Letter

Accelerated motional cooling with deep reinforcement learning

Bijita Sarma,* Sangkha Borah, A Kani, and Jason Twamley
Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan

(Received 28 January 2022; accepted 1 November 2022; published 29 November 2022)

Achieving fast cooling of motional modes is a prerequisite for leveraging such bosonic quanta for high-speed quantum information processing. In this Letter, we address the aspect of reducing the time limit for cooling, below that constrained by the conventional sideband cooling techniques, and propose a scheme to apply deep reinforcement learning (DRL) to achieve this. In particular, we have numerically demonstrated how the scheme can be used effectively to accelerate the dynamic motional cooling of a macroscopic magnonic sphere, and how it can be uniformly extended to more complex systems, for example, a tripartite opto-magno-mechanical system, to obtain cooling of the motional mode below the time bound of coherent cooling. While conventional sideband cooling methods do not work beyond the well-known rotating wave approximation (RWA) regimes, our proposed DRL scheme can be applied uniformly to regimes operating within and beyond the RWA, and thus, this offers a new and complete toolkit for rapid control and generation of macroscopic quantum states for application in quantum technologies.

DOI: 10.1103/PhysRevResearch.4.L042038

Introduction.— Fast cooling of bosonic mechanical modes of macroscopic systems is a primary objective of the ongoing efforts in quantum technology [1–5], and is a prerequisite for diverse prospective applications, such as the realization of macroscopic superposition states [6], gravitational tests of decoherence [7,8], ultraprecise measurements and sensing [9,10], and bosonic quantum computing [11]. Macroscopic yttrium-iron-garnet (YIG, $Y_3Fe_5O_{12}$) magnets have recently attracted strong interest towards such applications given the versatility of such systems in coupling to other modes falling in a wide frequency spectrum, e.g., with optical, microwave, and acoustic modes, as well as to superconducting qubits [12–17]. In addition, highly polished YIG spheres feature high magnonic Q factor and exhibit large frequency tunability properties due to the magnetic field dependence of the excited magnon modes. Considering, in particular, the cooling of motional modes of such objects, the usual method of sideband cooling based on weak magnomechanical interaction, operates on a time scale longer than the mechanical period of oscillation and depends on the relaxation dynamics of the subsystems. There have been a few proposals to improve the performance of optomechanical cooling of a mechanical resonator with squeezed light that works in the sideband-unresolved limit [18–20]. Cooling of mechanical modes in a timescale less than the mode frequency is highly advantageous for quantum computation and bosonic error correction [11,21]. By going over to the strong coupling regimes in magnon-phonon interactions, the speed of motional cooling can be highly enhanced giving rise to accelerated cooling. However, in such strong coupling regimes, the energy-nonconserving dynamics prevails because of the simultaneous presence of counter-rotating interactions, which makes it impossible to use sideband motional cooling techniques in this regime. In this work, we explore the usefulness of a machine learning based approach to address the aspect of reducing the time limit for cooling below that constrained by the conventional cooling techniques.

Recently, various machine learning (ML) approaches, aided with artificial neural networks as function approximators, have found widespread technological applications [22]. Among the various ML approaches, reinforcement learning (RL) [23] is considered to exhibit the closest resemblance to a humanlike learning approach, in that the RL-agent tries to gather experience on its own by interacting with its environment in a trial and error approach. In RL terminology, the *environment* describes the virtual/realworld surrounding the agent, with all the physics hardcoded into it along with a reward function based on which the agent can classify its good moves from the bad ones. RL, when operated in combination with artificial neural networks, is known as deep reinforcement learning (DRL). DRL has become crucial in many industrial and engineering applications, primarily after recent seminal works by Google DeepMind researchers [24,25]. Following these developments, there have been several fascinating applications of DRL in various fundamental domains of science, including some in quantum physics in areas of quantum error correction, [26–28] quantum control [29], multimode cooling [30], and state engineering [31–39].

In this Letter, we propose a DRL-based control scheme for accelerated motional cooling, that works for a generalized parameter setting of the coupling strength between the subsystems. In particular, we use the protocol to cool the acoustic

*bijita.sarma@oist.jp

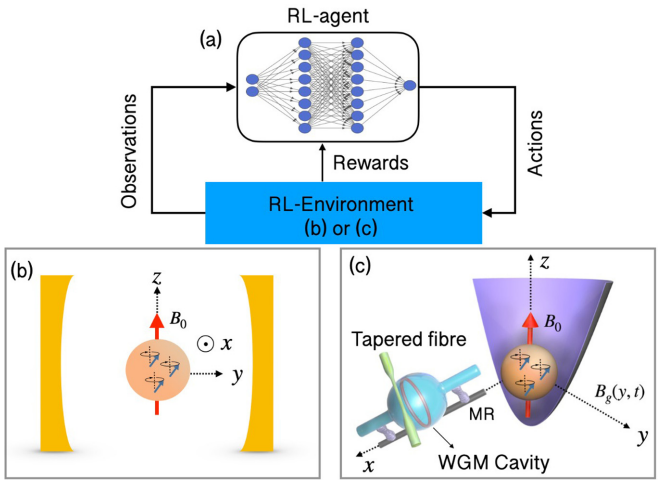

FIG. 1. (a) The sche matic workflow of the DRL protocol is shown, where the RLenvironment is either the bipartite magno-mechanical system (b), or the tripartite opto-magno-mechanical system (c). See text for further detail on DRL and the physical models.

phonon modes of a YIG sphere with a magno-mechanical interaction, and demonstrate numerically how it works efficiently in the strong coupling regime, where other methods such as sideband cooling fail. Also, we show how going over to the strong coupling regime is particularly advantageous, as it lowers the cooling time well below the phonon oscillation period and two orders of magnitude below the sideband cooling time limit. We demonstrate the usefulness and generalizability of our DRL cooling protocol by extending its application to a tripartite system of a trapped YIG magnet with its magnonic modes coupled to the center-of-mass (COM) mode in the trap and an optical cavity mode, and show that, despite the system being in the ultrastrong coupling regime, our DRL scheme can reveal nontrivial coupling modulations to cool the motional mode, which is usually not possible with coherent counter-intuitive protocols.

Bipartite magno-mechanical model.— We consider a setup as shown in Fig. 1(b), where a highly polished YIG sphere is placed in a microwave cavity. With a large external homogeneous magnetic field, B_0, applied in the $\hat{z}$ direction, the YIG sphere is magnetized to its saturation magnetization, $M_s = 5.87 \times 10^5\ \mathrm{Am}^{-1}$, which gives rise to collective spin wave magnon modes [16,40]. The frequency of the Kittel mode, which is the uniform magnon mode in the YIG sphere, is given by $\omega_m = \gamma B_0$, where $\gamma/2\pi = 28$ GHz/T is the gyromagnetic ratio. Due to the magnetostriction properties, YIG spheres also exhibit high- Q acoustic modes which are coupled to the magnon modes [12,15], and by driving the magnon modes with MW fields, the magno-mechanical coupling can be tuned and controlled [15,17]. In the limit of adiabatic elimination for a low-Q cavity, the bipartite magno-mechanical Hamiltonian is given by (see Supplemental Material [41] for detail)

$$\tilde{\mathcal{H}}/\hbar = \Delta_m m^\dagger m + \omega_b b^\dagger b + (\tilde{G} m + \tilde{G}^* m^\dagger)(b + b^\dagger), \quad (1)$$

where $m(m^\dagger)$ and $b(b^\dagger)$ are the magnonic and acoustic mode annihilation (creation) operators, ω_b is the resonance frequency of the acoustic mode, $\Delta_m = \omega_m - \omega_d$ (drive frequency ω_d) is the detuning, and $\tilde{G}$ is the magno-mechanical coupling strength. Note that, in such a bipartite system, while the beam-splitter interaction $\tilde{G} m b^\dagger + \tilde{G}^* m^\dagger b$, is valid for weak magno-mechanical coupling and favors mechanical cooling at the red sideband $\Delta_m = \omega_b$, the full coupling interaction accounting for the strong/ultrastrong coupling regimes is not favorable in the usual sideband cooling approach. Sideband cooling works through anti-Stokes scattering of the excitation from the thermally populated mode, b to the mode at zero entropy, m. However, such cooling needs constant driving as it is a steady-state process that takes a duration of the order of the relaxation dynamics of the subsystems (see Supplemental Material [41]). If one can access the strong coupling regime and manage to tame the counter-rotating interactions therein, there is a possibility of getting faster cooling than this limit. In the following, we design an algorithm based on DRL to model a dynamic variation of coupling, $\tilde{G}(t)$, to get faster cooling of the acoustic mode that operates within and beyond the weak coupling regime.

The schematic workflow of the DRL scheme applied to the physical model (RLenvironment) is shown in Fig. 1(a). The RLagent consists of a neural network model that is optimized for selected choices of actions that lead to desirable changes, and to net maximum rewards. The RLagent is modelled using the recently proposed Soft Actor-Critic (SAC) [42] algorithm that is based on the maximization of the entropy, $\mathcal{H}(\pi_\theta(\cdot|s_t))$ of the policy, π_θ as well as the long-term discounted cumulative return $r(s_t, a_t)$, i.e., $\max \mathbb{E}[\sum_t \gamma_t(r(s_t, a_t) + \alpha\mathcal{H}(\pi_\theta(\cdot|s_t)))]$, where θ denotes the optimizable weights of the neural networks, γ is the discount factor, and α is the regularization coefficient that determines the stochasticity of the policy. The policy $\pi_\theta(\cdot|s_t)$ sets the rules for the particular actions to be applied on the RLenvironment and essentially maps the observations (s_t) to particular actions (a_t) (see Supplemental Material [41] for further details).

The dynamics of the system is described by the quantum master equation (QME) for the density matrix ρ with the Hamiltonian $\tilde{\mathcal{H}}$ as

$$d\rho(t) = -\frac{i}{\hbar}[\tilde{\mathcal{H}}, \rho]dt + \sum_j [\kappa_j(\bar{n}_{c_j} + 1)\mathcal{L}[c_j]\rho + \kappa_j \bar{n}_{c_j}\mathcal{L}[c_j^\dagger]\rho]dt, \quad (2)$$

with dissipation and thermal fluctuations given by the Lindblad superoperators, $\mathcal{L}[c_j]\rho \equiv c_j \rho c_j^\dagger - \frac{1}{2}\{c_j^\dagger c_j, \rho\}$ where κ_j's are the damping rates of the modes, and the thermal occupation of each bosonic mode is given by $\bar{n}_{c_j} = [\exp(\hbar\omega_j/k_B T) - 1]^{-1}$, where T is the bath temperature and k_B is the Boltzmann constant. Solving the full QME to obtain the mean occupancy is a computationally intensive task. DRL typically requires several thousands of episodes of training, and solving the full QME within each episode is too resource sensitive for complex systems such as the ones we consider in this work. We employ an alternative approach to compute the mean occupancy in each mode using a set of linear differential equations for the second-order moments obtained from the QME, given by $\partial_t\langle\hat{o}_i\hat{o}_j\rangle = \mathrm{Tr}(\dot{\rho}\hat{o}_i\hat{o}_j) = \sum_{m,n}\zeta_{mn}\langle\hat{o}_m\hat{o}_n\rangle$, where $\hat{o}_i$, $\hat{o}_j$, $\hat{o}_m$, and $\hat{o}_n$ are one of the operators $(c_j{}^\dagger,\ c_j)$, and ζ_{mn} are the corresponding coefficients. Instantaneous solutions of

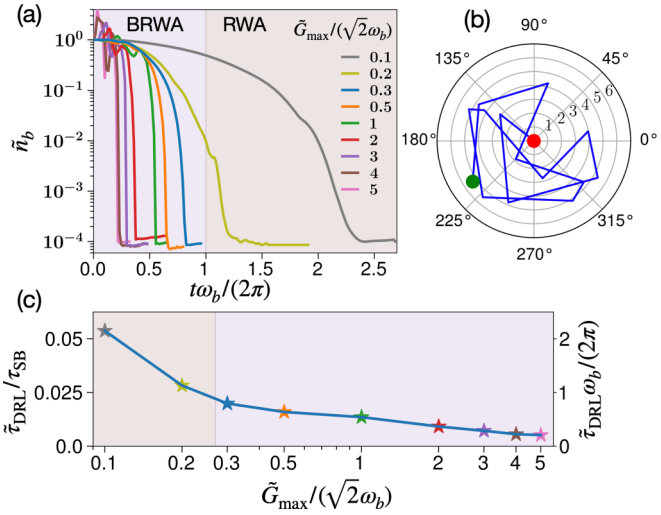

FIG. 2. (a) Accelerated DRL-based cooling of the phonon mode shown with the cooling quotient, $\tilde{n}_b$, of the bipartite system as shown in Fig. 1(b) as a function of $\tilde{G}_{\max}$, the maximum allowed control [with the values under RWA or beyond-RWA (BRWA)]. (b) The DRL-optimized complex pulse sequence is shown as a polar plot for $\tilde{G}_{\max}/(\sqrt{2}\omega_b) = 5$. The green dot shows the starting point while the red dot shows the final point. (c) The time required for cooling with the DRLscheme, $\tilde{\tau}_{\rm DRL}$ (right $y$axis), for $\tilde{n}_b \lesssim 10^{-4}$, is compared with the time limit for effective sideband cooling, $\tau_{\rm SB}$ (left $y$axis).

these equations, i.e., the second-order moments and the controls, $\tilde{G}(t)$, are used as the observations, s_t, for the RLagent in Fig. 1(a), and the reward function is chosen as, $r(t) = 1/\tilde{n}_b(t)$. Here $\tilde{n}_b(t) = \langle b^\dagger b\rangle(t)/n_b^T$ is the cooling quotient of the phonon number with respect to thermal occupancy, n_b^T at temperature T, where $\langle b^\dagger b\rangle$ represents the mean value of the phonon population. Further details of the DRL controller can be found in the Supplemental Material.

In Fig. 2(a), we show the cooling quotient $\tilde{n}_b$ for the DRL-optimized controls, as a function of $\tilde{G}_{\max}$ (the maximum of the coupling parameter). We consider $\Delta_m = \omega_b$, and damping rates as, $\kappa_b/\omega_b = 10^{-5}$ and $\kappa_m/\omega_b = 0.1$. It is found that as the coupling is increased towards the ultrastrong coupling regime ($\tilde{G} \gtrsim \omega_b$), the cooling becomes much faster. The DRL-optimized complex pulse sequence is shown as a polar plot for $\tilde{G}_{\max}/(\sqrt{2}\omega_b) = 5$ in Fig. 2(b). We denote the minimum time for cooling achieved by our method as $\tilde{\tau}_{\rm DRL}$. Figure 2(c) compares this time as a function of $\tilde{G}_{\max}$ with respect to the sideband cooling time limit, $\tau_{\rm SB}$, which represents the shortest cooling time limit possible with these methods, that work only when the rotating wave approximation (RWA) is applicable (see Supplemetal Material). With the DRL-based coupling modulations, we can achieve very low limits of cooling time compared to sideband cooling techniques, showing a lowering of approximately two orders of magnitude, which for the ultrastrong coupling regime, is further lowered.

Tripartite opto-magno-mechanical model.— Now we show that the proposed scheme can be extended effectively to more complex systems, for example, a higher-order tripartite opto-magno-mechanical system where we intend to cool the motional mode through nontrivial three-mode interactions. For this, we consider a system comprising a levitated YIG sphere in a harmonic trap [43–45], along with a driven WGM optical micr oresonator placed along the $\hat{x}$ direction with a magnetostrictive rod (MR) attached to it, as depicted in Fig. 1(c), which will be used as the RLenvironment in Fig. 1(a) in the DRL model. In a large external homogeneous magnetic field, B_0, applied in the $\hat{z}$ direction, the YIG sphere is magnetized to the saturation magnetization, M_s, and the homogeneously magnetized fundamental magnon mode (Kittel mode) produces a change in the axial length of the MR, which modulates the WGM optical mode frequency [2,46–48]. This gives rise to a coupling between the WGM optical mode (a) and the magnon mode (m) of the form $\Omega_S a^\dagger a(m + m^\dagger)$, where $\Omega_S = \Delta\omega$ is the optical frequency shift. The magnon mode can also be coupled to the COM motion of the YIG sphere by applying a spatially inhomogeneous external time-dependent magnetic field, $\mathbf{H}_g(y, t)$ [49,50], which satisfies the weak driving, $|\mathbf{H}_g(y, t)| \ll H_0$, and small-curl, $|\nabla \times \mathbf{H}_g(y, t)\|\mathbf{r}| \ll |\mathbf{H}_g(y, t)|$ conditions (see the Supplemental Material [41] for more information). Considering a time-varying gradient magnetic field of the form $\mathbf{H}_g(\mathbf{y}, t) = \frac{b_g(t)}{\mu_0} y\hat{y}$, ($b_g$ in units of [T/m]), the interaction Hamiltonian for the COM motion in the $\hat{y}$-directi on (frequency ω_b) and the magnon mode is given by $\mathcal{H}_{mb}(t) = \tilde{\Omega}_P(\hat{b} + \hat{b}^\dagger)(\hat{m} + \hat{m}^\dagger)$, with $\tilde{\Omega}_P = \frac{b_g}{4}\sqrt{\frac{|\gamma|M_S}{\rho\omega_b}}$, where $M_s = 5.87 \times 10^5 \mathrm{Am}^{-1}$ is the saturation magnetization, and $\rho = 5170$ kg/m^3 is the mass density of YIG. In the rotating frame of the cavity drive and the displacement picture of the average field in each mode [51], the complete Hamiltonian is described by

$$\tilde{\mathcal{H}}/\hbar = \Delta_a a^\dagger a + \omega_m m^\dagger m + \omega_b b^\dagger b + \tilde{\Omega}_S(a + a^\dagger)(m + m^\dagger) + \tilde{\Omega}_P(m + m^\dagger)(b + b^\dagger), \quad (3)$$

where Δ_a is the cavity detuning, and $\tilde{\Omega}_S$ is the driving-enhanced optomagnonic coupling rate. One can modulate $\tilde{\Omega}_S$ via the external drive, whereas the phonon-magnon coupling $\tilde{\Omega}_P$ can be modulated using the time-varying magnetic field gradient. In order to cool the COM motion, we intend to transfer the phonon population from the COM mode to the optical mode without populating the magnon mode. The damping rates of the cavity, magnon, and COM modes are given by, κ_i's, and the corresponding thermal populations at temperature T are n_i^T, with $i \in \{a, m, b\}$. Since the optical cavity mode oscillates at high frequency, its corresponding thermal bath, even at room temperature, yields zero thermal occupancy, however, the phonon and magnon baths are occupied.

Similar to the bipartite system discussed above, we next use the second-order moment equations to solve the dynamics of the system and the DRL scheme is used to optimize the controls $\tilde{\Omega}_{P/S}$ by maximizing the net reward signal per episode, $r(t) = 1/\tilde{n}_b - \lambda\,\langle m^\dagger m\rangle(t)$, where λ is a constant chosen such that the magnon mode does not get populated. Given the fact that the COM mode frequencies of the YIG are of the order of $\omega_b/2\pi \sim 10 - 100$'s of kHz, and the magnon frequency is $\omega_m/2\pi \sim 10$ GHz, the ideal choice of ω_m for $\omega_b/2\pi =$ 100 kHz is $\omega_m = 10^5\,\omega_b$. With such a high frequency difference with the intermediate magnon mode at $\omega_m = 10^5\,\omega_b$, this constitutes a largely detuned system ($\omega_m \gg \sqrt{\tilde{\Omega}_P^2 + \tilde{\Omega}_S^2}$). In such a system, while the magnon mode is usually decoupled,

the ideal time limit to obtain swap between mechanical quanta and optical mode with the ideal Raman pulses is given by $\tilde{\tau}_{\rm lim} = \pi\omega_m/(2\tilde{\Omega}_P\tilde{\Omega}_S)$ (see the Supplemental Material [41]). However, this is the limit for the situation as long as the RWA is valid. It is also noted that the effectiveness of this kind of cooling is highly reduced in the presence of damping, and going beyond the RWA to decrease the cooling time limit is not possible because of the counter-rotating dynamics. We apply the DRL strategy to work in a regime where $\tilde{\Omega}_i$'s are sufficiently high to access the cooling limit not obtainable by these conventional means, and also keep the counter-rotating dynamics in control. While the use of the method of coupled second order moments reduces the computation resources drastically, simulating the dynamics with the choice of the realistic parameter $\omega_m = 10^5\,\omega_b$ with high coupling strengths, for example $\tilde{\Omega}_S^{\rm max} = \tilde{\Omega}_P^{\rm max} = 100\,\omega_b$, turns out to be a computationally highly intensive problem, due to the very stiff solution of the set of differential equations. Hence, we adopt a two-step training procedure for the problem. In this protocol, we first use an *auxiliary* system with $\omega_m = 10^3\omega_b$ and $\tilde{\Omega}_S^{\rm max} = \tilde{\Omega}_P^{\rm max} = 10\,\omega_b$, for which the solution of the set of equations can be obtained without much computational effort. The trained auxiliary model is then used as a supervisor/teacher for the actual system, that we call *primary*, with $\omega_m = 10^5\omega_b$ and $\tilde{\Omega}_S^{\rm max} = \tilde{\Omega}_P^{\rm max} = 100\,\omega_b$. Training the primary system for a few hundred episodes with periodic evaluation of the RLagent yields the best trained model. In the literature of RL, such a scheme is known as *imitation learning*, and is a feature of generalizability of RL-trained models.

In Fig. 3(a), the cooling quotients of the photon, magnon, and phonon modes, $\tilde{n}_i = n_i/n_T$, are shown, where n_i's are the mean occupancies and n_T is the phonon thermal bath population. The cooling time limit for this system is given by $\tilde{\tau}_{\rm lim}\omega_b/(2\pi) \sim 5$ (see the Supplemental Material [41]). The plot shows the population dynamics in the three modes for the DRL-based coupling parameters with a maximum allowed value, $\tilde{\Omega}_i^{\rm max}/\omega_b = 100$. One can see that while the magnon mode is kept at a constant population, there is a scattering between the mechanical and optical mode that gives rise to five orders of cooling in the mechanical mode. This draws an analogy to the stimulated Raman adiabatic passage (STIRAP) techniques for three-mode systems. However, it is well-known that such adiabatic techniques require longer time and only works ideally as long as the counter-rotating terms are not present (see the Supplemental Material [41]). On the contrary, the coupling parameters found by our DRL protocol are nontrivial, which are shown in Fig. 3(b), and the overall time required for cooling is reduced below the adiabatic limit even for high values of coupling. In the bottom panel of Fig. 3(a) we show the cooling time required by our DRL method for even larger values, and higher coupling parameters yield even lower time limits.

Finally, we provide a critical understanding of the implications of the current work. The control of open quantum systems with large Hilbert space dimensionality faces several limitations, specifically, due to the huge numerical resources required to simulate and store such complex systems, e.g., the systems with infinite dimensionality with energy-nonconserving dynamics as considered in this work. Numerical differentiation has always been the bottleneck in

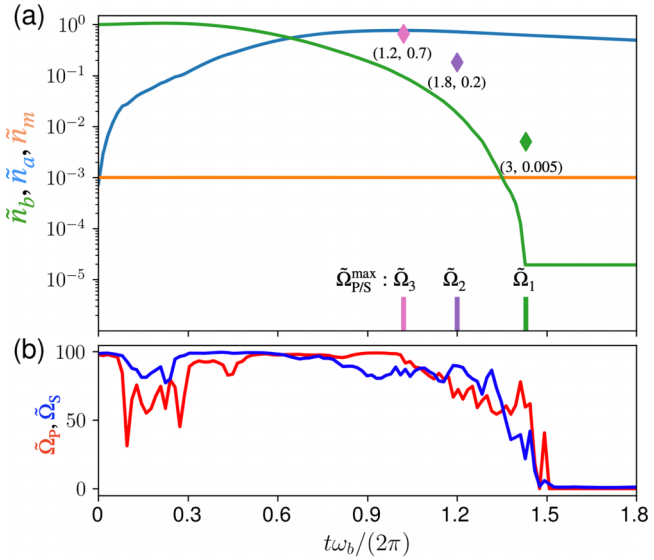

FIG. 3. (a) The DRL-scheme based mean occupancies of the phonon, magnon, and photon modes are shown in terms of the cooling quotient, $\tilde{n}_i$ (in green, orange, and blue lines, respectively) for $\tilde{\Omega}_{\{\rm S,P\}}/\omega_b = 100$. The vertical lines show the time limits for the DRL-based cooling with $\tilde{\Omega}_i^{\rm max}/\omega_b = \{100, 120, 150\}$, respectively (in green, purple, and pink). While the DRL-derived scheme gives a cooling quotient of $\tilde{n}_b < 10^{-4}$, the corresponding Raman pulses with the same amount of maximum allowed coupling strength gives lower values, and the cooling time limits for these are far higher than those obtained with the DRLscheme. The markers show the cooling results obtained for the conventional counter-intuitive pulses, with the corresponding cooling quotients and time limits shown with each point in the order $(\tilde{\tau}_{\rm lim}, \tilde{n}_b)$. (b) The corresponding DRL-optimized coupling parameters are shown for the case of maximum allowed value of $\tilde{\Omega}_{\{\rm S,P\}}/\omega_b = 100$.

such contexts. In the protocol we have outlined, such complexities and constraints can be eliminated since one no longer needs to deal with density matrices of order $\mathcal{N}^n$, where $\mathcal{N}$ is the dimension of the Hilbert space and n is the number of subsystems in the system. Instead, solving a set of $n(n+1)$ ordinary differential equations (ODE) is sufficient, which is an immense reduction in computational time and resources. A particular advantage of using RL-optimized controls in such context is that the trained model can be generalized to many control parameters depending on the problem of interest.

Conclusion.— In this Letter, we address the aspect of reducing the time required for cooling bosonic motional modes below the time limit accessible by well-known cooling methods with scattering techniques. While such conventional methods of cooling do not work in strong and ultrastrong coupling regimes where the RWA is not valid, we design a DRL-based algorithm that works within and beyond such regimes, and show that by accessing the ultrastrong coupling limit with DRL-designed pulses the cooling limit can be broken resulting in accelerated cooling. We further show how the protocol can be adapted for cooling in a tripartite system with an opto-magno-mechanical interaction, that represents more complexity for the DRLcontrol owing to the huge dimensionality in the Hilbert space, and find nontrivial

three-mode interactions leading to accelerated cooling breaking the coherent cooling time limits. Thus, this study outlines a comprehensive toolbox for application of DRL for fast and efficient quantum control in a magnonic system within and beyond RWA restrictions, which can be adapted to other quantum systems of interest. For future and ongoing efforts in quantum technology, this is expected to play a pivotal role, especially in conjunction with various laboratory experiments.

Acknowledgments. The authors thank Okinawa Institute of Science and Technology (OIST) Graduate University for the super-computing facilities provided by the Scientific Computing and Data Analysis section of the Research Support Division, and for financial support. The authors are grateful for the help and support provided by Jeffery Prine from the Digital Content, Brand, and Design Section of the Communication and Public Relations Division at OIST.

[1] C. Schäfermeier, H. Kerdoncuff, U. B. Hoff, H. Fu, A. Huck, J. Bilek, G. I. Harris, W. P. Bowen, T. Gehring, and U. L. Andersen, Quantum enhanced feedback cooling of a mechanical oscillator using nonclassical light, Nat. Commun. **7**, 13628 (2016).

[2] J. Guo, R. Norte, and S. Gröblacher, Feedback Cooling of a Room Temperature Mechanical Oscillator close to its Motional Ground State, Phys. Rev. Lett. **123**, 223602 (2019).

[3] Y.-S. Park and H. Wang, Resolved-sideband and cryogenic cooling of an optomechanical resonator - Nature Physics, Nat. Phys. **5**, 489 (2009).

[4] J. B. Clark, F. Lecocq, R. W. Simmonds, J. Aumentado, and J. D. Teufel, Sideband cooling beyond the quantum backaction limit with squeezed light - Nature, Nature (London) **541**, 191 (2017).

[5] M. Frimmer, J. Gieseler, and L. Novotny, Cooling Mechanical Oscillators by Coherent Control, Phys. Rev. Lett. **117**, 163601 (2016).

[6] M. Abdi, P. Degenfeld-Schonburg, M. Sameti, C. Navarrete-Benlloch, and M. J. Hartmann, Dissipative Optomechanical Preparation of Macroscopic Quantum Superposition States, Phys. Rev. Lett. **116**, 233604 (2016).

[7] S. Bose, K. Jacobs, and P. L. Knight, Scheme to probe the decoherence of a macroscopic object, Phys. Rev. A **59**, 3204 (1999).

[8] W. Marshall, C. Simon, R. Penrose, and D. Bouwmeester, Towards Quantum Superpositions of a Mirror, Phys. Rev. Lett. **91**, 130401 (2003).

[9] A. Schliesser, O. Arcizet, R. Rivière, G. Anetsberger, and T. J. Kippenberg, Resolved-sideband cooling and position measurement of a micromechanical oscillator close to the Heisenberg uncertainty limit - Nature Physics, Nat. Phys. **5**, 509 (2009).

[10] J. Manley, M. D. Chowdhury, D. Grin, S. Singh, and D. J. Wilson, Searching for Vector Dark Matter with an Optomechanical Accelerometer, Phys. Rev. Lett. **126**, 061301 (2021).

[11] J. E. Bourassa, N. Quesada, I. Tzitrin, A. Száva, T. Isacsson, J. Izaac, K. K. Sabapathy, G. Dauphinais, and I. Dhand, Fast simulation of bosonic qubits via gaussian functions in phase space, PRX Quantum **2**, 040315 (2021).

[12] X. Zhang, C.-L. Zou, L. Jiang, and H. X. Tang, Cavity magnomechanics, Sci. Adv. **2**, e1501286 (2016).

[13] D. Lachance-Quirion, Y. Tabuchi, A. Gloppe, K. Usami, and Y. Nakamura, Hybrid quantum systems based on magnonics, Appl. Phys. Express **12**, 070101 (2019).

[14] Y.-P. Wang and C.-M. Hu, Dissipative couplings in cavity magnonics, J. Appl. Phys. **127**, 130901 (2020).

[15] J. Li, S.-Y. Zhu, and G. S. Agarwal, Magnon-Photon-Phonon Entanglement in Cavity Magnomechanics, Phys. Rev. Lett. **121**, 203601 (2018).

[16] X. Zhang, C.-L. Zou, L. Jiang, and H. X. Tang, Strongly Coupled Magnons and Cavity Microwave Photons, Phys. Rev. Lett. **113**, 156401 (2014).

[17] Y.-P. Wang, G.-Q. Zhang, D. Zhang, T.-F. Li, C.-M. Hu, and J. Q. You, Bistability of Cavity Magnon Polaritons, Phys. Rev. Lett. **120**, 057202 (2018).

[18] M. Asjad, N. E. Abari, S. Zippilli, and D. Vitali, Optomechanical cooling with intracavity squeezed light, Opt. Express **27**, 32427 (2019).

[19] J.-H. Gan, Y.-C. Liu, C. Lu, X. Wang, M. K. Tey, and L. You, Intracavity-squeezed optomechanical cooling, Laser Photonics Rev. **13**, 1900120 (2019).

[20] H.-K. Lau and A. A. Clerk, Ground-State Cooling and High-Fidelity Quantum Transduction via Parametrically Driven Bad-Cavity Optomechanics, Phys. Rev. Lett. **124**, 103602 (2020).

[21] A. Joshi, K. Noh, and Y. Y. Gao, Quantum information processing with bosonic qubits in circuit qed, Quantum Sci. Technol. **6**, 033001 (2021).

[22] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, (MIT Press, Cambridge, 2016), Vol 1.

[23] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 2 edition, 2018.

[24] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, Mastering the game of Go with deep neural networks and tree search, Nature (London) **529**, 484 (2016).

[25] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, Mastering the game of Go without human knowledge, Nature (London) **550**, 354 (2017).

[26] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[27] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Reinforcement Learning in Different Phases of Quantum Control, Phys. Rev. X **8**, 031086 (2018).

[28] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement Learning with Neural Networks for Quantum Feedback, Phys. Rev. X **8**, 031084 (2018).

[29] S. Borah, B. Sarma, M. Kewming, G. J. Milburn, and J. Twamley, Measurement-Based Feedback Quantum Control with Deep Reinforcement Learning for a Double-Well Nonlinear Potential, Phys. Rev. Lett. **127**, 190403 (2021).

[30] C. Sommer, M. Asjad, and C. Genes, Prospects of reinforcement learning for the simultaneous damping of many mechanical modes, Sci. Rep. **10**, 1 (2020).

[31] Z. T. Wang, Y. Ashida, and M. Ueda, Deep Reinforcement Learning Control of Quantum Cartpoles, Phys. Rev. Lett. **125**, 100401 (2020).

[32] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf. **5**, 33 (2019).

[33] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, When does reinforcement learning stand out in quantum control? A comparative study on state preparation, npj Quantum Inf. **5**, 85 (2019).

[34] H. Xu, L. Wang, H. Yuan, and X. Wang, Generalizable control for multiparameter quantum metrology, Phys. Rev. A **103**, 042615 (2021).

[35] J. Mackeprang, D. B. R. Dasari, and J. Wrachtrup, A reinforcement learning approach for quantum state engineering, Quantum Mach. Intell. **2**, 5 (2020).

[36] T. Haug, W.-K. Mok, J.-B. You, W. Zhang, C. E. Png, and L.-C. Kwek, Classifying global state preparation via deep reinforcement learning, Mach. Learn.: Sci. Technol. **2**, 01LT02 (2021).

[37] S.-F. Guo, F. Chen, Q. Liu, M. Xue, J.-J. Chen, J.-H. Cao, T.-W. Mao, M. K. Tey, and L. You, Faster State Preparation across Quantum Phase Transition Assisted by Reinforcement Learning, Phys. Rev. Lett. **126**, 060401 (2021).

[38] M. Bilkis, M. Rosati, R. M. Yepes, and J. Calsamiglia, Real-time calibration of coherent-state receivers: Learning by trial and error, Phys. Rev. Res. **2**, 033295 (2020).

[39] R. Porotti, D. Tamascelli, M. Restelli, and E. Prati, Coherent transport of quantum states by deep reinforcement learning, Commun. Phys. **2**, 61 (2019).

[40] Y. Tabuchi, S. Ishino, T. Ishikawa, R. Yamazaki, K. Usami, and Y. Nakamura, Hybridizing Ferromagnetic Magnons and Microwave Photons in the Quantum Limit, Phys. Rev. Lett. **113**, 083603 (2014).

[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L042038 for a brief theory of reinforcement learning, technical details of the implementation and additional results, which includes Refs. [12,15–17,23–25,40,42,49,50,52–58].

[42] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, arXiv:1801.01290 (2018).

[43] T. Seberson, T. Seberson, P. Ju, J. Ahn, J. Bang, T. Li, T. Li, T. Li, T. Li, F. Robicheaux, and F. Robicheaux, Simulation of sympathetic cooling an optically levitated magnetic nanoparticle via coupling to a cold atomic gas, J. Opt. Soc. Am. B **37**, 3714 (2020).

[44] C. C. Rusconi, V. Pöchhacker, K. Kustura, J. I. Cirac, and O. Romero-Isart, Quantum Spin Stabilized Magnetic Levitation, Phys. Rev. Lett. **119**, 167202 (2017).

[45] P. Huillery, T. Delord, L. Nicolas, M. Van Den Bossche, M. Perdriat, and G. Hétet, Spin mechanics with levitating ferromagnetic particles, Phys. Rev. B **101**, 134415 (2020).

[46] S. Forstner, E. Sheridan, J. Knittel, C. L. Humphreys, G. A. Brawley, H. Rubinsztein-Dunlop, and W. P. Bowen, Ultrasensitive optomechanical magnetometry, Adv. Mater. **26**, 6348 (2014).

[47] K. Xia, M. R. Vanner, and J. Twamley, An opto-magneto-mechanical quantum interface between distant superconducting qubits, Sci. Rep. **4**, 5571 (2015).

[48] C. Yu, J. Janousek, E. Sheridan, D. L. McAuslan, H. Rubinsztein-Dunlop, P. K. Lam, Y. Zhang, and W. P. Bowen, Optomechanical Magnetometry with a Macroscopic Resonator, Phys. Rev. Appl. **5**, 044007 (2016).

[49] T. M. Hoang, J. Ahn, J. Bang, and T. Li, Electron spin control of optically levitated nanodiamonds in vacuum, Nat. Commun. **7**, 12250 (2016).

[50] T. Delord, P. Huillery, L. Schwab, L. Nicolas, L. Lecordier, and G. Hétet, Ramsey Interferences and Spin Echoes from Electron Spins Inside a Levitating Macroscopic Particle, Phys. Rev. Lett. **121**, 053602 (2018).

[51] K. E. Khosla, G. A. Brawley, M. R. Vanner, and W. P. Bowen, Quantum optomechanics beyond the quantum coherent oscillation regime, Optica **4**, 1382 (2017).

[52] J. Achiam, Spinning Up in Deep Reinforcement Learning, 2021, [Online; accessed 15. Jul. 2021].

[53] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, OpenAI Gym, arXiv:1606.01540 (2016).

[54] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, and N. Dormann, Stable baselines3, https://github.com/DLR-RM/stable-baselines3, 2019.

[55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, F. d' Alché-Buc, and editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[56] P. C. Fletcher and R. O. Bell, Ferrimagnetic resonance modes in spheres, J. Appl. Phys. **30**, 687 (1959).

[57] Y.-D. Wang and A. A. Clerk, Using Interference for High Fidelity Quantum State Transfer in Optomechanics, Phys. Rev. Lett. **108**, 153603 (2012).

[58] K. Bergmann, H. Theuer, and B. W. Shore, Coherent population transfer among quantum states of atoms and molecules, Rev. Mod. Phys. **70**, 1003 (1998).