PAPER

Efficient and robust entanglement generation with deep reinforcement learning for quantum metrology

Yuxiang Qiu\textsuperscript{1,2}, Min Zhuang\textsuperscript{1,2}, Jiahao Huang\textsuperscript{1,2,*} and Chaohong Lee\textsuperscript{1,2,*}

\textsuperscript{1} Guangdong Provincial Key Laboratory of Quantum Metrology and Sensing & School of Physics and Astronomy, Sun Yat-Sen University (Zhuhai Campus), Zhuhai 519082, People’s Republic of China

\textsuperscript{2} State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-Sen University (Guangzhou Campus), Guangzhou 510275, People’s Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: hjiahao@mail2.sysu.edu.cn and lichaoh2@mail.sysu.edu.cn

Keywords: quantum metrology, entanglement generation, reinforcement learning

Abstract

Quantum metrology exploits quantum resources and strategies to improve measurement precision of unknown parameters. One crucial issue is how to prepare a quantum entangled state suitable for high-precision measurement beyond the standard quantum limit. Here, we propose a scheme to optimize the state preparation pulse sequence to accelerate the one-axis twisting dynamics for entanglement generation with the aid of deep reinforcement learning (DRL). We consider the pulse train as a sequence of $\pi/2$ pulses along one axis or two orthogonal axes, and the operation is determined by maximizing the quantum Fisher information using DRL. Within a limited evolution time, the ultimate precision bounds of the prepared entangled states follow the Heisenberg-limited scalings. These states can also be used as the input states for Ramsey interferometry and the final measurement precisions still follow the Heisenberg-limited scalings. While the pulse train along only one axis is more simple and efficient, the scheme using pulse sequence along two orthogonal axes show better robustness against atom number difference between simulation and experiment. Our protocol with DRL is efficient and easy to be implemented in state-of-the-art experiments.

1. Introduction

Quantum metrology studies how to exploit quantum resources and strategies to improve the estimation precision of unknown parameters [1, 2]. Generally, the information of an unknown parameter is encoded into a phase which can be precisely estimated via interferometric techniques in experiments [3–5]. For interferometry with individual atoms, the sensitivity of the estimated phase can reach the so-called standard quantum limit (SQL) [6], i.e., $\Delta \phi = O(N^{-1/2})$ with $N$ the atom number. However, this bound is not fundamental and can be surpassed by using multi-particle entanglement [7–10]. Recent developments in quantum metrology focus on how to generate metrologically useful quantum entangled states and utilize them for phase estimation.

One kind of representative entangled quantum states that can provide sub-SQL phase sensitivity is spin-squeezed state [11]. Spin squeezed states can be prepared through the one-axis twisting (OAT) interaction, which is widely realized by light-mediated interactions [12–15] or atom–atom interaction within Bose condensed atoms [3, 16–19] and the phase sensitivity can be scaled as $\Delta \phi = O(N^{-2/3})$ [5, 20]. Apart from OAT, spin squeezing can be generated by two-axis counter-twisting (TACT) interaction, and the phase sensitivity can be improved to the Heisenberg limit, $\Delta \phi = O(N^{-1})$. However, this kind of spin squeezing is challenging to realize in experiments. In addition to spin squeezed states, non-Gaussian entangled states such as twin Fock state and spin cat state are also promising candidates for achieving Heisenberg-limited phase sensitivity [1, 20, 21].

© 2022 Published by IOP Publishing Ltd on behalf of the Institute of Physics and Deutsche Physikalische Gesellschaft
The main obstacle against the applications of quantum entangled states in practice is the entanglement generation in realistic experiments. Several theoretical schemes for preparing quantum entangled states such as adiabatic sweeping [8, 22–24], shortcut to adiabaticity [25–27] and optimal controls [23, 28–30] are developed. However, the schemes are either time-consuming or too complicated to be implemented, which are hard to realize in state-of-the-art experiments. Hence, developing fast and effective approaches for creating quantum entanglement is of great importance.

One promising way is to make use of machine learning, which has already attracted much attention [31]. In particular, deep reinforcement learning (DRL) [31, 32] which can provide optimal decision strategies or policies based upon a well-defined target, is gradually applied in quantum physics [31, 33–40]. It can provide a machine learning (ML) model, often neural networks that is capable of optimizing a certain objective function by providing a well-designed time sequence of control procedures [41, 42]. It is particularly suitable for seeking the optimal preparation of desired quantum states [43–51]. Recently, it is proposed that extreme spin squeezing can be achieved with OAT interaction using a sequence of rotation pulses designed via DRL [44]. Although spin squeezing is a good metrological quantum resource, the most metrologically useful one is usually characterized by the quantum Fisher information (QFI) $F_Q$ [52, 53]. Can we find out an experimentally feasible scheme to prepare the optimal quantum entangled state that maximizing $F_Q$ via DRL? Can the prepared quantum entangled state be suitable for practical quantum phase estimation?

In this work, we propose a scheme for preparing metrological useful entangled states based on OAT interaction with a sequence of rotation pulses designed via DRL. In our scheme, the OAT interaction which is the key for entanglement generation, exists persistently during the state preparation. Our scheme is inspired by the so-called twist and turn dynamics [30, 54] that is capable of generating spin squeezing efficiently. We design to apply a train of $\pi/2$ pulses within a limited time $T$ for state preparation. The time sequence of pulse train is optimized by maximizing $F_Q$ with the aid of DRL.

When considering $\pi/2$ pulses only along one axis, we find that only a few number of pulses can drive to a highly entangled state which enables the Heisenberg-limited scaling. However, this protocol is sensitive to the atom number difference between simulation and experiment. In experiment, the atom number may not be well-prepared and will have a difference in the atom number used in the DRL algorithm for designing the pulses. This kind of atom number difference may deviate the prepared state from the optimized results, hence degrade the ultimate measurement precision scaling. To strengthen the robustness, we consider $\pi/2$ pulses along two orthogonal axes. We find that although more pulses are required, it is more robust against atom number difference. To validate our scheme for phase estimation, we use the entangled states obtained by DRL as the input state to perform the Ramsey interferometry. The associated phase measurement precision $\Delta \phi$ can still display the Heisenberg-limited scaling. Besides, the scheme with $\pi/2$ pulses along two axes can also provide better robustness against the atom number difference. Our scheme via DRL provides a straightforward and efficient way to optimize the entangled state preparation for quantum metrology, and the robustness against the atom number difference makes it feasible in realistic experiments.

2. Entanglement generation via deep reinforcement learning

2.1. Preparation of quantum entangled state

We consider an ensemble of $N$ two-level identical atoms whose Hamiltonian ($\hbar = 1$) is given by
\[
H = \chi J_y^2 + \Omega J_z + \delta J_z.
\]
Here, $\hat{J}_y = \sum_{\alpha} \hat{\sigma}^{(\alpha)}_y$ and $\hat{J}_z = \sum_{\alpha} \hat{\sigma}^{(\alpha)}_z$ are the collective spin operators with the Pauli matrices $\sigma^{(\alpha)}_y$ for the $\alpha$th atom [3]. The system state can be expanded in the Dicke basis $|l,m\rangle = |m \rangle$ with $m = -N/2, -N/2 + 1, \ldots, N/2$. The Hamiltonian contains three terms. The first term $\chi J_y^2$ denotes the atom–atom interaction, which is the key for realizing OAT dynamics [3, 19]. The second term $\Omega J_z$ is the coupling between the two atomic levels. The third term $\delta J_z$ is the bias or detuning. The Hamiltonian $H$ can be applied to Bose condensed atoms occupying two hyperfine states [55, 56] or a single-component condensate trapped in a double-well potential [57–59]. The parameters $\chi$, $\Omega$ and $\delta$ can be well controlled via external fields in experiments [5, 60].

The first and significant step for quantum metrology is the entangled state preparation. Initially, the system state is usually prepared in a coherent spin state (CSS) [61, 62]
\[
|\psi\rangle_0 = e^{i?/\hbar}|\pi, 0\rangle_{CSS},
\]
which is rotated by a $\pi/2$ pulse along the $y$ axis [5, 63] from the state $|\pi, 0\rangle_{CSS} = |\uparrow\rangle^N$ with all $N$ atoms in $|\uparrow\rangle$. The OAT dynamics can squeeze the CSS to a spin squeezed state. There exists an optimal evolution time $T_{\text{opt}}$ that extreme spin squeezing can be achieved [44]. Apart from spin squeezing, the metrological
ability of a quantum state can also be characterized by QFI. Generally, maximizing $F_Q$ can obtain the optimized input state for attaining the best precision bounds [2, 10, 64]. Thus, we use QFI as a metric to evaluate the states prepared by the optimized pulse trains. For an pure input state $\ket{\psi}$, the QFI for phase estimation can be defined as [5]

$$F_Q = 4\left|\langle \psi'(\theta) | \psi'(\theta) \rangle - \langle \psi'(\theta) | \psi(\theta) \rangle \right|^2,$$

(2)

where $\ket{\psi(\theta)} = e^{-i\theta}\ket{\psi}$ and $\ket{\psi'(\theta)} = -i\Omega e^{i\theta}\ket{\psi}$, and the ultimate precision bound can be given by $F_Q^{-1/2}$ [52, 53]. The above process can be applied to various atom interferometers. Since Ramsey interferometry is commonly used in quantum metrology, in this work, we will use Ramsey interferometry to validate the quantum states prepared by the optimized pulse trains in section 3.

In the stage of state preparation, a pulse train is applied and the Hamiltonian of the system can be expressed as:

$$H = \chi \hat{J}_x^2 + \Omega_x(t) \hat{J}_x + \Omega_y(t) \hat{J}_y,$$

(3)

where $\Omega_x(t)$ and $\Omega_y(t)$ are time-dependent functions describing the applied pulses. Consider the total evolution time $T$ is around $T_{\text{tot}}$, and we divide $T$ equally into $n_t$ intervals and each interval length $\delta t = T/n_t$. At each interval, one can choose to apply a $\pi/2$ pulse along x or y axis with $\int_{t}^{t+\delta t}\Omega_{xy}(t')dt' = \pi/2$, or turn off the coupling $\Omega_{xy} = 0$ to let the state evolve barely under OA T interaction.

From the initial CSS $\ket{\psi}_0$, our goal is find an optimized pulse train to generate the input state $\ket{\psi}$ by maximizing $F_Q$ within $T$. To accomplish this goal, we adopt the technique from ML. The optimization process will be guided by an ML model obtained from DRL. In the following, we will introduce the DRL algorithm and show the optimization results in detail.

2.2. DRL algorithm

To obtain the optimal control, the optimization process will be guided by a DRL algorithm. Briefly, the DRL algorithm requests certain information about the current state $\ket{\psi}_t$ for the $t$th time step ($t \in [1, n_t]$), and determines the evolution happening in the next ($t+1$)th time step with an optimal policy. As one of the DRL algorithms, here we adopt the so-called asynchronous advantage actor–critic (A3C) algorithm [65] to accomplish our goal. It is based on a common actor–critic algorithm while designed in an asynchronous structure, as sketched in figure 1(a). Generally it uses neural networks to find an appropriate decision. The network parameters are updated via adaptive momentum gradient decent method (ADAM) [66]. The asynchronous structure of A3C is beneficial for the stability of the learning process and makes it fast to converge. The learning process also becomes more efficient because the network architecture is naturally parallel processing which can take full advantages of the multiple process units in the computing hardware.

Next, we show how to find the optimized pulse train in the framework of DRL algorithm. As shown in figure 1(b), at every time step $t$ the algorithmic state $s_t$ needed to know and feed into the algorithm is some expectation values of the evolved quantum state $\ket{\psi}_t$. $s_t$ can be encoded in a tuple with the following six expectation:

$$\langle \hat{J}_x(t) \rangle, \langle \hat{J}_y(t) \rangle, \langle \hat{J}_z(t) \rangle, \langle \hat{J}_x^2(t) \rangle, \langle \hat{J}_y^2(t) \rangle, \langle \hat{J}_z^2(t) \rangle.$$

It should be mention that, these six expectation quantities are the intermediate variables in the algorithm. They are only calculated numerically [44] and do not need to be measured in experiments. Then the action $a_t$ is obtained after receiving $s_t$, which is an evolution operator $U_t$ chosen from the action pool containing three candidates:

$$\hat{U}_0 = e^{-i\hat{J}_x \delta t}, \quad \hat{U}_1 = e^{-i\hat{J}_y \delta t}, \quad \hat{U}_2 = e^{-i\hat{J}_z \delta t}.$$

(4)

Finally a reward $r_t$ related to the QFI of evolved state $F_Q^{(t)}$ is calculated. The reward will be described later.

In this work, we consider two schemes, ‘only-\hat{J}_x’ and ‘both-\hat{J}_x, \hat{J}_y’; The former one only using $\pi/2$ pulses along x axis, in which $U_t$ is chosen only from $U_0$ and $U_1$. While for the latter one, $\pi/2$ pulses along x and y axis are both considered, i.e., $U_t \in \{U_0, U_1, U_2\}$. Then, the unitary evolution $\ket{\psi}_{t+1} = U_t \ket{\psi}_t$ is performed, and the consequent state $\ket{\psi}_{t+1}$ will participate the evolution at the next time step $t+1$ sequentially. Thus, the final prepared state can be written as

$$\ket{\psi}_T = U \ket{\psi}_0 = \prod_{t=1}^{n_t} U_t \ket{\psi}_0,$$

(5)

where the initial state is given by equation (1). To maximize $F_Q$ of $\ket{\psi}_T$, in each step we numerically calculate the QFI $F_Q^{(t)}$ for $\ket{\psi}_t$ to obtain the reward $r_t$ of the $t$th step. The calculation of the total reward $R_{\text{tot}}$ is then made after $n_t$ evolution steps. Finally, a specific pulse sequence $(U_1, U_2, \ldots, U_{n_t})$ can be generated from the optimal policy within the DRL algorithm.
Figure 1. (a) The sketch of A3C algorithm, featuring local networks design and asynchronous updating of network parameters. (b) The sketch of the quantum state preparation process guided by A3C algorithm. In the $t$th step, the trained network receives current state $s_t$ and then provides a certain action $a_t$, representing an operator $U_t$ participating in the next step. (c) The total evolution time $T$ versus atom number $N$ in our numerical calculations. A fitting function (the blue dashed line) is added, roughly showing an exponential relationship between $T$ and $N$. (d) The learning curves for $N = 100$ and 1000, including the results using only-$J_x$ and both-$J_x,J_y$ schemes. The slightly noisy lines are obtained by averaging from every 10 trails, and the smooth lines are obtained by running average. The convergent behaviors suggest the effectiveness of the whole learning process. The second row displays the results obtained by using only-$J_x$ scheme, i.e. the actions pool in (b) only contains $U_0$, $U_1$. (e) The optimized pulse trains. Blue histograms are placed at the time step when a $\pi/2$ pulse along $x$ axis is applied. (f) The evolution of $F_Q$ during the state preparation process. (g) The optimized states with maximized $F_Q$. The corresponding Husimi functions on Bloch spheres are shown in the insets. (h) The scaling of $F_Q^{-1}$ versus the atom number $N$. Here, we fit the points by least square method and the fitting line is denoted by an orange line. The green dash line stands for the exact Heisenberg limit. The bottom row displays the results obtained by using both-$J_x$, $J_y$ schemes, i.e. the actions pool in (b) contains $U_0$, $U_1$, and $U_2$. (i) The optimized pulse trains. Blue and orange histograms represent the $\pi/2$ pulse along $x$ and $y$ axis, respectively. (j) The evolution of $F_Q$ during the state preparation process. (k) The optimized states. (l) The scaling of $F_Q^{-1}$ versus the atom number $N$.

The total reward $R_{tot}$ is originally the accumulated reward of $n_t$ time steps as $R_{tot} = \sum_{t=0}^{n_t} r_t$ [44], while in our DRL algorithm the $n_t$ rewards are requested all in once after total evolution time $T$, by denoting the reward of the $t$th step as the largest reward among the rest steps after time $t$, as:

$$r_t = \max_{t' < T \leq n_t} F_Q^{(t')}.$$  

(6)

This non-step-wise design of reward allows us to denote every $r_t$ after knowing $F_Q^{(0-n_t)}$, which is beneficial for the training stability, efficiency and capability of convergent. Another advantage of this definition (6) is that in each training epoch the DRL algorithm can somewhat comprehend that the optimization task is fulfilled within $n_t$ steps so that the ML model can reach similar optimum once $n_t$ is large enough, see figure 2. In addition, we use two separated neural networks as actor and critic network. The benefit of this separation is that different quantities of $F_Q$ from different atom numbers $N$ can be greatly balanced. The parameters of our algorithm, including structure of the neural networks and the learning rate, do not need to be adjusted in the face of different atom number situations and can achieve convergence at the same rate, see figure 1(d).

One of the advantages of using neural network is that it has more capacity of high nonlinearity models than traditional optimization methods.
Although it seems to be complicated, the programming is quite convenient because there are many open-source ML tools and packages, which are optimized for computer hardware. In our case the majority of calculations comes from the simulations of quantum state evolutions. The time consumed in running the algorithm is closely related to the atom number $N$. We only use a single CPU Intel Core i7 8700 on a consumer class computer to run the algorithm. It costs less than 1 h for 8000 epochs of training when $N = 10^4$, it needs 32 GB of RAM to calculate the operators.

### 2.3. Results with DRL

In our numerical simulations, we choose $\chi = 1$ and $n_i = 50$. The total evolution time $T$ is chosen near the optimal squeezing time, which can be determined numerically. The relation between $T$ and $N$ is shown in figure 1(c), roughly an exponential dependence. For example, for $N = 100$ and 1000 we have $T = 0.13$ and 0.015, respectively. Starting from an initial $|\psi\rangle_0$ with a fixed $N$, we can obtain the maximized $F_Q$ and the corresponding prepared quantum state $|\psi\rangle_T$ with the help of DRL. Here, we display results of two representative cases ($N = 100$ and 1000) using only-$J_x$ scheme and both-$J_x, J_y$ scheme, see figures 1(e–h) and (i–l), respectively. In figure 1(d), the learning curves of DRL for both schemes with $N = 100$ and 1000 are given. It is shown that, after 8000 trails of learning the $F_Q$ of the final states $|\psi\rangle_T$ are optimized and converge to saturated values, indicating a successful optimization.

The associated pulse trains optimized by our DRL algorithm for $N = 100$ and 1000 are shown as histograms in figure 1(e) for only-$J_x$ scheme and in figure 1(i) for both-$J_x, J_y$ scheme, where blue and orange histograms stand for $\pi/2$ pulses along $x$ and $y$ axis, respectively. The corresponding time-evolutions of the $F_Q$ are shown in figures 1(f) and (j). The $F_Q$ of the final states $|\psi\rangle_T$ are highlighted by red dots, and the associated distributions of $|\psi\rangle_T$ are shown in in figures 1(g) and (k).

The optimized $F_Q$ of the prepared states using only-$J_x$ scheme and both-$J_x, J_y$ scheme are nearly the same, with the latter mostly being a little larger than the former. The final prepared states $|\psi\rangle_T$ become non-Gaussian with two humps appear near $|m = \pm N/2\rangle$, see the Husimi distribution on the generalized Bloch sphere and the probability distribution. However, the probability distribution of $|\psi\rangle_T$ using both-$J_x, J_y$ scheme is more rugged than the one using only-$J_x$ scheme. We also simulated cases of other atom numbers up to $10^4$.

Simulations of even more atom number cases are not implemented because of the constraint of random access memory in our computer. Essentially, we find that the scaling of $F_Q$ versus $N$ of the two schemes can both approach the Heisenberg limit. Here, we use least square method to fit the results and the fitting formula are displayed in the legends. Similarly, the both-$J_x, J_y$ scheme outperforms the only-$J_x$ scheme with a slightly smaller constant. Results of our simulations show that the method with DRL algorithm is promising for developing Heisenberg-limited metrology protocols within a wide range of atom number $N$.

On the other hand, the optimized pulses trains for these two schemes are much different. We can see that, for both $N = 100$ and 1000, only four $\pi/2$ pulses along $x$ axis is needed. With a final pulse applying at the final time step, the state can abruptly evolve to the one with desired value of $F_Q$. While for both-$J_x, J_y$
scheme, more $\pi/2$ pulses along $x$ axis with few $\pi/2$ pulses along $y$ axis are needed. Thus, the pulse trains for only-$J_x$ scheme is much sparse and simple, which will be more feasible in realistic experiments. For a fixed $N$, whatever by using only-$J_x$ scheme or both-$J_x, J_y$ scheme, we can find the optimal control for preparing the optimal state within $T$ with the help of DRL algorithm. However, the optimized pulse trains are always discrepant with different $N$ and $T$. Thus, we need to know the atom number $N$ roughly in advance to design the corresponding optimal pulse sequence.

The interval number $n_t$ we divide the total evolution time $T$ may slightly influence the optimization results. The resultant $F_Q$ of the final states $F_Q$ with different $n_t$ are shown in figure 2. It is shown that more pulses enable to push the optimization even better but the growth decreases when $n_t > 50$, especially for large $N$. Thus, we find that $n_t = 50$ is a balanced choice in condition that the structure of the two networks and hyperparameters in our DRL algorithm also remain unchanged. Despite that with increasing $n_t$, the $F_Q$ of the prepared state may be slightly larger, it requires more carefully designed algorithm parameters and increases operation complexity.

3. Phase estimation via time-reversal Ramsey interferometry

Generally, QFI only sets the ultimate measurement precision bound, but it may not always be attained. To validate metrological usefulness of the prepared states via DRL, we simulated the Ramsey interferometry for phase estimation [3, 4, 67] by inputting the prepared states $|\psi\rangle_T$.

For a conventional Ramsey interferometry, the whole process consists of a phase accumulation sandwiched by two $\pi/2$ pulses [10, 68]. Since we start from an initial CSS, it is suitable to use time-reversal protocol. Here, we consider a time-reversal protocol: a disentangling operation $U^\dagger$ after the phase accumulation process [69, 70], which is implemented by a reverse of $U$ in equation (5). As sketched in figure 3(a), the final state after Ramsey interferometry is thus:

$$|\psi\rangle_0 = e^{-i\hat{J}_z \phi} U^\dagger e^{-\phi \hat{J}_x} |\psi\rangle_T. \quad (7)$$

The time-reversal operation can be achieved by changing the sign of the entangling Hamiltonian [70]. This can be realized in various synthetic quantum systems, such as atom-cavity system [71] and cold atom system [72].

Then the measurement precision of $\phi$ can be calculated by using error propagation formula [60]:

$$\Delta \phi = \frac{\langle \Delta \hat{J}_z \rangle_0}{|\partial \langle \hat{J}_z \rangle_0 / \partial \phi|}, \quad (8)$$

where $\langle \Delta \hat{J}_z \rangle_0 = \sqrt{\langle \hat{J}_z^2 \rangle_0 - \langle \hat{J}_z \rangle_0^2}$, the subscript $\phi$ indicates the expectation with respect to $|\psi\rangle_0$. Here, we consider the estimated phase is tiny which is in the vicinity of $\phi = 0$.

The corresponding scalings of measurement precision versus $N$ are shown in figures 3(b) and (c). The resultant phase measurement precisions are given as blue (only-$J_x$ scheme) and red points (both-$J_x, J_y$ scheme), respectively. Despite the scaling is a bit deviated from the ultimate bounds of $F_Q$ in figures 1(h) and (l), the estimated phase measurement precision for only-$J_x$ and both-$J_x, J_y$ schemes still show Heisenberg-limited scaling as expected. This suggests the optimized entangled state we prepare by using

![Figure 3](image-url)
Figure 4. The robustness against atom number difference between simulation and experiment. Here, the pulse trains are obtained by DRL with (a) $N = 100$, (b) $N = 500$, (c) $N = 1000$, and (d) $N = 5000$, respectively. The ultimate precision bounds are obtained by the same pulse train in condition of different atom number $N$. The blue points connected by blue dashed lines are results obtained by only-$J_x$ scheme, and green points connected by green dashed lines are results obtained by both-$J_x$, $J_y$ scheme. The black dashed lines stands for the exact Heisenberg limit $1/N$, and the red dashed lines is the Heisenberg-limited scaling $\propto 1/N$ passing through the point of the $F_Q^{1/2}$ of original values of $N$ that is highlighted by red circles. (e)–(h) Phase measurement precision $\Delta \phi$ with time-reversal Ramsey interferometry for the same situations and plotted in the same manner with (a)–(d), respectively. Despite the absolute value using both-$J_x$, $J_y$ scheme is mostly a little worse, the both-$J_x$, $J_y$ scheme displays better robustness against atom number difference.

DRL algorithm also has great potential for Heisenberg-limited phase estimation with Ramsey interferometry.

The only-$J_x$ scheme shows a smoother scaling and closer to the Heisenberg limit, $2.0/N$ compared to $3.7/N$ that obtained by both-$J_x$, $J_y$ scheme. This may result from the addition of $U_2$ pulses in equation (4), while in the next section we will see that the participation of $U_2$ can contributes to a better robustness against the difference of atom number between simulation and experiment.

4. Robustness against the atom number difference between simulation and experiment

Finally, we discuss the robustness of our schemes against the atom number difference between simulation and experiment. As it is mentioned in section 2, the optimal pulse sequence obtained by DRL depends on the atom number $N$ and total evolution time $T$. In our numerical calculations, the initial state $|\psi_0\rangle$ is assumed to be a pure state with a well-defined atom number $N$. In practical experiments, $T$ can be precisely controlled but the estimation of atom number $N$ may be inaccurate. The atom number in experiment may not be the same as expected. There may be a difference between the atom number in experiment and the one set in the DRL algorithm for designing the pulses. Therefore, it is necessary to figure out the robustness of our scheme when this kind of difference exists.

We perform the robustness evaluation by applying the optimized pulse train of atom number $N$ to the situation with other atom number in the range of $[0.8N, 1.2N]$, i.e., the difference of atom number is assumed up to $\pm 20\%$. The results with $N = 100, 500, 1000$ and 5000 are shown in figure 4, including the $F_Q$ and phase measurement precision $\Delta \phi$ via time-reversal Ramsey interferometry, using only-$J_x$ scheme and both-$J_x$, $J_y$ scheme. The red dashed lines are added for reference, representing the Heisenberg-limited scalings passing the results of only-$J_x$ scheme cases without differences. Ideally, the results should be close to the red dashed lines.

It turns out that the difference of $N$ makes the resultant ultimate precision bound $F_Q^{1/2}$ degraded, and the results of $\Delta \phi$ also become worse. Compared with only-$J_x$ scheme, both-$J_x$, $J_y$ scheme show better robustness against atom number difference. As it is shown in figures 4(a)–(d), the $F_Q$ keeps in the same level when there is no difference of $N$, and the $F_Q$ using both-$J_x$, $J_y$ decrease much less than those using only-$J_x$ scheme. The cases of $\Delta \phi$ is shown in figures 4(e)–(h), showing the same manner of degradation with these two schemes. Although the phase measurement precision using both-$J_x$, $J_y$ scheme is worse than those using only-$J_x$ scheme for most $N$ as shown in figures 3(b) and (c), the robustness of the former scheme is better than the latter.
It suggests that the pulse trains optimized by the DRL algorithm is practicable even though the atom number $N$ of the system cannot be estimated accurately. If the atom number difference is small in experiment, one may give priority to use the only-$J_x$ scheme for phase estimation. Otherwise, the both-$J_x, J_y$ scheme which can show better robustness against atom number difference, may become favorable.

5. Conclusion and discussion

We have presented an efficient and robust scheme for preparing entangled state with DRL algorithm and their metrological usefulness with the Ramsey interferometry for phase estimation. The numerical simulations of quantum state preparation include only-$J_x$ scheme or both-$J_x, J_y$ scheme, referring to the OAT dynamics with pulse sequence along only one axis or along two orthogonal axes, respectively. The system starts from a CSS, then reaches an optimized entangled state under a pulse train obtained by DRL algorithm. The quantum state preparation process is accomplished within a short time duration and the ultimate precision bounds exhibit the Heisenberg-limited scaling. Further, the Heisenberg-limited scaling can be maintained by performing the Ramsey interferometry, which verify the usefulness of our schemes in experiments. We use the A3C algorithm [65] whose actor and critic networks are separately established. It makes our algorithm equally effective and efficient for different atom number cases from $N = 10$ to $10,000$ without reforming the neural networks and parameters of the DRL algorithm. Besides a non-step-wise reward design makes the training process feasible and stable, similarly successful when the total number of pulses $n_t$ is sufficient.

The only-$J_x$ scheme and both-$J_x, J_y$ scheme have different advantages. On one hand, the pulse trains of only-$J_x$ scheme provided by DRL algorithm is much more simple, and the scaling of phase measurement precision is better than that of both-$J_x, J_y$ scheme. On the other hand, we find that the entangled states prepared by both-$J_x, J_y$ scheme have better robustness against atom number difference between simulation and experiment. Therefore only-$J_x$ scheme can be used when one wants to simplify the process of state preparation and the difference of atom number can be well controlled, while the both-$J_x, J_y$ scheme is considerable when the robustness against atom number difference matters more.

Our algorithm can be used as an offline optimization for quantum entangled state preparation in synthetic many-body quantum systems, such as cold atoms [3, 19], and trapped ions [73]. Online optimization is also feasible when the QFI is extractable [74] while accompanying a huge consumption of time, which might be solved by starting from results provided by sufficient offline optimizations. In the future, the effects of decoherence and imperfect pulse shape can also be taken into account, which will be more feasible for practical experiments.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (12025509, 11874434), the Key-Area Research and Development Program of GuangDong Province (2019B030330001), and the Science and Technology Program of Guangzhou (201904020024). MZ is partially supported by the National Natural Science Foundation of China (12047563). JH is partially supported by the Guangzhou Science and Technology Projects (202002030459).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/Aburiann/DRL_OFI_optimize [76].

Appendix A. Pseudo-code of the DRL algorithm

Here we show the pseudo-code of the DRL algorithm used in our work, as algorithms 1 and 2 in the following. Algorithm 1 is the algorithm running in each local network shown in figure 1(a). It includes the quantum state evolutions and calculations of gradients of the local networks. Algorithm 2 is the algorithm describing the behavior of the global network. It firstly distributes the network parameters to the local networks, and then receives the gradients calculated by local networks and updates it the parameters. After that it distributes its network parameters again to start the next epoch of training.

We have uploaded and shared the code and results on GitHub.com, including the programs that can directly provide the results of $N = 100$ and $N = 1000$ cases in figures 1(e–g) and (i–k), and also the saved weights of neural networks. One can regain the neural networks by loading the weights from our saved data
Algorithm 1. A3C for individual local network.

**Input:** particle number \( N \), pulse number \( n_e \), total evolution time \( T \), pulse duration \( \Delta t = T/n_e \), discounting parameter \( \gamma = 0.99 \)

**Output:** gradients for updating of the global network \( \theta \)

**Initialize:** copy the \( \text{Actor}(\cdot) \) and \( \text{Critic}(\cdot) \) networks from the global networks, \( |\psi\rangle_0 = e^{-iH_\text{CSS}}|\pi, 0\rangle_\text{CSS} \)

[Realize a trajectory];

for \( t = 1 \) to \( n_e \) do

load the previous state:

\[ s_{t-1} \leftarrow (|J_z\rangle_{t-1}, |J_x\rangle_{t-1}, |J_y\rangle_{t-1}, |J_z\rangle_{t-1}, |J_x\rangle_{t-1}, |J_y\rangle_{t-1}) \]

Feed \( s_t \) into Actor and Critic network;

distribution of \( a_t \) = Softmax(\( \text{Actor}(s_t) \));

sample an action \( a_t \) according to \( p_{\pi,a_t} \), do the corresponding evolution:

\[ |\psi\rangle_t = U(a_t)|\psi\rangle_{t-1} \]

load the QFI of current quantum state:

\[ \text{QFI}_t \leftarrow 4|\langle \psi(\theta)|\psi(\theta)\rangle - |\langle \psi(\theta)|\psi(\theta)\rangle|_t|^2 \]

end

[Build the Replay Buffer];

for \( t = 1 \) to \( n_e \) do

restore memories \( \text{mem}_t = (s_t, a_t, r_t = \text{max}_{z \in \{0,1,2\}} \text{QFI}_t) \) to Replay Buffer

end

[Compute the loss];

for \( t = 1 \) to \( n_e \) do

memories \( (s_t, a_t, r_t) \) randomly sampled from the Replay Buffer;

Discounting reward at time \( t \):

\[ R_t = \sum_{i=t}^{n_e} \gamma^{i-t} r_i \]

value function of \( s_t \):

\[ V(s_t) \leftarrow \text{Critic}(s_t) \]

Advantage function of \( s_t \):

\[ A_t = R_t - V(s_t) \]

\[ \mathcal{L}_t \leftarrow \text{CrossEntropy}(a_t, \text{Actor}(s_t)) \]

end

loss of the Critic network:

\[ \mathcal{L}_\text{Critic} \leftarrow \sum_{i=t}^{n_e} A_i^2 \]

loss of the Actor network:

\[ \mathcal{L}_\text{Actor} \leftarrow \sum_{i=t}^{n_e} \mathcal{L}_A + 0.01R_t \]

total loss:

\[ \mathcal{L} \leftarrow 0.5\mathcal{L}_\text{Critic} + \mathcal{L}_\text{Actor} \]

gradient of network parameters:

\[ \frac{\partial \mathcal{L}}{\partial \theta} \]

end

Algorithm 2. A3C for the global network.

**Input:** learning rate \( lr = 10^{-3} \), total learning epochs \( \text{epoch}_\text{max} \)

**Output:** learned network \( \text{Actor}(\cdot) \) and \( \text{Critic}(\cdot) \)

**Initialize:** parameter \( \theta \) of the global network \( \text{Actor}(\cdot) \) and \( \text{Critic}(\cdot) \) using HE normal method

[The following iterations is performed asynchronously at different workers];

for \( ep = 1 \) to \( \text{epoch}_\text{max} \) do

distribute parameter \( \theta \) to the local networks and run algorithm 1, get the gradient \( \partial \mathcal{L}/\partial \theta \);

Apply the gradient \( \partial \mathcal{L}/\partial \theta \) to the global networks using ADAM optimizer;

end

and immediately obtain the results shown in figure 1, or get a brand new result just by changing the value of \( N \) and \( T \). For capacity constraints of a free account of GitHub, data of other cases of \( N \) will not be uploaded and is available from the corresponding author upon reasonable request.

**Appendix B. Comparison to OAT and TACT approaches**

Here we compare the resultant QFI of our work to these obtained by OAT and TACT approaches. The corresponding Hamiltonian can be expressed by:

\[ H_{\text{OAT}} = \chi J_z^2 \quad \text{(B.1)} \]

and:

\[ H_{\text{TACT}} = \chi (J_z^2 - J_y^2)/3. \quad \text{(B.2)} \]

The initial state is the same as equation (1) that we used in our work, and the final states are thus \( |\psi\rangle_\text{OAT} = e^{-iH_{\text{OAT}}T}|\psi\rangle_0 \) and \( |\psi\rangle_\text{TACT} = e^{-iH_{\text{TACT}}T}|\psi\rangle_0 \). We denote \( \hbar = \chi = 1 \) in all simulations. The total evolution time \( T \) is also the same as those used in our algorithm for different cases of \( N \). The QFI is also calculated by equation (2) to get a value that is meaningful in the situation of Ramsey interferometry.

For the case of OAT, the computation of QFI is implemented after a rotation operation introduced by \( e^{-i\frac{\pi}{2}J_y} \)
Figure B1. Evolutions of QFI $F_Q$ versus evolution time $t$. Orange lines are results from OAT evolutions and green lines are results from TACT evolutions. Blue dotted lines are results from our DRL algorithm, evolutions under the optimized pulse trains, displayed in the same way as in figures 1(f) and (j). Four columns of subfigures are results from $N = 50, 100, 500$ and 1000 cases. The upper line of subfigures are results from ‘only-$J_x$’ scheme, and the lower line of subfigures are results from ‘both-$J_x, J_y$’ scheme.

The results are shown in figure B1 including cases of $N = 50, 100, 500$ and 1000. Our results outperform the conventional OAT and TACT schemes within the same $T$. The comparison shows that our algorithm provides easy implementation and efficiency of the pulse trains, while the evaluation of TACT Hamiltonian equation (B.2) is hard to realized in experiments.

Appendix C. QFI under uncertainties of atom number

In section 4, we have discussed the robustness of our schemes against the atom number difference between simulation and experiment. Here we consider the situation that the resultant states are mixed states containing many states of different atom number $N$.

Without loss of generality, we set a Gaussian noise on the atom number $N$. For a certain desired atom number $N_0$, the distribution of $N$ obeys a Gaussian distribution as $N \sim \text{Normal}(N_0, \gamma^2 N_0)$, where the standard deviation is $\gamma \sqrt{N_0}$ and we use $\gamma$ to represent the noise strength. The probability of a certain state of atom number $N$ is $p_N$, and probabilities of the $N_0 = 50$ case is plotted in figure C1(a) for $\gamma = 0, 0.5, 1, 1.5$ and 2. Then we calculate the QFI for the mixed state, whose density operator can be expressed as:

$$
\rho = \sum_{N} p_N |\psi_N\rangle \langle \psi_N|.
$$

(C.1)

Each state $|\psi_N\rangle$ evolves from the initial state (equation (1)) of atom number $N$, while under the pulse train from the simulation of $N_0$. The QFI $F_Q$ is calculated by [75]:

$$
F_Q = 2 \sum_{N_j, N_k} \left(\frac{p_{N_j} - p_{N_k}}{p_{N_j} + p_{N_k}}\right)^2 |\langle \psi_{N_j}| \hat{J}_z |\psi_{N_k}\rangle|^2,
$$

where we also use operator $\hat{J}_z$ to derive the QFI that is meaningful in the situation of Ramsey interferometry.

In figure C1(b) we plot the resultant $F_Q$ in the case of $N_0 = 50$ under different noise strength. It is clear that when the value of $\gamma$ increases, the corresponding $F_Q$ decreases, while the decrease of the results from both-$J_x, J_y$ scheme is more gentle.

Then, to find out how the noise of $N$ affects the scaling of $F_Q$ that shows in figures 1(h) and (l), we simulate the case of different $N_0$, including $N_0 = 50, 100, 500, 1000$ and 2000. The corresponding results are shown in figures C1(c) and (d), including those from two schemes and under different strength of noise. It shows that the noise of $N$ has little change to the scaling of $F_Q$. The $F_Q$ decrease a small amount while the decrement rate also decreases when the value of $N_0$ increase. Since a factor of $\gamma = 2$ is quite large, our results shows good robustness against noise of atom number $N$, and states prepared by both-$J_x, J_y$ scheme also have better performance.
Figure C1. (a) Histograms are the Gaussian distributions of probability $p_j$ in the case of $N_0 = 50$. Distributions under different noise strength $\gamma = 0, 0.5, 1, 1.5, 2$ from top to low. (b) Blue dotted dashed lines are the resultant $P_0$ from the mixed state under a noisy distribution of $N$, also in the $N_0 = 50$ case. The upper subfigure is the results from ‘only-$j_x$’ scheme and the lower subfigure is those from ‘both-$j_x,j_y$’ scheme.

ORCID iDs

Yuxiang Qiu  
https://orcid.org/0000-0002-5279-1524
Min Zhuang  
https://orcid.org/0000-0002-3685-8207
Chaohong Lee  
https://orcid.org/0000-0001-9883-5900

References

[1] Giovannetti V, Lloyd S and Maccone L 2006 Phys. Rev. Lett. 96 010401
[2] Giovannetti V, Lloyd S and Maccone L 2011 Nat. Photon. 5 222–9
[3] Gross C, Zibidlo T, Nicklas E, Estève J and Oberthaler M K 2010 Nature 464 1165–9
[4] Lücke B et al 2011 Science 334 773–6
[5] Pezzè L, Smerzi A, Oberthaler M K, Schmied R and Treutlein P 2018 Rev. Mod. Phys. 90 035005
[6] Caves C M 1981 Phys. Rev. D 23 1683–708
[7] Sørensen A S and Mølmer K 2001 Phys. Rev. Lett. 86 2304–7
[8] Xu H, Li J, Liu L, Wang Y, Yuan H and Wang X 2019 npj Quantum Inf. 5 82
[9] Schuster F, Kloc M, Dauert L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 Rev. Mod. Phys. 91 045002
[10] Sutton R S and Barto A G 2018 Reinforcement Learning: An Introduction (Cambridge, MA: MIT Press)
[11] Sutton R S and Barto A G 2018 Rep. Prog. Phys. 81 074001
[12] Palittapongarnpim P, Wittek P, Zadehnejad E, Vedaie S and Sanders B C 2017 Neurocomputing 268 116–26
[13] Fösel T, Tighineanu P, Weiss T and Marquardt F 2018 Phys. Rev. X 8 031084
[14] Wallnofer J, Melnikov A A, Dür W and Briegel H J 2020 PRX Quantum 1 010301
[15] Chen Y Q, Chen Y, Lee C K, Zhang S and Hsieh C Y 2020 arXiv:2004.02836
[16] Schäfer F, Kloc M, Bruder C and Lörch N 2020 Mach. Learn.: Sci. Technol. 1 035009
[17] Saito H 2020 J. Phys. Soc. Japan 89 074006
[18] Rose D C, Mair J F and Garrahan J P 2021 New J. Phys. 23 013013
[19] Fallani A, Rossi M A C, Támascelli D and Genoni M G 2022 PRX Quantum 3 020310
[20] Erdman P A and Noé F 2022 npj Quantum Inf. 8 1
[21] Xu H, Li J, Liu L, Wang Y, Yuan H and Wang X 2019 npj Quantum Inf. 5 82
[44] Chen F, Chen J J, Wu L N, Liu Y C and You L 2019 Phys. Rev. A 100 041801
[45] Predko A, Albarelli F and Serafini A 2020 Phys. Lett. A 384 126268
[46] Haug T, Mok W K, You J H, Mao T W, Png C E and Kwek L C 2020 Mach. Learn.: Sci. Technol. 2 01LT02
[47] Xu H, Wang L, Yuan H and Wang X 2020 arXiv:2012.13577
[48] Schuff J, Fiderer I J and Braun D 2020 New J. Phys. 22 035001
[49] An Z, Song H J, Hu Q K and Zhou D L 2021 Phys. Rev. A 103 012404
[50] Fiderer I J, Schuff J and Braun D 2021 PRX Quantum 2 020303
[51] Guo S F, Chen F, Liu Q, Xue M, Chen J J, Cao J H, Mao T W and You L 2021 Phys. Rev. A 103 012404
[52] Braunstein S L and Caves C M 1994 Phys. Rev. Lett. 72 3439–43
[53] Braunstein S L, Caves C M and Milburn G 1996 Ann. Phys., NY 247 135–73
[54] Muessel W, Strobel H, Linnemann D, Zibold T, Julíà-Díaz B and Oberthaler M K 2015 Phys. Rev. A 92 023603
[55] Hall D S, Matthews M R, Ensher J R, Wieman C E and Cornell E A 1998 Phys. Rev. Lett. 81 1539–42
[56] Cirac J I, Lewenstein M, Mølmer K and Zoller P 1998 Phys. Rev. A 57 1208–18
[57] Schumm T, Hofferberth S, Andersson L M, Wildermuth S, Groth S, Bar-Joseph I, Schmiedmayer J and Krüger P 2005 Nat. Phys. 1 57–62
[58] Hall B V, Whitlock S, Anderson R, Hannaford P and Sidorov A I 2007 Phys. Rev. Lett. 98 030402
[59] Jo G B, Shin Y, Will S, Pasquini T A, Saba M, Ketterle W, Pritchard D E, Vengalattore M and Prentiss M 2007 Phys. Rev. Lett. 98 030407
[60] Gross C 2012 J. Phys. B: At. Mol. Opt. Phys. 45 103001
[61] Radcliffe J M 1971 J. Phys. A: Gen. Phys. 4 313–23
[62] Arecchi F T, Courtens E, Gilmore R and Thomas H 1972 Phys. Rev. A 6 2211–37
[63] Zhang W M, Feng D H and Gilmore R 1990 Rev. Mod. Phys. 62 867–927
[64] Paris M G A 2009 Int. J. Quantum Inf. 07 125–37
[65] Mnihi V, Badia A, Mirza M, Graves A, Lillicrap T, Harley T, Silver D and Kavukcuoglu K 2016 arXiv:1602.01783v2
[66] Paszke A et al 2017 NIPS-W https://openreview.net/forum?id=BJJsrmfCZ
[67] Ramsey N F 1963 Molecular Beams (Oxford: Oxford University Press)
[68] Demkovicz-Dobrzanski R, Jarzyna M and Kolodyński J 2015 Chapter Four—Quantum Limits in Optical Interferometry vol 60 (Amsterdam: Elsevier)
[69] Frowis F, Sekatski P and Dürr W 2016 Phys. Rev. Lett. 116 090801
[70] Davis E, Bentsen G and Schleier-Smith M 2016 Phys. Rev. Lett. 116 053601
[71] Colombo S, Pedrozo-Peaief L, Adiyatullin A F, Li Z, Mendez E, Shu C and Vuletic V 2021 Time-reversal-based quantum metrology with many-body entangled states (arXiv:2106.03754)
[72] Linnemann D, Strobel H, Muesse W, Schulz J, Lewis-Swan R J, Kheruntsyan K V and Oberthaler M K 2016 Phys. Rev. Lett. 117 030001
[73] Gilmore K A, Affolter M, Lewis-Swan R J, Barberena D, Jordan E, Rey A M and Bollinger J J 2021 Science 373 673–8
[74] Strobel H, Muesse W, Linnemann D, Zibold T, Hume D B, Pezzé I, Smerzi A and Oberthaler M K 2014 Science 345 424–7
[75] Dorner U 2012 New J. Phys. 14 043011
[76] Qiu Y 2022 DRL_OFI_optimize https://github.com/Aburiann/DRL_OFI_optimize