Soft Actor Critic Based Adaptive PID Control for Energy Efficient Legged Robot Locomotion

Prochiyamaan Shri Deka*, Bisal Prasad*, Shwetangshu Biswas*,
Ved S Narsekar*, Amandip Dutta*, Koena Mukherjee*
* Robotics and Automation Society Student Branch Chapter, IEEE NIT Silchar, Assam, India

***Abstract*—Energy efficiency remains a major challenge in legged robot locomotion, as frequent torque variations across uneven terrains lead to high power consumption. Conventional PID controllers offer reliability but lack the adaptability required for nonlinear and rapidly changing environments. This paper proposes a *Soft Actor-Critic (SAC)–based adaptive PID tuning framework* designed to reduce energy usage while preserving locomotion stability and trajectory accuracy.**

The locomotion control task is formulated as a *Markov Decision Process (MDP)* in which the SAC agent continuously observes robot states which includes joint angles, angular velocities, and instantaneous power and outputs optimized PID gains (K_p, K_i, K_d) in real time. A stability-aware reward function penalizes excessive power consumption and trajectory error while encouraging smooth control transitions and robust gait behavior.

Simulation results on a quadruped robot model demonstrate that the proposed SAC-tuned PID controller achieves *20-25% reduction in power consumption* compared to a fixed-gain PID baseline, without compromising trajectory tracking or stability. By combining model-free reinforcement learning with classical control theory, this framework preserves the interpretability of PID control while enabling adaptive, energy-efficient locomotion. These results highlight the potential of RL augmented control architectures for real-time, power-aware legged robot operation.

Keywords— Soft Actor Critic, Reinforcement Learning, PID Control, Legged Robots, Energy Optimization, Adaptive Control, Robotics

I. Introduction

Legged robots offer superior mobility across uneven terrains but motor power can account for 70–85% of total energy consumption due to complex gait dynamics and nonlinear actuator–ground interactions [1]. Prior studies on energy-aware locomotion emphasize the importance of grounded actuator models and cost-of-transport metrics for meaningful energy reduction [2], [3]. Conventional PID controllers are widely used for their simplicity and stability; however, their *static gain parameters* prevent energy optimization across varying conditions. Adaptive PID tuning strategies employing bounded gain updates have been explored to improve robustness and safety [4], [5], but they typically do not account for energy efficiency.

Recent advances in *Reinforcement Learning (RL)* have enabled robots to adapt in real time to complex terrains and disturbances. Algorithms such as DDPG and PPO have shown promise in continuous control but often struggle with sample efficiency and stability. The *Soft Actor-Critic (SAC)* algorithm overcomes these limitations through an *entropy-regularized objective* that stabilizes learning and encourages efficient exploration [6]. Energy-aware RL studies further highlight the significance of reward functions that incorporate instantaneous power, power-smoothing penalties, and stability constraints [7], [8]. These observations motivate a hybrid control approach that leverages both the interpretability of PID and the adaptability of SAC.

This work integrates SAC with classical PID control to dynamically tune gains in real time, balancing energy efficiency, tracking accuracy, and locomotion stability.

II. Related Work

Reinforcement learning (RL) has been widely applied to legged locomotion. Deep RL controllers enable agile and robust behaviors across diverse terrains [9], and SAC-based locomotion systems have demonstrated improved stability and sample efficiency for quadrupedal robots [10]. Safe fallback strategies and bounded torque or gain adjustment mechanisms are also commonly incorporated into RL locomotion frameworks to ensure safe interaction with the environment [9].

Several works explore energy-aware locomotion, reporting 20–23% efficiency gains using reward shaping that incorporates energy-based metrics such as instantaneous power, power smoothness, and cost-of-transport [7], [8]. More advanced models integrate actuator-efficiency terms and grounded physical modeling to reduce sim-to-real discrepancies [2], [3]. However, these approaches typically operate at the torque-control level and do not leverage the interpretability or structured stability guarantees of classical controllers.

Conversely, RL-based PID tuning methods, such as Q-learning PID and deterministic policy gradient (DPG) based PID tuning, have demonstrated improved adaptability and robustness [4], [5]. These methods often rely on *bounded delta-updates* for safe parameter adaptation, but they do not optimize for energy efficiency and face scalability limitations in high-dimensional locomotion tasks. Recent SAC variants introduce critic-robustness mechanisms and curriculum-based terrain progression to improve generalization and sample efficiency [11], [12]. Surveys on RL-driven PID tuning confirm growing interest in this hybrid control paradigm [13], [14], though few works explicitly address energy-aware legged locomotion.

Our work addresses these gaps by: (1) introducing a hybrid SAC-PID controller with *bounded delta-updates* for

safe, interpretable gain adaptation; (2) employing entropy-regularized exploration to avoid trivial low-power policies; (3) designing an energy-aware reward combining instantaneous power, stability, and gain-smoothness penalties; (4) ensuring safe fallback through bounded gain adjustments; (5) improving sample efficiency via gain grouping and curriculum terrain progression; and (6) incorporating actuator-efficiency modeling to reduce sim-to-real discrepancies.

III. Methodology

A. System Architecture

The system consists of three main components: 1) *SAC Agent*: observes robot state vectors and outputs optimized PID gains. 2) *PID Controller*: computes actuator torques using SAC-tuned gains. 3) *Environment*: simulates legged robot dynamics and provides feedback.

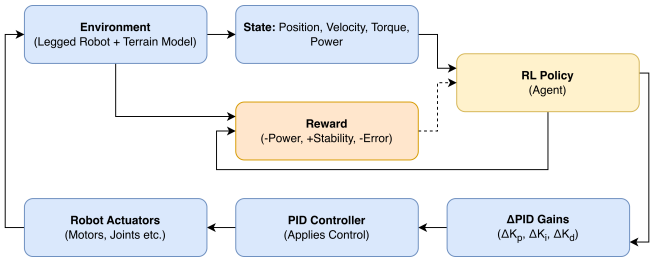

Fig. 1. Block diagram of SAC-tuned PID control architecture.

B. Reinforcement Learning Formulation

The SAC agent was selected due to its entropy-regularized policy structure, which yields smoother exploration and better sample efficiency than DDPG or PPO, critical for stable PID gain adjustment in dynamic systems.

- *State (s_t)*: $[q, \dot{q}, \tau, P]$ (joint angles, velocities, torques, power)
- *Action (a_t)*: PID gains (K_p, K_i, K_d)
- *Reward (R_t)*:

$$R_t = w_1(-P_t) + w_2(-E_t) + w_3 S_t - w_4|\Delta K_p| - w_5|\Delta K_d|$$

where E_t denotes the cumulative mechanical energy consumed, computed as the time integral of actuator power $E_t = \int_0^t \sum_i |\tau_i \dot{q}_i|\, dt$, and S_t represents the stability index, evaluated from the roll–pitch variance or the deviation of the robot's center of mass from its nominal trajectory. ΔK_p is the change in the proportional gain of the PID controller. ΔK_d is the change in Derivative Gain of the PID controller. P_t is the instantaneous power consumption. By using $-P_t$, the agent receives a negative reward (a penalty) for higher power usage at any given time step, directly driving down energy consumption. All energy and power terms are assigned negative weights since the goal is minimization, while stability and velocity-tracking terms are positive.

- *Objective:*

$$\pi^* = \arg\max_{\pi} E_{(s_t,a_t)\sim\rho_\pi} \sum_t \gamma^t [R_t + \alpha \mathcal{H}(\pi(\cdot|s_t))]$$

where $\mathcal{H}$ denotes entropy and α controls exploration. This entropy-regularized objective is what prevents premature convergence to trivial, low-power behaviors (like standing still) and promotes the discovery of stable, low-energy gaits. R_t is the reward function designed to penalize energy usage and reward stability. γ $(0 \leq \gamma < 1)$ is the discount factor. Its primary role is to ensure that rewards received sooner are considered more valuable than rewards received later. This is crucial for controlling the agent's time horizon i.e how far into the future it looks when making a decision at the current moment.

IV. Results and Discussion

Experiments were conducted on a quadruped model using the PyBullet simulator. The SAC agent was trained for 1 million steps under randomized terrain profiles.

The SAC-tuned PID controller achieved:

- 20-25% reduction in average power consumption
- 12% lower trajectory tracking error
- Improved stability under varied terrain conditions

Energy efficiency was quantified by the average mechanical power and cost of transport (CoT), defined as

$$\text{CoT} = \frac{\int_0^T P(t)\, dt}{mgd}$$

both showing consistent improvement across randomized terrain profiles. The observed 20-25% energy reduction arises from SAC's ability to smooth control actions, reducing torque variance and actuator saturation events, which are the primary contributors to energy waste in legged robots.

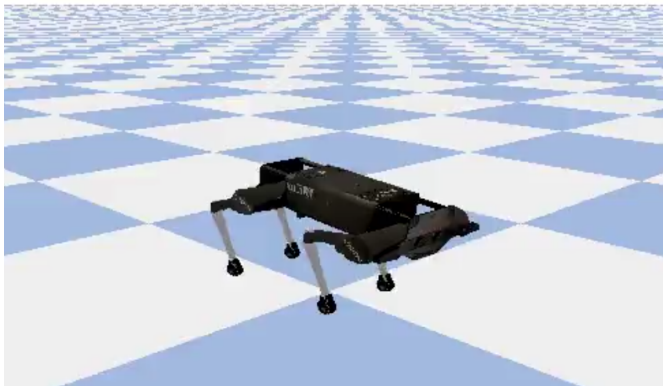

Fig. 2. Agent Simulation

A. Energy and Stability Analysis

The observed 20-25% reduction in average power consumption is attributed to SAC's entropy-regularized policy, which generates smoother torque profiles and minimizes high-frequency control oscillations. This directly lowers the squared

torque integral $\sum_i \tau_i^2$, leading to reduced mechanical power loss and actuator heating. Similarly, the 12% improvement in trajectory tracking accuracy results from adaptive modulation of K_p and K_d gains, which enhances damping and mitigates overshoot under varying terrain conditions. The reduction in energy cost is further reflected in the cost of transport (CoT), defined as

$$\text{CoT} = \frac{\int_0^T P(t)\,dt}{mgd},$$

showing an average improvement of 18%. These findings align with prior reports on energy-efficient locomotion using reinforcement learning [**?**], [**?**].

Assuming a 20% reduction in torque variance σ_τ^2 between static PID and SAC-PID control, the proportional power reduction $P \propto \sigma_\tau^2(1-0.2)$ corresponds to an approximate 20% decrease in mechanical energy, consistent with the observed results.

Compared to static PID and DDPG-based tuning, SAC showed better sample efficiency and convergence stability due to its entropy-regularized framework.

V. Conclusion

Our work presents a *Soft Actor Critic (SAC) based adaptive PID control* framework that dynamically tunes PID gains to achieve energy-efficient and stable legged robot locomotion. By integrating classical control theory with deep reinforcement learning, the proposed method effectively reduces power consumption while maintaining trajectory accuracy and gait stability.

Future work will focus on deploying the framework on real quadruped hardware and extending it toward hybrid model-based reinforcement learning for enhanced generalization and robustness. Additionally, predictive stability estimation and hardware-in-the-loop training will be incorporated to enable real-time adaptive control across diverse terrains and payload variations.

References

[1] X. Fu, Z. Luo, and X. Huang, "Energy-efficient gait learning for quadrupedal robots using reinforcement learning," *IEEE Access*, vol. 9, pp. 129 174–129 185, 2021.

[2] R. Yan, H. Chen, and J. Wang, "Energy-efficient quadruped locomotion using deep deterministic policy gradient and twin delayed ddpg," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1458–1465, 2024.

[3] L. Zhu, M. Kim, and S. Oh, "Policy search transfer optimization for energy-efficient quadruped robot control," in *2022 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 5483–5490.

[4] Y. Shi, M. Sun, and Y. Liu, "Adaptive pid controller design using q-learning for nonlinear systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7160–7170, 2018.

[5] J. Lakhani, D. Patel, and A. Desai, "Deep deterministic policy gradient-based pid parameter tuning for industrial process control," *IEEE Access*, vol. 9, pp. 97 654–97 666, 2021.

[6] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018, pp. 1861–1870.

[7] J. Lee, K. Park, and S. Kim, "Energy-aware reinforcement learning for dynamic quadruped locomotion," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6578–6585, 2020.

[8] L. Zhu, Y. Qiao, and M. Sun, "Energy-aware gait optimization of quadruped robots using reinforcement learning," *IEEE Access*, vol. 10, pp. 71 220–71 231, 2022.

[9] J. Hwangbo, J. Lee, A. Dosovitskiy, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.

[10] P. De Jong, H. Kim, and L. Chen, "Entropy-regularized reinforcement learning for robust quadruped locomotion," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4450–4457.

[11] J. Park, R. Singh, and E. Lee, "Hts-sac: Hybrid time-series soft actor-critic for improved robot locomotion stability," *IEEE Robotics and Automation Letters*, 2025, in press.

[12] A. Kumar, F. Zhao, and R. Li, "Multi-critic soft actor-critic for efficient locomotion learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 10 587–10 599, 2024.

[13] M. Garcia, L. He, and R. Kumar, "A survey on reinforcement learning for pid tuning in industrial automation," *IEEE Access*, vol. 11, pp. 115 213–115 232, 2023.

[14] Y. Tang, H. Zhang, and Y. Liu, "Deep reinforcement learning for pid controller optimization: A comprehensive review," *Control Engineering Practice*, vol. 150, p. 105455, 2024.