Towards Low-Gravity Planetary Exploration: Reinforcement Learning for Quadrupedal Walking, Jumping, and Attitude Control

Jørgen Anker Olsen and Kostas Alexis

***Abstract*— This paper presents reinforcement learning (RL) policies for dynamic quadrupedal locomotion in planetary exploration scenarios. Building on the quadruped Olympus, we develop RL policies for walking, vertical jumping, forward jumping, and in-flight attitude control, explicitly tailored for the reduced gravity on Mars. These policies jointly enable such robots to overcome obstacles much larger than themselves through coordinated jumping and in-flight reorientation for safe landings. We demonstrate the attitude control policy on the Olympus quadruped through reorientation tests, while all locomotion policies are validated in simulation. A complete Mars exploration mission scenario demonstrates coordinated policy deployment across challenging terrain.**

I. INTRODUCTION

Traditional space exploration has been dominated by the rover and lander form factor due to their historic success in returning scientific images and data when deployed on the Moon and Mars [1], [2]. Their wheeled design allows for efficient exploration of the flatter parts of lunar and planetary surfaces, but they often struggle on steep slopes, loose soil, and rough terrain with large obstacles [3]. However, many scientifically interesting areas are located in harder-to-reach places that rovers would struggle to access. An important example is lava tubes, considered to bring together the trifecta of science, exploration, and resources [4], [5]. This limitation heavily motivates the use of alternative robotic form factors [3]. Beyond proven platforms like rovers and helicopters [6], [7], proposed systems for planetary exploration include pit-bots [8] and legged robots [3], [9]. Legged robots present a promising solution, having demonstrated significant improvements in capabilities and robustness in recent years [10]. The reduced gravity environments of Mars ($3.71\,\mathrm{m/s^2}$) and the Moon ($1.62\,\mathrm{m/s^2}$) particularly favor dynamic locomotion: jumping maneuvers that would be challenging on Earth become feasible, enabling robots to overcome obstacles significantly larger than their body size [11]. Some concepts also investigate continuous jumping as a primary mode of locomotion [12].

However, controlling these dynamic behaviors poses significant challenges. Jumping requires precise coordination during takeoff, in-flight attitude control, and coordinated landing, all while adapting to uneven terrain and possible loose soil. The complexity of attitude reorientation during flight and uncertain contact schedules makes classical control approaches difficult, motivating the use of RL [12], [11]. RL has emerged as a powerful approach for controlling complex quadruped behaviors [13], with recent advances yielding highly robust policies for walking [14], jumping [15], [16], and attitude control [12], [17].

The authors are with the Autonomous Robots Lab, NTNU, O.S. Bragstads Plass 2D, 7034, Trondheim, Norway. jorgen.a.olsen@gmail.com, Project page: https://ntnu-arl.github.io/olympus/

Fig. 1. Olympus standing in a Mars analog environment.

Building upon previous work with Olympus (Fig. 1), a quadruped platform optimized for powerful jumps and in-flight attitude control in low-gravity environments [18], we investigate the use of deep reinforcement learning (DRL) to enable highly dynamic quadrupedal locomotion on Martian terrain. This work presents the use of a series of controllers to enable highly dynamic locomotion. Contributions include:

- A DRL-based attitude control policy capable of rapid in-flight reorientation to ensure safe landings, achieving 90° rotations in 2.5 s during experimental validation on robot hardware.
- DRL policies for walking, vertical jumping, and forward jumping trained specifically for Martian gravity conditions. Simulations demonstrate vertical jumps up to 3.1 m and horizontal jumps up to 3.9 m.
- A planetary exploration pipeline demonstrating the traversal of challenging terrain and large obstacles through coordinated multi-policy execution.

II. THE OLYMPUS PLATFORM AND RL CONTROL

This section details the quadrupedal hardware and the reinforcement learning control architecture designed to enable dynamic locomotion in low-gravity environments.

A. Hardware Morphology

The quadrupedal robot Olympus (Fig. 1 and 2) was specifically designed for dynamic locomotion in Martian gravity. The robot has a mass of 14.5 kg and a total body length of 0.67 m. It employs three degrees of freedom per leg, utilizing

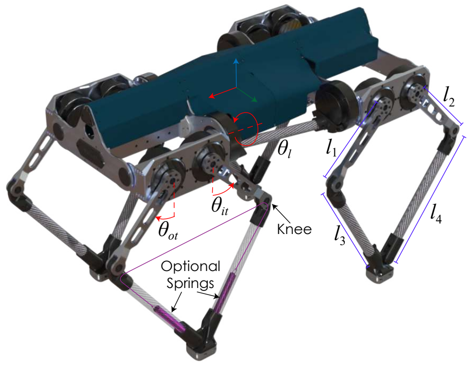


Fig. 2. Olympus design with annotated leg parameters.

a 5-bar leg design that provides both a large workspace for in-flight attitude control and significant jumping capabilities. For actuation, it uses twelve torque-controlled brushless DC motors, providing a maximum torque of 18.0 N m for lateral movement and 24.8 N m for transversal movement.

Previous work optimized the robot morphology via grid search to maximize vertical jump height, horizontal jump distance, and angular reorientation rates while maintaining dynamic walking capabilities [11]. The 5-bar linkage also allows for the optional use of integrated parallel springs, though the default configuration utilized in this work operates without springs.

B. Control Architecture and Policy Training

The control architecture inputs the task-specific observations into the RL policies, which output motor target positions that are then rescaled and offset. To ensure safe operation, these targets pass through a safety filter to enforce joint limits and task-specific torque saturation before being tracked by PD motor controllers. During hardware deployment, policy inference runs at 60 Hz on an NVIDIA Jetson Orin NX onboard computer, receiving body orientation and angular velocity from a motion capture system alongside direct motor states.

Training was done using IsaacLab [19] and the RL Games PPO implementation [20]. To enable a unified exploration pipeline, four distinct policies were developed:

Attitude Control: The objective of this policy is to control the orientation of the robot in the flight phase by using its legs as reaction masses. The reward structure primarily minimizes quaternion error relative to a target orientation while secondary rewards encourage the legs to assume a default landing configuration once the orientation error falls below 5°.

Walking: Trained directly in simulated Mars gravity, this policy tracks commanded linear velocity and yaw rate while penalizing vertical velocity and excessive body tilt.

Vertical and Forward Jumping: These policies track commanded jump heights and xy-plane target positions respectively. To accelerate the learning of jumping behavior, a curriculum-based reference state initialization scheme is used. Agents are initialized at various stages of the jumping maneuver (squatting, in-flight, and pre-touchdown). During flight, the reward structure relies on projectile-based estimates to provide continuous feedback before landing, densifying the reward. The driving reward for vertical jumping is reaching a commanded height after a jump command is given, and for horizontal jumping it is jumping to a commanded jump target in the plane in front of the robot.

III. Experimental and Simulation Validation

To evaluate the trained policies, we conducted extensive simulations for all the policies under Martian gravity conditions and hardware experiments for the attitude control policy to verify Sim2Real transfer.

A. Locomotion Performance in Low Gravity

The vertical and forward jumping policies were evaluated across hundreds of jumps with varying targets in simulation (Fig. 4). The vertical jumping policy achieved a mean absolute error of 0.123 m and an 88.9% success rate (apogee within 0.2 m of target), demonstrating reliable vertical jumps up to 3.1 m. The forward jumping policy exhibited a mean absolute error of 0.208 m and an 80.7% success rate, achieving forward jumps up to 3.9 m. Evaluation targets span the training distribution and extend to the robot's physical jumping capabilities. Failures primarily occur on out-of-distribution samples where landing stability reaches the edge of the platform's capabilities. Furthermore, simulating the optional parallel knee springs (stiffness of 800 N/m) yielded an approximate 21% increase in jump height and distance, highlighting the platform's morphological versatility.

B. Attitude Control: Simulation and Hardware

The attitude control policy was first tested in simulated zero-g conditions. The policy was evaluated through single-axis tests and a complex 3D reorientation maneuver, commanding a simultaneous multi-axis rotation from $[-90°, 90°, 90°]$ (roll, pitch, yaw) to $[0°, 0°, 0°]$. The policy successfully reached the target orientation within 2.45 s, demonstrating highly coordinated multi-axis control.

The policy was then validated on robot hardware using a custom rotating rod test stand with a floating air-bearing platform to simulate these free-flight dynamics (Fig. 5). State estimation was provided by a motion capture system, and motor torques were limited to 12 N m. Due to the instability of the mounting platform, leg motion speeds were safely reduced using a 20-sample moving average filter.

Despite the hardware safety filter, the policy demonstrated rapid convergence across multiple tests, including 90° steps, 180° steps, and continuous multi-step sequences. In single-axis 90° step response tests, the robot successfully reoriented the roll axis in 2.6 s, with pitch and yaw stabilizing in 4.2 s and 3.9 s respectively. Unfiltered simulations indicate the policy is capable of even faster responses (e.g., 90° roll in 0.96 s, see Table I) when physical test stand constraints are removed.

IV. Planetary Exploration Pipeline Testing

To validate the use of highly dynamic maneuvers in a Mars exploration setting, we set up a comprehensive exploration

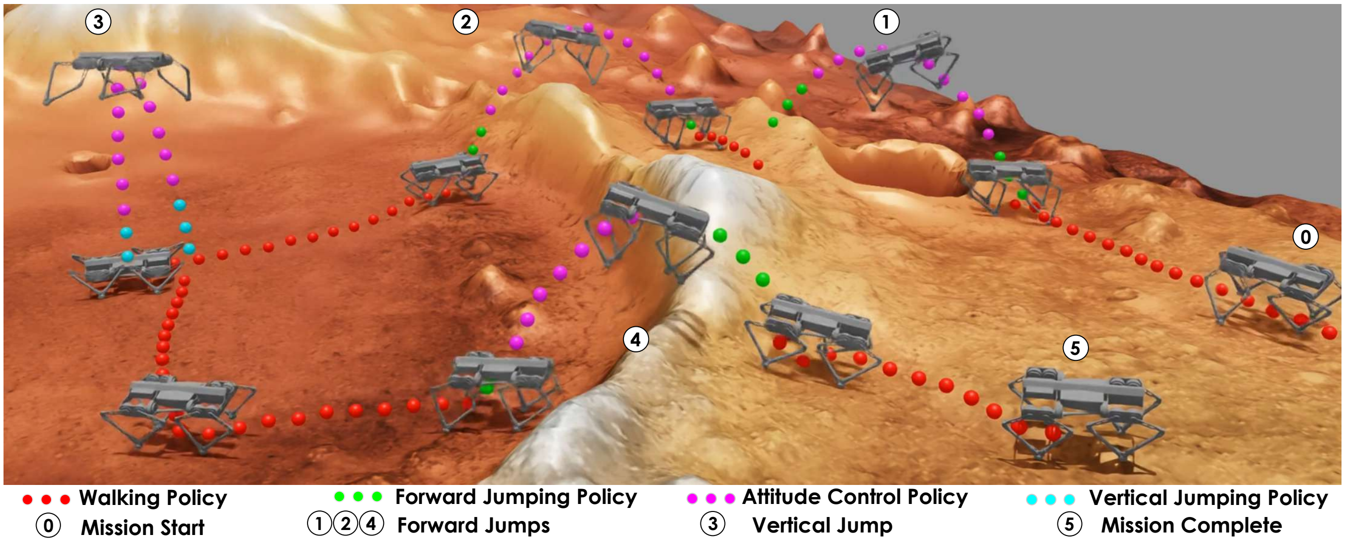
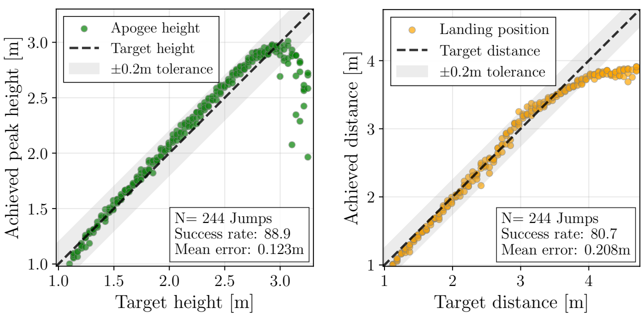


Fig. 3. Sequential frames from the integrated planetary exploration mission simulation. The dot colors represent the active task-specific policy. Olympus successfully navigates rough terrain, a 2.1 m crater, a 3.5 m forward jump from a ledge, and performs a 2.6 m vertical reconnaissance jump, with attitude control ensuring safe landings.



(a) Vertical jumps (b) Forward jumps

Fig. 4. Jump target tracking performance in simulated Martian gravity.

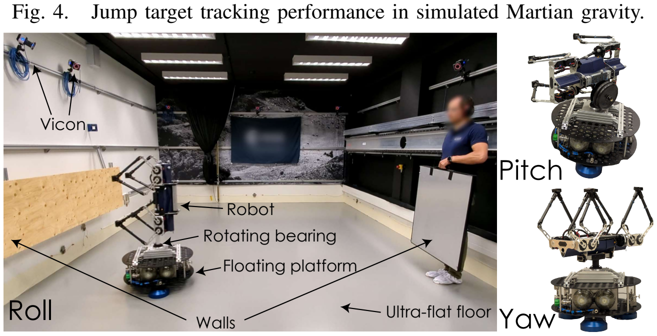


Fig. 5. Hardware attitude control test setup utilizing a floating platform on a flat floor to simulate free-flight dynamics.

mission in a simulated Martian environment. This pipeline demonstrates the coordinated deployment of the walking, jumping, and attitude control policies to traverse terrain features that would be impossible for traditional rovers to overcome.

A waypoint-based hierarchical controller coordinates task execution. The controller tracks position and orientation errors, commanding the walking policy to walk over rough terrain. The walking policy maintains the important ability to remain fully stationary under zero velocity commands, ensuring a consistent policy handover.

Upon reaching a jump waypoint, the system transitions

TABLE I

ATTITUDE CONTROL REORIENTATION TIMES

Test	Roll		Pitch		Yaw	
	Sim	Real	Sim	Real	Sim	Real
90°	0.96 s	2.6 s	1.08 s	4.2 s	1.4 s	3.9 s
180°	1.9 s	4.6 s	2.3 s	8.4 s	2.4 s	7.1 s

through a predefined state sequence: 1) The robot stabilizes in a stance for 1 s. 2) The selected jumping policy takes over and initiates takeoff. 3) Once the feet lose ground contact and altitude exceeds 0.6 m with positive vertical velocity, the attitude control policy engages to control the body orientation during flight. 4) As the robot descends below 0.9 m, control reverts to the jumping policy to brace for impact. 5) Shortly after ground contact, the walking policy takes over, utilizing its recovery control capabilities to stabilize the momentum carryover from the aerial phase, if needed.

During the simulated mission, this integrated system successfully traversed a highly complex terrain (Fig. 3). The sequence included walking in rough terrain, jumping over a 2.1 m wide crater, executing a 3.5 m forward jump from a 1.1 m ledge, performing a 2.6 m vertical reconnaissance jump, and clearing a 1.1 m ledge with challenging landing conditions.

V. CONCLUSION

This extended abstract presented a multi-policy hierarchical control framework for dynamic quadrupedal locomotion in planetary exploration. By combining task-specific policies for walking, jumping, and in-flight attitude control, the Olympus platform successfully navigated challenging simulated Martian terrain, including craters and steep ledges. Hardware validation of the attitude control policy further demonstrated the feasibility of rapid in-flight reorientation. These results illustrate that jumping legged robots, guided by multi-policy RL frameworks, offer a highly capable alternative to traditional rovers for exploring extreme planetary environments such as Martian lava tubes.

References

[1] S. T. Arzo, D. Sikeridis, M. Devetsikiotis, F. Granelli, R. Fierro, M. Esmaeili, and Z. Akhavan, "Essential technologies and concepts for massive space exploration: Challenges and opportunities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 1, pp. 3–29, 2022.

[2] A. R. Vasavada, "Mission overview and scientific contributions from the mars science laboratory curiosity rover after eight years of surface operations," *Space Science Reviews*, vol. 218, no. 3, p. 14, 2022.

[3] P. Arm, G. Waibel, J. Preisig, T. Tuna, R. Zhou, V. Bickel, G. Ligeza, T. Miki, F. Kehl, H. Kolvenbach *et al.*, "Scientific exploration of challenging planetary analog environments with a team of legged robots," *Science robotics*, vol. 8, no. 80, p. eade9548, 2023.

[4] F. Sauro, R. Pozzobon, M. Massironi, P. De Berardinis, T. Santagata, and J. De Waele, "Lava tubes on earth, moon and mars: A review on their size and morphology revealed by comparative planetology," *Earth-Science Reviews*, vol. 209, p. 103288, 2020.

[5] H. Kolvenbach *et al.*, "Lunarleaper-a mission concept to explore the lunar subsurface with a small-scale legged robot," in *IAC 2024 Conference Proceedings*.

[6] A. Patel, S. Karlsson, B. Lindqvist, C. Kanellakis, A.-A. Agha-Mohammadi, and G. Nikolakopoulos, "Towards energy efficient autonomous exploration of mars lava tube with a martian coaxial quadrotor," *Advances in Space Research*, vol. 71, no. 9, pp. 3837–3854, 2023.

[7] F. Mier-Hicks *et al.*, "Sample Recovery Helicopter," in *2023 IEEE Aerospace Conference*, Mar. 2023, pp. 1–11, iSSN: 1095-323X. [Online]. Available: https://ieeexplore.ieee.org/document/10115951

[8] J. Thangavelautham, M. S. Robinson, A. Taits, T. McKinney, S. Amidan, and A. Polak, "Flying, hopping pit-bots for cave and lava tube exploration on the moon and mars," *arXiv preprint arXiv:1701.07799*, 2017.

[9] R. Doyle, T. Kubota, M. Picard, B. Sommer, H. Ueno, G. Visentin, and R. Volpe, "Recent research and development activities on space robotics and ai," *Advanced Robotics*, vol. 35, no. 21-22, pp. 1244–1264, 2021.

[10] M. Tranzatto, T. Miki, M. Dharmadhikari, L. Bernreiter, M. Kulkarni, F. Mascarich, O. Andersson, S. Khattak, M. Hutter, R. Siegwart *et al.*, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, p. eabp9742, 2022.

[11] J. A. Olsen, G. Malczyk, and K. Alexis, "Olympus: A jumping quadruped for planetary exploration utilizing reinforcement learning for in-flight attitude control," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 4366–4372.

[12] N. Rudin, H. Kolvenbach, V. Tsounis, and M. Hutter, "Cat-like jumping and landing of legged robots in low gravity using deep reinforcement learning," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 317–328, 2022.

[13] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," *arXiv preprint arXiv:1804.10332*, 2018.

[14] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[15] J. A. Olsen, L. R. Pettersen, and K. Alexis, "Towards quadrupedal jumping and walking for planetary exploration using reinforcement learning," *arXiv preprint arXiv:2510.24584*, 2025.

[16] V. Atanassov, J. Ding, J. Kober, I. Havoutis, and C. D. Santina, "Curriculum-based reinforcement learning for quadrupedal jumping: A reference-free design," *IEEE Robotics and Automation Magazine*, vol. 32, no. 2, pp. 35–48, 2025.

[17] T. El-Agroudi *et al.*, "In-flight attitude control of a quadruped using deep reinforcement learning," in *8th Annual Conference on Robot Learning*, 2024.

[18] J. A. Olsen and K. Alexis, "Design and experimental verification of a jumping legged robot for martian lava tube exploration," in *2023 21st International Conference on Advanced Robotics (ICAR)*. IEEE, 2023, pp. 452–459.

[19] M. Mittal, C. Yu, Q. Yu, J. Liu, N. Rudin, D. Hoeller, J. L. Yuan, R. Singh, Y. Guo, H. Mazhar, A. Mandlekar, B. Babich, G. State, M. Hutter, and A. Garg, "Orbit: A unified simulation framework for interactive robot learning environments," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3740–3747, 2023.

[20] D. Makoviichuk and V. Makoviychuk, "rl-games: A high-performance framework for reinforcement learning," https://github.com/Denys88/rl_games, May 2021.