Creation and manipulation of quantized vortices in Bose-Einstein condensates using reinforcement learning

Hiroki Saito
Department of Engineering Science, University of Electro-Communications, Tokyo 182-8585, Japan

We apply the technique of reinforcement learning to the control of nonlinear matter waves. In this method, an agent controls the position, strength, and shape of an external Gaussian potential to create and manipulate quantized vortices in a Bose-Einstein condensate (BEC) trapped in a harmonic potential. The density and velocity distributions of the BEC at each moment obtained by the Gross-Pitaevskii evolution are directly input into a convolutional neural network to determine the next action of the agent. We demonstrate that a stationary single-vortex state can be produced in a two-dimensional system, and a stationary vortex-ring state can be produced in a three-dimensional system.

1. Introduction

Recent developments in machine learning have been remarkable. A standout example is the computer program “AlphaGo”,\(^1, 2\) which defeated the best human players in the game of Go. In AlphaGo, reinforcement learning with deep neural networks (called deep-Q learning) is used to evaluate the situation of the game and determine the next action. Another interesting example of the use of deep-Q learning was reported in Ref. 3, in which a computer agent playing video games is trained to get higher scores. It was demonstrated that the computer agent outperforms human players after training, without prior knowledge about the games.

In this study, we focus on the control of quantum systems by reinforcement learning.\(^4–17\) Controlling quantum systems and producing desired quantum states are important in a variety of areas in quantum physics. Reinforcement learning consists of an agent and the environment. In this case, the quantum system is regarded as the environment. An easily accessible initial state, such as the ground state, is first prepared, and during the time evolution, the agent makes decisions to control the time-dependent parameters in the Hamiltonian. The quantum state develops depending on the parameters determined by the agent, and the information of the quantum state is fed back to the agent. Depending on the quantum state, a reward is also given to the agent, and the agent is trained to maximize the reward. The reward is, for example, the overlap or fidelity between the controlling state and the target state.

In this paper, we consider a Bose-Einstein condensate (BEC) of an atomic gas as a quantum system to be controlled. Optimal control of a BEC has been studied for a variety of purposes, such as spatial transport,\(^18–20\) splitting into two BECs,\(^18, 21, 22\) squeezing of quantum fluctuations,\(^23, 24\) excitation to specific states,\(^21, 25–29\) changing trap geometry,\(^22, 30\) improving interferometry,\(^31\) and creating a self-bound droplet.\(^32\) In these studies, quantum control theories, such as GRAPE\(^33\) and CRAB\(^34, 35\) are used. Recently, machine learning techniques have been used for the optimized creation of a BEC\(^36–39\) as well as the transport and decompression\(^39\) of a BEC.

Figure 1 illustrates reinforcement learning to control a BEC. We aim to create and manipulate quantized vortices in a BEC by changing a Gaussian external potential. Experimentally, such a potential can be produced by a nonresonant Gaussian laser beam applied to the BEC. On the environmental side, the dynamics of the BEC with a time-dependent Gaussian potential are obtained by numerically solving the Gross-Pitaevskii (GP) equation, and the state of the BEC and the reward are given to the agent. The agent is trained to get a higher reward, which leads to the production of a desired state of the BEC.
locity of the BEC are directly input into the CNN, giving the next action as an output. At each time step, the agent also receives reward from the environment. The reward is higher when the state of the BEC is closer to the target state. The CNN of the agent is trained to maximize the total amount of rewards, which enables the agent to discover an optimal control of the Gaussian potential.

Using the above scheme, we demonstrate that quantized vortices can be created and manipulated in two-dimensional (2D) and three-dimensional (3D) BECs. In a 2D system, the target state is set to be the state with a singly-quantized vortex at the center. Although both vortices and antivortices are created in pairs by a simple translation of the Gaussian potential, the agent finds a clever way to generate a single vortex state to the agent. The environment also evaluates whether the action of the agent was a good action, and gives a reward to the agent.

The remainder of the paper is organized as follows. Section 2 explains our method, Sec. 3 shows the numerical results, and Sec. 4 gives the conclusions to this study.

2. Method

2.1 Environment

We begin by describing the environment of the reinforcement learning system. As described above, the environment receives the action from the agent and evolves to the next time step depending on the action. The environment then relays its state to the agent. The environment also evaluates whether the action of the agent was a good action, and gives a reward to the agent.

As an environment, we consider a BEC of bosonic atoms with mass $m$ confined in a harmonic potential. In the mean-field approximation at zero temperature, the macroscopic wave function $\psi(r, t)$ obeys the GP equation given by

$$i \hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V_{\text{traps}}(r)\psi + V_G(r, t)\psi + g|\psi|^2\psi,$$  

where $V_{\text{traps}} = m[\omega_r^2(x^2+y^2)+\omega_z^2z^2]/2$ is the harmonic trap potential with $\omega_r$ and $\omega_z$ being the radial and axial trap frequencies, $V_G$ is the Gaussian external potential, and $g = 4\pi\hbar^2a/m$ is the interaction coefficient with $a$ being the s-wave scattering length of the atom. The wave function is normalized as $\int |\psi|^2 d\mathbf{r} = N_{\text{atom}}$, where $N_{\text{atom}}$ is the number of atoms.

In the next section, we will investigate both 2D and 3D systems. For the 3D system, for simplicity, the harmonic trap potential is assumed to be isotropic, i.e., $\omega_r = \omega_z \equiv \omega$. The external Gaussian potential for the 3D problem has the form

$$V_G(x, y, t) = A(t) \exp \left\{ \frac{\xi^2}{d_y^2(t)} + \frac{(z - \zeta(t))^2}{d_z^2} \right\},$$

where the strength $A(t) > 0$, width $d_y(t) > 0$, and position $\zeta(t)$ are control parameters, and the width $d_z$ is a constant. Such an external potential can be produced by a blue-detuned Gaussian laser beam propagating in the $z$ direction.

The 2D system is experimentally realized by tight confine-
A greedy policy:
observes the new state (100 episodes), and a constant value of $\epsilon$ environment, where the Gaussian external potential is changed. $\epsilon$ indicates the value of $\epsilon$.

The original CNN to generate value, where $\hat{Q}$ is expressed by $32 \times 32 \times 32 \times 32 \times 3$.

Structures of the CNNs for the 2D and 3D cases. The density and flux distributions of the BEC and the Gaussian potential distribution are in-

Algorithm 1

Initialize deep-Q network $Q$

for episode $= 1$, $N_{episode}$ do

Initialize environment and get initial observation $s_0$

for $t = 0$, $T$ end do

Select action $a_t = \text{argmax}_a Q(s_t, a)$ with $\epsilon$-greedy exploration

Execute action $a_t$ on environment and get reward $r_t$ and next state $s_{t+\Delta t}$

Store $(s_t, a_t, r_t, s_{t+\Delta t})$ in replay memory

Sample minibatch of $(s_t, a_t, r_t, s_{t+\Delta t})$ from replay memory

Set $y_t = r_t + \gamma \hat{Q}(s_{t+\Delta t}, a')$, where $a' = \text{argmax}_a Q(s_{t+\Delta t}, a)$

Train network using gradient of $L(y_t - Q(s_t, a_t))$

Copy $\hat{Q} = Q$ every $C$ steps

end for

end for

3. Results

We numerically demonstrate that our method can create and manipulate vortices in BECs. In the following, we normalize the length, time, and energy by $|\hbar/(m\omega)|^{3/2}$, $\omega^{1/2}$, and $\hbar\omega$, respectively. The density $|\psi|^2$ is normalized by $N_{\text{atom}}/(m\omega)$ in 2D, and $|\psi|^2$ is normalized by $N_{\text{atom}}(m\omega)/(h^3)$ in 3D. For 2D calculations, the normalized interaction coefficient $\tilde{g}_\perp = N_{\text{atom}}(8\pi m\omega)/(h^3)$ is taken to be 1000. For example, for $^{87}$Rb atoms in a quasi-2D trap with $\omega = 2\pi \times 100$ Hz and $\omega$ = $2\pi \times 10$ kHz, $\tilde{g}_\perp = 1000$ corresponds to $N_{\text{atom}} \approx 4.1 \times 10^3$. For 3D calculations, the normalize-
ized interaction coefficient \( \tilde{g} = 4\pi N_{\text{atom}}(\hbar/m\omega)^{1/2} \) is taken to be 6000. For a 3D isotropic trap with \( \omega = 2\pi \times 100 \text{ Hz} \), \( \tilde{g} = 6000 \) corresponds to \( N_{\text{atom}} = 9.7 \times 10^4 \). The spatial discretization size is 0.15 with a 128×128 mesh for the 2D case, and 0.25 with a 64×64×64 mesh for the 3D case. These spatial data are averaged and reduced to 64×64 for the 2D case and 32×32×32 for the 3D case to be input into the CNN in Fig. 2. The time interval for numerical integration is \( \Delta t = 0.002 \), and the time step for the reinforcement learning is \( \Delta t = 50\Delta t = 0.1 \).

### 3.1 Two-dimensional system

We first consider the 2D system, which obeys the GP equation in Eq. (3). The initial parameters of the Gaussian potential in each episode are \( \xi(0) = 0 \), \( \eta(0) = 2 \), and \( A_\perp(0) = 20 \). The initial state of the BEC is the ground state for these parameters, as shown in Fig. 3(a), where the density dip is due to the Gaussian potential. For this initial state, therefore, the rotational symmetry of the problem is broken. Here we set our target state \( \psi_{\text{target}}(x, y) \) to the lowest-energy stationary state having a singly-quantized counterclockwise vortex at the center, as shown in Fig. 3(f). Numerically, such a wave function can be obtained by phase imprinting followed by imaginary-time evolution. Starting from the initial state without a vortex, the agent tries to produce \( \psi_{\text{target}}(x, y) \) by controlling the position and strength of the Gaussian potential in Eq. (4).

We specify the action and reward. Depending on the six actions of the agent, \( a = 0, 1, \cdots, 5 \), the parameters in Eq. (4) are changed in each time step as

\[
\begin{align*}
    a = 0 & : \xi \to \xi + 0.15, \\
    a = 1 & : \xi \to \xi - 0.15, \\
    a = 2 & : \eta \to \eta + 0.15, \\
    a = 3 & : \eta \to \eta - 0.15, \\
    a = 4 & : A_\perp \to A_\perp + 2, \\
    a = 5 & : A_\perp \to A_\perp - 2,
\end{align*}
\]

where the displacement 0.15 is identical with the numerical mesh size. Since the strength \( A_\perp \) of the Gaussian potential produced by a blue-detuned laser beam should be nonnegative, the actions \( a = 5 \) is ignored when \( A_\perp \) becomes negative. The fidelity between the wave function \( \psi_\perp(x, y, t) \) at time \( t \) and the target wave function \( \psi_{\text{target}}(x, y) \) is defined by

\[
F(t) = \left| \int \psi_{\text{target}}(x, y)\psi_\perp(x, y, t)dx dy \right|^2.
\]

Since the present purpose is to increase the fidelity as much as possible, the reward \( r_t \) given to the agent should be taken to be a monotonically increasing function of the fidelity. Here, to enhance the increase of the fidelity near \( F = 1 \), we take the reward as

\[
r_t = F(t) + 8[F(t)]^{1.6}.
\]

As shown in the inset of Fig. 3(g), this function steeply rises for \( F \approx 0.9 \).

We performed the reinforcement learning in Algorithm 1 for \( N_{\text{episode}} = 5000 \) episodes. The total reward obtained in each episode is defined as

\[
R_{\text{total}} = \sum_{n=0}^{N_{\text{episode}}} r_n \Delta t,
\]

where the episode terminates at \( T_{\text{end}} \). The episode with the largest \( R_{\text{total}} \) among 5000 episodes is shown in Fig. 3. First, the Gaussian potential is moved leftward (Fig. 3(b)), and at the edge of the BEC a clockwise vortex is released to the periphery of the BEC, leaving a counterclockwise vortex in the Gaussian potential (Fig. 3(c)). The Gaussian potential then carries the counterclockwise vortex to the center of the BEC (Fig. 3(d)), at which point the strength of the Gaussian potential is reduced to produce the desired state (Fig. 3(e)). This process finishes at \( t \approx 130 \), and after that the vortex is kept at the center. It is numerically predicted and experimentally verified that the simple translation of a circular potential in a BEC produces a pair of clockwise and counterclockwise vortices simultaneously. In the above dynamics, in contrast, the strong anisotropy at the edge of the condensate is used to release only a single vortex from the potential.
A snapshot of the density profile for the center, even if \( t \approx 4 \), which helps fix the vortex to the center. It was confirmed that the vortex stays almost at the center, even if \( A(t) \) vanishes for \( t \geq 130 \).

Figure 4(a) shows the total reward \( R_{\text{total}} \) obtained in each episode. The total reward increases as the agent experiences more episodes, and finally reaches \( R_{\text{total}} \approx 2000 \). The total reward first increases to \( R_{\text{total}} \approx 400 \), then remains on a plateau until the \( \approx 2000 \) episodes and then increases again. This behavior is related to the form of the reward in the inset in Fig. 3(g), because the total reward saturates at \( R_{\text{total}} \approx 400 \) if we replace the reward in Eq. (10) with

\[
\Delta t = F(t).
\]

Figure 4(b) shows the time evolution of the fidelity \( F(t) \) for the best episode, which indicates that the fidelity reaches \( F \approx 0.97 \) using the reward in Eq. (10), while \( F \approx 0.9 \) for the reward in Eq. (12). A snapshot of the density profile for the

\[
y_t = r_t + \gamma Q(s_{t+\Delta t}, \arg\max_a Q(s_{t+\Delta t}, a))
\]

instead of Eq. (6). Without the target network \( \hat{Q} \), the learning process does not proceed, which indicates the validity of the target network.

### 3.2 Three-dimensional system

Next, we consider a 3D system obeying Eq. (1), in which a BEC is confined in an isotropic harmonic potential, and a Gaussian laser beam is applied from the \( x \) direction, producing the potential in Eq. (2). The initial parameters in Eq. (2) are set to \( A(0) = 20, d_r(0) = 1 \), and \( \zeta(0) = 0 \), and \( d_r \) is fixed. The initial wave function is the ground state for these parameters, as shown in Fig. 5(a). Here, our target state \( \psi_{\text{target}}(r) \) is the stationary state containing a vortex ring with axial symmetry about the \( z \) axis, as shown in Fig. 5(f). Although a vortex ring in a uniform system travels in the direction of the symmetry axis, the vortex ring in Fig. 5(f) is stationary due to the inhomogeneity. Numerically, this target state is generated by imprinting the phase \( e^{i\phi} \) with \( \phi = \tan^{-1}[z/(r_0 - r)] - \tan^{-1}[z/(r_0 + r)] \) followed by the imaginary-time evolution, where \( r_0 \) and the duration of the imaginary-time evolution are chosen appropriately.

As in the 2D case, the agent chooses one of six actions, \( a = 0, 1, \cdots, 5 \), according to which the parameters of the Gaussian potential are modified as

\[
a = 0: \quad a = 0 \quad \zeta \rightarrow \zeta + 0.15,
\]
\[
a = 1: \quad a = 1 \quad \zeta \rightarrow \zeta - 0.15,
\]
\[
a = 2: \quad a = 2 \quad d_r \rightarrow d_r + 0.2,
\]
\[
a = 3: \quad a = 3 \quad d_r \rightarrow d_r - 0.2,
\]
\[
a = 4: \quad a = 4 \quad A \rightarrow A + 2,
\]
\[
a = 5: \quad a = 5 \quad A \rightarrow A - 2,
\]

in each time step. The actions \( a = 3 \) and 5 are ignored when \( d_r \) and \( A \) become negative, respectively. The reward has the same form as in Eq. (10) with the fidelity being given by

\[
F(t) = \left| \int \psi_{\text{target}}(r) \psi(r, t) dr \right|^2.
\]
Fig. 5. (a)-(e) Dynamics of a 3D BEC with the best total reward. The BEC is confined in an isotropic harmonic trap with an external Gaussian potential in Eq. (2). (f) Target state containing a stationary vortex ring. The arrows indicate the directions of circulation. Surfaces of equal density in Eq. (2). (f) Target state containing a stationary vortex ring. The arrows indicate the directions of circulation. Surfaces of equal density, as shown in Fig. 5(c), which bend the vortex lines and forms a ring.

- The time evolution of the fidelity $F(t)$ is shown in Fig. 5(g). Unlike the 2D case in Fig. 4(b), the fidelity in Fig. 5(g) exhibits nonadiabatic oscillation, which may be unavoidable for the present condition. Figure 5(h) shows the total reward $R_{\text{total}}$ obtained in each episode. The total reward first increases gradually for $\leq 1000$th step and then suddenly rises to $R_{\text{total}} = 800$. This behavior is similar to that in the 2D case shown in Fig. 4(a).

4. Conclusions and discussions

We have applied reinforcement learning to the control of nonlinear matter waves. The agent in the reinforcement learning is implemented using a deep convolutional neural network (CNN), and the state of the Bose-Einstein condensate (BEC) is input into the CNN to determine the next action of the agent. According to the action of the agent, the position and shape of the external Gaussian potential applied to the BEC are controlled. The agent is trained so that the state of the BEC approaches the prescribed target state.

Using this method, we demonstrated that quantized vortices can be created and manipulated in 2D and 3D systems. In the 2D system, the target state was set to be the single-vortex state at the center. Although vortices and antivortices are always created in pairs, the agent found a way to expel one of them to leave only a single vortex at the center of the BEC (Fig. 3). In the 3D system, the target state was set to be the stationary vortex-ring state. Although the Gaussian potential is 2D (uniform in the $x$ direction), the agent found a way to produce such a 3D structure (Fig. 5).

The two examples demonstrated in the present paper are interesting, but rather simple; the control parameters found in Figs. 3(g) and 5(g) for creating the target states are very simple functions. Other methods have been developed to control quantum systems\(^{33-35}\) that may find similar optimal results.

Furthermore, the simple vortex states used for our target states can also be created by other methods, such as methods using Rabi transitions.\(^{52,53}\) In this sense, our method might be considered overcomplicated for the present examples. However, reinforcement learning is very versatile and may prove to be an effective approach for handling more complicated problems (e.g., manipulation of multiple vortex states, control of quantum turbulence, creation of nontrivial topological excitations, etc.). This will be the topic of future studies.

Another challenging extension of this study will be the real-time control of BECs in experiments, where a series of nondestructive imaging data is given to the agent, and the reward is based on the measurement data. Reinforcement learning may be suitable for such measurement-based control which intro-
ducers errors and noise, since the method has already been suc-

cessfully applied to control real-world objects, such as robots.

This research was supported by JSPS KAKENHI Grant

Numbers JP17K05595 and JP17K05596.

1) D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driess-

cche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S.
Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillic-

crap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, Nature 529,
484 (2016).

2) D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A.
Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F.
Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, Nature 550,
354 (2017).

3) V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Belle-
mare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen,
C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra,
S. Legg, and D. Hassabis, Nature 518, 529 (2015).

4) C. Chen, D. Dong, H.-X. Li, J. Chu, and T.-T. Tarn, IEEE Trans. Neural
Netw. Learn. Syst. 25, 920 (2014).

5) M. Bakov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P.
Meha, Phys. Rev. X 8, 031086 (2018).

6) M. Bakov, Phys. Rev. B 98, 224305 (2018).

7) F. Albarrán-Ariagada, J. C. Retamal, E. Solano, and L. Lamata, Phys.
Rev. A 98, 042315 (2018).

8) S. Yu, F. Albarrán-Ariagada, J. C. Retamal, Y.-T. Wang, W. Liu, Z.-J. Ke,
Y. Meng, Z.-P. Li, J.-S. Tang, E. Solano, L. Lamata, C.-F. Li, and G.-C.
Guo, Adv. Quantum Technol. 2, 1800074 (2019).

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

10) X.-M. Zhang, Z.-W. Cui, X. Wang, and M.-H. Yung, Phys. Rev. A 97,
052333 (2018).

11) P. Andreassen, J. Johansson, S. Liljestrand, and M. Granath, Quantum 3,
183 (2019).

12) F. Chen, J.-J. Chen, L.-N. Wu, Y.-C. Liu, and L. You, Phys. Rev. A 100,
041801(R) (2019).

13) M. August, J. M. Hernández-Lobato, arXiv:1802.04063.

14) M. Y. Niu, S. Boixo, V. Smelyanskiy, and H. Neven, arXiv:1803.01857.

15) J.-J. Chen and M. Xue, arXiv:1901.08748.

16) Z. T. Wang, Y. Ashida, and M. Ueda, arXiv:1910.09200.

17) J. Mackeprang, D. Dasari, and J. Wachtrop, arXiv:1908.05981.

18) U. Hohenester, P. K. Rekdal, A. Borzi, and J. Schmiedmayer, Phys.
Rev. A 75, 023602 (2007).

19) X. Chen, E. Torrontegui, D. Stefanatos, J.-S. Li, and J. G. Muga, Phys.
Rev. A 84, 043415 (2011).

20) S. Amiri, R. Corgier, D. Sugny, E. M. Rasel, N. Gaaloul, and E. Charron,
Sci. Rep. 9, 5346 (2019).

21) G. Jäger, D. M. Reich, M. H. Goerz, C. P. Koch, and U. Hohenester, Phys.
Rev. A 90, 033628 (2019).

22) J.-F. Mennemann, D. Matthes, R.-M. Weishäupl, and T. Langen, New J.
Phys. 17, 113027 (2015).

23) J. Grond, J. Schmiedmayer, and U. Hohenester, Phys. Rev. A 79,
021603(R) (2009).

24) G. Jäger, T. Berrada, J. Schmiedmayer, T. Schumm, and U. Hohenester,
Phys. Rev. A 92, 053632 (2015).

25) R. Bücker, T. Berrada, S. van Frank, J.-F. Schaaff, T. Schumm, J. Schmied-
mayer, G. Jäger, J. Grond, and U. Hohenester, J. Phys. B 46, 104012
(2013).

26) S. van Frank, M. Bonneau, J. Schmiedmayer, S. Hild, C. Gross, M. Ch-
eeneu, I. Bloch, T. Pichler, A. Negretti, T. Calarco, and S. Montangero,
Sci. Rep. 6, 34187 (2016).

27) D. Hocker, J. Yan, and H. Rabitz, Phys. Rev. A 93, 053612 (2016).

28) C. A. Weidner and D. Z. Anderson, Phys. Rev. Lett. 120, 263201 (2018).