New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
11 September 2019

REVISED
13 January 2020

ACCEPTED FOR PUBLICATION
23 January 2020

PUBLISHED
6 March 2020

PAPER

Improving the dynamics of quantum sensors with reinforcement learning

Jonas Schuff, Lukas J Fiderer and Daniel Braun

Institute for Theoretical Physics, University of Tübingen, Auf der Morgenstelle 14, D-72076 Tübingen, Germany

E-mail: jonas.schuff@student.uni-tuebingen.de

Keywords: quantum metrology, machine learning, reinforcement learning, quantum-chaotic sensors, control theory, spin squeezing, superradiance master equation

Abstract

Recently proposed *quantum-chaotic sensors* achieve quantum enhancements in measurement precision by applying nonlinear control pulses to the dynamics of the quantum sensor while using classical initial states that are easy to prepare. Here, we use the cross-entropy method of reinforcement learning (RL) to optimize the strength and position of control pulses. Compared to the quantum-chaotic sensors with periodic control pulses in the presence of superradiant damping, we find that decoherence can be fought even better and measurement precision can be enhanced further by optimizing the control. In some examples, we find enhancements in sensitivity by more than an order of magnitude. By visualizing the evolution of the quantum state, the mechanism exploited by the RL method is identified as a kind of spin-squeezing strategy that is adapted to the superradiant damping.

1. Introduction

The rise of machine learning [1] has led to intense interest in using machine learning in physics, and in particular in combining it with quantum information technology [2, 3]. Recent success stories include discriminating phases of matter [4–6] and efficient representation of many-body quantum states [7–9].

In physics, many problems can be described within control theory which is concerned with finding a way to steer a system to achieve a goal [10]. The search for optimal control can naturally be formulated as reinforcement learning (RL) [11–19], a discipline of machine learning. RL has been used in the context of quantum control [17], to design experiments in quantum optics [20], and to automatically generate sequences of gates and measurements for quantum error correction [16, 21, 22].

RL has also been applied to control problems in quantum metrology [2]: in the context of global parameter estimation, i.e. when the parameter is *a priori* unknown, the problem of optimizing single-photon adaptive phase-estimation was investigated [23–25]. There, the goal is to estimate an unknown phase difference between the two arms of a Mach–Zehnder interferometer. After each measurement, an additional controllable phase in the interferometer can be adjusted dependent on the already acquired measurement outcomes. The optimization with respect to policies, i.e. mappings from measurement outcomes to controlled phase shifts, can be formulated as a RL problem and tackled with particle swarm [23, 24, 26, 27] or differential evolution [25, 28] algorithms, where the results of the former were recently applied in an experiment [29].

Also in the regime of local parameter estimation, where the parameter is already known to high precision (typically from previous measurements), actor-critic and proximal-policy-optimization RL algorithms were used to find policies to control the dynamics of quantum sensors [30–32]. There, the estimation of the precession frequency of a dissipative spin-$\frac{1}{2}$ particle was improved by adding a linear control to the dynamics in form of an additional controlled magnetic field [32].

Recently it was shown theoretically that the sensitivity (in the regime of local parameter estimation) of existing quantum sensors based on precession dynamics, such as spin-precession magnetometers, can be increased by adding nonlinear control to their dynamics in such a way that the dynamics becomes non-regular or (quantum-)chaotic [33, 34]. The nonlinear kicks (described by a 'nonlinear' Hamiltonian $\propto J_y^2$ compared to

the 'linear' precession Hamiltonian $\propto J_z$ where J_x, J_y, J_z are the spin angular momentum operators) lead to a torsion, a precession with rotation angle depending on the state of the spins.

Adding nonlinear kicks to the otherwise regular dynamics comes along with a large number of new degrees of freedom that remained so far unexplored: rather than kicking the system periodically with always the same strength and with the same preferred axis as in [33], one can try to optimize each kick individually, i.e. vary its timing, strength, or rotation axis. The number of parameters increases linearly with the total measurement time (assuming a fixed upper bound of kicks per unit time), and becomes rapidly too large for brute-force optimization.

In this work, we use cross-entropy RL to optimize the kicking strengths and times in order to maximize the quantum Fisher information (QFI), whose inverse constitutes a lower bound on the measurement precision. The cross-entropy method is used to train a neural network that takes the current state as input and gives an action on the current state (the nonlinear kicks) as output. In this way, the neural network generates a sequence of kicks that represents the policy for steering the dynamics.

This represents an offline, model-free approach which is aimed at long-term performance, i.e. the optimization is done based on numerical simulations, without being restricted to a specific class of policies, and with the goal of maximizing the QFI only after a given time and not, as it would be the case for greedy algorithms, for each time step. We show that this can lead to largely enhanced sensitivity even compared to the already enhanced sensitivity of the quantum-chaotic sensor with constant periodic kicks [33].

2. Quantum metrology

The standard tool for evaluating the sensitivity with which a parameter can be measured is the quantum Cramér–Rao bound [35–37]. It gives the smallest uncertainty with which a parameter ω encoded in a quantum state (density matrix) ρ_ω can be estimated. The bound is optimized over all possible (POVM = positive operator valued measure) measurements (including but not limited to standard projective von-Neumann measurements of quantum observables), and all possible data-analysis schemes in the sense of using arbitrary unbiased estimator functions $\hat{\omega}$ of the obtained measurement results. It can be saturated in the limit of a large number M of measurements, and hence gives the ultimate sensitivity that can be reached once technical noise has been eliminated and only the intrinsic fluctuations due to the quantum state itself remain.

The quantum Cramér–Rao bound for the smallest possible variance of the estimate $\hat{\omega}$ reads

$$\mathrm{Var}(\hat{\omega}) \geqslant \frac{1}{MI_\omega}. \tag{1}$$

For a state given in diagonalized form, $\rho_\omega := \sum_{\ell=1}^{d} p_\ell |\psi_\ell\rangle\langle\psi_\ell|$, where d is the dimension of the Hilbert space, the QFI is given by [38]

$$I_\omega = 2\sum_{\ell,m=1}^{d} \frac{|\langle\psi_\ell|\partial_\omega\rho_\omega|\psi_m\rangle|^2}{(p_\ell + p_m)^2}, \tag{2}$$

where the sum runs over all ℓ, m such that $p_\ell + p_m \neq 0$, and $\partial_\omega\rho_\omega := \frac{\partial\rho_\omega}{\partial\omega}$.

3. The system

We consider a spin model based on the angular momentum algebra, with spin operators $\mathbf{J} = (J_x, J_y, J_z)$, $J_z|jm\rangle = \hbar j|jm\rangle$ and $\mathbf{J}^2|j, m\rangle = \hbar^2 j(j+1)|j, m\rangle$, where j and m are angular momentum quantum numbers. Note that the model can be implemented not only with physical spins but with any physical system with quantum mechanical operators that fulfill the angular momentum algebra. The Hamiltonian of our model is given by

$$\mathcal{H}_{\mathrm{KT}}(t) = \omega J_z + \frac{J_y^2}{(2j+1)\hbar}\sum_{\ell=-\infty}^{\infty}\kappa_\ell\tau\delta(t - t_\ell). \tag{3}$$

The first summand describes a precession about the z-axis with precession frequency ω. The second summand describes the nonlinear kicks, i.e. a torsion about the y-axis, see figure 1. This corresponds to a precession about the y-axis with a precession angle proportional to the y-component. The time τ defines a time scale such that t and t_ℓ measure time in units of τ. The ℓth kick is applied at time t_ℓ where κ_ℓ quantifies its kicking strength (in units of a frequency).

In an atomic spin-precession magnetometer, as discussed in [33], the first summand corresponds to a Larmor precession characterized by the Larmor frequency $\omega = g\mu_B B/\hbar$ with Landé g-factor g, Bohr magneton μ_B, and magnetic field strength B, which is the parameter that one wants to estimate. The nonlinear kicks can, for

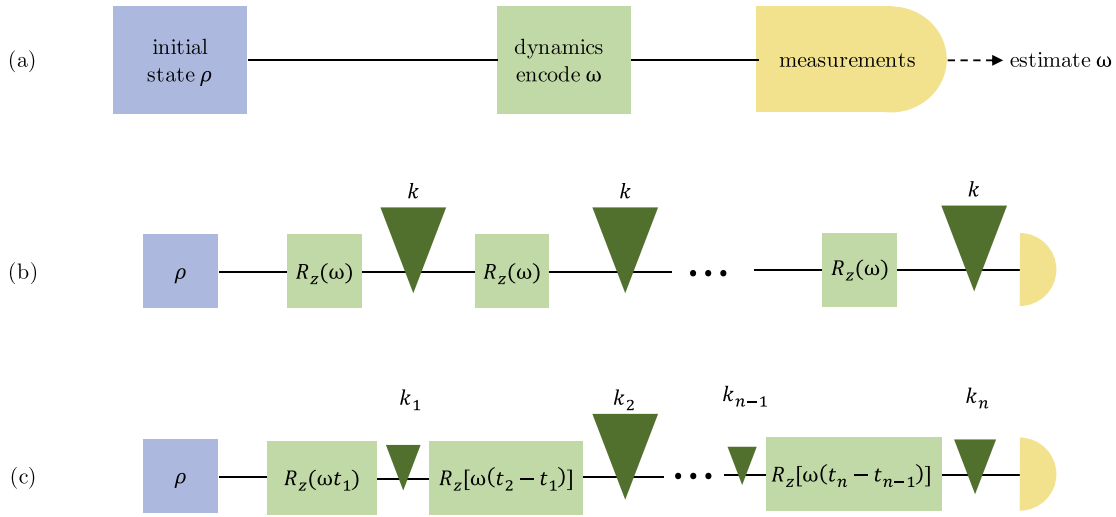

Figure 1. Schematic representation of parameter encoding in quantum metrology. Panel (a) shows the standard protocol: the parameter ω is encoded in the initial state ρ through the dynamics, the resulting state is measured, and the parameter is inferred by (classical) post processing of the measurement outcomes. In panel (b), the dynamics is given by the kicked top model: the encoding of the parameter ω through linear precession $R_z(\omega)$ about the z-axis is periodically disrupted through parameter-independent, nonlinear, controlled kicks (green triangles) with kicking strength k that can render the dynamics chaotic. In panel (c), the dynamics is given by a generalized kicked top model: the kicking strengths k_ℓ and times t_ℓ between kicks are optimized in order to maximize the sensitivity with which ω can be inferred (varying k_ℓ are indicated by different sizes of the green triangles). Variation of the kicking axis is possible but beyond the scope of this work.

example, be generated with off-resonant light pulses exploiting the ac Stark effect. We introduce a dimensionless kicking strength as $k_\ell := \kappa_\ell \tau$ and, for the sake of simplicity, we set $\tau = 1$ and $\hbar = 1$.

For a pure state, the unitary time evolution of the system between kicks at time $t_{\ell-1}$ and t_ℓ is given by

$$|\psi_\omega(t_\ell)\rangle = U_\omega(k_\ell)|\psi(t_{\ell-1})\rangle, \tag{4}$$

where the unitary transformation $U_\omega(k_\ell)$ propagates the state according to the Hamiltonian (3), from time $t_{\ell-1}$ [directly after the $(\ell-1)$ th kick] to t_ℓ [directly after the ℓth kick], as indicated by the index ℓ [in order to simplify notation, the index ℓ of k not only labels the kicking strength at time t_ℓ but also refers to the propagation from $t_{\ell-1}$ to t_ℓ of $U_\omega(k_\ell)$]. We have

$$U_\omega(k_\ell) = \mathcal{T}\exp\left[-\mathrm{i}\int_{t_{\ell-1}}^{t_\ell} \mathrm{d}t' \mathcal{H}_{\mathrm{KT}}(t')\right], \tag{5}$$

where $\mathcal{T}$ denotes time-ordering. Since the kicks are assumed to be instantaneous, this leads to

$$U_\omega(k_\ell) = \exp\left[-\mathrm{i}\, k_\ell \frac{J_y^2}{(2j+1)}\right] \exp\left[-\mathrm{i}\,\omega(t_\ell - t_{\ell-1})J_z\right], \tag{6}$$

i.e. a precession for time $t_\ell - t_{\ell-1}$ followed by a kick of strength k_ℓ. The kick occurs at the end of the time interval $[t_{\ell-1}, t_\ell]$.

For the standard kicked top (KT), see figure 1, the kicking strengths are constant, $k_\ell = k$, and kicking times are given by $t_\ell = \ell\tau = \ell$, with $\ell \in \mathbb{N}$. Dynamics of the standard KT is non-integrable for $k > 0$ and has a well defined classical limit that shows a transition from regular to chaotic dynamics when k is increased. In [33] the behavior of the QFI for regular and chaotic dynamics was studied in this transition regime (for $k = 3$ and $\omega = \pi/2$) which manifests itself by a mixed classical phase space between regular and chaotic dynamics. Quantum chaos is defined as quantum dynamics that becomes chaotic in the classical limit. In contrast to classical chaos, quantum chaos does not exhibit exponential sensitivity to changes of initial conditions due to the properties of unitary quantum evolution, but can be very sensitive to parameters of the evolution [39]. The KT has been realized with atomic spins in a cold gas [40] and with a pair of spin-$\frac{1}{2}$ nuclei using NMR techniques [41]. Here, we generalize the standard KT to kicks of strength k_ℓ at arbitrary times t_ℓ as given in equation (6), see also figure 1.

Any new quantum metrology method needs to demonstrate its viability in the presence of noise and decoherence. We study two different versions of the KT which differ in the decoherence model used: phase damping and superradiant damping. Both can be described by Markovian master equations and are well studied models for open quantum systems [42–45]. While phase damping conserves the energy and only leads to decoherence in the $|j, m\rangle$ basis, superradiant damping leads in addition to a relaxation to the ground state

$|j, -j\rangle$. Its combination with periodic kicking in the chaotic regimes is known to give rise to a non-equilibrium steady state in the form of a smeared-out strange attractor [45] that still conserves information about the parameter ω, whereas without the kicking the system in presence of superradiant damping simply decays to the ground state. The master equations for both processes have the Kossakowski–Lindblad form [46, 47], with

$$\dot{\rho}(t) = \gamma_{\mathrm{pd}}([J_z, \rho(t)J_z] + \mathrm{h.c.}) \tag{7}$$

for phase damping, where $\dot{\rho}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\rho(t)$, and

$$\dot{\rho}(t) = \gamma_{\mathrm{sr}}([J_-, \rho(t)J_+] + \mathrm{h.c.}) \tag{8}$$

for superradiant damping, where $J_\pm := J_x \pm \mathrm{i}J_y$ are the ladder operators, and γ_{pd} and γ_{sr} denote the decoherence rates. With the generator Λ, defined by $\dot{\rho}(t) = \Lambda\rho(t)$, one has in both cases the formal solution $\rho(t_n) = D(t_n - t_{n-1})\rho(t_{n-1})$ with the continuous-time propagator $D(t) := \mathrm{e}^{\Lambda t}$. The solution of equation (7) in the $|j, m\rangle$ basis, where $\rho(t) = \sum_{m,m'=-j}^{j} \rho_{m,m'}(t)|j, m\rangle\langle j, m'|$, is immediate,

$$\rho_{m,m'}(t) = \rho_{m,m'}(0)\exp[-\gamma_{\mathrm{pd}}t(m - m')^2]. \tag{9}$$

Also for equation (8) a formally exact solution has been found [48] and efficient semiclassical (for large j) expressions are available [49, 50]. For our purposes it was the simplest to solve equation (8) numerically by diagonalization of Λ. Combining these decoherence mechanisms with the unitary evolution the transformation $\rho(t_{\ell-1}) \to \rho(t_\ell)$ reads

$$\rho(t_\ell) = U_\omega(k_\ell)[D(t_\ell - t_{\ell-1})\rho(t_{\ell-1})]\,U_\omega(k_\ell)^\dagger, \tag{10}$$

because in both cases the dissipative generator Λ commutes with the precession.

As initial state we use an SU(2) coherent state, which can be seen as the most classical state of a spin [51, 52], and is usually easy to prepare (for instance by optically polarizing the atomic spins in a SERF magnetometer). Also, it is equivalent to a symmetric state of $2j$ spin-$\frac{1}{2}$ pointing all in the same direction. With respect to the $|j, m\rangle$ basis it reads

$$|j, \theta, \phi\rangle = \sum_{m=-j}^{j} \sqrt{\binom{2j}{j-m}} \sin\left(\frac{\theta}{2}\right)^{j-m} \cos\left(\frac{\theta}{2}\right)^{j+m} \mathrm{e}^{\mathrm{i}(j-m)\phi}|j, m\rangle. \tag{11}$$

We choose $\theta = \frac{\pi}{2}$, $\phi = \frac{\pi}{2}$.

4. Optimizing the KT

4.1. The KT as a control problem

We consider the KT as a control problem and discretize the kicking strengths k_ℓ and times t_ℓ. The precise parameters of the discretized control problem vary between the following examples and are summarized in appendix A. In the following, t_{step} denotes a discrete time step (measured in units of $\tau = 1$), k_{step} is a discrete step of kicking strength, the RL agent optimizes the QFI at time T_{opt}, and we bound the total accumulated kicking strength $\sum_\ell k_\ell < 15000$ which is never reached in optimized policies though. The frequency ω, that we want to estimate, is set to induce a rotation of the state by $t\pi/2$ (t is measured in units of $\tau = 1$).

Possible control policies are simply given by a vector of kicking strengths $\boldsymbol{k} = (k_1,\ldots,k_N) \in \mathbb{R}^N$ with $k_\ell \in \{qk_{\mathrm{step}}: q = 0, 1, 2,\ldots\}$. To each policy corresponds a QFI value, calculated from the resulting state $\rho(T_{\mathrm{opt}})$, which quantifies how well the policy performs. To tackle this type of problem, various numerical algorithms are available, each with its own advantages and drawbacks [2, 3, 15]. We pursue the relatively unexplored (in the context of physics) route of cross-entropy RL.

4.2. Reinforcement learning

Figure 2 shows the typical way we imagine RL. There is an *agent* that interacts with an *environment* by choosing *actions* and receiving an *observation*and a *reward* from the environment. One cycle of action and observation/reward is called a *step*.

In general, the idea of RL is to reinforce behavior that leads to high rewards. The precise mechanism depends on the used RL algorithm.

4.3. The KT as a RL problem

The system, the generalized KT as introduced in section 3, represents the RL environment. The agent can choose between only two actions: (i) increase the kicking strength (by k_{step}) or (ii) go on from the current position in time ℓt_{step} to $(\ell + 1)t_{\mathrm{step}}$. In this way, the vector $\boldsymbol{k}$ is built up step by step. After each action, the agent obtains an

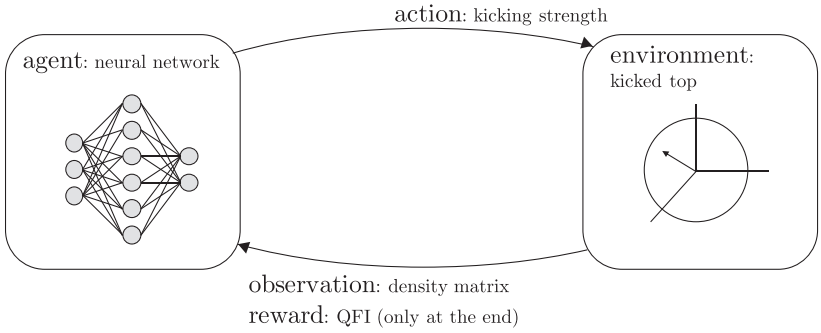

Figure 2. Typical setup of reinforcement learning: the RL agent acts upon the environment which in return gives the RL agent an observation and a reward. In our case the RL agent is a neural network and the environment is the generalized kicked top.

observation given by the full density matrix of the current state of the environment. Since we simulate the evolution of the environment, the density matrix is readily available.

Only after the total time T_{opt}, a reward [the QFI of $\rho(T_{\text{opt}})$] is given to the agent. This concludes one *episode*, and the resulting vector $\boldsymbol{k}$ represents a policy. Then, the environment is reset (i.e. the spin is reinitialized with the coherent state at $\theta = \frac{\pi}{2}$, $\phi = \frac{\pi}{2}$, see equation (11)), and the next episode starts.

A neural network represents the RL agent: the observation is given to the neural network's input neurons while each output neuron represents one possible action, i.e. we have two output neurons for 'kick' and 'go on'. The activation of these output neurons determines the probability of executing that action. The policy, however, is not given by the neural network directly. Since the environment is deterministic (i.e. the state evolves deterministically for a given policy $\boldsymbol{k}$ of kicking strengths) there is no point in choosing a stochastic policy such as a neural network. Instead, a single choice of kicking strengths $\boldsymbol{k}$ represents the policy. We obtain this by first training the neural networks using the cross-entropy method, then generating a few episodes with the trained neural network, and then picking the episode with the largest QFI. The kicking strengths applied in that episode represent the policy[1].

4.4. Cross-entropy method

The RL cross-entropy method [53] we use works as follows: we first produce a set of episodes (i.e. we obtain several vectors $\boldsymbol{k}$) with a neural network that is initialized randomly. Then, we rank those episodes according to their reward[2]. We select the best 10% of episodes (with highest reward) for further computations. Every episode can be split into several pairs of action and observation and we use those pairs to train the neural network with the stochastic gradient descent method called *Adam* [54]. As a result of this training, the weights of the neural network are adjusted, i.e. the agent learns from its experience. Future actions taken by the agent are influenced not only by randomness but also by this experience. One run of producing episodes, ranking them, and using the best 10% to train the neural network is called an *iteration*. Training a neural network consists of several iterations. See appendix C for pseudocode of this algorithm. For the parameters of the training process see appendix B. In appendix D we study the learning success for different numbers of episodes and iterations.

5. Results

We compare the QFI for different models: (i) the top (simple precession without kicks), (ii) the standard KT, as studied in [33], with periodic kicks (period $\tau = 1$, i.e. a precession angle of $\pi/2$ for one period, and kicking strength $k = 30$), and (iii) the generalized KT optimized with RL. In case of superradiance damping (phase damping) we denote the top by SR-T (PD-T), the standard KT by SR-KT (PD-KT) and the RL-optimized generalized KT by SR-GKT (PD-GKT). Details on the training and the optimization of the RL results are provided in appendix B.

Let us first consider superradiant damping with results presented in figure 3. The QFI for the SR-T exhibits a characteristic growth quadratic in time. However, due to decoherence, the QFI does not maintain this growth

[1] In comparison, Sanders *et al* [23–25] restricted their policy search for adaptive single-photon interferometry in such a way that their search space corresponds to points in $\mathbb{R}^N$, making it similar to our problem. However, in their case the observations from the environment are probabilistic measurement outcomes while in our case the observation is the deterministic state ρ.

[2] We do not give an immediate reward at every step but only at the very end of an episode, and the reward is not reduced with the number of steps (i.e. the discount factor is 1).

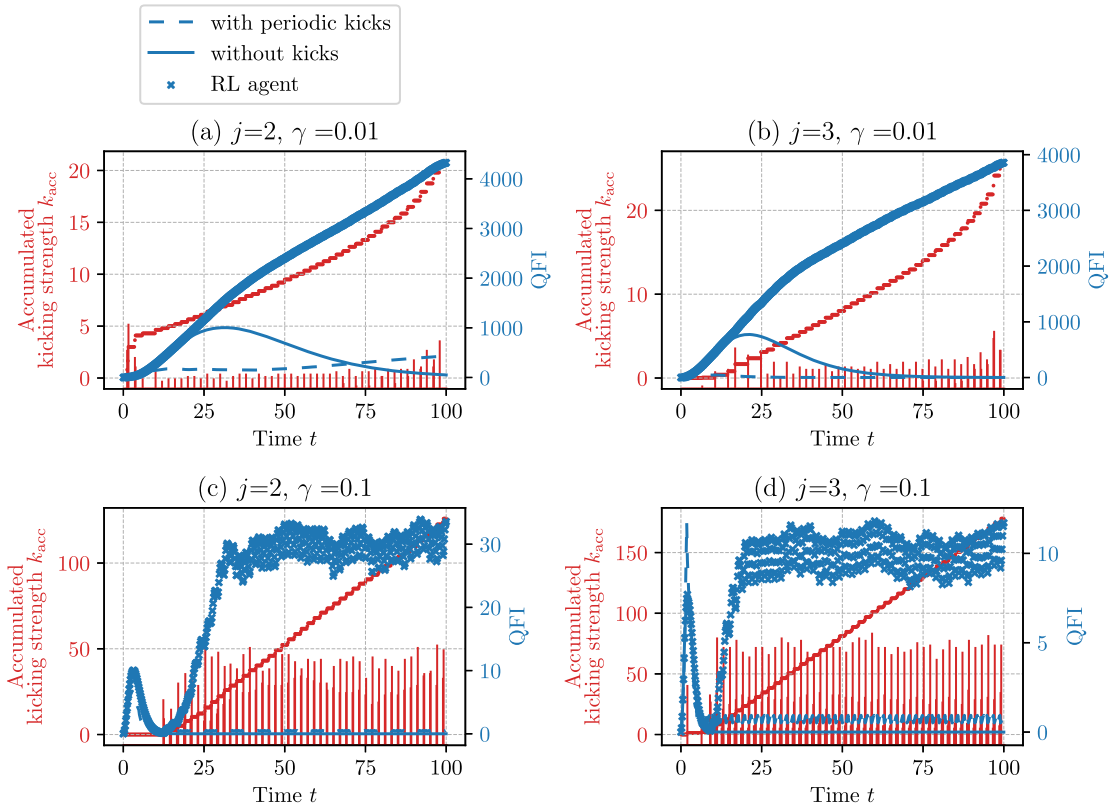

Figure 3. Examples for the policy adopted by the RL agent for superradiant damping. We plot the accumulated kicking strength $k_{\mathrm{acc}}(t)$ (integrating over the kicks from time 0 to t) on the left axes as red dots and on the right axes in blue the quantum Fisher information for the top (solid line), the periodically kicked top with $k = 30$ chosen as in [33] (dashed line) and the QFI that corresponds to the policy of the RL agent (crosses). We additionally plot red vertical lines in the places, where the RL agent decides to set a kick. The height of the lines correspond to the kicking strength in arbitrary units and are not on the scale of the left axis. There is a regime where the RL agent manages to increase the QFI with each time step (panels (a) and (b)), and a regime where the RL agent makes the QFI oscillate (panels (c) and (d)).

but starts to decay rapidly towards zero. The time when the QFI reaches its maximum was found to decay roughly as $1/(\gamma_{\mathrm{sr}} j)$ with spin size j and damping rate γ_{sr} [33].

The situation changes with the introduction of nonlinear kicks. There, the QFI for the SR-KT shows the interesting behavior of not decaying to zero for large times. Instead it reaches a plateau value which was found to take surprisingly high values for specific choices of j and dissipation rates [33], in particular, for $j = 2$. The system looses energy through superradiant damping while the nonlinear kicks add energy. This prevents the state from decaying to the ground state, which is an eigenstate of the precession and would lead to a vanishing QFI. From this perspective, the plateau results from a dynamical equilibrium established by the interplay of superradiant damping and kicks.

However, the full potential of exploiting such effects and increasing the QFI with the help of nonlinear kicks is not achieved with constant periodic kicks. Instead, the RL agent[3] finds policies to make the QFI of the SR-GKT increase further even when the QFI of the SR-T decayed already to zero and the QFI of the SR-KT reached its plateau value.

Examples for $j = 2$ and $j = 3$ are presented in figure 3. The QFI of the SR-GKT is optimized for a total time T_{opt} which is the largest time plotted in each example. At T_{opt}, the plateau value of the SR-KT for $j = 3$ is relatively low and the RL-optimized policy achieves an improvement in sensitivity (associated with $1/\sqrt{I_\omega}$) of more than an order of magnitude. Panels (a) and (b) show continuous growth of the QFI through an optimized kicking policy. Only if the time T_{opt} (the QFI is optimized to be maximal at T_{opt}) is increased further, the impressive growth of the QFI finally breaks down. Instead of increasing T_{opt}, we choose to increase superradiant damping while keeping T_{opt} constant, which has a similar effect. In that case, see panels (c) and (d), the RL agent chooses a policy which makes the QFI oscillate at a relatively high level before the time T_{opt} is reached.

The superiority of the policies found by the RL agent can be understood by taking a look at the evolution of the quantum state, see figure 4: we represent the quantum state in the space of $\boldsymbol{r} = (x, y, z) = (\langle J_x\rangle, \langle J_y\rangle, \langle J_z\rangle)$ where $\langle J_\ell\rangle := \mathrm{tr}(\rho J_\ell)$ and, due to the conservation of angular momentum, $|\boldsymbol{r}| = 1$ which restricts the space to a

[3] The training of one RL agent takes about eight hours on a desktop computer.

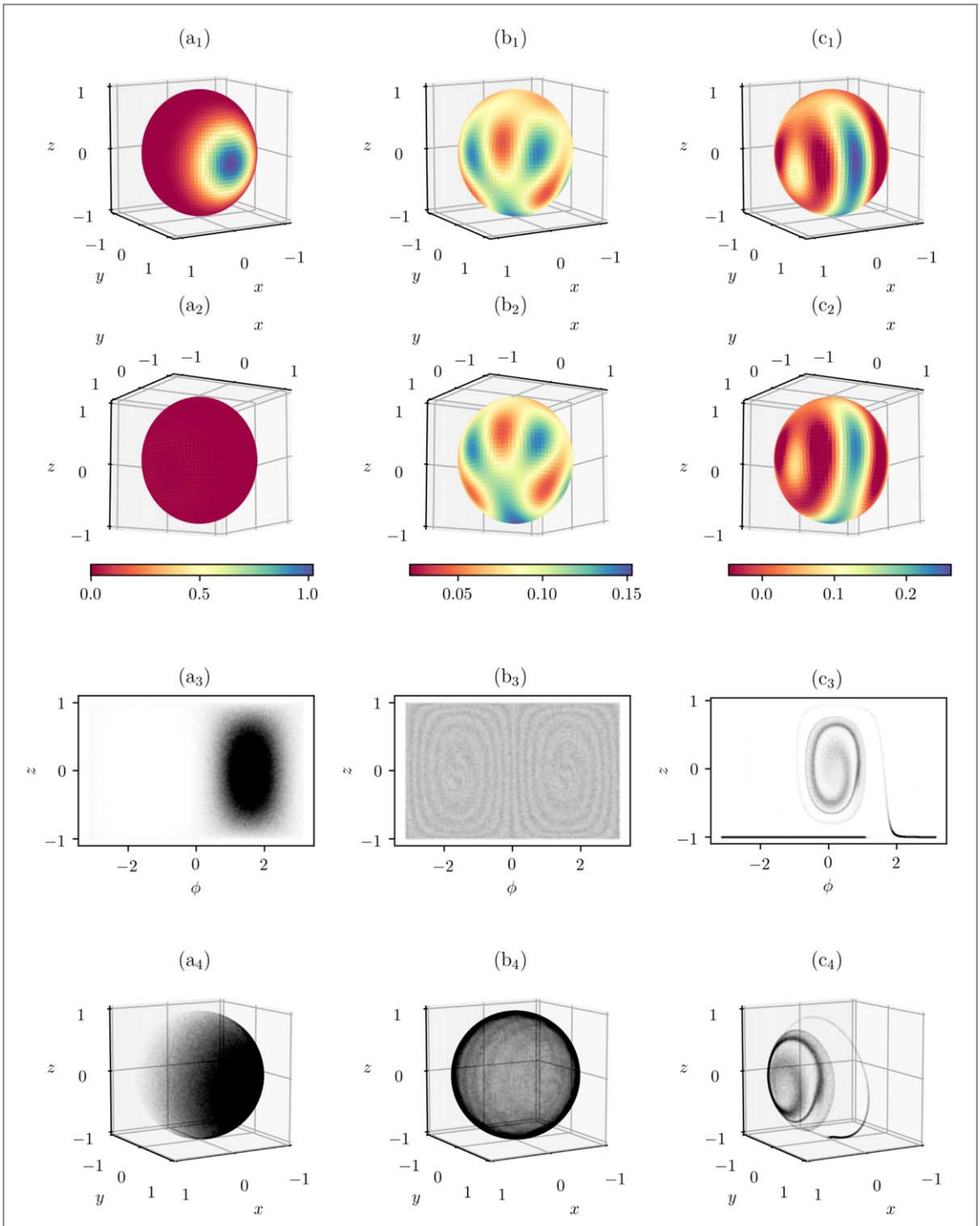

Figure 4. Illustration of kicked superradiant dynamics with Wigner functions and its classical limit. The spin size is $j = 3$ and the dissipation rate is $\gamma_{sr} = 0.01$. Panels in the left column (a) correspond to the initial spin coherent state at $\theta = \phi = \pi/2$. The middle and right columns correspond to the state at time T_{opt} generated with periodic kicks (middle column (b), $k = 30$) and with kicks optimized with reinforcement learning (right column (c), the corresponding QFI is shown in panel (b) of figure 3). The top two rows show the Wigner functions of the density matrix, the bottom two rows show the classical phase space, populated by 10^6 points initially distributed according to the Husimi distribution of the initial spin coherent state and then propagated according to the classical equations of motion.

sphere. This is represented in figure 4 with either a sphere parameterized with x, y, and z, or in a plane (the phase space) spanned by the z-coordinate and the azimuthal angle $\phi \in (-\pi, \pi]$ such that $\phi = z = 0$ corresponds to the positive x-axis, $\phi = \pi/2, z = 0$ to the positive y-axis, and $z = \pm 1$ with arbitrary ϕ to the positive (negative) z-axis.

The quantum state can be represented in the phase space with the help of the Husimi or the Wigner distributions which are quasi probability distributions of the quantum state. The first two rows of panels in figure 4 depict the Wigner distribution of the initial quantum state (left column) and the quantum states of the

SR-KT (middle column, with kicking strength $k = 30$) and SR-GKT (right column) evolved for a time T_{opt} with damping rate $\gamma_{\mathrm{sr}} = 0.01$. The plotted cases for the SR-KT and SR-GKT correspond to the QFI given in panel (b) of figure 3, where one can also see the corresponding RL-optimized distribution of kicks.

Due to the small spin size of $j = 3$, we are deep in the quantum mechanical regime which manifests itself in an uncertainty of the initial spin coherent state that is relatively large compared to total size of the phase space. The distribution of the states evolved under dissipative dynamics exhibit remarkable differences for periodic and RL-optimized kicks.

In case of periodic kicks, we find that the initially localized distribution gets distributed over the phase space. It exhibits a maximum on the negative z-axis, see panels (b_1) and (b_2) in figure 4. This is reminiscent of the dissipative evolution in the absence of kicks, where the state is driven towards the ground state $|j, -j\rangle$ which is centered around $z = -1$. The ground state $|j, -j\rangle$ is an eigenstate of the precession and, thus, insensitive to changes in the frequency ω we want to estimate. Similarly, we interpret the part of the state distribution of the SR-KT that is centered around negative z-axis as insensitive. However, the distribution also exhibits non-vanishing parts distributed over the remainder of the phase space that can be understood as being sensitive to changes of ω and therefore explain the non-zero QFI of the SR-KT.

The state corresponding to RL-optimized kicks looks like a strongly squeezed state that almost encircles the whole sphere. Similar to spin squeezing, which is typically applied to the initial state as a part of the state preparation, we interpret the squeezed distribution as particularly sensitive with respect to the precession dynamics. This is due to the reduced uncertainty along the precession trajectories, i.e. with respect to the ϕ coordinate. We provide clips of the evolution over time of the state distributions that illustrate how the RL agent generates the squeezed state[4]. In particular, the squeezed state distribution can be seen as a feature the RL agent is aiming for with its policy. The distribution of RL-optimized kicks is shown in figure 3 (in appendix F, we provide a finer resolution of the distribution of kicks): it is roughly periodic with period corresponding to a precession angle of π. Also note that for the SR-GKT the Wigner distribution has negative contributions which is associated with non-classicality of the quantum state [55].

An advantage of the superradiant dynamics lies in its well-defined simple classical limit [45], see also appendix E. The lower two rows of panels in figure 4 depict the corresponding classical limit where the quantum state is represented by a cloud of phase space points (distributed according to the Husimi distribution of the initial spin coherent state) that are propagated according to the classical equations of motion. One of the reasons why the evolved classical distributions differ from the Wigner distributions is the absence of quantum uncertainty in the classical dynamics; in principle, over the course of the dynamics all classical phase space points can be concentrated to an arbitrarily small region of the phase space. In case of the SR-KT, the phase space points are distributed over the whole phase space, reminiscent of classical chaos. However, the distribution is not completely uniform but it exhibits a spiral density inhomogeneity. The plots as in figure 4 but for $j = 2$ are shown in the appendix F.

Figure 5 shows the gains of the RL-optimized SR-GKT over the SR-T. The gain is defined as the ratio of the RL-optimized QFI at time T_{opt} and the maximum QFI for the SR-T. A broad damping regime is found where gains can be achieved: in the regime of small decoherence rates γ_{sr}, the RL agent can fight decoherence in such a way that the QFI exhibits a continuous growth over the total time T_{opt} (see panels (a) and (b) in figure 3). In comparison with the SR-T, the RL agent benefits of stronger damping in this regime and, therefore, the gain increases with the dissipation rate γ_{sr}. For larger decoherence rates, the RL agent can no longer fight decoherence in the same manner (see panels (c) and (d) in figure 3), which manifests itself in a reduction of gains for large decoherence rates. In panel (b) of figure 5, we can see the (even larger) gain in QFI compared to the plateau value reached by the SR-KT.

The RL-optimized QFI is associated with a lower bound on the sensitivity (see equation (1)) for a given measurement time T_{opt}. If the measurement time can be chosen arbitrarily, sensitivity is associated with $\max_t I_\omega(t)/t$ [33]. This sensitivity represents the standard quantity reported for experimental parameter estimation because it takes time into account as a valuable resource; sensitivity is given in units of the parameter to be estimated per square root of Hertz. With RL we try to maximize $\max_t I_\omega(t)/t$ with respect to policies.

Figure 6 compares the SR-T with the SR-GKT where the latter was optimized with RL in order to maximize the rescaled QFI. Note, that the initial spin coherent state is centered around the positive y-axis, which means it is an eigenstate of the nonlinear kicks; kicks cannot induce spin squeezing at the very beginning of the dynamics. This changes when the spin precesses away from the y-axis. Therefore, it makes sense that the RL agent applies the strongest kick only after a precession by about $\pi/2$. The actions that the RL agent takes after the rescaled QFI reached its maximum are irrelevant and can be attributed to random noise generated by the RL algorithm.

As we have seen, the interplay of nonlinear kicks and superradiant damping is very special. However, also for other decoherence models the QFI can be increased significantly, for instance in case of a alkali-vapor

[4] The clips are available at https://doi.org/10.6084/m9.figshare.c.4640051.v3.

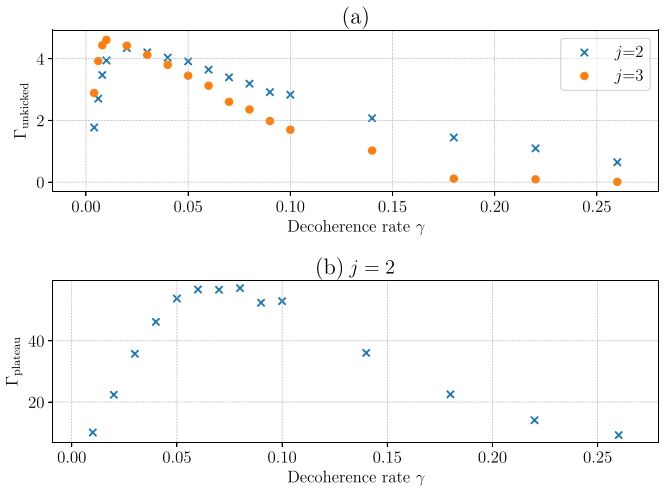

Figure 5. Improvement in the quantum Fisher information due to reinforcement learning for superradiant damping. The improvement in panel (a) is the ratio Γ_{unkicked} of quantum Fisher information at time T_{opt} (100 discretized time steps) optimized with reinforcement learning and the maximum QFI of the top (no kicks). In panel (b) we plot the ratio Γ_{plateau} of the QFI optimized with reinforcement learning and the plateau values achieved by periodic kicking for spin size $j = 2$ and kicking strength $k = 30$. In panel (b), the case of $j = 3$ is omitted due to the very small plateau values in that case. The discretization is coarser than in previous examples: $t_{\text{step}} = 1$ (i.e. a precession angle of $\pi/2$ per time step) and $k_{\text{step}} = 0.1$.

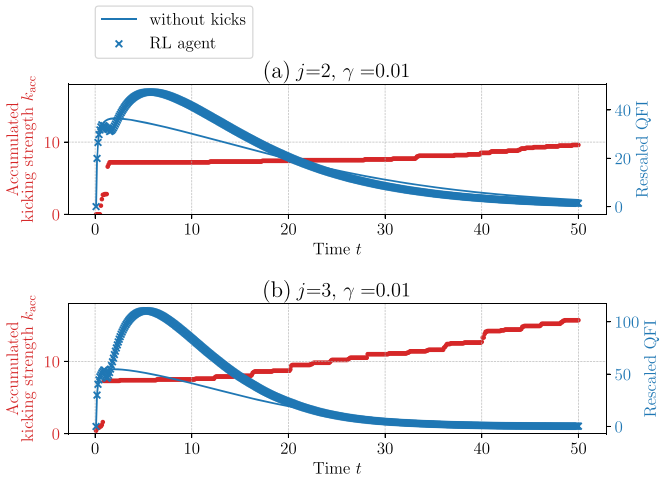

Figure 6. Examples for the policy adopted by the RL agent for maximizing the rescaled quantum Fisher information with superradiant damping. We plot the accumulated kicking strength $k_{\text{acc}}(t)$ (integrating over the kicks from time 0 to t) on the left axis as red dots and on the right axis the rescaled quantum Fisher information for the top (blue solid line) and for the generalized kicked top optimized with reinforcement learning (blue crosses). In case of $j = 2$ ($j = 3$) the strongest kick is applied after an initial rotation angle of $13\pi/20$ ($9\pi/20$).

magnetometer [33]. To demonstrate the performance of the RL agent in connection with another decoherence model, we take a look at phase damping, see figure 7. The behavior of the QFI of the PD-T is qualitatively similar to superradiant damping. The introduction of kicks, however, has a qualitatively different effect on the QFI. The RL agent can achieve improvements of the QFI for the PD-GKT at time T_{opt} (the highest time plotted in each panel of figure 7) compared with the QFI of the PD-T at the same time. Compared to the superradiant case, improvements are rather small. Notably, the policies applied by the RL agent are also different from superradiant damping; for instance, the RL agent avoids kicks for large parts of the dynamics.

6. Discussion

This work builds on recent results on quantum-chaotic sensors [33]. Our aim is to optimize the dynamical control that was used in [33] to render the sensor dynamics chaotic. Due to the high dimensionality of the problem we use techniques from RL. The control policies found with RL are tailored to boundary conditions such as the initial state, the targeted measurement time, and the decoherence model under consideration. At the example of

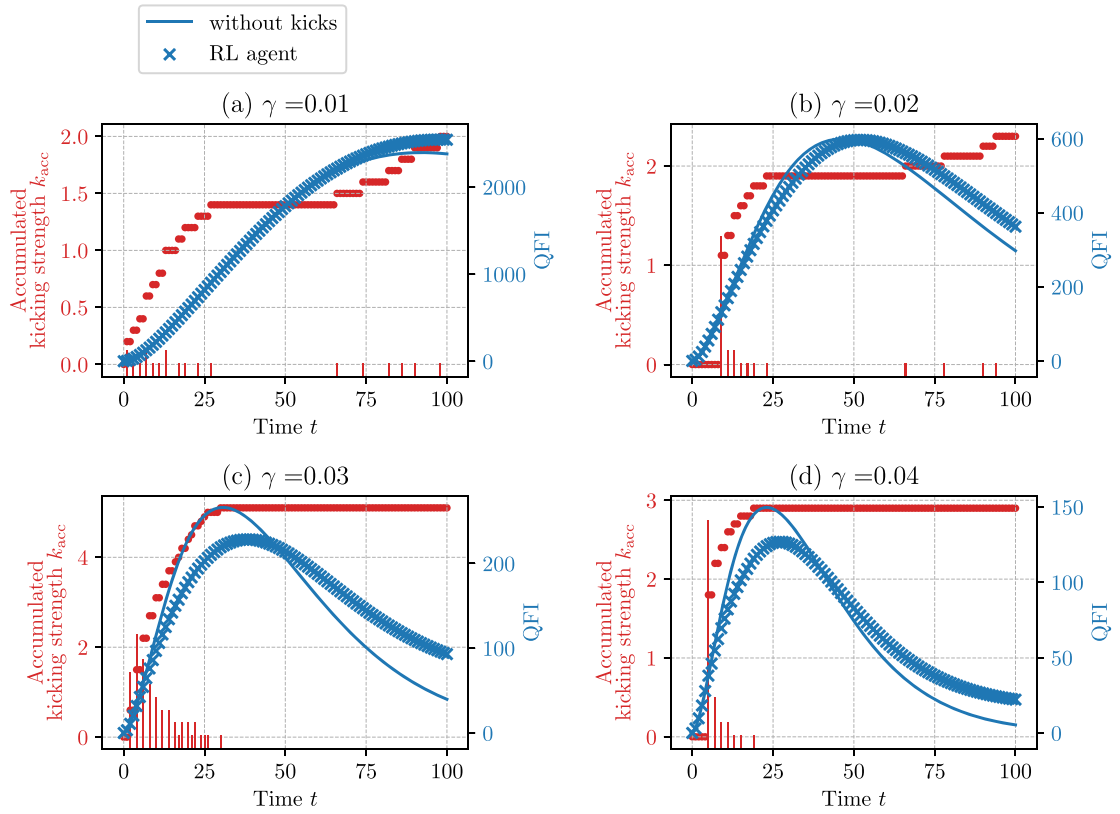

Figure 7. Examples for the strategy adopted by the RL agent for phase damping. All data is for spin $j = 2$ with increasing damping rates γ_{pd} from panel (a) to (d). We plot the accumulated kicking strength $k_{acc}(t)$ (integrating over the kicks from time 0 to t) on the left axis as red dots and on the right axis the quantum Fisher information for the top (blue solid line) and for the generalized kicked top optimized with reinforcement learning (blue crosses). We additionally plot red vertical lines at times when the RL agent sets a kick. The length of the lines corresponds to the kicking strength in arbitrary units (independent of the scale of the left axis). Note that the RL agent aims to maximize the QFI for $T_{opt} = 100$ and outperforms the top in all examples.

superradiant damping we demonstrate improvements in measurement precision and an improved robustness with respect to decoherence. A drawback of RL often lies in the expensive hyperparameter tuning of the algorithm. However, here we show that a basic RL algorithm (the cross-entropy method) can be used for several choices of boundary conditions with practically no hyperparameter tuning (there was no hyperparameter search necessary, solely parameters that directly influence the computation time were chosen conveniently).

In the example of superradiant damping, we unveil the approach taken by RL by visualizing the quantum dynamics with the help of the Wigner distribution of the quantum state. This reveals that RL favors a policy that is reminiscent of spin squeezing. However, instead of squeezing the state only at the beginning of the dynamics, the squeezing is refreshed and enhanced in roughly periodic cycles in order to fight against the superradiant damping.

In the spirit of [33], these findings emphasize the potential that lies in the optimization of the measurement dynamics. We are optimistic that RL can be used to tackle other problems in quantum metrological settings in order to achieve maximum measurement precision with limited quantum resources.

Acknowledgments

LJF and DB acknowledge support from the Deutsche Forschungsgemeinschaft (DFG), Grant No. BR 5221/1-1. We also acknowledge support from the Open Access Publishing Fund of the University of Tübingen.

Appendix A. Control problem and optimization parameters of the examples

Table A1 shows the parameters of the control problem and for the optimization used in each example. We train n_{agents} RL agents for $n_{iterations}$ iterations with $n_{episodes}$ episodes in each iteration. Each episode is simulated until a total time T_{opt} is reached. Then we produce $n_{samples}$ sample episodes of each trained RL agent and choose the best episode to plot the sample policies and gains.

Table A1. Hyperparameters used for the examples in the main text.

Figure	n_{agents}	$n_{\text{iterations}}$	n_{episodes}	n_{samples}	t_{step}	k_{step}	T_{opt}
Samples with superradiant damping (figure 3)	5	500	50	20	0.2	0.05	100
Gains of superradiant damping (figure 5)	20	300	40	20	1.0	0.10	100
Samples of rescaled QFI (figure 6)	2	500	50	20	0.1	0.10	50
Samples with phase damping (figure 7)	1	1000	100	1	1.0	0.10	100

Appendix B. Hyperparameters of RL

Here we give further information on the neural network and the hyperparameters of the algorithm.

The input layer of the neural network is defined by the observation. The output layer is determined by the number of actions (two) and we choose 300 neurons in the hidden layer. The layers are fully connected. The hidden layer has the rectified linear unit as its activation function and the output layer has the softmax function as its activation function [56]. As a cost function we choose the categorical cross entropy [56]. The share of best episodes σ_{share} is always 10%. The number of iterations and number of episodes vary for different settings, see table A1 for detailed information. For training we use the Adam optimizer [54] with learning rate 0.001.

Appendix C. Pseudocode for cross-entropy RL

This is the pseudocode for the cross-entropy method with discrete actions.

Algorithm: Cross entropy method

Inputs:
Number of iterations $n_{\text{iterations}}$
Number of episodes n_{episodes}
Share of best episodes σ_{share}

Other variables:
Total Reward R
Current Reward r
Observations o
Actions a
Training set S (consists of observations as inputs and actions as labels)

```
for 1 to n_iterations:
    for 1 to n_episodes:
        R, o, a ← Play Game
    end for
    sort episodes according to R
    S ← best σ_share episodes
    train neural network with S
end for

Function Play Game():
    while episode not finished do:
        put observation into neural network and receive probabilities of action as output
        choose action according to probability
        add action and observation to a, o
        tell the environment the action choice and receive a new observation o and reward r
        R ← R + r
    end while
    return R, o, a
```

The code implementation is based on an example by Jan Schaffranek[5].

Appendix D. Learning curve and stability of the algorithm

At the example of the superradiance decoherence model, we study the learning behavior of the cross-entropy RL algorithm for different training lengths (i.e. number of iterations) and different numbers of episodes per iteration. The results are summarized in figure D1. Spin size is $j = 2$ and dissipation rate is $\gamma_{\text{sr}} = 0.02$.

[5] https://udemy.com/artificial-intelligence-und-reinforcement-learning-in-python.

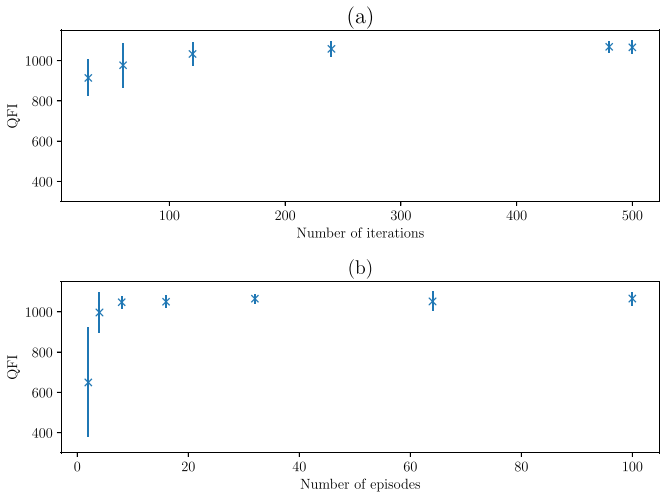

Figure D1. Learning behavior of the algorithm at the example of superradiant damping with $j = 2, \gamma_{sr} = 0.02, k_{step} = 0.1, t_{step} = 1$, and $T_{opt} = 100$. Panel (a) shows how the mean QFI and its standard deviation with respect to different runs of the algorithm behaves for various numbers of iterations and fixed number of episodes fixed to 100. In Panel (b) the number of episodes is varied and number of iterations fixed to 500.

In order to see the influence of the number of iterations, we set the number of episodes to 100 and let 20 different RL agents (with different random seeds) train for various numbers of iterations. The training of a single RL agent takes about one hour at most (for the higher number of iterations) on a desktop computer. We then use each RL agent to produce 20 episodes, giving us 400 episodes for each data point in figure D1. We used those episodes to calculate mean and standard deviation of the reward. The results are shown in the panel (a) of figure D1. In order to see the influence of the number of episode in each iteration, we fix the number of iterations to 500 and do the same procedure as before. The results are shown in panel (b) of figure D1.

We can see that the standard deviation over policies decreases with the number of iterations while the mean QFI increases. The same is true for the number of episodes (panel (b)), where for 32 episodes a stable plateau of the QFI is reached such that increasing the number of episodes does not achieve any further improvements. Overall, these results demonstrate the stability of the algorithm if the number of episodes and iterations is chosen sufficiently large.

Appendix E. Classical equations of motion

The KT with superradiant damping has a well defined classical limit. It is obtained from the quantum equations of motion by taking the limit $j \to \infty$ where $\hbar = 1$ and $\tau = 1$. The rescaled angular momentum operator $2\boldsymbol{J}/(2j+1) = 2(J_x, J_y, J_z)/(2j+1)$ then becomes the classical coordinate vector $\boldsymbol{r} = (x, y, z)$ and with $\lim_{j\to\infty}\left(\frac{2J}{2j+1}\right)^2 = 1$ the unit sphere becomes the classical phase space with azimuthal angle ϕ and z-coordinate as canonical variables. The equations of motions $\boldsymbol{r} \to \tilde{\boldsymbol{r}}$ are found to be [45]

$$\tilde{x} = x\cos(\alpha) - y\sin(\alpha), \tag{E1}$$

$$\tilde{y} = x\sin(\alpha) + y\cos(\alpha), \tag{E2}$$

$$\tilde{z} = z, \tag{E3}$$

for the precession about the z-axis by an angle α

$$\tilde{x} = z\sin(ky) + x\cos(ky), \tag{E4}$$

$$\tilde{y} = y, \tag{E5}$$

$$\tilde{z} = z\cos(ky) - x\sin(ky), \tag{E6}$$

for the kicks about the y-axis with kicking strength k, and, with azimuthal angle ϕ (see main text)

$$\tilde{\theta} = \arccos\left(\frac{1 - \left(\frac{1-z}{1+z}\exp(2\tau)\right)}{1 + \left(\frac{1-z}{1+z}\right)\exp(2\tau)}\right), \tag{E7}$$

$$\tilde{x} = \sin(\tilde{\theta})\cos(\phi), \tag{E8}$$

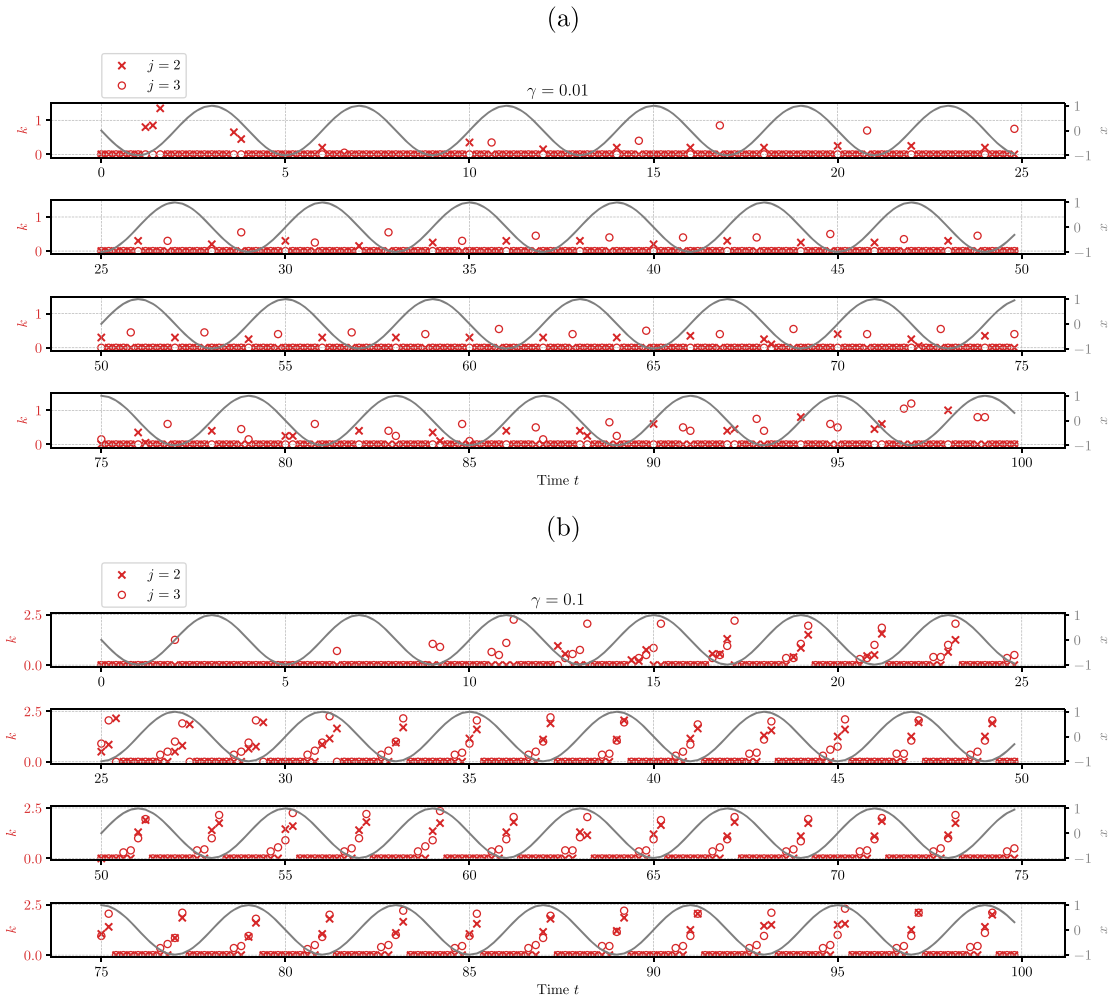

Figure F1. Kicks set by the RL agent for the SR-GKT. Panel (a) shows the case for $\gamma_{sr} = 0.01$ and panel (b) for $\gamma_{sr} = 0.1$. In red on the left axis are the kicking strengths k for $j = 2$ (crosses) and $j = 3$ (circles). To show the precession, we plot on the right axis in grey the x component of an unkicked spin coherent state without decoherence.

$$\tilde{y} = \sin(\tilde{\theta})\sin(\phi), \tag{E9}$$

$$\tilde{z} = \cos(\tilde{\theta}), \tag{E10}$$

for the superradiant damping, where

$$\tau = (2j + 1)\gamma_{sr} t, \tag{E11}$$

for a time t, spin size j, and superradiant decoherence rate γ_{sr}.

Appendix F. A closer look at the kicks set by the RL agent

Here we take a closer look at the kicks chosen by the RL agent in the examples with superradiant damping, considered in figure 3 in the main text.

In case of $\gamma_{sr} = 0.01$, for both, $j = 2$ and $j = 3$, we find relatively similar distribution of kicks, see panel (a) in figure F1. The most striking difference between the two policies for $j = 2$ and $j = 3$ are the comparatively strong kicks in the beginning of the sequence. By observing the time evolution of the Wigner function (see footnote 4), we find that these kicks basically rotate the state by an additional angle $\pi/2$ about the z-axis. This leads to a phase shift of $\pi/2$ between the two policies (see panels (d_3) and (d_4) of figure F2) compared to the initial state (see panels (a_3) and (a_4) of figure F2).

For $\gamma_{sr} = 0.1$ the policies are even more similar with several kicks increasing in strength with a period length of π, see panel (b) in figure F1.

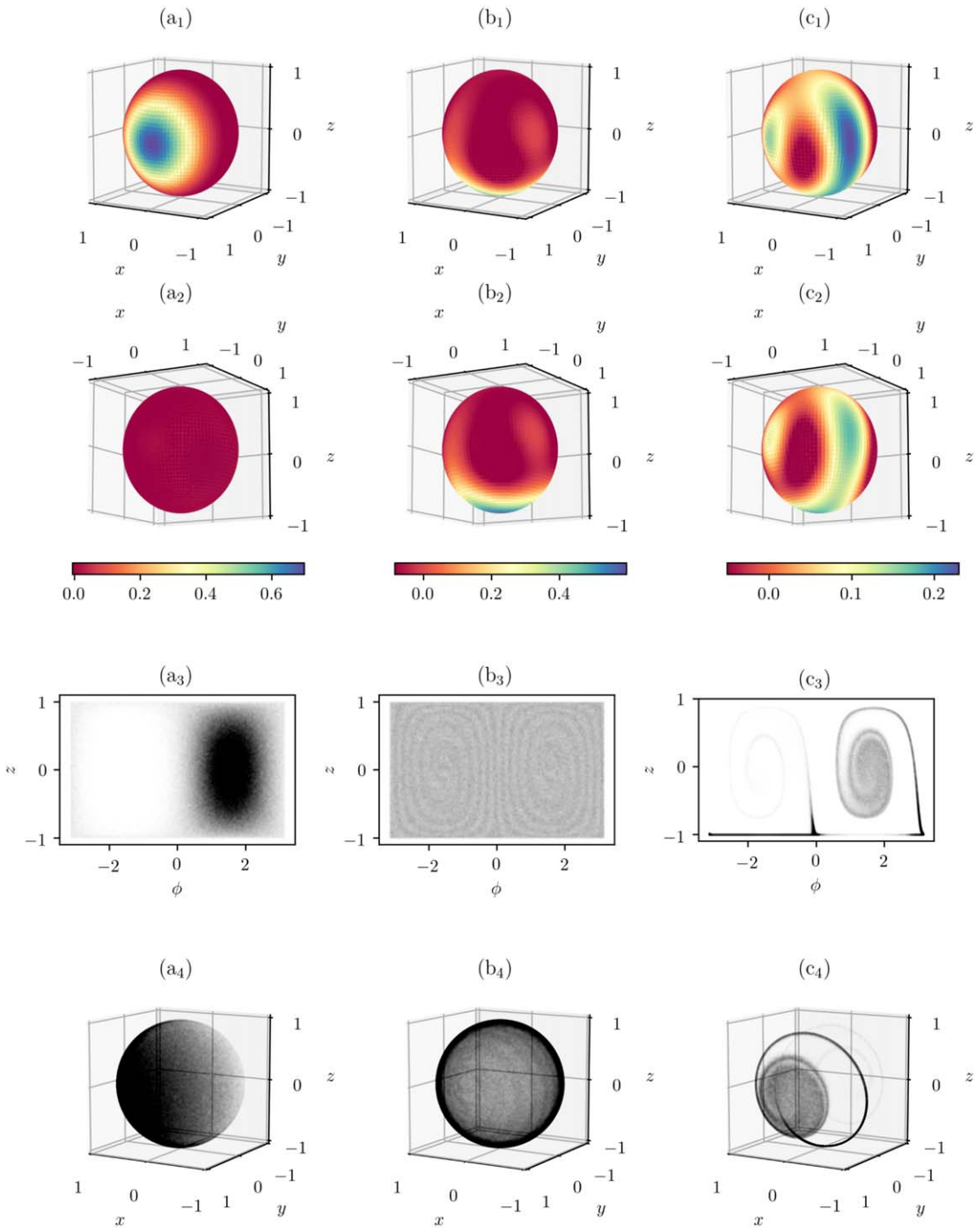

Figure F2. Shows the corresponding data as in figure 4 but for $j = 2$ instead of $j = 3$: illustration of kicked superradiant dynamics with Wigner functions and its classical limit. The dissipation rate is $\gamma_{sr} = 0.01$. Panels in the left column (a) correspond to the initial spin coherent state at $\theta = \phi = \pi/2$. The middle and right columns correspond to the state at time T_{opt} generated with periodic kicks (middle column (b), $k = 30$) and with kicks optimized with reinforcement learning (right column (c), the corresponding QFI is shown in panel (b) of figure 3). The top two rows show the Wigner functions of the density matrix, the bottom two rows show the classical phase space, populated by 10^6 points initially distributed according to the Husimi distribution of the initial spin coherent state and then propagated according to the classical equations of motion.

Figure F2 is analog to figure 4 in the main text but for $j = 2$ instead of $j = 3$. The only qualitative difference compared to the $j = 3$, is the periodically KT: the combination of periodic kicks with $k = 30$ and $j = 2$ seems to be a special configuration. The classical phase space is comparable with the $j = 3$ case, but there is much less structure in the Wigner function. Instead, the state concentrates on the south pole and exhibits a slightly squeezed shape (this is difficult to judge from figure F2 though). The rather high value of the QFI for $k = 30$ and $j = 2$, is best explained by this squeezing. When choosing other kicking strength, we observed a Wigner function similar to the case of $j = 3$.

ORCID iDs

Daniel Braun https://orcid.org/0000-0001-8598-2039

References

[1] Murphy K P 2012 *Machine Learning: A Probabilistic Perspective* (Cambridge, MA: MIT Press)
[2] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys.* **81** 074001
[3] Mehta P, Bukov M, Wang C-H, Day A G, Richardson C, Fisher C K and Schwab D J 2019 *Phys. Rep.* **810** 1–124
[4] Carrasquilla J and Melko R G 2017 *Nat. Phys.* **13** 431
[5] Broecker P, Assaad F F and Trebst S 2017 arXiv:1707.00663
[6] Van Nieuwenburg E P, Liu Y-H and Huber S D 2017 *Nat. Phys.* **13** 435
[7] Carleo G and Troyer M 2017 *Science* **355** 602
[8] Carleo G, Nomura Y and Imada M 2018 *Nat. Commun.* **9** 5322
[9] Gao X and Duan L-M 2017 *Nat. Commun.* **8** 662
[10] Leigh J R 2004 *Control Theory* (London: Institution of Electrical Engineers)
[11] Kaelbling L P, Littman M L and Moore A W 1996 *J. Artif. Intell. Res.* **4** 237
[12] Sutton R S and Barto A G 2018 *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press)
[13] Sutton R S, Barto A G and Williams R J 1992 *IEEE Control Syst. Mag.* **12** 19
[14] Chen C, Dong D, Li H-X, Chu J and Tarn T-J 2013 *IEEE Trans. Neural Netw. Learn. Syst.* **25** 920
[15] Palittapongarnpim P, Wittek P, Zahedinejad E, Vedaie S and Sanders B C 2017 *Neurocomputing* **268** 116
[16] Fösel T, Tighineanu P, Weiss T and Marquardt F 2018 *Phys. Rev.* X **8** 031084
[17] Bukov M, Day A G, Sels D, Weinberg P, Polkovnikov A and Mehta P 2018 *Phys. Rev.* X **8** 031086
[18] Albarrán-Arriagada F, Retamal J C, Solano E and Lamata L 2018 *Phys. Rev.* A **98** 042315
[19] Niu M Y, Boixo S, Smelyanskiy V N and Neven H 2019 Universal quantum control through deep reinforcement learning *npj Quantum Inf.* **5** 33
[20] Melnikov A A, Nautrup H P, Krenn M, Dunjko V, Tiersch M, Zeilinger A and Briegel H J 2018 *Proc. Natl Acad. Sci.* **115** 1221
[21] Sweke R, Kesselring M S, van Nieuwenburg E P and Eisert J 2018 arXiv:1810.07207
[22] Andreasson P, Johansson J, Liljestrand S and Granath M 2019 *Quantum* **3** 183
[23] Hentschel A and Sanders B C 2010 *2010 7th Int. Conf. on Information Technology: New Generations* (IEEE) pp 506–11
[24] Hentschel A and Sanders B C 2011 *Phys. Rev. Lett.* **107** 233601
[25] Lovett N B, Crosnier C, Perarnau-Llobet M and Sanders B C 2013 *Phys. Rev. Lett.* **110** 220501
[26] Sergeevich A and Bartlett S D 2012 *2012 IEEE Congress on Evolutionary Computation* (IEEE) pp 1–3
[27] Stenberg M P, Köhn O and Wilhelm F K 2016 *Phys. Rev.* A **93** 012122
[28] Palittapongarnpim P, Wittek P and Sanders B C 2016 *24th European Symp. on Artificial Neural Networks (Bruges, 27–29 April, 2016)* pp 327–32
[29] Lumino A, Polino E, Rab A S, Milani G, Spagnolo N, Wiebe N and Sciarrino F 2018 *Phys. Rev. Appl.* **10** 044033
[30] Liu J and Yuan H 2017 *Phys. Rev.* A **96** 012117
[31] Liu J and Yuan H 2017 *Phys. Rev.* A **96** 042114
[32] Xu H, Li J, Liu L, Wang Y, Yuan H and Wang X 2019 *npj Quantum Inf.* **5** 82
[33] Fiderer L J and Braun D 2018 *Nat. Commun.* **9** 1351
[34] Fiderer L J and Braun D 2019 *Optical, Opto-Atomic, and Entanglement-Enhanced Precision Metrology* vol 10934 (Bellingham, WA: SPIE) p 109342S
[35] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[36] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[37] Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
[38] Paris M G A 2009 *Int. J. Quantum Inf.* **7** 125
[39] Peres A 1993 *Quantum Theory: Concepts and Methods* vol 57 (Dordecht: Kluwer)
[40] Chaudhury S, Smith A, Anderson B, Ghose S and Jessen P S 2009 *Nature* **461** 768
[41] Krithika V, Anjusha V, Bhosale U T and Mahesh T 2019 *Phys. Rev.* E **99** 032219
[42] Dicke R H 1954 *Phys. Rev.* **93** 99
[43] Gross M, Fabre C, Pillet P and Haroche S 1976 *Phys. Rev. Lett.* **36** 1035
[44] Gross M and Haroche S 1982 *Phys. Rep.* **93** 301
[45] Braun D 2001 *Dissipative Quantum Chaos and Decoherence* (*Tracts in Modern Physics* vol 172) (Berlin: Springer)
[46] Kossakowski A 1972 *Rep. Math. Phys.* **3** 247
[47] Lindblad G 1976 *Math. Phys.* **48** 119
[48] Bonifacio R, Schwendimann P and Haake F 1971 *Phys. Rev.* A **4** 302
[49] Braun P A, Braun D and Haake F 1998 *Eur. Phys. J.* D **3** 1
[50] Braun P A, Braun D, Haake F and Weber J 1998 *Eur. Phys. J.* D **2** 165
[51] Giraud O, Braun P and Braun D 2008 *Phys. Rev.* A **78** 042112
[52] Giraud O, Braun P and Braun D 2010 *New J. Phys.* **12** 063005
[53] De Boer P-T, Kroese D P, Mannor S and Rubinstein R Y 2005 *Ann. Oper. Res.* **134** 19
[54] Kingma D P and Ba J 2014 arXiv:1412.6980
[55] Agarwal G S 2012 *Quantum Optics* (Cambridge: Cambridge University Press)
[56] Nielsen M A 2015 *Neural Networks and Deep Learning* vol 25 (San Francisco, CA: Determination Press)