Smart Magnetic Microrobots Learn to Roll with Reinforcement Learning

D. González-Calatayud[1, *] and J. L. Aragonés[1]

[1] *Department of Theoretical Condensed Matter Physics, Condensed Matter Physics Center (IFIMAC) and Instituto Nicolás Cabrera, Universidad Autónoma de Madrid, 28049, Madrid, Spain*

Accurate control of micron-scale robots, or microrobots, requires addressing unique challenges, including the difficulties associated with achieving motion in a non-inertial fluid dynamics regime, the development of real-time control strategies, and the presence of complex and dynamic environments. Unlike their macroscopic counterparts, microrobots are subject to Brownian motion, which randomizes their position and orientation. Reinforcement learning is a promising tool for autonomously developing robust controllers to create smart microrobots, providing them with the capability to adapt their behavior to operate in uncharacterized environments without the need to model system dynamics. In this work, we report the development of smart magnetic-driven spherical particles that harness lubrication forces to roll on a substrate. We use real-time control of these rollers to train a deep reinforcement learning model that generates an actuation policy to direct their motion in a simple standard navigation problem whose optimal solution can be derived from the underlying physics. The model's performance is analyzed under different sets of physical-informed system states. As a hallmark of the microscopic world, we characterize the influence of Brownian motion on the learning process by comparing learning rate at different Peclet numbers. Although the field of microrobotics has not yet reached the evolutionarily honed refinement of microscopic living organisms, deep reinforcement learning is a promising approach that is likely to enhance the capabilities of the next generation of microrobots.

I. INTRODUCTION

The development of microscale robots for medical applications presents unique control challenges that conventional engineering approaches struggle to address. At macroscopic scale, we have made significant advances in achieving a high degree of control over transport, employing diverse sensor data and complex decision-making algorithms to navigate through ever-changing conditions. However, controlled transport at microscopic scale, such as biological cells, nanoparticles, or microrobots, in complex environments represents a new and challenging frontier. Microrobots operating remotely inside the human body have potential to enable minimally invasive medical procedures including targeted drug, cell, or other cargo delivery, and blood clot removal [1–3]. Controlling these miniature devices within complex and dynamic environments such as the human body can present a significant engineering challenge. Additionally, the environmental dynamics encountered by a biomedical microrobot inside the human body may be variable, complex, and poorly characterized.

As one potential approach to overcoming these obstacles, we can examine and adopt the strategies of natural biological agents which have evolved to function in complex, unpredictable settings. Many biological systems have the capability to adapt by learning new behaviors from experiences, allowing them to succeed in diverse and changing environmental conditions by adjusting their behavior to fit the situation. Systems that learn adaptive behavioral patterns from historical events are widespread in nature and exist at all levels of biological organization. Reinforcement learning (RL) is a biomimetic machine learning optimization method inspired by the adaptive behaviors of real-world organisms that enables artificial systems to learn behaviors [4]. In RL, an agent observes the state of its environment and selects actions to achieve a goal defined by a reward signal. This reward signal teaches the agent to perform actions that maximize future rewards, thereby enhancing the agent's ability to perform the task based on previous experiences.

Controlled transport at the microscale is challenging not only because viscous forces dominate over inertial forces, and thus directed motion is difficult to achieve, but also because the environment is very complex and dynamic. Without inertial forces, the propulsion mechanism relies on breaking the flow field symmetry around the moving object, which results in a net translational motion. In nature, all motile microorganisms have developed the capability of motion and navigation in these complex environments [5]. These microorganisms break the flow field symmetry around their bodies by rotating or bending appendages such as cilia or flagella in a non-reciprocal manner. Inspired by nature, synthetic microrobots capable of propelling themselves by disrupting the symmetry of the flow fields have been developed [6]. The symmetry is either disrupted by the robot's shape or by utilizing nearby boundaries.

In this work, we focus on magnetic microrollers, spherical magnetic particles that are actuated by homogeneous rotating magnetic fields. In the absence of solid boundaries, the particle rotates in place, dissipating the input energy into the surrounding fluid, generating a rotational velocity field. When the particle settles near a surface, this symmetry breaks, and a lubrication force emerges

* david.gonzalez-calatayud@estudiante.uam.es

between the particle and the wall, coupling rotational and translational motion to enable directed rolling. However, controlling these surface microrollers presents multiple challenges: thermal fluctuations significantly affect their dynamics at this scale, and their motion can be dramatically altered by environmental obstacles, potentially leading to reverse motion or trapping.

Recent work has shown RL's potential for microrobot control in various contexts [7–9]. Here, we implement a deep learning approach to RL (deep RL) that can effectively control magnetic microrollers without explicit physical modeling, while naturally handling environmental uncertainties and obstacles. By choosing a simple directional control task with a known physics-based solution, we can rigorously evaluate the effectiveness of the RL approach. Our results show that the learned control policies independently converge to physics-informed optimal solutions, suggesting deep RL's broader applicability to more complex microrobotic systems where analytical solutions may be intractable.

II. RESULTS

A. Magnetic microrobots

As the actuator for our microrobot system, we developed a three-axis magnetic coil setup (Magneto). This setup uses six Helmholtz coils, designed and arranged in parallel pairs in Helmholtz configuration, to generate the rotating magnetic fields necessary for actuating our rolling microrobots, as shown in Fig. 1. The applied magnetic field, $\mathbf{B}$, imparting forces and torques on the robot. For a microrobot with a magnetic moment $\mathbf{m}$, the robot experiences a force according to

$$\mathbf{F} = \boldsymbol{\nabla}(\mathbf{m} \cdot \mathbf{B}). \quad (1)$$

In a nonuniform magnetic field, a ferromagnetic microrobot feels force in the direction of increasing magnetic gradient. However, due to the design of our system, the applied magnetic field is spatially constant (i.e., homogeneous) across the whole observation area, and varies only with time. Additionally, the magnetic microrobot also experiences a torque given by

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B}, \quad (2)$$

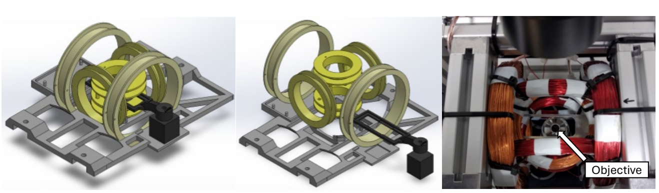

FIG. 1. Design schematics and image of the 3D Helmholtz-based System.

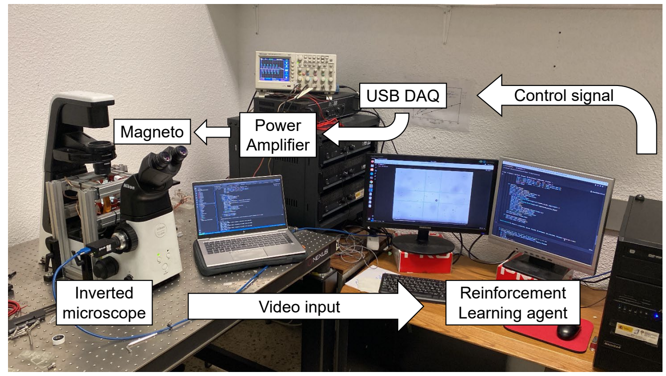

FIG. 2. Hardware for real-world control of magnetic microrobots using RL.

which acts to align the magnetic moment of the microrobot with the direction of the magnetic field. When the magnetic field rotates so that the direction of $\mathbf{B}$ is constantly changing, it is possible to use this torque to impart spin to the microrobot at the frequency of the rotating magnetic field. The whole experimental setup is depicted in Fig. 2.

B. RL implementation

The control problem for our microrobotic system was formulated as an episodic, discrete-time problem with a continuous action space and a continuous state space. The state space contains the two-dimensional (2D) coordinates of the microrobot in the present time step, and the 2D coordinates of the previous step. While the current waveforms could theoretically take on an infinite number of shapes, we chose to define the applied waveforms as sinusoids to bound the space of possible actions that the agent could take, and fixed for each training their frequency to f. The action space consisted of five continuous actions, which controlled the magnitudes (M_x, M_y, M_z) and phase angles (φ_x, φ_y) for the sinusoidal currents. The equations that describe the currents in the electromagnetic coils that generate the magnetic field, which places a torque on the roller, are as follows:

$$\begin{cases} I_x = M_x \sin(2\pi f t + \varphi_x) \\ I_y = M_y \sin(2\pi f t + \varphi_y) \\ I_z = M_z \sin(2\pi f t) \end{cases} \quad (3)$$

The goal of the control system for our remotely actuated microrobot was to manipulate the shape and magnitude of the actuating magnetic field to move the microrobot and achieve the intended dynamic behavior. The RL agent started without any prior information about the task and had to learn to perform it by sampling actions from the space of all possible actions and identifying which actions resulted in rewarded behavior. Developing the RL controller required us to formulate the task and define the associated reward signal. The objective of the RL agent was to develop an action policy π that maximized the total value of the rewards it would receive by following that policy in the future. When the environment was in state s, the agent chose an action a from the policy according to $a \sim \pi(\cdot \mid s)$, probabilistically selecting from a distribution of possible actions available in that state. As a first approach to the problem, we decided that our RL agent should derive a policy to move the roller in the positive x direction towards a goal target x_g. The agent received a reward when it selected actions that moved the microrobot into increasing x and a negative reward when it moved the roller in any other direction. The reward function selected was

$$r(s,a) = \Delta x_r - 0.5|\Delta y_r| - 1 + (1000 \quad \text{if} \quad x_r > x_g) \quad (4)$$

where $\mathbf{r} = (x_r, y_r, z_r)$ is the position of the roller, x_g the position of the goal, and $\Delta\mathbf{r} = (\Delta x_r, \Delta y_r, \Delta z_r)$ the change in the position of the roller as a result of taking action a in state s. Within RL, the selection of an appropriate reward function can significantly impact the system's performance.

We designed the reward to encourage the agent to reach the goal as quickly as possible by adding a -1 penalty for every extra step the roller took to reach the objective and a final bonus reward of 1000 when it reached the goal. To rigorously evaluate the effectiveness of our RL approach, we deliberately chose this simple directional control task as it has a known physics-informed (PI) solution. In particular, optimal motion in the positive $\hat{\mathbf{x}}$ direction can be achieved using sinusoidal magnetic fields in the $\hat{\mathbf{z}}$ and $\hat{\mathbf{x}}$ directions shifted by $\pi/2$, causing the microrobot to rotate around the y-axis. The optimal action values for this motion are $M_y = 0$, $|M_x/M_z| = 1$, $\varphi_x = \pi/2$, and φ_y any value. This known solution provides a benchmark against which we can evaluate our RL agent's ability to discover optimal control strategies through experience alone.

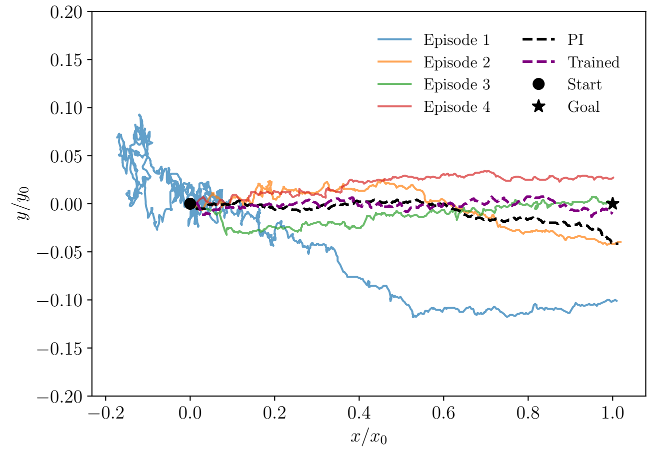

FIG. 3. Microrobot trajectories for $f = 8$ Hz.

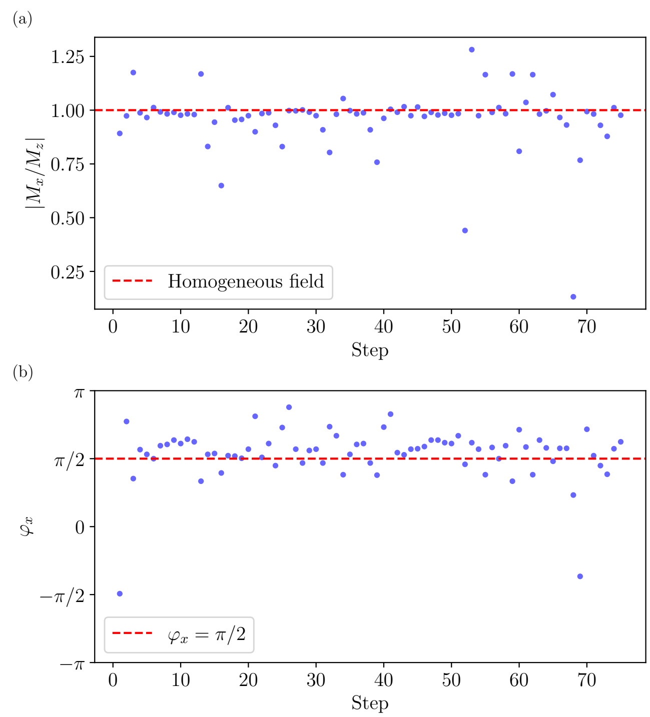

FIG. 4. Evolution of action parameters for the trained model. (a) Ratio between magnetic fields ($|M_y/M_z|$) stabilizes around 1, indicating the agent learned to maintain a homogeneous rotating field. (b) Phase angle φ_x is near $\pi/2$, corresponding to optimal rolling conditions. The dashed red lines show the theoretically optimal values from the PI analysis.

C. Learning process

Fig. 3 shows how our microrobots learned to navigate over time. We tracked a 2.7 μm diameter microrobot actuated at 8 Hz through multiple training episodes, comparing its path to the PI trajectory. Early in training (Episode 1, blue line), the microrobot moved erratically with large sideways motions, often drifting away from the target. During Episode 2 (orange line), the motion became more focused but still showed significant deviations from the ideal path. By Episodes 3 (green line) and 4 (red line), the microrobot achieved consistent forward motion with only minor lateral drift. The final trained policy (purple dashed line) matches well with the PI trajectory (black dashed line). This shows the RL agent learned efficient rolling behavior without being explicitly programmed with the underlying physics. Small oscillations remain visible in all paths due, in part, to Brownian motion, which continuously affects the microrobot at this scale. As it can be seen, even the PI trajectory exhibit the characteristic *jiggling* motion associated with Brownian motion.

To better understand the learned behavior, we an-

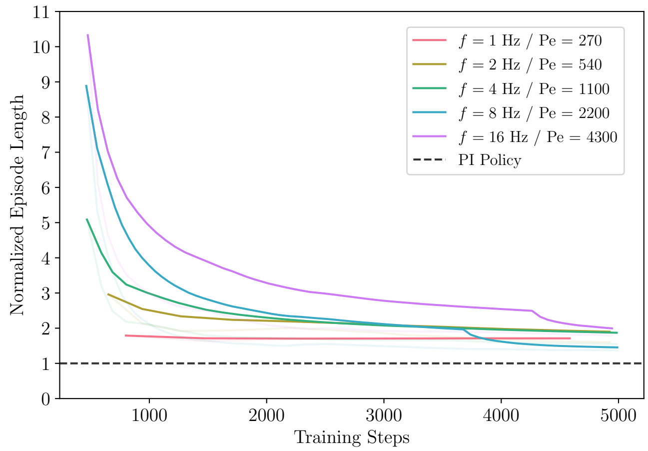

FIG. 5. Training performance at different actuation frequencies. The episode length is normalized by dividing by the length of the optimal path determined by the PI policy, shown as a dashed line.

alyzed the evolution of the control parameters in the trained policy. Fig. 4 shows how the key action parameters converge to PI optimal values. The trained agent independently discovered the optimal relationship between magnetic fields that was previously derived from the underlying physics. The ratio $|M_y/M_z|$ oscillates around 1 (Fig. 4a), matching the requirement for maintaining a homogeneous rotating field. Similarly, the phase angle φ_x converges to values near $\pi/2$ (Fig. 4b), the theoretically optimal phase shift for maximizing forward rolling motion. This convergence to PI values is particularly interesting as it emerged purely through trial, error and learning, without any explicit knowledge of the underlying physics being programmed into the agent. The RL algorithm discovered these relationships by maximizing the reward signal, which was designed only to encourage forward motion and goal-reaching behavior.

D. Influence of thermal fluctuations in the learning process

In the field of RL, artificial noise sources are commonly introduced to enhance exploration of the action and parameter landscape. In our microrobotic system, this noise emerges naturally from Brownian motion, raising the question of whether these thermal fluctuations could improve learning efficiency. The significance of noise in Brownian systems is often quantified using the Peclet number, defined as $\mathrm{Pe} = rv/2D$. This dimensionless parameter compares the product of the particle radius r and the deterministic particle displacement $v\Delta t$ to the mean square displacement caused by Brownian motion, $2D\Delta t$. We explored the influence of the noise strength by modifying the rotation frequency f of the magnetic field, changing the speed v of the rollers, while keeping constant the diffusion coefficient D.

Fig. 5 shows the normalized episode length versus training steps for different actuation frequencies of the microrollers. The episode length is normalized by dividing it by the length of the optimal path determined by the PI policy, shown as a dashed line. Training was conducted at five frequencies: 1, 2, 4, 8, and 16 Hz. Higher normalized episode lengths indicate longer, less efficient paths to reach the target. The convergence behavior exhibits clear frequency dependence. At 16 Hz, the initial normalized path length is approximately 10 times longer than optimal, while at 1 Hz it starts around 2 times longer. All frequencies show eventual convergence toward the PI policy baseline, with higher frequencies requiring more training steps but ultimately achieving comparable performance to lower frequencies by 5000 training steps. This figure demonstrates that while Brownian motion (more significant at lower frequencies) may aid initial learning, it does not significantly impact final performance after sufficient training.

This behavior can be understood by examining the Pe regimes. At high frequencies ($\mathrm{Pe} > 2000$), deterministic motion dominates, leading to consistent but potentially suboptimal paths as the agent has limited natural exploration. At low frequencies ($\mathrm{Pe} < 500$), Brownian motion provides significant exploration, facilitating rapid initial learning but potentially interfering with precise control. The intermediate regime ($\mathrm{Pe} = 500 - 2000$) appears optimal, balancing exploration and control. Our results indicate that while Brownian motion's influence is most evident during early training stages, long-term performance remains relatively independent of Pe in the studied range.

III. DISCUSSION

Our results demonstrate that RL can effectively control magnetic microrollers without explicit physical modeling, while independently discovering optimal control strategies that match theoretically-derived solutions. The RL agent's convergence to the PI optimal parameters, specifically the magnetic field ratio and phase relationships, validates this approach for microrobot control. This is particularly significant as it suggests RL could be valuable for controlling more complex microrobotic systems where analytical solutions are challenging to derive.

The influence of Brownian motion on the learning process reveals an interesting trade-off: while thermal fluctuations accelerate initial learning at low Pe numbers, they do not significantly affect final performance. This suggests that microrobotic systems may benefit from varying actuation frequencies during training: starting with lower frequencies to exploit natural exploration from Brownian motion, then transitioning to higher frequencies for precise control.

A key advantage of our RL approach is its adaptability to changing conditions. Unlike traditional control methods that require explicit system modeling, our RL controller learns directly from experience, potentially allowing it to adapt to variations in fluid properties, surface interactions, or mechanical wear. This adaptability will be crucial for real-world applications where environmental conditions may be unpredictable or time-varying. By demonstrating successful learning and characterizing the role of thermal fluctuations in a simple navigation task, this work establishes a foundation for applying reinforcement learning to more complex microrobotic challenges, including collective behavior and navigation through dynamic, obstacle-rich environments [10].

IV. METHODS

A. Experimental setup

To power the coils, three digital signals are sent from a Linux control computer into a USB DAQ (Measurement computing USB-1208HS), which converts these signals into analog form. The analog signals are then directed to three 500W power amplifiers (LD Systems), which subsequently drive the amplified signal to the three pairs of coils. Thus, generating an homogeneous, rotating magnetic field around any 3D axis of rotation with an intensity up to 10 mT. The frequency of the rotating field can be potentially varied from 0.5 to 100 Hz, being the upper limit given by the microrobot magnetic moment. The voltage sent to the coils is monitored using an oscilloscope (Tektronix TDS 2014B) connected to the setup. The coils are mounted on an inverted bright field optical microscope (Nikon ECLIPSE Ts2R), with a CCD camera (Grasshopper3 USB3, FLIR) through a 40X air objective (Nikon S PLAN FLUOR). The video input from the camera is sent back to the computer and the microrobot tracked using OpenCV [11].

B. Training method and procedure

We implemented our microrobot control system using the deep RL Soft Actor-Critic (SAC) algorithm [12], accessed through the Stable-Baselines3 library [13]. SAC is particularly well-suited for continuous control problems and offers sample efficiency crucial for real-world robotics applications. The environment was wrapped using an OpenAI Gym [14] for ease of integration with Stable-Baseline3 and training. The coil amplitude actions were scaled to $(-1, 1)$. Training was conducted on an RTX 1080 Ti with hyperparameters: learning rate $\alpha = 3 \times 10^{-3}$, discount factor $\gamma = 0.99$, target smoothing coefficient $\tau = 0.01$, and batch size of 256. The entropy coefficient was automatically tuned, and the policy was trained for 5000 timesteps with gradient updates performed at every step.

[1] S. Jeon, S. Kim, S. Ha, *et al.*, Magnetically actuated microrobots as a platform for stem cell transplantation, Science Robotics **4**, 1 (2019).

[2] S. Scheggi, K. K. T. Chandrasekar, C. Yoon, B. Sawaryn, G. Van De Steeg, D. H. Gracias, and S. Misra, Magnetic motion control and planning of untethered soft grippers using ultrasound image feedback, in *2017 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2017) pp. 6156–6161.

[3] A. Hosney, J. Abdalla, I. S. Amin, N. Hamdi, and I. S. Khalil, In vitro validation of clearing clogged vessels using microrobots, in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)* (IEEE, 2016) pp. 272–277.

[4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. (The MIT Press, 2018).

[5] E. M. Purcell, Life at low Reynolds number, American Journal of Physics **45**, 3 (1977).

[6] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, Rev. Mod. Phys. **88**, 045006 (2016).

[7] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow navigation by smart microswimmers via reinforcement learning, Phys. Rev. Lett. **118**, 158004 (2017).

[8] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, Science Robotics **6**, eabd9285 (2021).

[9] M. R. Behrens and W. C. Ruder, Smart magnetic microrobots learn to swim with deep reinforcement learning, Advanced Intelligent Systems **4**, 2200023 (2022).

[10] L. Yang, J. Jiang, F. Ji, Y. Li, K. Yung, A. Ferreira, and L. Zhang, Machine learning for micro- and nanorobots, Nature Machine Intelligence **6** (2024).

[11] G. Bradski, The opencv library, Dr. Dobb's Journal of Software Tools (2000).

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

[13] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, Stable-baselines3: Reliable reinforcement learning implementations, Journal of Machine Learning Research **22**, 1 (2021).

[14] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, Openai gym (2016), arXiv:1606.01540.