Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 14, Number 1, pages 74–76, January 2025

Quantum state generation using model-based deep reinforcement learning

1st Yasuhiro Kishida
Engineering Department
The University of Electro-Communications
Tokyo, Japan
k2333032@gl.cc.uec.ac.jp

2nd Ryosuke Koga
Engineering Department
The University of Electro-Communications
Tokyo, Japan
k2433050@gl.cc.uec.ac.jp

3rd Tomah Sogabe
Engineering Department
The University of Electro-Communications
Tokyo, Japan
sogabe@uec.ac.jp

***Abstract*—In the pursuit of realizing a quantum computer through quantum gating, generating a Schrödinger's cat state that is robust against noise and decoherence presents a significant challenge. Recent studies have explored deep reinforcement learning methods; however, the results often fail to accurately reflect the state of the quantum system due to quantum back-action during the observation process. In this study, we propose a novel approach for quantum state generation that integrates particle filters with deep reinforcement learning. The particle filter estimates the quantum system's state based on observed results and subsequently provides feedback to the reinforcement learning agent. We compare our current findings on generating Schrödinger's cat state with previous results derived from deep reinforcement learning techniques.**

Keywords—Quantum State Generation, Deep Reinforcement Learning, Particle Filter

I. Introduction

A. Background

The advancement of quantum computing based on quantum gates hinges on overcoming several critical challenges. One of the foremost requirements is the generation of a ground state that is resilient to noise and decoherence from the environment, serving as the initial state for computation [1-3]. However, in practical physical systems, noise and decoherence are inevitable. As a result, implementing effective feedback control based on real-time monitoring of quantum systems has become essential [4].

Traditional optimal control techniques have proven effective in linear, unitary, and deterministic systems. However, modeling nonlinear quantum systems remains a significant challenge, and currently, no general methods are available to address these complex dynamics [6,27,41-47]. Recently, model-free approaches based on deep reinforcement learning (DRL) have gained attention, as they allow control strategies to be learned directly from data patterns generated by the system, without the need for an explicit physical mode [14].

Nevertheless, conventional DRL frameworks face considerable difficulties when applied to quantum systems, as these systems exhibit nonlinear and stochastic time evolution due to quantum back-action induced by continuous observation.

To address these challenges, this study combines DRL with particle filtering to achieve the ground state within a double-well potential system. Our approach overcomes these difficulties by employing particle filtering to estimate the quantum system's state in real time, followed by the heuristic application of DRL to develop an optimal control policy that maintains the desired state.

B. Related Works

Numerous applications of deep reinforcement learning (DRL) have emerged in the quantum domain, encompassing areas such as quantum control [7,15-17], quantum state preparation and engineering [18-22], state transitions [23-25], and quantum error correction [26,27]. While research utilizing DRL is on the rise, very few studies have explicitly considered the use of continuous measurement outcomes for training DRL agents [7,19,26]. A notable contribution that addresses this gap is the work by Borah et al. [14], which investigates measurement-based feedback quantum control through deep reinforcement learning in a double-well nonlinear potential. This study underscores the effectiveness of continuous measurement results in enhancing the training of DRL agents for quantum control applications [14].

C. Our Results

In this study, we develop and analyze a deep reinforcement learning (DRL) framework integrated with a particle filter to achieve high-fidelity quantum state control under continuous measurement feedback within a nonlinear double-well potential system. In this framework, we choose the coherence state with an inverse temperature parameter $\beta = 1$ as the initial state, and the DRL agent learns an optimal control strategy to generate the ground state, specifically targeting the Schrödinger's cat state. By varying the particle count in the particle filter, we investigate the impact of particle number on the accuracy and stability of quantum state estimation and control.

Our extensive simulations demonstrate that a higher particle count significantly enhances both fidelity and reward stability, enabling the DRL agent to effectively learn and maintain a robust control strategy. Specifically, a particle count of $N = 40$ consistently achieves stable, high-fidelity control, whereas lower particle counts ($N = 10$ and $N = 20$) lead to noticeable performance fluctuations, highlighting challenges in achieving reliable state estimation and control with fewer particles.

Future work is anticipated to explore more complex initial states to broaden the applicability of this approach. Our findings emphasize the critical role of particle count in quantum control systems.

II. Methods

In this study, we model the quantum evolution of a quantum particle within a double-well (DW) potential using a stochastic master equation (SME) to describe the dynamics. Continuous measurement at a rate Γ on the operator x^2 is applied to prevent the wave function from localizing in either well due to even parity [28]. This continuous measurement induces quantum back-action, introducing noise that affects both the conditioned quantum dynamics and the observed measurement data, ultimately preventing an accurate reflection of the quantum system's state ($\rho(t+1)_{inc.}$, $dQ(\rho(t+1)_{inc.})$). To obtain a precise state $\rho(t+1)_{corr.}$ of the quantum system, we estimate the state using a particle filter ($\rho(t+1)^{\mathrm{est}}_{corr.}$), and feedback this estimated information ($dQ(\rho(t+1)^{\mathrm{est}}_{corr.})$) to a deep reinforcement learning (DRL) agent. The DRL agent controls the quantum dynamics through modulation of the Hamiltonian $H'(t) = H + F(t)$, where $F(t) = \mathcal{A}(t)(xp + px)$, with x and p being dimensionless canonical operators, and $\mathcal{A}(t)$ representing the strength of the squeezing operator modulated by the DRL agent. The DRL agent is trained via the continuous measurement current and acts on the system in real time through $F(t)$. A diagram of the DRL model used in this study is shown in Figure 1(a).

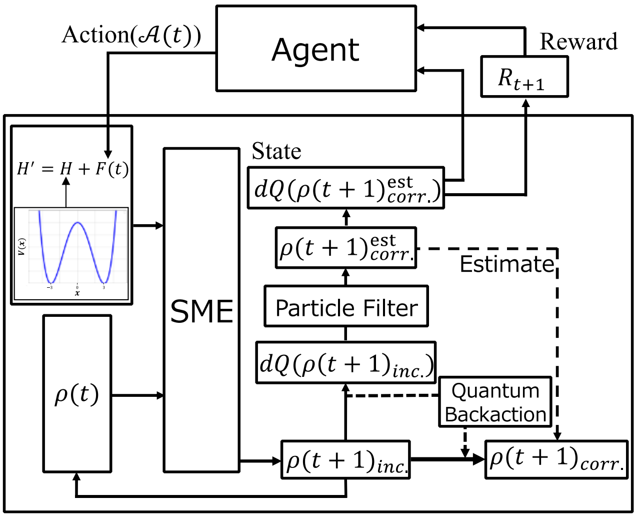

Fig. 1. Overview of the DRL-Based Quantum Control Framework

A. Environment

This model uses dimensionless position and momentum variables (x, p), which relate to physical position and momentum as $x = Q/Q_0$ and $p = P/P_0$, where Q_0 and P_0 are scaling factors for position and momentum. The canonical commutation relation is $[\hat{Q}, \hat{P}] = i\hbar$, resulting in $[\hat{x}, \hat{p}] = i\bar{k}$, where the dimensionless Planck constant is given by $\bar{k} = \hbar/(Q_0 P_0)$. The DW potential we consider is aligned along the x-axis and characterized by the quantum particle's Hamiltonian:

$$H = \frac{p^2}{2} + \frac{h}{b^4}[(x-a)^2 - b^2]^2 \qquad (1)$$

where b indicates the position of the well minima, h is the height of the barrier between the wells, and a is the offset along x. Parameters are set as $a = 0$, $b = 3$, and $h = 5$, creating a symmetric potential around the origin at $x = 0$. The ground state represented by this Hamiltonian H is an even-parity "cat state" due to symmetry in x and p, while the first excited state has odd paritytem's conditioned density operator, $\rho(t)$, evolves over time according to a quantum stochastic master equation(SME) [14] stochastic measurement record up to time t:

$$\text{SME: } d\rho(t) = -i[H', \rho]dt + \mathcal{D}[A]\rho dt + \mathcal{H}[A]\rho dW(t) \quad (2)$$

In this study, the initial state $\rho(0)$ was set to a coherence state with an inverse temperature parameter $\beta = 1$. At the beginning of each episode, the environment is reset to the initial state, with the initial density matrix defined as $\rho(0)$.

Here, A is a Hermitian observable operator under continuous measurement (known as the measurement operator), and $\mathcal{H}[A]$ and $\mathcal{D}[A]$ are super-operators given by:

$$\mathcal{H}[A]\rho(t) = [\{A, \rho(t)\} - tr(\{A, \rho(t)\})]\rho(t) \qquad (3)$$

$$\mathcal{D}[A]\rho(t) = \frac{1}{2}[2A\rho(t)A^\dagger - \{\rho(t), A^\dagger A\}] \qquad (4)$$

In Equation (3), $dW(t)$ represents the Wiener increment, a noise term introduced by continuous measurement with a mean of zero and variance dt. The measurement current $dQ(t)$ is a classical stochastic process that satisfies an Itô stochastic differential equation:

$$dQ(\rho(t+1)_{inc.}) =$$

$$\gamma g \left(tr(A(t) \cdot \rho(t+1)_{inc.})dt + \frac{1}{\sqrt{4\Gamma}} dW(t) \right) \qquad (5)$$

where g is a gain coefficient, inverse in units to A, and $dQ(t)$ has units of frequency. Here, $A = \sqrt{\Gamma}x^2$, with Γ as the measurement rate, quantifying the quality of measurement. Since x and p are dimensionless, we set $\gamma g = 1$. For continuous cooling of the system to the ground state (cat state), we select the stochastic operator A in Equation (3) as $\sqrt{\Gamma}x^2$.

The time density operator calculated using Equations (2) and (5), along with weak measurement values, does not fully reflect the quantum state due to quantum back-action and noise ($\rho(t+1)_{inc.}$, $dQ(\rho(t+1)_{inc.})$). Therefore, in this study, we estimate the precise state of the quantum system $\rho(t+1)_{corr.}$ using a particle filter ($\rho(t+1)^{\mathrm{est}}_{corr.}$), and feedback this information ($dQ(\rho(t+1)^{\mathrm{est}}_{corr.}$) to the DRL agent. The particle filter can adapt to nonlinear state-space models and non-Gaussian noise. By taking the difference between the filter particle state inferred from previous measurements and the current measurement values, we create a distribution of the noise present at the current time.

$$dW^{(i)}(t) = \sqrt{4\Gamma}\left(\frac{dQ(\rho(t+1)_{inc.})}{\gamma g} - \langle A(t)\rangle_c^{(i)} dt\right) \quad (7)$$

Using the noise derived from Equation (7), we evolve the conditioned density operator over time according to Equation (3) and apply weights based on measurement values to estimate the quantum system state.

$$\alpha^{(i)}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{dQ(t) - \left(\langle A(t)\rangle_c^{(i)} dt + \frac{1}{\sqrt{0.4}} dW^{(i)}(t)\right)}{2\sigma^2}\right] \tag{8}$$

$$\rho(t+1)^{\text{est}}_{corr.} = \frac{\sum_i \rho'^{(i)}(t)\alpha^{(i)}(t)}{\alpha^{(i)}(t)} \tag{9}$$

This expectation is then fed back to the DRL agent.

$$dQ(\rho(t+1)^{\text{est}}_{corr.}) = tr(A(t) \cdot \rho(t+1)^{\text{est}}_{corr.}) \tag{10}$$

$$R_{t+1} = -|dQ(\rho(t+1)^{\text{est}}_{corr.}) - 3^2| \tag{11}$$

B. *DRL Agent*

In this study, we used a DRL agent based on Proximal Policy Optimization (PPO) combined with the Advantage Actor-Critic (A2C) framework. The PPO objective is optimizing a clipped surrogate loss function,

$$\mathcal{L}(\theta) = \hat{\mathbb{E}}_t\left(\min\left\{r_t(\theta)\hat{A}_t, \text{clip}[r_t(\theta), 1-\epsilon, 1+\epsilon]\right\}\right) \tag{12}$$

which is representing the probability ratio between current and previous policy actions, and $\hat{A}_t$ is the advantage function calculated by the critic. Limiting the clipping range to $\epsilon = 0.2$ prevents excessive updates, while the A2C framework enables parallel training environments, accelerating learning.

Recent studies have investigated the application of deep reinforcement learning, especially robust algorithms like PPO for continuous state-action spaces, in quantum state generation and control. For instance, Borah et al. (2021) investigated continuous measurement-based feedback quantum control in double-well nonlinear potentials using PPO [14]. Kuo et al. (2021) introduced a quantum architecture search framework using PPO to generate gate sequences for multi-qubit GHZ states without prior knowledge of quantum physics [29]. Additionally, Zhu & Hou (2023) improved this approach with a trust region-based PPO method, achieving better policy performance and reduced execution time [30]. Chen & Xue (2019) demonstrated the effectiveness of PPO in spin-squeezed state preparation in a spin-1 atomic system, both in mean-field and quantum systems [31]. Moro et al. (2021) optimized digital pulse sequences for stimulated Raman adiabatic passage (STIRAP), achieving fast and flexible solutions for integer and fractional STIRAP [32]. These studies underscore the potential of PPO in quantum state generation, architecture search, and control optimization.

In each training episode, the DRL interacts with the environment over 1000 discrete time steps at intervals of $dt = 0.001$, applying actions each time.

During each interaction, the PPO agent adds the squeezing term $F(t) = \mathcal{A}(t)(xp + px)$ to the Hamiltonian [Equation (2)] and attempts to adjust $\mathcal{A}(t)$ in the range $\mathcal{A}(t) \in [-5, 5]$ to maximize the reward given by Equation (11). This choice of feedback is motivated by the physics of the problem and analyzed in detail through Bayesian control using the conditional mean of the measurement record, following Stockton et al. [13]. The xp-type Hamiltonian terms can bed, for example, via quantum particle motion in a magnetic field [33].

III. Results and Discussions

This section discusses the simulation outcomes, focusing on the effectiveness of the proposed deep reinforcement learning (DRL) framework for quantum state estimation and control under various particle counts in the p article filter. The system's performance was evaluated using fidelity and reward metrics, which provide insight into the control strategy's accuracy and efficiency.

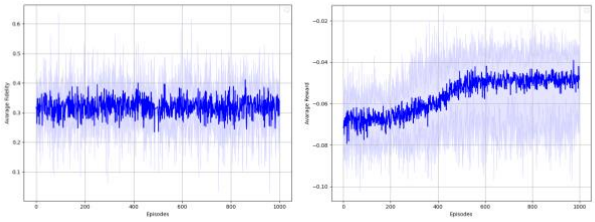

Fig. 2. Results with N=40 number of particles
(a). Average Fidelity Over Episodes
(b). Average Reward Over Episodes

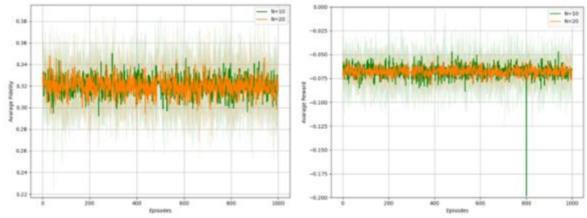

Fig. 3. Results with N=10 and 20 number of particles
(a). Average Fidelity Over Episodes
(b). Average Reward Over Episodes

Figure 2 shows the simulation results when the particle count N is set to 40. This high particle count enables a fine-grained approximation of the quantum state, supporting the DRL agent's learning and control capabilities. In Figure 2(a), the average fidelity remains around 0.3 across episodes, exhibiting limited improvement over time. This suggests that while the DRL framework can stabilize the quantum state to a certain degree, it faces challenges in achieving higher fidelity necessary for precise ground-state preparation. The relatively constant fidelity indicates that the current control strategy maintains basic alignment with the target state but lacks the refinements needed to enhance state fidelity further. In contrast, Figure 2(b) shows a more pronounced improvement in the reward curve. The average reward increases steadily from approximately −0.06 to −0.04 during the initial 500 episodes before stabilizing. This increase in reward reflects the DRL agent's learning process as it optimizes control actions based on continuous feedback, effectively balancing quantum back-action and measurement noise. However, the divergence between the stabilized reward and the relatively static fidelity suggests that while the DRL framework succeeds in maintaining system stability, additional modifications may be required to enhance the fidelity metric, aligning the controlled state more closely with the desired quantum ground state.

Figure 3 shows the simulation results when the particle count N is set to 10 and 20. In Figure 3(a), we see the fidelity curves for both particle counts, with $N = 10$ in green and $N = 20$ in orange. Unlike the results with $N = 40$, fidelity remains relatively low (around 0.3 to 0.34) and fluctuates significantly throughout the episodes, failing to reach a stable high-fidelity state. This instability suggests that lower particle counts limit the particle filter’s ability to accurately estimate the quantum state, thereby hindering the DRL agent's learning process and reducing control precision. Similarly, Figure 3(b) illustrates the reward curves for $N = 10$ and $N = 20$. Both curves show considerable volatility, with rewards remaining within the range of -0.075 to -0.05 throughout the episodes, without a clear upward trend or stable plateau. This result implies that the DRL agent struggles to maintain an effective control strategy with fewer particles due to lower estimation accuracy, as the particle filter provides insufficient information for the agent to learn an optimal policy.

These findings highlight the critical role of particle count in achieving stable learning and effective control. Lower particle counts lead to unstable and suboptimal fidelity and reward, indicating that the DRL agent requires a higher particle count for accurate state estimation and reliable control.

The results underscore both the potential and limitations of the proposed DRL-based quantum control framework. While increasing the particle count enhances reward stability and reduces fluctuations, the lack of improvement in fidelity suggests that further optimizations are necessary. This indicates that while the framework is effective in mitigating quantum back-action and noise, achieving high-fidelity state control may require adaptive filtering techniques or modifications to the DRL structure. Such enhancements could be critical for advancing the framework's precision in aligning the controlled state with the desired ground-state configuration.

IV. Conslusion

In this study, we introduced a DRL-based quantum control framework combined with particle filtering to address the challenge of quantum state control within a nonlinear double-well potential system under continuous measurement feedback. By varying the particle count in the particle filter, we demonstrated that a higher count (specifically $N = 40$) improves the stability of the reward, allowing the DRL agent to maintain a consistent control strategy over time. However, the fidelity metric showed limited improvement, indicating that while the framework can stabilize the control process, achieving precise ground-state preparation remains a challenge.

Our findings highlight the critical influence of particle count on the stability and performance of DRL-based quantum control systems. However, the limited increase in fidelity underscores the need for further refinement of the control methodology. Future research should explore adaptive particle filtering techniques and structural adjustments to the DRL framework to improve fidelity alignment with the target ground state. Additionally, investigating more complex initial states beyond the coherence state with an inverse temperature parameter $\beta = 1$ could extend the applicability of this approach, offering a more comprehensive understanding of the framework’s capabilities. Such advancements have the potential to enhance both computational efficiency and control accuracy in DRL-based quantum systems, paving the way for more robust applications in quantum computing and control.

References

[1] F. Verstraete, M. M. Wolf, and J. Ignacio Cirac, Nat. Phys. 5, 633 (2009).*)*

[2] M. Motta, C. Sun, A. T. K. Tan, M. J. O’Rourke, E. Ye, A. J.Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Nat. Phys. 16, 205 (2020).

[3] P. J. Love, Nat. Phys. 16, 130 (2020).

[4] H. M. Wiseman and G. J. Milburn, Quantum Measurement and Control (Cambridge University Press, Cambridge, 2009).

[5] J. Zhang, Y.-x. Liu, R.-B. Wu, K. Jacobs, and F. Nori, Phys. Rep. 679, 1 (2017).

[6] Z. T. Wang, Y. Ashida, and M. Ueda, Phys. Rev. Lett. 125, 100401 (2020).

[7] J. Werschnik and E. K. U. Gross, J. Phys. B 40, R175 (2007).

[8] A. P. Peirce, M. A. Dahleh, and H. Rabitz, Phys. Rev. A 37, 4950 (1988).

[9] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. 106, 190501 (2011).

[10] E. Zahedinejad, S. Schirmer, and B. C. Sanders, Phys. Rev. A 90, 032310 (2014).

[11] A. C. Doherty and K. Jacobs, Phys. Rev. A 60, 2700 (1999).

[12] A. C. Doherty, S. Habib, K. Jacobs, H. Mabuchi, and S. M. Tan, Phys. Rev. A 62, 012105 (2000).

[13] J. K. Stockton, R. van Handel, and H. Mabuchi, Phys. Rev. A 70, 022106 (2004).

[14] S. Borah, Phys. Rev. Lett., 127, 190403 (2021).

[15] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, npj Quantum Inf. 5, 33 (2019).

[16] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, npj Quantum Inf. 5, 85 (2019).

[17] H. Xu, L. Wang, H. Yuan, and X. Wang, Phys. Rev. A 103, 042615 (2021).

[18] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, npj Quantum Inf. 5, 85 (2019).

[19] J. Mackeprang, D. B. R. Dasari, and J. Wrachtrup, Quantum Mach. Intell. 2, 5 (2020).

[20] T. Haug, W.-K. Mok, J.-B. You, W. Zhang, C. E. Png, and L.-C. Kwek, Mach. Learn. 2, 01LT02 (2020).

[21] S.-F. Guo, F. Chen, Q. Liu, M. Xue, J.-J. Chen, J.-H. Cao, T.-W. Mao, M. K. Tey, and L. You, Phys. Rev. Lett. 126, 060401 (2021).

[22] M. Bilkis, M. Rosati, R. M. Yepes, and J. Calsamiglia, Phys. Rev. Research 2, 033295 (2020).

[23] R. Porotti, D. Tamascelli, M. Restelli, and E. Prati, Commun. Phys. 2, 1 (2019).

[24] Y. Ding, Y. Ban, J. D. Martín-Guerrero, E. Solano, J. Casanova, and X. Chen, Phys. Rev. A 103, L040401 (2021).

[25] I. Paparelle, L. Moro, and E. Prati, Phys. Lett. A 384, 126266 (2020).

[26] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Phys. Rev. X 8, 031084 (2018).

[27] H. P. Nautrup, N. Delfosse, V. Dunjko, H. J. Briegel, and N. Friis, Quantum 3, 215 (2019).

[28] K. Jacobs, L. Tian, and J. Finn, Phys. Rev. Lett. 102, 057208 (2009).

[29] E.-J. Kuo, Y.-L. L. Fang, and S. Y.-C. Chen, arXiv preprint arXiv:2104.07394 (2021).

[30] X. Zhu and X. Hou, Sci. Rep. 13, 32349 (2023).

[31] J.-J. Chen and M. Xue, arXiv preprint arXiv:1901.10351 (2019).

[32] L. Moro, I. Paparelle, and E. Prati, Int. J. Quantum Inf. 19, 2141002 (2021).

[33] E. Romero-Sánchez, W. P. Bowen, M. R. Vanner, K. Xia, and J. Twamley, Phys. Rev. B 97, 024109 (2018).