.1103/PhysRevLett.107.233601.

[17] D. S. Newman, "On the probability distribution of a filtered random telegraph signal," *Ann. Math. Stat.*, vol. 39, no. 3, pp. 890–896, 1968.

[18] A. Azzalini and A. Capitanio, "The skew-normal distribution: probability," in *The Skew-Normal and Related Families*, ser. Institute of Mathematical Statistics Monograph, D. R. Cox, A. Agresti, B. Hambly, S. Holmes, and X.-L. Meng, Eds. New York: Cambridge University Press, 2014, ch. 2, pp. 24 – 56.

[19] S. Foss, D. Korshunov, and S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions*, ser. Operations Research and Financial Engineering, T. Mikosch, S. I. Resnick, and S. M. Robinson, Eds. New York: Springer, 2011.

[20] R. D. Nowak, "The geometry of generalized binary search," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7893–7906, Dec 2011, doi:10.1109/TIT.2011.2169298.

[21] 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, May 2001, doi:10.1103/PhysRevA.63.053804.

[22] A. Hentschel and B. C. Sanders, "Ordered measurements of permutationally-symmetric qubit strings," *J. Phys. A: Math. Theor.*, vol. 44, no. 11, Feb 2011, doi:10.1088/1751-8113/44/11/115301.

[23] G. Summy and D. Pegg, "Phase optimized quantum states of light," *Opt. Commun.*, vol. 77, no. 1, pp. 75 – 79, Jun 1990, doi:10.1016/0030-4018(90)90464-5.

[24] 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, Dec 2000, doi:10.1103/PhysRevLett.85.5098.

[25] K. J. Astrom, "Optimal control of Markov decision processes with incomplete state estimation," *J. Math. Anal. Appl.*, vol. 10, no. 1, pp. 174–205, 1965.

[26] A. Barto and T. Dietterich, "Reinforcement learning and its relationship to supervised learning," in *Handbook of Learning and Approximate Dynamic Programming*, ser. Compuational Intelligence, J. Si, A. Barto, W. Powell, and D. Wunsch, Eds. New Jersey: IEEE, 2004, ch. 2, pp. 47–60.

[27] R. S. Sutton and A. G. Barto, "Planning and learning," in *Reinforcement Learning : An Introduction.*, ser. Adaptive Computation and Machine Learning. Massachusetts: A Bradford Book, 1998, ch. 9, pp. 227–254.

[28] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992, doi:10.1023/A:1022676722315.

[29] R. Storn and K. Price, "Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997, doi:10.1023/A:1008202821328.

[30] N. B. Lovett, C. Crosnier, M. Perarnau-Llobet, and B. C. Sanders, "Differential evolution for many-particle adaptive quantum metrology," *Phys. Rev. Lett.*, vol. 110, no. 22, May 2013, doi:10.1103/PhysRevLett.110.220501.

[31] S. Das, A. Konar, and U. K. Chakraborty, "Improved differential evolution algorithms for handling noisy optimization problems," in *Proc. of CEC-05, 7th Congress on Evolutionary Computation*, vol. 2, 2005, pp. 1691–1698, doi:10.1109/CEC.2005.1554892.

[32] P. Palittapongarnpim, P. Wittek, and B. C. Sanders, "Controlling adaptive quantum-phase estimation with scalable reinforcement learning," in *Proc. of ESANN-16, 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2016, pp. 327–332.

[33] A. Capitanio, "On the approximation of the tail probability of the scalar skew-normal distribution," *METRON*, vol. 68, no. 3, pp. 299–308, Dec 2010, doi:10.1007/BF03263541.

[34] D. Ghorbanzadeh, L. Jaupi, and P. Durand, "A method to simulate the skew normal distribution," *Appl. Math.*, vol. 5, no. 13, pp. 2073–2076, 2014, doi:10.4236/am.2014.513201.

[35] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience*, vol. 51, no. 5, pp. 341–352, 2001, doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

[36] B. D. Flury, "Acceptance-rejection sampling made easy," *SIAM Rev.*, vol. 32, no. 3, pp. 474–476, Sep 1990.