

Quadrupedal gait on oscillating surfaces

Vierbeiniger Gang auf oszillierenden Oberflächen

Master thesis in the department of Computer Science by Arne Bick

Date of submission: May 12, 2025

1. Review: Maximilian Stasica
 2. Review: Nico Bohlinger
 3. Review: Jan Peters
- Darmstadt



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Erklärung zur Abschlussarbeit gemäß § 22 Abs. 7 APB TU Darmstadt

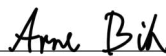
Hiermit erkläre ich, Arne Bick, dass ich die vorliegende Arbeit gemäß § 22 Abs. 7 APB der TU Darmstadt selbstständig, ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe. Ich habe mit Ausnahme der zitierten Literatur und anderer in der Arbeit genannter Quellen keine fremden Hilfsmittel benutzt. Die von mir bei der Anfertigung dieser wissenschaftlichen Arbeit wörtlich oder inhaltlich benutzte Literatur und alle anderen Quellen habe ich im Text deutlich gekennzeichnet und gesondert aufgeführt. Dies gilt auch für Quellen oder Hilfsmittel aus dem Internet.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Falle eines Plagiats (§ 38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei einer Thesis des Fachbereichs Architektur entspricht die eingereichte elektronische Fassung dem vorgestellten Modell und den vorgelegten Plänen.

Darmstadt, 12. Mai 2025


A. Bick

Abstract

Legged robots, especially quadrupeds, are increasingly deployed in unstructured environments where terrain instability is common. While reinforcement learning (RL) has enabled robust locomotion over uneven and laterally perturbed surfaces, the effects of vertical ground motion—a frequent real-world disturbance in settings such as bridges, platforms, and maritime environments—remain underexplored. This thesis addresses this gap by evaluating quadruped locomotion under vertical oscillations using a real-world experimental bridge with a dominant eigenfrequency of 2 Hz, designed to perturb locomotion. We use the Unitree Go2 quadruped robot, trained in simulation with the Proximal Policy Optimization (PPO) algorithm and the MuJoCo physics engine, to develop 18 locomotion policies. These policies span six gait patterns (default, trot, pace, bound, pronk, free) and three training regimes: a static surface and two oscillating bridge setups differing in their height regulation strategy (relative to ground or bridge surface). Domain randomization enables zero-shot policy transfer to the physical system. Our results show that policies trained under oscillating conditions exhibit significantly improved stability and robustness during real-world deployment. Furthermore, we introduce set of evaluation metrics beyond reward functions, focusing on height control, gait consistency, and adaptability. The findings underscore the importance of including dynamic terrain perturbations during training and highlight the potential of RL-based controllers for resilient locomotion in vertically unstable environments.

Zusammenfassung

Roboter mit Beinen, insbesondere Vierbeiner, werden zunehmend in unstrukturierten Umgebungen eingesetzt, in denen instabiles Terrain häufig vorkommt. Während Reinforcement Learning (RL) bereits robuste Fortbewegung auf unebenen und lateral gestörten Oberflächen ermöglicht hat, sind die Auswirkungen vertikaler Bodenbewegungen – eine häufige reale Störung in Umgebungen wie Brücken, Plattformen und maritimen Bereichen – noch wenig erforscht. In dieser Arbeit wird diese Lücke behandelt, indem die Fortbewegung von Vierbeinern unter dem Einfluss vertikaler Schwingungen untersucht wird. Hierfür wird eine reale experimentelle Brücke mit einer dominanten Eigenfrequenz von 2 Hz verwendet, die zur Störung der Fortbewegung entwickelt wurde. Für die Experimente kommt der vierbeinige Roboter Unitree Go2 zum Einsatz, der in Simulation mit dem Proximal Policy Optimization (PPO) Algorithmus in der MuJoCo-Physik-Engine trainiert wurde, um 18 Fortbewegungsstrategien zu entwickeln. Diese umfassen sechs Gangarten (default, trot, pace, bound, pronk, free) sowie drei Trainingsstile: eine statische Oberfläche und zwei oszillierende Brückenkonfigurationen, die sich in ihrer Höhenregulierungsstrategie (relativ zum Boden oder zur Brückenoberfläche) unterscheiden. Die Randomisierung der Trainingsumgebungen ermöglicht einen Zero-Shot-Transfer der Strategien auf das physikalische System. Unsere Ergebnisse zeigen, dass die unter oszillierenden Bedingungen trainierten Policies eine deutlich verbesserte Stabilität und Robustheit im realen Einsatz aufweisen. Darüber hinaus führen wir einige Bewertungsmetriken ein, die über die Reward-Funktionen hinausgehen und sich auf Höhenkontrolle, Ganggleichförmigkeit und Anpassungsfähigkeit konzentrieren. Die Resultate verdeutlichen die Bedeutung der Einbeziehung dynamischer Geländestörungen im Training und unterstreichen das Potenzial von RL-basierten Steuerungen für robuste Fortbewegung in vertikal instabilen Umgebungen.

Contents

1. Introduction	1
1.1. Motivation	2
2. Foundations	4
2.1. Locomotion Learning	4
2.2. Gaits	5
2.3. MuJoCo	6
2.4. Reinforcement Learning	7
2.5. HUMVIB Bridge	9
2.6. Harmonic Oscillator	10
2.7. Unitree Go2	11
3. Methodology	12
3.1. Training Setup	12
3.2. Bridge Model	13
3.3. Learning Distinct Gaits	14
3.4. Gait Verification and Simulation Testing	17
3.5. Real-World Experiment	18
4. Results	20
4.1. Evaluation of Gaits	20
4.2. Real-World Experiment	26
5. Discussion	30
5.1. Gait Learnability and Training Biases	30
5.2. Policy Performance in Simulation	31
5.3. Evaluation of Real-World Experiments	32

6. Conclusion	35
6.1. Outlook	35
A. Appendix	43

Figures, Tables and Abbreviations

List of Figures

2.1. The agent–environment interaction in a reinforcement learning setup. The agent receives observations, selects actions, and receives rewards from the environment.	7
2.2. Schematic of the Unitree Go2 quadruped crossing the HUMVIB bridge, which measures 13.24 m in length and 2.5 m in width. The coordinate systems used for both the bridge and the robot’s motion are also shown. .	10
2.3. Schematic of the Unitree Go2	11
3.1. Screenshot of the bridge model in MuJoCo during testing, featuring the Unitree Go2 walking on the oscillating surface.	14
3.2. Example of characteristic footfall patterns for the <i>trot</i> , <i>pace</i> , <i>bound</i> and <i>pronk</i> gaits that impose no penalties. The robot’s feet are labeled as front left (<i>FL</i>), front right (<i>FR</i>), rear left (<i>RL</i>) and rear right (<i>RR</i>).	16
3.3. Unitree Go2 traversing the HUMVIB bridge	18
4.1. Average episode return for the different gaits (left) and height regulation styles (right) during training.	21
4.2. Average episode return over the command velocity on idle (top) and on the oscillating bridge (bottom) for the different gaits (left) and training conditions (right) of the policies during evaluation.	22

4.3. Footfall pattern of the different gaits using the <i>nos</i> style on the idle bridge with a target speed $v_x = 0.5$ m/s. The feet of the robot are denoted with front left (<i>FL</i>), front right (<i>FR</i>), rear left (<i>RL</i>) and rear right (<i>RR</i>). A characteristic gait stance phase is detected when all four feet simultaneously exhibit one of the ideal contact combinations defined in Figure 3.2.	23
4.4. Average and standard deviation of the robot’s CoM height for all gait–style combinations on the idle bridge (top) and the oscillating bridge (bottom), using HUMVIB test conditions.	25
4.5. Means and standard deviations of the force readouts of the foot sensors of the robot over gaits, styles and bridge-setting in the real world.	27
4.6. PSD of the <i>default eb</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	28
A.1. Episode return over the command velocity on idle bridge for all gait-style combinations.	46
A.2. Episode return over the command velocity on oscillating bridge, using HUMVIB settings, for all gait-style combinations.	47
A.3. PSD of the <i>default eb</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	51
A.4. PSD of the <i>default eg</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	52
A.5. PSD of the <i>default nos</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	53

A.6. PSD of the <i>trot eb</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	54
A.7. PSD of the <i>trot eg</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	55
A.8. PSD of the <i>trot nos</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	56
A.9. PSD of the <i>bound eb</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	57
A.10. PSD of the <i>bound nos</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	58
A.11. PSD of the <i>free eb</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	59
A.12. PSD of the <i>free eg</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	60
A.13. PSD of the <i>free nos</i> policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.	61

List of Tables

3.1. Reward terms and coefficients that make up the reward function. This symmetry penalty is applied only with the <i>default</i> gait.	15
4.1. Percentage of time the <i>nos</i> policies exhibited their characteristic gait phases for each defined gait on rigid ground. The Free (%) gait check is omitted, as it imposes no symmetry constraints. Pronk_g (%) and Pronk_a (%) indicate the proportion of time all feet were on the ground and all feet were in the air, respectively.	24
4.2. Mean estimated power averaged over gaits, styles and bridge-setting . . .	26
A.1. The PPO hyperparameters used for training.	44
A.2. Domain randomization parameters and ranges	45
A.3. Gait detection percentages across all policies on oscillating and idle bridge.	48
A.4. Complete statistics of power usage, standard deviation (Std) for each gait, style, and setting (Set)	49
A.5. Complete statistics of average combined foot force, standard deviation (Std) for each gait, style, and setting (Set)	50

List of Abbreviations

Notation	Description
CoM	Center of Mass
DRL	Deep Reinforcement Learning
FFT	Fast Fourier Transform
FT	Fourier Transform
GRF	Ground Reaction Force
HUMVIB	HUMAN-structure interaction and gait adaptation during locomotion on VIBrating structures
IMU	Inertial Measurement Unit
MDP	Markov Decision Process
MuJoCo	Multi-Joint dynamics with Contact
PPO	Proximal Policy Optimization
PSD	Power Spectral Density
RL	Reinforcement Learning
ROS2	Robot Operating System

1. Introduction

Legged robots and quadrupeds in particular, have emerged as promising systems for navigating unstructured and complex environments, including rough outdoor terrain, disaster zones and human-built spaces. Their ability to maintain mobility where wheeled or tracked robots fail makes them ideal for applications such as exploration, inspection and rescue missions. Recent advancements have demonstrated remarkable agility in handling obstacles, stairs and irregular terrain [1, 2, 3]. These developments have been fueled by improvements in both mechanical design and control algorithms, allowing robots to perform highly dynamic motions.

Despite these successes, the majority of current locomotion research focuses on scenarios where the ground is either static or changes in a predictable way. Environments with active ground perturbations—such as vibrating platforms, swinging bridges, or machinery-mounted surfaces—pose a different challenge. These situations are common in real-world applications, yet remain underexplored in robotics. While some studies incorporate variations in ground stiffness [4] or surface friction [5, 6], few consider the implications of vertical ground movements. The same is true in the field of human locomotion, where such perturbations are known to significantly affect gait dynamics [7], yet are difficult to replicate in controlled experimental settings. The scarcity of work in this area stems in part from the technical challenges of replicating vertical oscillations in a safe and repeatable way. Furthermore, conventional control methods, including those used in many commercial quadrupeds, are typically tuned for stable or slowly varying terrain and struggle to cope with sudden vertical displacements. Humans and some animals, in contrast, adjust almost reflexively to changes in ground level or compliance, employing strategies such as modulating leg stiffness and timing [8, 9, 10]. Replicating this kind of adaptive behavior in robots remains a major open problem.

In response, Reinforcement Learning (RL) has emerged as a powerful approach to acquire adaptive and robust controllers for legged locomotion. RL treats locomotion as a sequential decision-making problem, allowing robots to learn complex motor skills by interacting

with simulated environments. Algorithms such as Proximal Policy Optimization (PPO) [11] have been used to train policies capable of agile running [6], high jumps [12], parkour-like movements [13] and even acrobatics [14]. Domain randomization techniques further support the sim-to-real transfer by injecting variability in the training environment, covering factors such as sensor noise, latency and external perturbations [15, 16].

However, a shortcoming in most RL-based locomotion research is the omission of active ground dynamics. While many policies are trained to handle uneven terrain or lateral disturbances such as random side pushes [17, 18], vertical ground motion—a common and impactful real-world phenomenon—remains largely unaddressed in both training and evaluation. This omission is especially significant given that many real-world structures, such as bridges, elevated platforms and ships, frequently exhibit vertical oscillations caused by human activity, mechanical systems, or environmental forces [19, 20]. As a result, there is a substantial research gap in understanding how learned locomotion policies generalize to environments where vertical instability is a dominant factor, hindering their reliability in critical applications such as disaster response, industrial inspection and maritime operations.

1.1. Motivation

To address this, the present thesis explores how different quadruped gaits and training regimes perform under vertical ground oscillations. Specifically, we study the Unitree Go2 quadruped (Unitree, Hangzhou, China) walking on a purpose-built oscillating bridge, developed within the HUMan-structure interaction and gait adaptation during locomotion on VIBrating structures (HUMVIB) project. This custom structure, composed of steel beams and concrete slabs, has a dominant eigenfrequency of approximately 2 Hz, making it susceptible to resonance effects during locomotion [21, 22].

The primary aim is to evaluate how gait selection, timing and footfall patterns influence stability and performance under vertical excitation. Additionally, we assess whether the robot passively adapts its posture or if active height regulation is needed. Another goal is to identify training strategies that lead to more robust and transferable policies and to explore whether simulation-trained behaviors remain effective on real-world oscillating structures.

To support this, we developed methods for generating distinct gait types and evaluation metrics both in simulation and reality. We propose novel performance metrics tailored to

vertically dynamic terrain, moving beyond standard reward-based measures to capture height control and locomotion stability.

Ultimately, this work contributes to a deeper understanding of adaptive locomotion on unstable surfaces and offers practical strategies for improving real-world quadruped deployment in environments where traditional assumptions of ground stability no longer apply.

2. Foundations

This chapter lays the theoretical groundwork for our study by introducing key concepts in legged locomotion, control strategies and simulation tools. It begins with an overview of locomotion learning and gait types, then covers the reinforcement learning framework and simulation environment used to train policies. Finally, it introduces the experimental infrastructure, including the oscillating HUMVIB bridge, the harmonic oscillator model and the physical robot.

2.1. Locomotion Learning

Legged locomotion is a fundamental capability in mobile robotics, requiring coordination of multiple joints to produce stable and efficient movement. Traditional control approaches often rely on analytic models of dynamics and kinematics [23]. While these can work well in structured environments, they struggle with real-world complexity, such as uneven terrain or unexpected disturbances.

To address these limitations, biologically inspired methods have been explored. Animals exhibit remarkably adaptive and efficient locomotion across diverse conditions. Controllers based on central pattern generators, which mimic neural circuits producing rhythmic motion, have been widely used in robotics to generate stable gaits with minimal sensory input [24]. Similarly, reflex-based control and neuromechanical models offer reactive behavior and robustness, drawing on insights from biological locomotion systems [25].

Recent years have seen a growing shift towards learning-based approaches, which allows robots to autonomously discover control strategies through data. These include optimization-based methods [26], imitation learning [27] and increasingly RL, which frames control as a sequential decision-making problem. RL has shown success in learning

robust locomotion policies that generalize across tasks and environments without requiring precise system models [28, 29]. A notable line of work demonstrates the zero-shot transfer of proprioceptive policies, trained solely in simulation, that are capable of blindly traversing unstructured real-world terrains, such as snow, rubble and dense vegetation [30]. This is made possible by onboard terrain mapping based exclusively on inertial and kinematic sensing [31]. Together, these advances underscore the growing importance of RL and learning-based perception in enabling legged robots to operate autonomously in complex, unpredictable environments.

2.2. Gaits

Quadrupedal robots and animals utilize different gaits depending on their speed, terrain and stability requirements. Gaits are defined by the sequence and timing of footfalls, which influence the efficiency, stability and adaptability of movement. Among the common gaits used in both biological and robotic quadrupeds are the trot, pace, bound and pronk. Each of these gaits has distinct advantages depending on the context of locomotion. In robotics, selecting an appropriate gait involves trade-offs between speed, stability and energy efficiency, making gait optimization an essential aspect of quadrupedal robot control.

2.2.1. Trot

The trot is a diagonally synchronized gait where the front-left and rear-right legs move together, alternating with the front-right and rear-left legs. This gait is widely used for efficient locomotion at moderate speeds and provides good stability due to the alternating diagonal support. Trot is commonly observed in horses, dogs and robotic quadrupeds designed for stable traversal over varied terrain [32]. It minimizes vertical oscillations and is often preferred for energy-efficient locomotion in robotic applications [33].

2.2.2. Pace

In a pacing gait, legs on the same side move together, creating a lateral sway but enabling smoother motion [34]. Common in camels and some horse breeds, it suits fast, energy-efficient travel on flat terrain [35]. While pacing reduces vertical energy loss

and prevents hind-foot interference—a benefit for long-legged animals—it compromises stability, especially in short-legged ones. Not all long-legged animals pace, but all animals that do are long-legged [36].

2.2.3. Bound

Bounding is a high-speed gait where the forelimbs move together followed by the rear limbs moving together. This gait is typical of small mammals like rabbits and certain carnivores when accelerating or chasing prey [37]. In robotics, bounding is used for fast locomotion, particularly when stability is less critical than speed [32]. This gait is often employed in agile robotic systems designed for dynamic movement, such as navigating obstacles or rapid maneuvers in unstructured environments.

2.2.4. Pronk

The pronking gait involves all four legs leaving and touching the ground simultaneously. This gait is seen in animals such as gazelles and springboks, often as a display of strength or agility [38]. In robotics, pronking is useful for testing symmetrical force distribution and for applications requiring sudden jumps or bounces, such as traversing gaps or negotiating soft terrain.

2.3. MuJoCo

MuJoCo (Multi-Joint dynamics with Contact) is a physics engine designed for fast and accurate simulation of robotic systems. Its strength lies in simulating complex interactions between rigid bodies, particularly in tasks involving contact dynamics such as locomotion and manipulation. MuJoCo provides soft contact modeling, inverse dynamics, and real-time simulation, enabling researchers to learn and test control strategies before deploying them on real robots [39].

2.4. Reinforcement Learning

RL is a branch of machine learning concerned with how agents can learn to make optimal decisions, with respect to the rewards, by interacting with their environment to achieve long-term goals. This process is typically modeled as a Markov Decision Process (MDP), which provides a formal framework for decision-making under uncertainty, where outcomes are influenced both by the agent's actions and by chance.

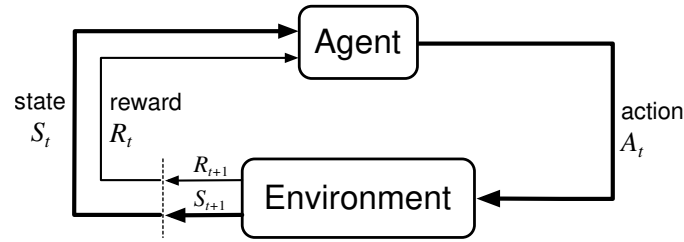


Figure 2.1.: The agent–environment interaction in a reinforcement learning setup. The agent receives observations, selects actions, and receives rewards from the environment.

As illustrated in Figure 2.1 and described by Sutton and Barto [40], the agent and environment interact over a sequence of discrete time steps t . At each time step, the agent observes the current state $S_t \in \mathcal{S}$, where \mathcal{S} is the set of all possible states. Based on this observation, it selects an action $A_t \in \mathcal{A}(S_t)$, drawn from the set of valid actions in that state $\mathcal{A}(S_t)$. As a result of the action, the agent receives a scalar reward $R_{t+1} \in \mathcal{R}$ and transitions to a new state S_{t+1} , completing one cycle of interaction.

The agent's behavior is defined by a policy $\pi_\theta(a|s)$, which maps states to probabilities of selecting each action. The goal of reinforcement learning is to find the policy that maximizes the expected return:

$$J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \right], \quad (2.1)$$

where $\gamma \in [0, 1)$ is a discount factor that determines the importance of future rewards. The optimal policy is then defined as:

$$\pi^* = \arg \max_{\pi} J(\pi). \quad (2.2)$$

Effective learning requires balancing exploration of unknown strategies with exploitation of known ones that yield high rewards.

In robotic locomotion, for example, RL enables agents to learn how coordinated joint movements impact stability, energy efficiency, or speed. These objectives are encoded in the reward function, guiding the robot to discover locomotion strategies that would be difficult to derive analytically.

2.4.1. Proximal Policy Optimization

To optimize the policy, this work uses the PPO algorithm [11]. PPO improves the stability of policy updates by introducing a clipped surrogate objective, which constrains the difference between new and old policies during training. The central idea is to avoid overly large policy updates that could destabilize learning, while still enabling improvement.

The standard objective in policy gradient methods is to maximize the expected advantage-weighted probability of actions under the current policy:

$$L^{\text{PG}}(\theta) = \mathbb{E}_t \left[\pi_\theta(a_t|s_t) \hat{A}_t \right], \quad (2.3)$$

where \hat{A}_t is an estimator of the advantage function at time step t . This advantage function measures how much better an action is compared to the average, helping the agent to favor more promising actions.

In practice, this expression is reformulated using importance sampling to correct for the fact that data was collected under an earlier policy $\pi_{\theta_{\text{old}}}$. The probability ratio

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (2.4)$$

is introduced to reweight the objective. The PPO algorithm then applies clipping to this ratio to constrain updates:

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right], \quad (2.5)$$

where ϵ is a small positive constant. This clipped objective discourages large policy updates, maintaining a balance between policy improvement and stability. By doing so, PPO achieves a strong performance across a variety of RL tasks, including high-dimensional control problems such as robotic locomotion.

2.4.2. RL-X

RL-X [41] is an open-source general-purpose Deep Reinforcement Learning (DRL) library used in this work for training locomotion policies. Originally developed for RoboCup simulation leagues, RL-X is designed to be modular making it suitable for a broad range of robot learning tasks. Each algorithm and environment is implemented in a self-contained directory, which promotes rapid prototyping, customization of robot models and easy experimentation with different reward structures. We are using RL-X in conjunction with the Multi-Joint dynamics with Contact (MuJoCo) physics engine to simulate dynamic locomotion across varied gaits and environments. The training relies on the PPO algorithm, with mini-batch gradient descent applied over trajectories collected through environment rollouts. RL-X also integrates with Weights & Biases, enabling comprehensive logging, visualization, and monitoring of experiment progress.

2.5. HUMVIB Bridge

Bridges, like all flexible structures, can oscillate when subjected to dynamic loading from environmental forces such as wind, traffic, or pedestrians. These oscillations occur in distinct mode shapes, each characterized by a specific frequency and deformation pattern. The most relevant types for pedestrian-induced vibrations are lateral bending, vertical bending and torsional modes. In lateral bending, the bridge sways horizontally side-to-side; in vertical bending, the deck moves up and down along its span; and in torsional modes, the structure twists around its longitudinal axis, with opposite sides of the deck moving in opposing vertical directions. The excitation of these modes—especially those with frequencies near human walking—can strongly influence gait and stability.

The HUMVIB bridge [42] at the Technical University of Darmstadt is a purpose-built foot-bridge designed to study such human–structure interactions. The bridge spans 13.24 m, weighs approximately 12.2 t and consists of two steel beams supporting a 13-part segmented concrete deck. The segments are separated by 2 cm gaps, allowing for significant structural flexibility. Its dynamic behavior has been thoroughly characterized and includes multiple modal responses: the first lateral bending mode at 1.63 Hz, the first vertical bending mode at 2.04 Hz and the first torsional mode at 4.09 Hz. These are followed by higher-order modes including the second lateral bending mode at 4.96 Hz, the second vertical bending mode at 7.80 Hz. Among these, the first vertical mode at 2.04 Hz is of particular importance, as it lies close to the natural frequency of human walking and is

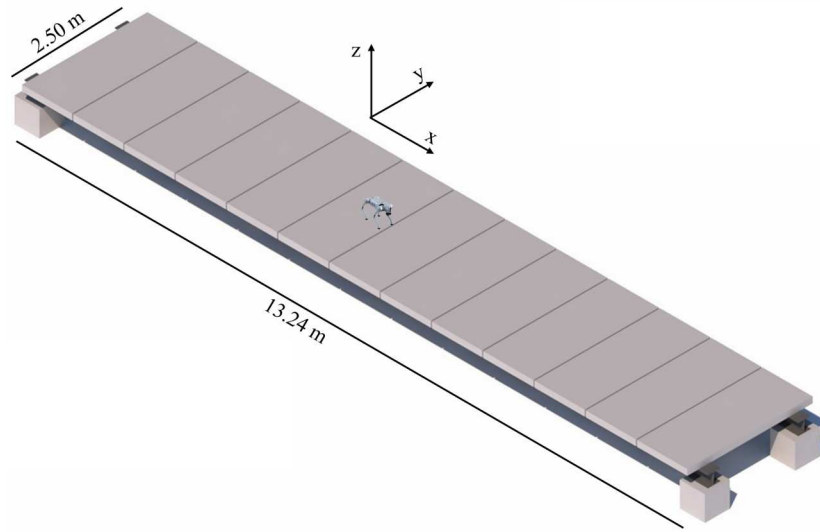


Figure 2.2.: Schematic of the Unitree Go2 quadruped crossing the HUMVIB bridge, which measures 13.24 m in length and 2.5 m in width. The coordinate systems used for both the bridge and the robot's motion are also shown.

thus the most likely to be excited by pedestrians. Equipped with comprehensive instrumentation to measure displacements, Ground Reaction Forces (GRFs), and accelerations with Inertial Measurement Units (IMUs), the HUMVIB bridge serves as an ideal platform for analyzing gait adaptation, stability, and control strategies under real-world oscillatory conditions.

2.6. Harmonic Oscillator

Harmonic oscillators provide a fundamental model for analyzing systems that exhibit periodic motion around an equilibrium point [43]. They are widely used to approximate various physical systems, including aspects of human gait [44, 45] and serve as a natural representation for the oscillatory behavior of flexible structures such as bridges.

An one dimensional mass-spring harmonic oscillator is a system that experiences a restoring force proportional to its displacement from equilibrium, resulting in sinusoidal motion governed by the system's mass m and stiffness k . In this common example the force follows

Hooke's Law [46] while damping or external forces influence its real-world behavior. The eigenfrequency f_n of such a system is determined by its mass and stiffness, with the angular frequency ω_n related to the frequency by $\omega_n = 2\pi f_n$ and to the system parameters by $\omega_n^2 = k/m$. From these relationships, the stiffness k can be calculated as:

$$k = (2\pi f_n)^2 m \quad (2.6)$$

Additionally, the maximum acceleration a_{\max} , which occurs at the equilibrium point, is determined by the amplitude A and angular frequency ω_n :

$$a_{\max} = A \omega_n^2 \quad (2.7)$$

Rearranging this expression allows the amplitude to be computed based on a known acceleration, mass and stiffness:

$$A = \frac{a_{\max} m}{k} \quad (2.8)$$

This set of equations enables the design of harmonic oscillators with specified dynamic characteristics.

2.7. Unitree Go2

The Unitree Go2 EDU (Unitree, Hangzhou, China), shown in Figure 2.3, is a versatile quadrupedal robot designed for research in locomotion, perception and autonomous navigation. Weighing approximately 15 kg and measuring 70 cm \times 31 cm \times 40 cm when standing, it features 12 high-performance actuators with integrated torque sensing, capable of delivering up to 45 N m of torque. The robot is equipped with an IMU, foot-end force sensors and a 3D LIDAR for perception. Its standard software supports multiple gait modes and enables dynamic locomotion at speeds up to 3.7 m/s. Its Robot Operating System (ROS2) compatible software architecture and integrated sensors make it a flexible platform for advanced robotics development.

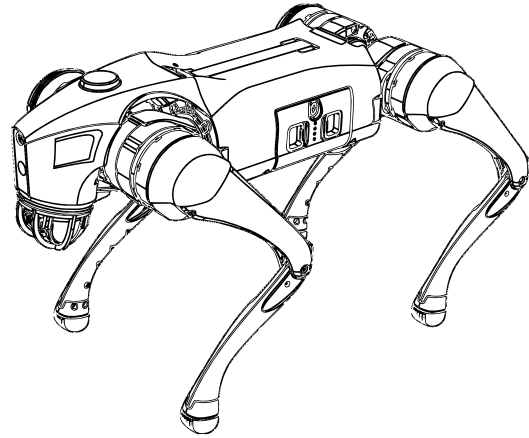


Figure 2.3.: Schematic of the Unitree Go2

3. Methodology

This chapter presents the methodology used to develop and evaluate robust locomotion policies for a quadruped robot operating on a vertically oscillating bridge. Our goal is to investigate how periodic vertical disturbances influence gait stability and to train control strategies that maintain reliable locomotion under such challenging conditions. We begin by detailing the RL setup and the simulation model of the bridge. Then, we explain how different gaits were trained under varying height regulation strategies and conclude with an overview of the real-world testing and evaluation procedure.

3.1. Training Setup

RL-X was used to train a range of distinct gait policies for the Unitree Go2, to which the framework had not previously been applied. Training was conducted in simulation using the CPU-based MuJoCo physics engine, with 48 parallel environments enabling efficient data collection. The PPO algorithm was used to learn locomotion policies, following established approaches in prior work [16, 5, 47]. The full table of hyperparameters is shown in Table A.1. The entire pipeline was implemented using the RL-X deep reinforcement learning framework, integrated with the MuJoCo simulation environment. The policies were trained to generate target joint positions that enable tracking of a commanded velocity vector $\bar{v} \in [-1.0, 1.0]^3$, representing desired forward, lateral, and yaw velocities in the robot’s local frame. The robot is controlled using a joint-level PD controller running at 50 Hz, with fixed gains $k_p = 20$ and $k_d = 0.5$. The controller sets the target joint velocity to zero and computes the target joint positions as $q^{\text{target}} = q^{\text{nominal}} + \sigma_a a$, where q^{nominal} is the nominal joint position, a is the action produced by the policy and $\sigma_a = 0.3$ is a scaling factor. The observation space is composed of three categories: joint-specific, feet-specific and general observations. Joint-specific inputs o_j include joint positions q , velocities \dot{q} and previous actions a_{t-1} . Feet-specific inputs o_f consist of binary contact states p_f and the

time since the last contact p_f^T for each foot. General observations o_g include the robot’s linear velocity v , angular velocity ω , command velocity \bar{v} , gravity vector orientation g , height above ground h , PD control parameters, scaling factor σ_a , total mass m and physical dimensions (length, width and height). All features are concatenated into a single observation vector $o = [o_j, o_f, o_g]$ and normalized to the range $[-1, 1]$. To facilitate zero-shot sim-to-real transfer, extensive domain randomization was applied throughout training. Parameters such as body mass, inertia, Center of Mass (CoM) location, actuator latency and dynamics, ground friction, compliance, sensor noise, and external perturbations were randomized to improve policy robustness under real-world variability. The whole table of randomized parameters and their ranges is shown in Table A.2. To assess the robustness of trained policies under challenging terrain, a simulation of a dynamically oscillating bridge was developed based on the HUMVIB bridge. Training scenarios included both rigid and vertically oscillating ground, with variations in oscillation amplitude and eigenfrequency, which was realized as a part of the domain randomization.

3.2. Bridge Model

The HUMVIB bridge is modeled in MuJoCo as a single-degree-of-freedom harmonic oscillator (Figure 3.1) that emulates the dynamics of its first vertical bending mode, as this mode is the most readily excited by human locomotion. The start position of the bridge surface is set to 1.05 m over the ground—the peak of the oscillation—while the equilibrium position can be adjusted downward to modify the oscillation amplitude. This allows to keep the robots starting position at a fixed point in space to always have the same start height over the surface. The downward disposition of the equilibrium position represents the maximum amplitude A . To fully describe the system we need the position in respect to the equilibrium the stiffness k and the mass of the bridge. To better approximate the dynamics of the HUMVIB bridge, the model mass m was set to half the actual bridge mass, resulting in a value of 6100 kg. This adjustment reflects the lower effective mass of the real structure—not all portion of the bridge are equally involved in the oscillatory motion—compared to an idealized harmonic oscillator. This behavior can be more accurately described using catenary theory as a basis for the bridge’s motion equation [48]. With these parameters, a wide set of different oscillations can be assigned to the bridge. For a chosen frequency the stiffness is calculated via Equation 2.6.

To emulate the HUMVIB bridge the stiffness is tuned to $k = 963\,273 \text{ N/m}$ such that the bridge exhibits an eigenfrequency of 2 Hz with an oscillation amplitude of $\pm 0.05 \text{ m}$. During

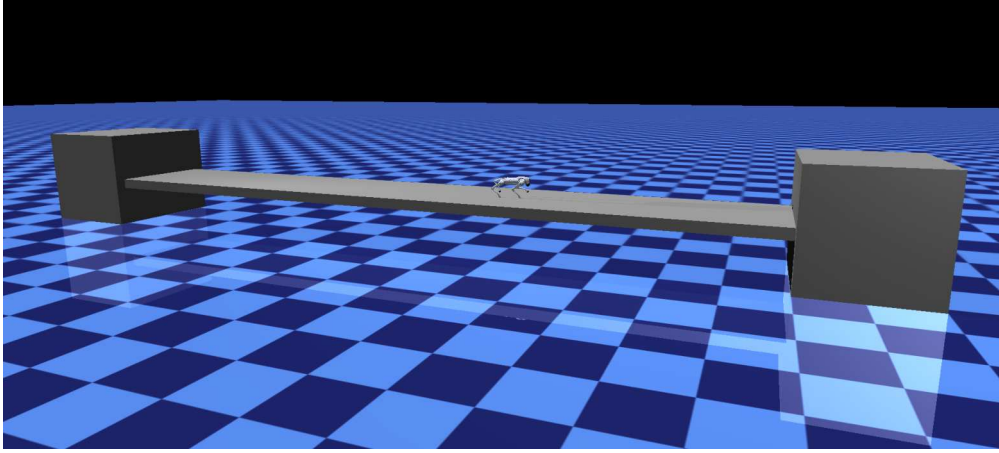


Figure 3.1.: Screenshot of the bridge model in MuJoCo during testing, featuring the Unitree Go2 walking on the oscillating surface.

training, we varied the eigenfrequency of the bridge between 0.75 Hz and 7.5 Hz and the amplitude between zero and a constrained maximum value that ensures the bridge’s acceleration remains below $a_{\max} = 9.81 \text{ m/s}^2$ which is derived via Equation 2.8 given the mass and stiffness. This constraint is necessary to prevent the robot from experiencing too much acceleration to become airborne.

3.3. Learning Distinct Gaits

The same reward terms, coefficients r_c , and training curriculum as in [47] were used to train the different gaits, while the learning progress was monitored using the wandb.com platform. In particular, the reward structure for the *default* gait, as shown in Table 3.1, was trained with minimal modifications—only those necessary to adapt to the altered training environment. The reward function consists of tracking terms that encourage following the commanded velocities \bar{v} , as well as multiple penalty terms to shape the gait behavior. All terms are scaled by individual coefficients, summed and clipped to ensure the total reward remains non-negative. The general reward structure consists of these components:

- **Tracking velocity rewards:** Encourage following the desired velocity commands for both linear (x and y) and angular (yaw) directions.

- **Joint penalties:** Penalize deviations from position limits, acceleration and torque.
- **Action rate penalty:** Penalizes rapid changes between consecutive actions.
- **Collision penalty:** Discourages collisions with the environment.
- **Base height penalty:** Encourages the robot’s trunk to stay close to the desired height.
- **Air time penalty:** Penalizes excessive air time, encouraging foot contact with the ground.
- **Symmetry penalty:** Discourages asymmetric foot placement during movement.

Term	Equation	Reward coefficient r_c
XY velocity tracking	$\exp(- v_{xy} - \bar{v}_{xy} ^2/0.25)$	2
Yaw velocity tracking	$\exp(- \omega_{yaw} - \bar{v}_{yaw} ^2/0.25)$	1
Z velocity penalty	$- v_z ^2$	2
Pitch-roll velocity penalty	$- \omega_{pitch, roll} ^2$	0.05
Pitch-roll position penalty	$- \theta_{pitch, roll} ^2$	0.2
Joint limits penalty	$-\mathbb{1}(0.9 q_{\min} < q < 0.9 q_{\max})$	10
Joint accelerations penalty	$- \ddot{q} ^2$	2.5e-7
Joint torques penalty	$- \tau ^2$	2e-4
Action rate penalty	$- \dot{a} ^2$	0.01
Base height penalty	$- h - h_{\text{nominal}} ^2$	30
Collisions penalty	$-n_{\text{collisions}}$	1
Air time penalty	$-\sum_f \mathbb{1}(p_f)(p_f^T - 0.5)$	0.1
Symmetry penalty	$-\sum_f \mathbb{1}(p_f^{\text{left}})\mathbb{1}(p_f^{\text{right}})$	0.5

Table 3.1.: Reward terms and coefficients that make up the reward function. This symmetry penalty is applied only with the *default* gait.

To encourage the emergence of distinct gaits—*trot*, *pace*, *bound* and *pronk*—the existing symmetry penalty term was adapted to penalize deviations from the characteristic stance phases associated with each gait, as illustrated in Figure 3.2. The reward function for the *default* gait applies fewer constraints, primarily encouraging at least two feet to remain

in contact with the ground while still discouraging bound behavior. Finally, setting the symmetry penalty to zero results in the unconstrained *free* gait.

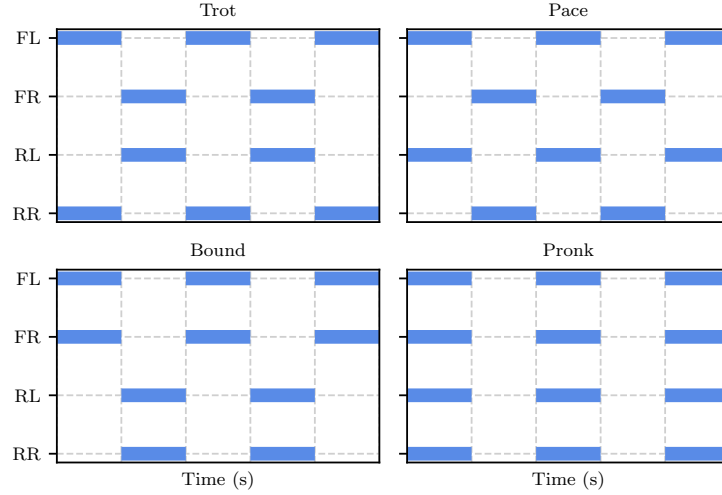


Figure 3.2.: Example of characteristic footfall patterns for the *trot*, *pace*, *bound* and *pronk* gaits that impose no penalties. The robot’s feet are labeled as front left (*FL*), front right (*FR*), rear left (*RL*) and rear right (*RR*).

Beyond gait-specific penalty terms, the training environment plays a crucial role in shaping the resulting locomotion behavior. For gaits trained on an oscillating bridge, two distinct policies were developed by modifying the base height penalty term, which encourages the robot to maintain a constant height of 0.325 m—its nominal trunk height in a standing posture—relative to the walking surface. In the *equidistant bridge* (*eb*) condition, the robot is rewarded for maintaining a constant height relative to the oscillating bridge surface. In contrast, the *equidistant ground* (*eg*) condition encourages the robot to maintain a constant height relative to the fixed ground beneath, which is expected to result in a more dampened gait. When training on a rigid, non-oscillating bridge, the height penalty has a single interpretation—height changes relative to the ground and the bridge surface are always equal—referred to as *no oscillation* (*nos*).

This setup yields 18 distinct policies, comprising six gaits (*default*, *trot*, *pace*, *bound*, *pronk*, *free*), each trained with two height regulation strategies on an oscillating bridge (*eb*, *eg*) and a baseline condition on a rigid bridge (*nos*).

The base height is formulated as a penalty term, which the learning algorithm seeks to

minimize. This penalty is defined by the deviation between the nominal trunk height $h_{\text{nominal}} = 0.325$ m and the vertical position of the robot's CoM, denoted as h . In simulation this vertical position is influenced by the bridge model and its oscillations, resulting in a trunk coordinate $\text{CoM}_{\text{Go}2_z}(t)$ at discrete time step t .

The three height regulation styles *eb*, *eg* and *nos* introduce slight variations in the computation of the base height h , depending on the modeled bridge. These variations are influenced by the initial bridge height of $b_0 = 1.05$ m, the instantaneous vertical bridge position $b_z(t)$ and physical parameters such as the gravitational constant g , the bridge's mass m , stiffness k and oscillation amplitude A . The new formulation of the base height are:

$$h_{eb}(t) = \text{CoM}_{\text{Go}2_z}(t) - b_0 - b_z(t) \quad (3.1)$$

$$h_{eg}(t) = \text{CoM}_{\text{Go}2_z}(t) - b_0 + \frac{g m}{k} - \frac{A}{2} \quad (3.2)$$

$$h_{nos}(t) = \text{CoM}_{\text{Go}2_z}(t) - b_0 \quad (3.3)$$

Embedding these base height terms within the reward functions is intended to encourage the development of distinct walking styles.

3.4. Gait Verification and Simulation Testing

The learning performance of the 18 gait styles was analyzed by reviewing the learning returns and verifying the emergence of the desired gait patterns in simulation. To assess gait correctness, footfall patterns were visualized for each learned policy.

In simulation, the return was evaluated using a reward function that omits the symmetry and base height terms for comparability. It was assessed under command velocities ranging from 0 m/s to 0.7 m/s in 0.05 m/s increments, across both rigid and oscillating surfaces, replicating the test conditions of the HUMVIB bridge. For a representative target velocity—which was later used in the real world experiment—of 0.5 m/s, the trajectory of the trunk's CoM was recorded to evaluate lateral drift, compliance with the commanded velocity and the ability to adapt to vertical oscillations.

Additionally, the conformity of the footfall pattern with the reward function used during training was quantified as a percentage at a commanded velocity of 0.5 m/s, both on rigid ground and on the oscillating bridge under HUMVIB settings. This measure indicates not

only whether a gait has been successfully learned, but also to what extent, providing a basis for assessing the similarity between different gaits.

3.5. Real-World Experiment

The real-world experiments were conducted on the HUMVIB bridge (Figure 3.3). To induce oscillations, the bridge was manually excited at 2 Hz by the experimenters until the desired amplitude was reached, after which it was maintained at a stable level throughout the trials. For each combination of gait and training style, the robot completed eight trials across both the pre-oscillated and idle bridge configurations. The command velocity \bar{v} , manually controlled by another experimenter, was limited to a maximum of 0.5 m/s in the x-direction to maintain a consistent forward speed. The operator could additionally adjust the y- and yaw-velocities as needed to compensate for lateral drift and keep the robot on track. Comprehensive onboard data was recorded using ROS2 for subsequent analysis. This included full IMU measurements, joint positions, velocities, and torques, as well as data from the foot pressure sensors. Joint velocities and torques were used to estimate power consumption over time, while foot pressure data were analyzed to determine which gait-style combinations exhibited softer contact with the bridge.



Figure 3.3.: Unitree Go2 traversing the HUMVIB bridge

3.5.1. Data processing and power estimation

The recorded sensor data was screened for anomalies and segmented into individual bridge traversal trials using the forward walking control signal as a temporal reference. To evaluate the interaction between the robot and the bridge structure, the mean GRFs at the feet were computed for each trial.

For a more detailed analysis, the estimated power exerted by all four legs was calculated per trial using:

$$\text{Estimated Power} = \sum |\tau| \cdot |\dot{q}|, \quad (3.4)$$

where τ denotes the joint torques and \dot{q} the corresponding joint velocities. The resulting power estimates were then averaged: first across gait types (*default*, *trot*, *bound*, *free*), then across training styles (*eb*, *eg*, *nos*) and finally across setting the trials were conducted in, either the idle or oscillating bridge.

3.5.2. Frequency analysis

The Fourier Transform (FT) is a fundamental tool for analyzing periodic signals by transforming them from the time domain to the frequency domain [49]. In practice, this is achieved using the Fast Fourier Transform (FFT), which efficiently reveals dominant frequency components [50]. The Power Spectral Density (PSD) provides a complementary view by indicating how signal power is distributed across frequencies [51]. These frequency-domain techniques are widely used in gait analysis to quantify rhythmic structure, regularity and dynamic stability. When applied to signals such as ground reaction forces [52, 53], vertical acceleration, or estimated joint torques and powers, they uncover subtle differences in locomotor behavior under varying surface conditions.

To examine the periodic nature of interactions between gait, style and bridge dynamics, time series data for estimated power and vertical IMU acceleration were extracted from real-world trials. Simulated trunk height trajectories (z) were included to enable comparison and validation. All signals were transformed into the frequency domain using Welch's method [54] to compute the PSD, which reduces noise and non-stationary effects through segment averaging, at the expense of some frequency resolution.

4. Results

This chapter presents the detailed results and evaluation of the learned locomotion policies by systematically comparing their performance across multiple gaits (*default*, *trot*, *pace*, *bound*, *pronk*, *free*), training styles (*eb*, *eg*, *nos*) and testing conditions (idle and oscillating bridge), including both simulation environments and real-world experiments on the vibrating HUMVIB bridge.

4.1. Evaluation of Gaits

4.1.1. Comparison of learning curves

As a first step, the impact of gait and style during training on the learning process was examined.

As shown in Figure 4.1, all learning curves are steep during the initial quarter of training and gradually flatten toward the end without reaching a clear plateau. The *free* and *default* gaits—both imposing minimal gait-specific constraints—consistently achieve the highest rewards. In contrast, the *trot*, *pronk*, *pace* and *bound* gaits incorporate strong penalty terms to enforce specific footfall patterns, manifest an initial decline in reward before gradual improvement. Initially, the learning curves for the *trot*, *pronk*, *pace* and *bound* gaits exhibit similar behavior. However, the *trot* policy eventually achieves higher returns, approaching the performance of the *free* and *default* policies. Among the four constrained gaits, *trot* is both visually and reward-wise the most similar to the *default* gait, whereas *pronk*, *pace* and *bound* are more distinct and ultimately yield lower returns.

With respect to training conditions, the *nos* policy performs best, benefiting from a rigid training surface and therefore experiencing fewer destabilizing oscillations. The *eb* and *eg* policies perform similarly to one another, although *eg* slightly outperforms *eb*.

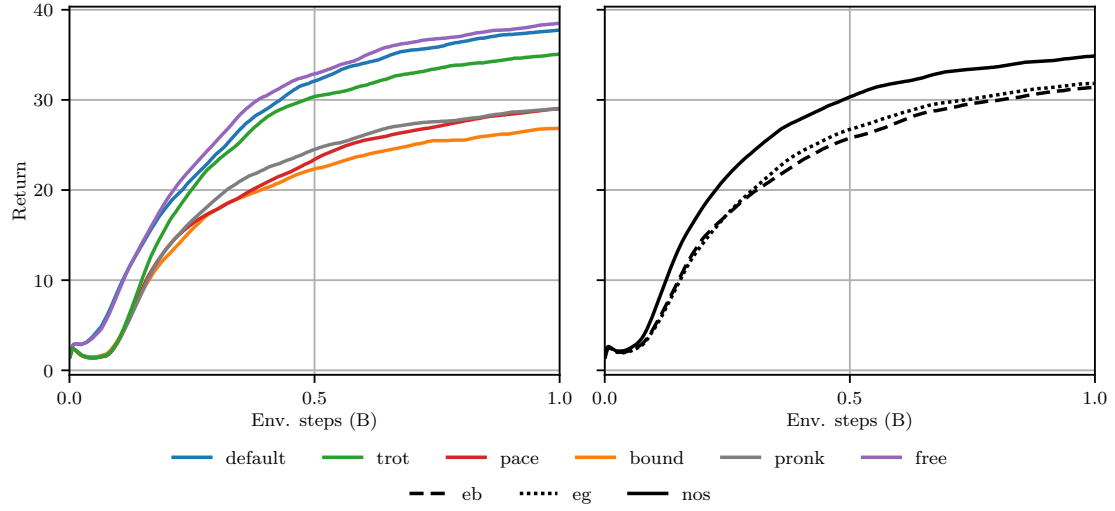


Figure 4.1.: Average episode return for the different gaits (left) and height regulation styles (right) during training.

4.1.2. Velocity simulation

Next, the performance of the policies was evaluated on both rigid ground and the oscillating bridge using the HUMVIB test conditions with varying command velocities. To ensure a fair comparison, episode returns were computed without including the gait-specific reward terms. As shown in Figure 4.2, a general trend of higher rewards at lower velocities can be observed. Tests conducted on the rigid ground yielded significantly higher returns compared to those on the oscillating bridge.

On the idle bridge, the *default*, *trot*, *pronk* and *free* gaits achieved very similar returns. The *bound* gait initially performed on par with these gaits when stationary, but its performance dropped to the level of the *pace* gait as velocity increased, with both remaining distinctly lower than the others. Comparing the training conditions on idle ground reveals that all three (*nos*, *eg* and *eb*) perform similarly, with *nos* slightly on top the others, followed closely by *eg* and then *eb*.

On the oscillating bridge, returns were generally about 10 points lower and exhibited greater variability across policies. The *trot* gait achieved the highest reward at velocities above 0.2 m/s, while *pace* consistently received the lowest returns. Notably, the *bound* gait closed the performance gap and even surpassed the *default* and *free* gaits at certain

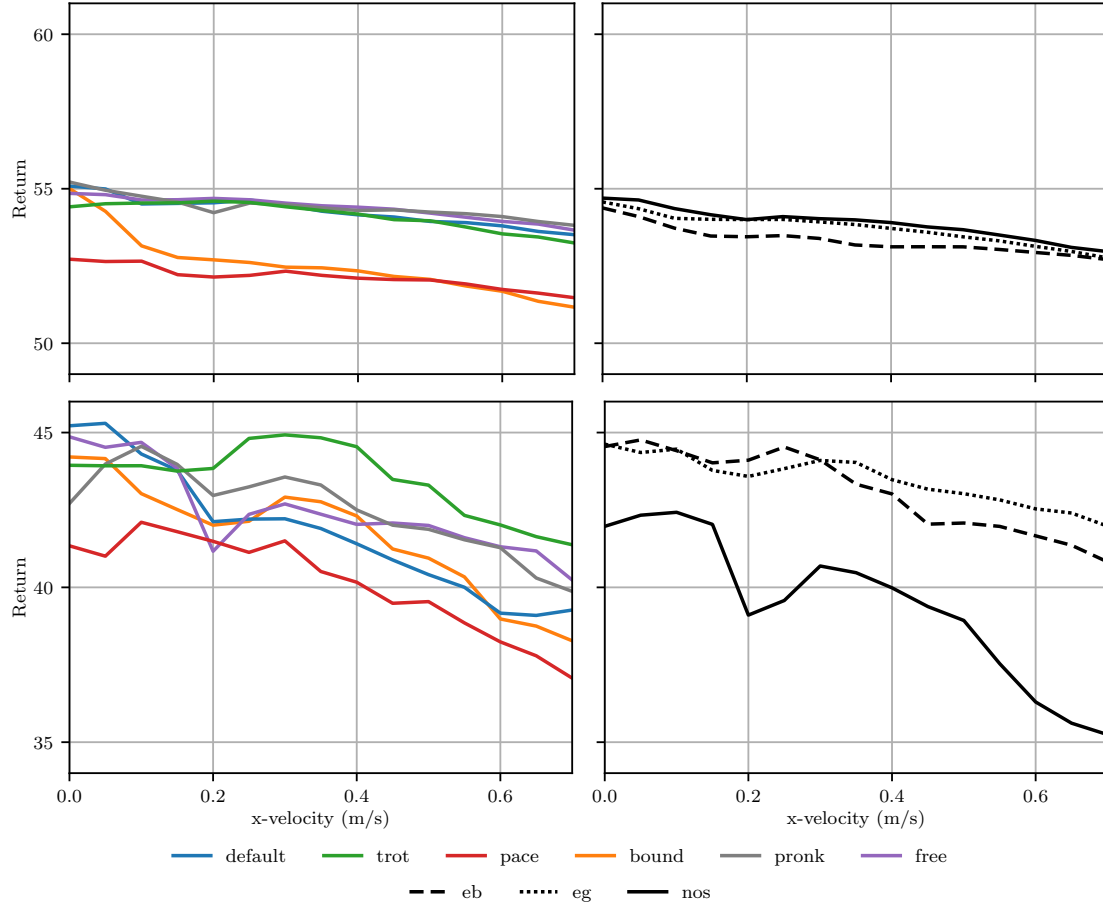


Figure 4.2.: Average episode return over the command velocity on idle (top) and on the oscillating bridge (bottom) for the different gaits (left) and training conditions (right) of the policies during evaluation.

velocities. Several gaits showed a local minimum around 0.2 m/s and a local maximum near 0.3 m/s.

Among the training styles, the *nos* policy consistently underperformed on the oscillating bridge, while the *eg* policy achieved the highest returns, closely followed by *eb*. This trend is also evident in the overall gait comparison (Figure A.1, Figure A.2), where *eg* generally outperforms *eb*.

4.1.3. Footfall pattern and gait analysis

To evaluate whether the desired gaits were successfully learned, the footfall patterns were analyzed, as illustrated in Figure 4.3. Table 4.1 shows the percentage of the gait cycle during which each *nos* policy maintained its characteristic stance phase on rigid ground.

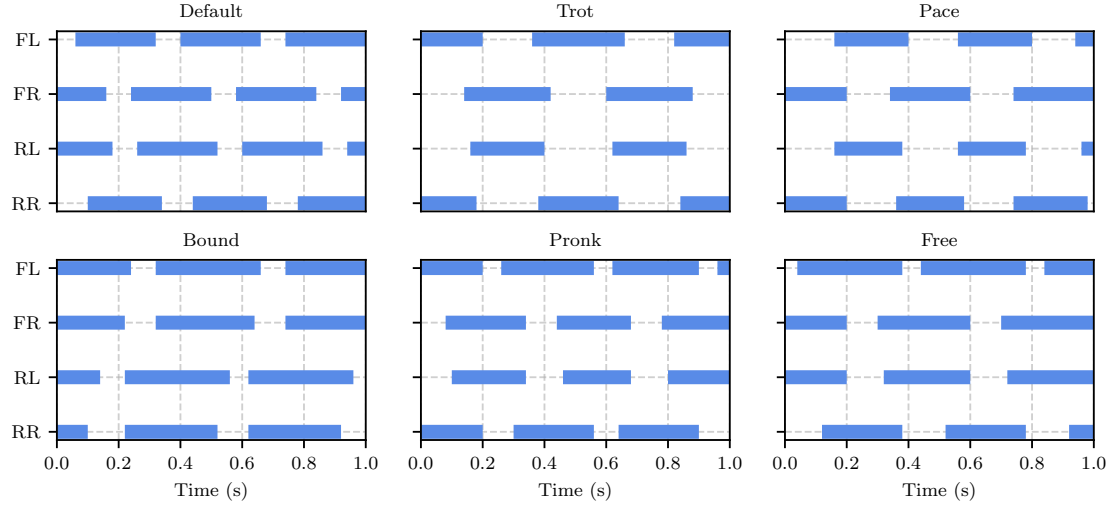


Figure 4.3.: Footfall pattern of the different gaits using the *nos* style on the idle bridge with a target speed $v_x = 0.5$ m/s. The feet of the robot are denoted with front left (FL), front right (FR), rear left (RL) and rear right (RR). A characteristic gait stance phase is detected when all four feet simultaneously exhibit one of the ideal contact combinations defined in Figure 3.2.

The *default* gait (100.0 %) adheres strictly to predefined contact patterns. The *free* gait, while unconstrained, exhibits the highest percentage similarity to the *default* gait, suggesting a natural tendency toward similar coordination even without enforced structure. The *trot* (74.3 %) and *pace* gaits (73.3 %) frequently exhibit their characteristic stance phases described in Figure 3.2. The *bound* gait (35.9 %), marked by both front feet being in the air while the rear feet are grounded—and vice versa—also emerges regularly. In contrast, the *pronk* gait only displayed the phase with all feet on the ground (40.1 %); no instance with all feet in the air (0.0 %) were observed. Although the resulting motion labeled as *pronk* does not appear qualitatively worse than the others, it fails to exhibit the intended airborne behavior. Therefore, it was excluded from further evaluation.

Gait & Style	Default (%)	Trot (%)	Pace (%)	Bound (%)	Pronk _g (%)	Pronk _a (%)
<i>default nos</i>	100.0	35.3	0.0	0.0	35.3	0.0
<i>trot nos</i>	100.0	74.3	0.0	0.0	8.6	0.0
<i>pace nos</i>	100.0	0.0	73.3	0.0	15.2	0.0
<i>bound nos</i>	64.1	0.0	0.0	35.9	48.3	0.0
<i>pronk nos</i>	100.0	44.7	0.0	0.0	40.1	0.0
<i>free nos</i>	100.0	40.1	0.0	0.0	33.7	0.0

Table 4.1.: Percentage of time the *nos* policies exhibited their characteristic gait phases for each defined gait on rigid ground. The Free (%) gait check is omitted, as it imposes no symmetry constraints. Pronk_g (%) and Pronk_a (%) indicate the proportion of time all feet were on the ground and all feet were in the air, respectively.

4.1.4. Walking height

To investigate how the different policies cope with the oscillating bridge, the movement of the robot’s CoM relative to the bridges equilibrium position was analyzed. Figure 4.4 presents the mean and standard deviation of the CoM height for all gait-style combinations on both the idle and oscillating bridge, using a fixed forward velocity command of $\bar{v}_x = 0.5 \text{ m/s}$.

Across all gaits, the highest average CoM height was consistently maintained by the *nos* style on both rigid and oscillating ground. Similarly, the largest standard deviations were generally observed in the *nos* policies, with the exception of *bound-nos* on the idle bridge.

Despite these variations, the average CoM heights for all gait–style combinations remained within a narrow range of 3 cm for both ground condition. When comparing the *eb* and *eg* styles across gaits, no clear trend was identified regarding which policy adopted a lower stance. However, on the oscillating bridge, slightly lower standard deviations were observed in the *eg* policies, suggesting more stable vertical motion.

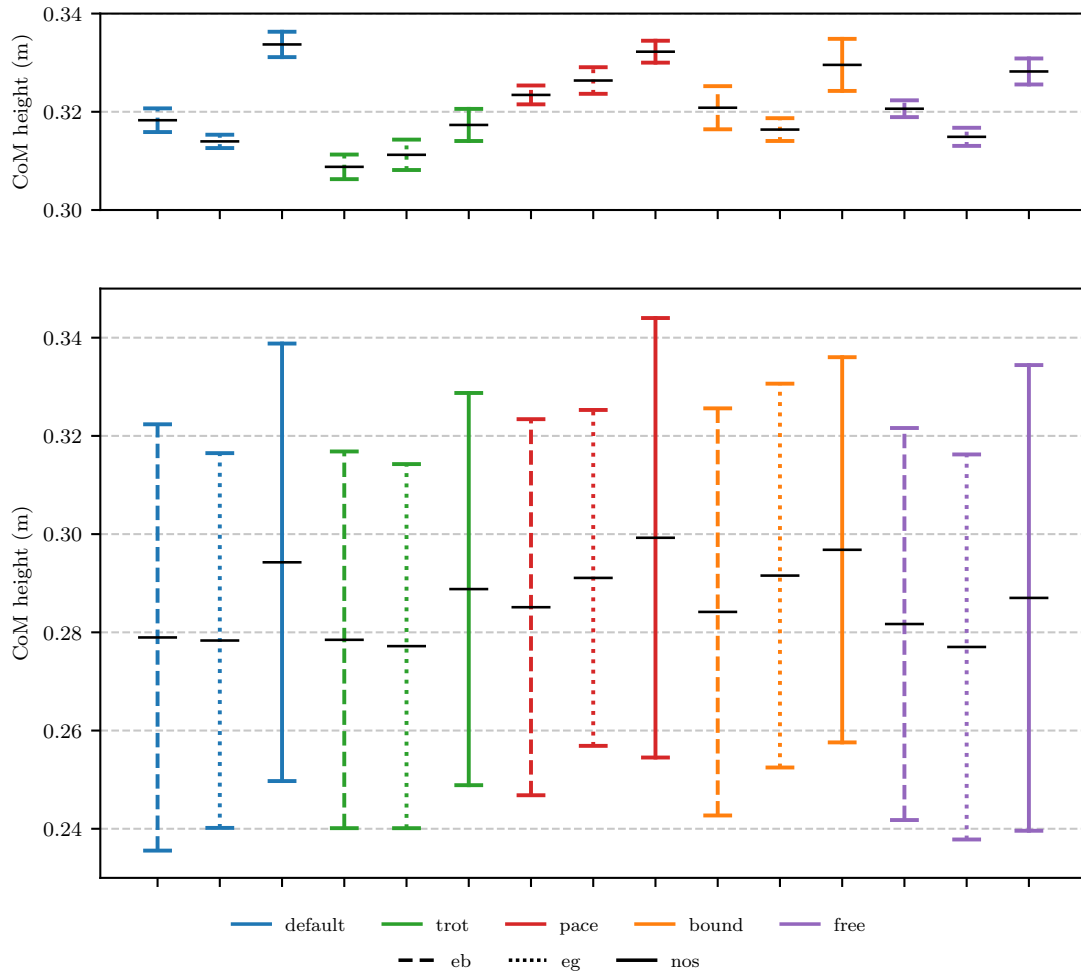


Figure 4.4.: Average and standard deviation of the robot's CoM height for all gait-style combinations on the idle bridge (top) and the oscillating bridge (bottom), using HUMVIB test conditions.

4.2. Real-World Experiment

In the real-world validation experiments, the policies were evaluated on the HUMVIB structure, with the exception of all *pace* gaits and the *bound eg* gait. These were excluded due to observed instability, which posed a heightened risk of the robot falling off the structure and potentially getting damaged. All remaining policies were successfully tested with a minimum of eight trials. A noticeable reduction in locomotion speed was consistently observed as the robot traversed the center of the bridge, where oscillatory motion was most pronounced. The *eb* and *eg* policies exhibited increased stability in both vertical and lateral CoM movement.

Interestingly, despite its comparatively poor performance in simulation, the *bound* gait was able to achieve stable locomotion on the oscillating bridge during real-world testing. While initial oscillations were manually introduced by human operators, the *trot eg* policy was observed to autonomously induce bridge oscillations through their inherent locomotion dynamics with a gait frequency of about 2 Hz.

Upon reviewing the collected data, anomalies in recording durations and sudden power spikes—often preceding unexpected robot behavior—were identified. These trials were removed from the dataset, after which a minimum of five valid trials remained for each gait-style-setting combination. Additionally, it was discovered that the rear-left foot pressure sensor was defective and its values had to be estimated for consistency in the analysis.

4.2.1. Power usage

The average estimated power use in Table 4.2 was computed with Equation 3.4 for each gait-style combination on both the pre-oscillated and idle bridge conditions. The *free* and *default* gaits demonstrated the lowest power usage, with values closely together. The *trot* gait exhibited moderate power use, while the *bound* gait required significantly more power than all others.

	default	trot	bound	free	eb	eg	nos	idle	pre
Power (W)	72.11	99.64	130.65	77.64	100.21	89.89	83.57	89.98	92.84

Table 4.2.: Mean estimated power averaged over gaits, styles and bridge-setting

Among the styles, *eg* is in the middle and nearer to *nos* than to *eb*. Notably, the differences in estimated power use between the idle and pre-oscillated bridge conditions were negligible across all configurations. The complete set of results is presented in Table A.4.

4.2.2. Foot force

The interaction force between the robot and the bridge was measured as the sum of the forces exerted by all four feet at each time step. Due to a malfunctioning rear right foot sensor, its data were estimated using measurements from the functional rear left foot sensor. The average forces across different gaits, styles and bridge conditions (oscillating vs. idle) are summarized in Figure 4.5. The *default* and *free* gaits exhibited the lowest contact forces, while the *bound* gait, characterized by its leaping motion, produced higher forces. The *trot* gait, featuring the shortest transition phases—with more than two feet in contact with the bridge—recorded the highest overall force values. Among the styles, *nos* policies led to the largest forces, whereas *eg* policies showed the lowest. Forces measured on the idle bridge were slightly reduced and exhibited less variance compared to the oscillating bridge, although the overall differences remained small.

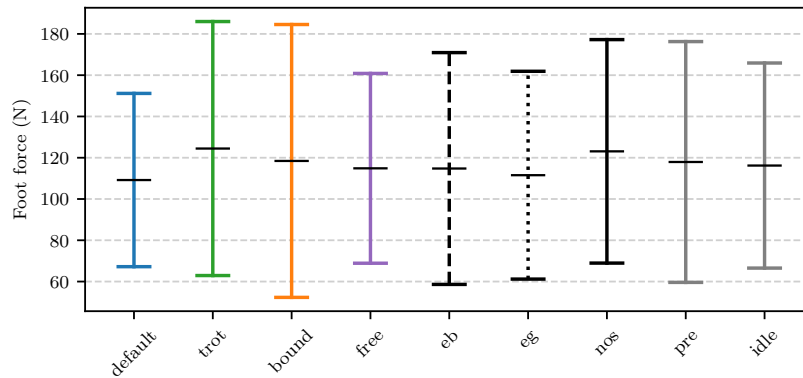


Figure 4.5.: Means and standard deviations of the force readouts of the foot sensors of the robot over gaits, styles and bridge-setting in the real world.

A detailed overview of all gait, style and bridge condition combinations is provided in Table A.5.

4.2.3. PSD analysis

The PSD of estimated power, IMU vertical acceleration and simulated trunk height were computed and normalized to the highest peak above 2.2 Hz, in order to reduce the influence of the dominant 2.0 Hz peak induced by the oscillating bridge (Figure 4.6). This normalization enabled clearer comparisons of gait-induced frequency content across different training styles and gaits.

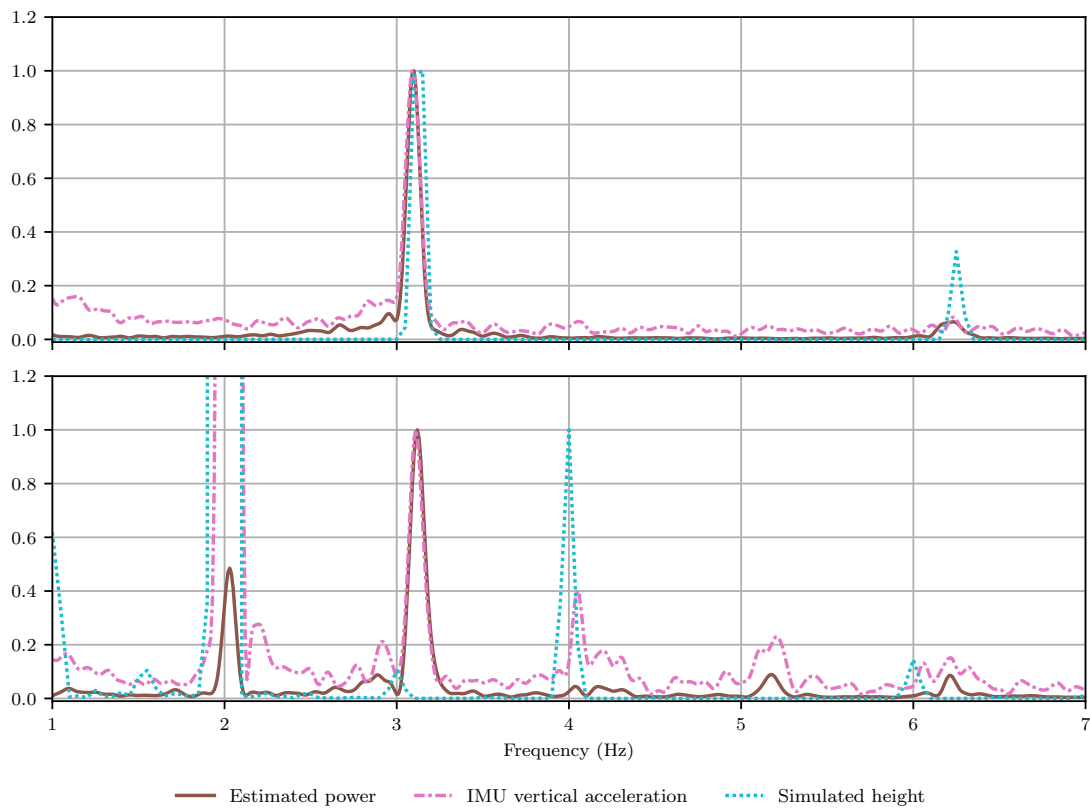


Figure 4.6.: PSD of the *default eb* policy on the idle bridge (top) and the oscillating bridge (bottom). The spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

The onboard data—estimated power and IMU vertical acceleration—showed strong alignment. The IMU signal exhibited a broader frequency distribution, while the simulated

trunk height featured more distinct and isolated peaks, separated by flatter regions.

In both the estimated power and IMU signals, bridge oscillation introduced a clear peak at 2.0 Hz along with harmonic components. This effect was more pronounced in the IMU data than in the estimated power, though the overall spectral shape remained similar to that of the idle bridge condition.

Across all gait-style combinations (Figure A.3 to A.13), no systematic shift of gait-related frequency peaks toward the bridge's natural frequency was observed.

For most gait-style combinations on the idle bridge, two main frequency peaks were visible: one between 2.0 Hz and 3.5 Hz, and another at its second harmonic, indicating consistent gait-related rhythmic patterns in the absence of perturbations.

5. Discussion

This chapter critically discusses the training and evaluation outcomes of various locomotion policies, highlighting how gait selection, training styles, and environmental conditions influence learning efficiency, robustness and real-world performance. It also proposes a new gait-style aimed at enhancing stability, while evaluating the effectiveness of metrics such as energy efficiency, force distribution, and gait robustness under oscillatory conditions.

5.1. Gait Learnability and Training Biases

During the learning phase, the *free* gait was the easiest for the RL agents to learn, consistently achieving the highest rewards. This outcome is expected, as the *free* gait imposes no constraints on footfall patterns, in contrast to the biologically inspired gaits that require strict temporal coordination. The lower performance of the *bound* and *pronk* policies is expected, as these gaits typically involve extended foot contact phases that deviate from the ideal footfall pattern (Figure 3.2)—penalized by the reward function—and inherently offer less stability than the *trot* and *default* gaits. Lower rewards implicate that these policies are less effective at tracking commanded velocities and are more susceptible to environmental disturbances encountered during training. The initial drop followed by a rapid improvement in the learning curve of the *trot* policy suggests that its gait-specific constraints are initially challenging but can be efficiently overcome during training.

Higher rewards observed for the *nos* policies can be attributed to the less disruptive nature of the rigid training surface. This also highlights that including training on oscillating ground—as done with the *eb* and *eg* styles—is a meaningful extension when learning policies for more challenging environments.

It is worth noting that each of the 18 policies evaluated in this study was trained with a single random seed, which limits statistical robustness. This limitation may also contribute to the failure of the *pronk* policy to reproduce its characteristic stride phase during training.

5.2. Policy Performance in Simulation

The evaluation across a range of command velocities on both rigid ground and the oscillating bridge revealed that gaits trained on oscillating terrain can perform on par with those trained on rigid ground; however, the opposite does not hold. Policies trained exclusively on rigid ground exhibited markedly reduced performance under oscillatory conditions. This highlights the importance of training locomotion policies with vertical ground perturbations to achieve robustness in dynamic environments.

On rigid terrain, increased command velocity correlated with lower episode returns. While a similar trend is observed on the oscillating bridge, several gaits exhibit distinct local peaks and troughs in performance, indicating interactions between gait dynamics and the bridge’s oscillatory motion. Although the chosen gait significantly influences the resulting gait frequency, the robot’s morphology—particularly leg length—also plays a critical role in determining feasible step lengths [55]. As a result, gait frequencies at given velocities are expected to lie within a similar range across all policies. The observed fluctuations in performance likely arise from constructive and destructive interaction of the robot’s gait frequency and the bridge’s oscillation. For instance, the *trot* policies exhibit a step frequency of approximately 2 Hz at a command velocity of 0.35 m/s.

Analysis of footfall patterns plays a key role in verifying whether the desired gait behaviors have been learned. Comparing short sequences of actual footfall data (Figure 4.3) with idealized references (Figure 3.2) allows for immediate visual assessment of gait quality. This is further quantified in Table 4.1, where the learned patterns for *trot*, *pace* and *bound* are shown to be highly distinct, with no overlap.

Interestingly, while both *default* and *free* gaits would theoretically allow for the emergence of *trot* and *pace*, only *trot* behavior emerged consistently, indicating it may be the most natural or easiest for the policy to converge on. The exclusivity of *default* and *bound* is clearly reflected in their opposing reward structures: when the *bound* policy is in compliance with its own criteria, it directly violates those of the *default* policy and vice versa. Moreover, the absence of *bound* behavior in the *free* gait policies suggests that the reward structure used in the *default* gait could be simplified. Instead of explicitly discouraging *bound*, it may suffice to enforce a constraint requiring at least two legs to remain in contact with the ground. Or it could be extended by an additional penalty on *pace* behavior to further accelerate convergence.

The distinction between ground and aerial phases was essential for evaluating the *prnck* gait, as no instance of all feet being simultaneously in the air was observed, neither within

the footfall pattern time frame nor during testing in simulation. Consequently, the *pronk* gait was omitted from further evaluation due to the failure to induce the desired behavior. It is possible that a different or more strongly weighted penalty term could successfully induce this gait.

The *trot* and *pace* gaits both exhibit their characteristic stance phases exceeding 70% of the gait cycle, which may contribute to reduced dynamic stability under varying terrain conditions.

A broader analysis of gait classification results, presented in Table A.3, shows that while testing policies on oscillating ground, they are more likely to deviate from their intended footfall patterns. This was especially observed in those trained using the *nos* style. Policies trained on oscillating terrain demonstrate greater consistency in preserving their learned gait patterns, even under challenging conditions. This finding supports the conclusion that training on oscillating ground enhances gait robustness.

An evaluation of the CoM height reveals that gaits trained on oscillating terrain adopt a lower posture on average, which likely aids in compensating for vertical oscillations. This consistently observed adaptation suggests that learning with a reduced nominal trunk height could improve policy performance when training new gaits for bridge environments. On oscillating ground, both *eb* and *eg* styles not only adopt a lower stance but also better maintain their posture, as indicated by lower standard deviations in CoM height—particularly in the case of the *eg* policies.

5.3. Evaluation of Real-World Experiments

Due to the absence of precise velocity measurements and the limited accuracy of the available IMU data, a reliable estimation of the robot’s velocity was not feasible. Consequently, no computation of the cost of transport was performed. In terms of energy efficiency, average power consumption analysis shows that the *default* and *free* gaits are the most economical. Policies trained with the *nos* style emerge as the most energy efficient overall. This may be due to their relatively higher base height—which reduces joint torques—and the lack of effort spent compensating for oscillations. Notably, the difference in energy expenditure between idle and oscillating terrain is minimal, suggesting that the oscillatory energy input either has little effect on forward locomotion or is passively absorbed and utilized through the system’s inherent joint stiffness—effectively canceling out its impact.


The lower average forces observed for default and free gaits indicate a more efficient and gentler interaction with the ground, potentially contributing to improved stability and reduced mechanical wear. The higher force peaks associated with the trot gait likely result from its rapid alternation pattern, which increases instantaneous loading and may elevate the risk of slippage or impact-related disturbances. Lower-posture styles such as *eg* appear to promote a more even distribution of force across steps, reducing peak stresses and possibly enhancing long-term durability of the robot's structure and sensors. The lack of change in average forces between the oscillating and idle bridge conditions suggests that the locomotion strategies are resilient to low-frequency vertical disturbances. However, the standard deviation under oscillation hints at occasional high-force interactions, potentially triggered by resonances or phase alignment between the robot and the moving bridge.

Based on the evidence, the most promising gait and training style for further exploration is the *default eg* gait. It proves to be the most robust against disturbances, as it frequently uses three feet on the ground, providing enhanced stability compared to *pace* and *trot*. Additionally, it offers effective height regulation, soft foot contacts and a balanced power consumption. These findings are particularly compelling given that the *default* gait has already been used to train robust policies. In contrast, *pace* is typically used by long-legged animals for fast movement on even terrain, while *bound* is specialized for sprinting.

The PSD of gait-style combinations tested on the idle bridge, the simulated height data typically provides a good estimate of the gait frequency. However, on the oscillating bridge, the dominant spectral peak at 2.0 Hz and its harmonics overwhelm the signal, making it difficult to distinguish gait-related frequencies. As a result, simulated height becomes less reliable for spectral analysis under oscillatory conditions.

In contrast, the PSD of the estimated power proves more robust than those of vertical acceleration and simulated height. It preserves the general spectral shape observed on the idle bridge while still reflecting the influence of the bridge's oscillation, allowing for a more nuanced interpretation of gait dynamics. Thus, estimated power is best suited for further investigation across varying bridge conditions.

The presence of a spectral peak between 2.0 Hz and 3.5 Hz in the estimated power can be attributed to the gait frequency. A second peak is often located at the second harmonic of this frequency and corresponds to the rate at which the synchronous legs contact the ground. Harmonics, in general, reflect non-linearities and distortions in locomotion dynamics, which can arise from gait asymmetries or environmental interactions. However, in quadruped locomotion, the relative strength of the second harmonic can also serve as an indicator of gait quality: a pronounced harmonic suggests a more synchronized and regular gait cycle. This arises because two by about π phase-shifted stride frequencies



interfere to dampen the gait frequency and produce a signal with twice the original frequency. This effectively highlights the coordination of limb pairs.

6. Conclusion

This study demonstrates that locomotion policies trained in simulation on an oscillating surface significantly outperform those trained on rigid terrain when deployed on the Unitree Go2 quadruped traversing the HUMVIB bridge. Using RL with the PPO algorithm, we trained 18 distinct policies across six gaits and three training conditions. Exposure to vertical ground perturbations during training improved both stability and adaptability, as validated through zero-shot transfer to real-world experiments.

High rewards achieved during training primarily reflect the compatibility of specific gaits with the reward function. In contrast, performance differences across height regulation styles and training environments highlight the difficulty of walking on oscillating ground without practicing on it beforehand, underscoring the importance of targeted exposure during training for robust generalization. Analyzing footfall patterns provided insight into gait acquisition, while simulation height analysis revealed that lower postures learned under unstable conditions promoted vertical stability. Estimates of foot force and power usage offered a broader understanding of the biomechanical trade-offs associated with different gaits and styles. Finally, frequency-domain analyses uncovered periodic features that are preserved from simulation to physical execution.

These insights suggest that an ideal future policy might be a more restrictive *default* gait, incorporating elements from the *eg* style—such as a lower trunk posture—and deployed selectively in response to detected ground oscillations using onboard IMU data. Such adaptive switching could enhance both robustness and energy efficiency.

6.1. Outlook

While the results show promising transferability, further optimization is needed for an exhaustive comparison of gait styles. Learning each gait—including the newly proposed

default—across multiple seeds and evaluating them both in simulation and on the real robot would help quantify their consistency and robustness. Exploring learning without any explicit height regulation on both stable and unstable terrain may also reveal whether posture adjustments emerge naturally from task dynamics. Additionally, simulating different bridge frequencies—particularly those close to the natural gait frequencies of each gait—may uncover resonance effects or gait-specific vulnerabilities that remained hidden when testing was limited to the bridge’s primary eigenfrequency of 2.0 Hz.

Most policies performed well within their respective training domains, but current black-box learning approaches do not generalize perfectly to unfamiliar settings. Integrating white-box models, incorporating physiological principles such as muscle properties [56] or sensor-mechanical couplings, could help overcome this limitation. Hybrid modeling may explain the remarkable versatility of animals traversing dynamic and uncertain environments by leveraging the inherent adaptability of neural control systems [57].

Alternative evaluation strategies, such as analyzing the PSD of simulated power usage instead of trunk height, could provide deeper insights into the mechanical implications of gait and height control strategies, offering better predictive power for sim-to-real transfer.

Future directions could include the combination of vertical perturbations with moving or adversarial obstacles, as well as the integration of high-level planning systems capable of navigating complex, multilayered terrain. Finally, although no biological quadrupeds have yet been tested on the HUMVIB bridge, future studies should aim to explore their strategies under vertical oscillations to complement robotic findings with biologically grounded insights [58].

Taking these next steps will deepen our understanding of gait-specific adaptation, sim-to-real transfer, and policy generalization for developing robots that are not only trainable but also robustly deployable in dynamic and unpredictable real-world environments.

Bibliography

- [1] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning robust perceptive locomotion for quadrupedal robots in the wild,” *Science robotics*, vol. 7, no. 62, p. eabk2822, 2022.
- [2] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, “Extreme parkour with legged robots,” in *RoboLetics: Workshop on Robot Learning in Athletics@ CoRL 2023*, 2023.
- [3] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, “Anymal parkour: Learning agile navigation for quadrupedal robots,” *Science Robotics*, vol. 9, no. 88, p. eadi7566, 2024.
- [4] S. Choi, G. Ji, J. Park, H. Kim, J. Mun, J. H. Lee, and J. Hwangbo, “Learning quadrupedal locomotion on deformable terrain,” *Science Robotics*, vol. 8, no. 74, p. eade2256, 2023.
- [5] G. Ji, J. Mun, H. Kim, and J. Hwangbo, “Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4630–4637, 2022.
- [6] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, “Rapid locomotion via reinforcement learning,” *The International Journal of Robotics Research*, vol. 43, no. 4, pp. 572–587, 2024.
- [7] D. Tokur, M. Grimmer, and A. Seyfarth, “Review of balance recovery in response to external perturbations during daily activities,” *Human movement science*, vol. 69, p. 102546, 2020.
- [8] T. A. McMahon and P. R. Greene, “Fast running tracks,” *Scientific American*, vol. 239, no. 6, p. 148, 1978.

-
-
- [9] D. P. Ferris, K. Liang, and C. T. Farley, “Runners adjust leg stiffness for their first step on a new running surface,” *Journal of biomechanics*, vol. 32, no. 8, pp. 787–794, 1999.
 - [10] M. A. Daley and A. A. Biewener, “Running over rough terrain reveals limb control for intrinsic stability,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 42, pp. 15681–15686, 2006.
 - [11] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
 - [12] K. Caluwaerts, A. Iscen, J. C. Kew, W. Yu, T. Zhang, D. Freeman, K.-H. Lee, L. Lee, S. Saliceti, V. Zhuang, *et al.*, “Barkour: Benchmarking animal-level agility with quadruped robots,” *arXiv preprint arXiv:2305.14654*, 2023.
 - [13] Z. Zhuang, Z. Fu, J. Wang, C. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, “Robot parkour learning,” in *Conference on Robot Learning (CoRL)*, 2023.
 - [14] D. Kim, H. Kwon, J. Kim, G. Lee, and S. Oh, “Stage-wise reward shaping for acrobatic robots: A constrained multi-objective reinforcement learning approach,” *arXiv preprint arXiv:2409.15755*, 2024.
 - [15] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 23–30, IEEE, 2017.
 - [16] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, “Learning to walk in minutes using massively parallel deep reinforcement learning,” in *Conference on Robot Learning*, pp. 91–100, PMLR, 2022.
 - [17] D. Vogel, R. Baines, J. Church, J. Lotzer, K. Werner, and M. Hutter, “Robust ladder climbing with a quadrupedal robot,” *arXiv preprint arXiv:2409.17731*, 2024.
 - [18] F. Shi, C. Zhang, T. Miki, J. Lee, M. Hutter, and S. Coros, “Rethinking robustness assessment: Adversarial attacks on learning-based quadrupedal locomotion controllers,” *arXiv preprint arXiv:2405.12424*, 2024.
 - [19] K. Liu, J. Gu, X. He, and L. Zhang, “Optimization algorithms for dynamic environmental sensing and motion planning of quadruped robots in complex environments on unmanned offshore platforms,” *Measurement Science and Technology*, vol. 36, no. 1, p. 015122, 2024.

-
-
- [20] S. Misenti, B. Hertel, B. Weng, R. Donald, A. Jawaji, M.-T. Kosoko-Thoroddsen, J. G. Trafton, A. Norton, R. Azadeh, and Y. Gu, “Experimental evaluation of commercial quadruped robots: stability and performance in non-inertial environments,” *International Journal of Intelligent Robotics and Applications*, pp. 1–26, 2025.
- [21] A. Firus, J. Schneider, W. Francke, and B. Kunkel, “Tu darmstadt humvib-bridge: Investigation of the human-structure interaction on a full-scale experimental pedestrian bridge,” *Bulletin of the Transilvania University of Brasov. Series I-Engineering Sciences*, pp. 63–72, 2016.
- [22] M. Fritzsche, H. Berthold, S. Lorenzen, J. Schneider, A. Firus, M. Stasica, G. Zhao, and A. Seyfarth, “Integrated measurement concept for identification of human-structure interaction of flexible structures for natural gait,” in *Bridge Safety, Maintenance, Management, Life-Cycle, Resilience and Sustainability*, pp. 713–720, CRC Press, 2022.
- [23] M. Raibert, “Legged robots that balance,” *MIT Press*, 1986.
- [24] A. Ijspeert, J. Nakanishi, P. Pastor, H. Hoffmann, and S. Schaal, “Central pattern generators for locomotion control in animals and robots: a review,” *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008.
- [25] P. Ramdya and A. J. Ijspeert, “The neuromechanics of animal locomotion: From biology to robotics and back,” *Science Robotics*, vol. 8, no. 78, p. eadg0279, 2023.
- [26] S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, and R. Tedrake, “Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot,” *Autonomous robots*, vol. 40, pp. 429–455, 2016.
- [27] J. Nakanishi, J. Morimoto, G. Endo, G. Cheng, S. Schaal, and M. Kawato, “Learning from demonstration and adaptation of biped locomotion,” *Robotics and Autonomous Systems*, vol. 47, no. 2, pp. 79–91, 2004. Robot Learning from Demonstration.
- [28] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Learning to walk via deep reinforcement learning,” *arXiv preprint arXiv:1905.01677*, 2019.
- [29] N. Heess, S. Nasiriany, D. Horgan, G. Dulac-Arnold, W. Zaremba, V. Mnih, and D. Silver, “Emergence of locomotion behaviors in rich environments,” *arXiv preprint arXiv:1707.02286*, 2017.
- [30] 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.

-
-
- [31] P. Fankhauser, M. Bloesch, and M. Hutter, “Probabilistic terrain mapping for mobile robots with uncertain localization,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3019–3026, 2018.
- [32] G. Bellegarda, M. Shafiee, and A. Ijspeert, “Allgaits: Learning all quadruped gaits and transitions,” *arXiv preprint arXiv:2411.04787*, 2024.
- [33] A. Luo, Q. Wan, Y. Meng, S. Kong, W. Chi, C. Zhang, S. Zhang, Y. Liu, Q. Zhu, and J. Yu, “Tfgait—stable and efficient adaptive gait planning with terrain recognition and froude number for quadruped robot,” *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 9218–9229, 2025.
- [34] T. Akbas, S. E. Eskimez, S. Ozel, O. K. Adak, K. C. Fidan, and K. Erbatur, “Zero moment point based pace reference generation for quadruped robots via preview control,” in *2012 12th IEEE International Workshop on Advanced Motion Control (AMC)*, pp. 1–7, 2012.
- [35] H. Kimura, I. Shimoyama, and H. M. and, “Dynamics in the dynamic walk of a quadruped robot,” *Advanced Robotics*, vol. 4, no. 3, pp. 283–301, 1989.
- [36] D. Kar, K. Kurien Issac, and K. Jayarajan, “Gaits and energetics in terrestrial legged locomotion,” *Mechanism and Machine Theory*, vol. 38, no. 4, pp. 355–366, 2003. NaCOMM99, the National Conference on Machines and Mechanisms.
- [37] J. P. van der Weele and E. J. Banning, “Mode interaction in horses, tea, and other nonlinear oscillators: The universal role of symmetry,” *American Journal of Physics*, vol. 69, no. 9, pp. 953–965, 2001.
- [38] M. M. Ankaralı, U. Saranlı, and A. Saranlı, “Control of underactuated planar hexapedal pronking through a dynamically embedded slip monopod,” in *2010 IEEE International Conference on Robotics and Automation*, pp. 4721–4727, 2010.
- [39] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ international conference on intelligent robots and systems*, pp. 5026–5033, IEEE, 2012.
- [40] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning*. MIT Press, 2 ed., 2012. (draft 2nd ed.).
- [41] N. Bohlinger and K. Dorer, “RL-x: A deep reinforcement learning library (not only) for robocup,” in *Robot World Cup*, pp. 228–239, Springer, 2023.

-
-
- [42] A. Firus, R. Kemmler, H. Berthold, S. Lorenzen, and J. Schneider, "A time domain method for reconstruction of pedestrian induced loads on vibrating structures," *Mechanical Systems and Signal Processing*, vol. 171, p. 108887, 2022.
- [43] L. I. Deych and L. I. Deych, "Harmonic oscillator models," *Advanced Undergraduate Quantum Mechanics: Methods and Applications*, pp. 201–254, 2018.
- [44] R. Blickhan, "The spring-mass model for running and hopping," *Journal of Biomechanics*, vol. 22, no. 11, pp. 1217–1227, 1989.
- [45] M. Ahmad Sharbafi, A. Mohammadi Nejad Rashty, C. Rode, and A. Seyfarth, "Reconstruction of human swing leg motion with passive biarticular muscle models," *Human Movement Science*, vol. 52, pp. 96–107, 2017.
- [46] J. O. Thompson, "Hooke's law," *Science*, vol. 64, no. 1656, pp. 298–299, 1926.
- [47] N. Bohlinger, G. Czechmanowski, M. Krupka, P. Kicki, K. Walas, J. Peters, and D. Tateo, "One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion," *Conference on Robot Learning*, 2024.
- [48] M. Gohnert and R. Bradley, "Reinforced concrete beam design using catenary theory," *International Journal of Civil Engineering*, vol. 21, no. 5, pp. 751–762, 2023.
- [49] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*, vol. 31999. McGraw-Hill New York, 1986.
- [50] P. Duhamel and M. Vetterli, "Fast fourier transforms: A tutorial review and a state of the art," *Signal Processing*, vol. 19, no. 4, pp. 259–299, 1990.
- [51] M. Tiboni, G. Incerti, C. Remino, and M. Lancini, "Comparison of signal processing techniques for condition monitoring based on artificial neural networks," in *Advances in Condition Monitoring of Machinery in Non-Stationary Operations* (A. Fernandez Del Rincon, F. Viadero Rueda, F. Chaari, R. Zimroz, and M. Haddar, eds.), (Cham), pp. 179–188, Springer International Publishing, 2019.
- [52] S. R. Wurdeman, J. M. Huisinga, M. Filipi, and N. Stergiou, "Multiple sclerosis affects the frequency content in the vertical ground reaction forces during walking," *Clinical Biomechanics*, vol. 26, no. 2, pp. 207–212, 2011.
- [53] K. Van Nimmen, G. Zhao, A. Seyfarth, and P. Van den Broeck, "A robust methodology for the reconstruction of the vertical pedestrian-induced load from the registered body motion," *Vibration*, vol. 1, no. 2, pp. 250–268, 2018.

-
-
- [54] O. M. Solomon, Jr, "Psd computations using welch's method. [power spectral density (psd)]," tech. rep., Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 12 1991.
- [55] R. C. Gonzalez, D. Alvarez, A. M. Lopez, and J. C. Alvarez, "Modified pendulum model for mean step length estimation," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1371–1374, 2007.
- [56] C. Schumacher and A. Seyfarth, "Sensor-motor maps for describing linear reflex composition in hopping," *Frontiers in computational neuroscience*, vol. 11, p. 108, 2017.
- [57] M. M. Van Der Krogt, W. W. De Graaf, C. T. Farley, C. T. Moritz, L. Richard Casius, and M. F. Bobbert, "Robust passive dynamics of the musculoskeletal system compensate for unexpected surface changes during human hopping," *Journal of applied physiology*, vol. 107, no. 3, pp. 801–808, 2009.
- [58] K. T. Kalveram and A. Seyfarth, "Inverse biomimetics: How robots can help to verify concepts concerning sensorimotor control of human arm and leg movements," *Journal of Physiology-Paris*, vol. 103, no. 3-5, pp. 232–243, 2009.



A. Appendix

The following plots and tables present extended results for those gait, style and oscillatory conditions that were only partially shown in the main text, where selected examples or averages were displayed.

Hyperparameter	Value
Total timesteps	1000000000
Batch size	130560 (48 envs * 2720 steps)
Mini-batch size	32640
Nr. epochs	5
Initial and final learning rate	0.0004, 0.0
Entropy coefficient	0.0
Discount factor	0.99
GAE λ	0.9
Clip range	0.1
Max gradient norm	5.0
Initial action standard deviation	1.0
Clip range action standard deviation	1e-8, 2.0
Clip range action mean	-10.0, 10.0 (before applying σ_a)

Table A.1.: The PPO hyperparameters used for training.

Parameter	Range	Description
Control Noise (HardDomainControl)		
motor strength	[0.5, 1.5]	Global motor strength scaling
p gain factor	[0.5, 1.5]	Proportional gain factor
d gain factor	[0.5, 1.5]	Derivative gain factor
asymmetric factor	[0.95, 1.05]	Joint-wise multiplicative asymmetry
position offset	[−0.05, 0.05]	Joint-wise position offset (rad)
Physical Parameters (HardDomainMuJoCoModel)		
friction (tangential)	[0.001, 2.0]	Surface tangential friction
friction (torsional, ground)	[0.00001, 0.01]	Ground torsional friction
friction (torsional, feet)	[0.00001, 0.04]	Feet torsional friction
friction (rolling, ground)	[0.00001, 0.0002]	Ground rolling friction
friction (rolling, feet)	[0.00001, 0.02]	Feet rolling friction
contact damping	[30, 130]	Contact damping
contact stiffness	[500, 1500]	Contact stiffness
gravity	[8.81, 10.81]	Gravity (m/s ²)
added trunk mass	[−2.0, 2.0]	Additional trunk mass (kg)
trunk COM displacement	[−0.01, 0.01]	Trunk CoM offset (m)
foot size	[0.020, 0.024]	Foot size (m)
joint damping	[0.0, 2.0]	Joint damping coefficient
joint armature	[0.008, 0.05]	Joint armature
joint stiffness	[0.0, 2.0]	Joint stiffness
joint friction loss	[0.0, 1.0]	Joint friction loss
External Perturbations (HardDomainPerturbation)		
push velocity (x)	[−1.0, 1.0]	Push velocity in x-direction (m/s)
push velocity (y)	[−1.0, 1.0]	Push velocity in y-direction (m/s)
push velocity (z)	[−1.0, 1.0]	Push velocity in z-direction (m/s)

Table A.2.: Domain randomization parameters and ranges

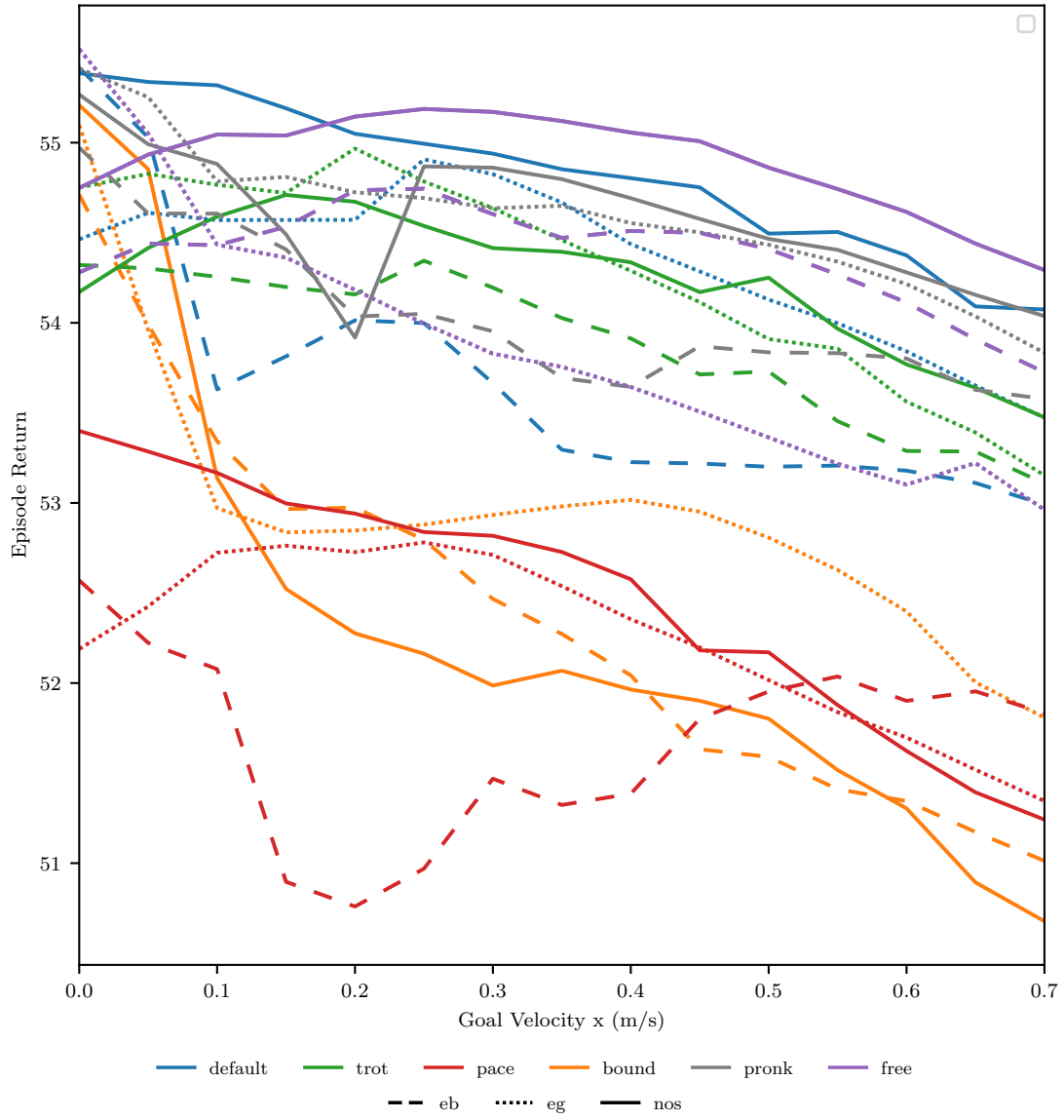


Figure A.1.: Episode return over the command velocity on idle bridge for all gait-style combinations.

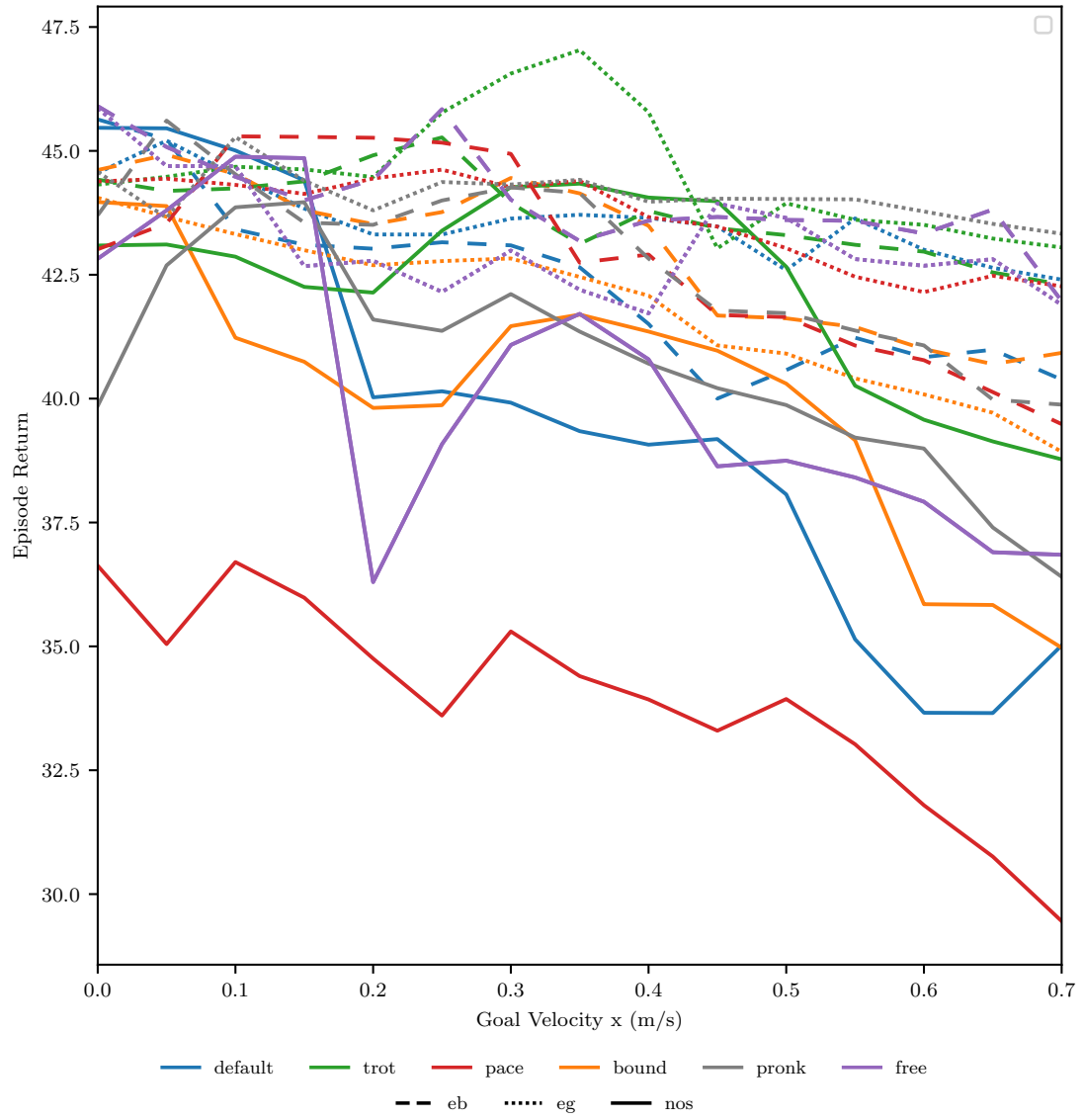


Figure A.2.: Episode return over the command velocity on oscillating bridge, using HUMVIB settings, for all gait-style combinations.

Gait	Style	Set	Default (%)	Trot (%)	Pace (%)	Bound (%)	Pronk _g (%)	Pronk _a (%)
default	eb	pre	100.0	29.9	2.6	0.0	41.9	0.0
default	eb	idle	100.0	31.3	0.0	0.0	43.7	0.0
default	eg	pre	100.0	32.9	0.2	0.0	46.9	0.0
default	eg	idle	100.0	22.8	0.0	0.0	44.1	0.0
default	nos	pre	91.6	38.3	2.2	3.6	34.1	2.2
default	nos	idle	100.0	35.3	0.0	0.0	35.3	0.0
trot	eb	pre	99.2	73.7	0.0	0.6	15.8	0.0
trot	eb	idle	100.0	70.1	0.0	0.0	15.0	0.0
trot	eg	pre	99.4	77.6	0.0	0.4	14.0	0.0
trot	eg	idle	100.0	75.6	0.0	0.0	13.0	0.0
trot	nos	pre	100.0	74.5	0.0	0.0	17.6	0.0
trot	nos	idle	100.0	74.3	0.0	0.0	8.6	0.0
pace	eb	pre	99.2	0.0	56.3	0.4	29.1	0.0
pace	eb	idle	100.0	0.0	57.9	0.0	26.3	0.0
pace	eg	pre	100.0	0.4	48.7	0.0	17.6	0.0
pace	eg	idle	100.0	0.0	60.3	0.0	12.8	0.0
pace	nos	pre	97.4	1.0	74.3	1.0	19.0	1.0
pace	nos	idle	100.0	0.0	73.3	0.0	15.2	0.0
bound	eb	pre	67.3	0.8	0.0	32.1	48.1	0.0
bound	eb	idle	74.1	1.0	0.0	25.9	47.5	0.0
bound	eg	pre	89.2	1.2	0.0	10.4	42.1	0.0
bound	eg	idle	90.0	0.0	0.0	10.0	55.1	0.0
bound	nos	pre	47.9	0.0	0.0	52.1	47.9	0.0
bound	nos	idle	64.1	0.0	0.0	35.9	48.3	0.0
pronk	eb	pre	96.4	19.8	3.0	0.0	48.7	0.0
pronk	eb	idle	100.0	5.8	0.0	0.0	35.3	0.0
pronk	eg	pre	99.6	29.5	0.8	0.0	28.9	0.0
pronk	eg	idle	100.0	33.1	0.0	0.0	31.7	0.0
pronk	nos	pre	95.6	42.7	1.0	1.4	34.7	0.8
pronk	nos	idle	100.0	44.7	0.0	0.0	40.1	0.0
free	eb	pre	99.4	31.7	0.2	0.6	43.5	0.0
free	eb	idle	100.0	41.7	0.0	0.0	44.5	0.0
free	eg	pre	99.4	31.3	0.4	0.6	40.9	0.0
free	eg	idle	100.0	14.4	0.0	0.0	43.1	0.0
free	nos	pre	96.0	41.5	0.8	0.8	42.7	0.4
free	nos	idle	100.0	40.1	0.0	0.0	33.7	0.0

Table A.3.: Gait detection percentages across all policies on oscillating and idle bridge.

Gait	Style	Set	Count	Mean power (W)	Std	Min	Max
default	eb	pre	7	81.39	2.44	77.04	84.36
default	eb	idle	9	79.06	1.34	76.69	80.59
default	eg	pre	8	76.55	1.99	73.38	78.52
default	eg	idle	6	78.88	1.27	77.37	80.47
default	nos	pre	10	63.53	0.62	62.73	64.70
default	nos	idle	10	60.35	0.85	59.36	61.58
trot	eb	pre	10	118.35	1.34	116.58	121.25
trot	eb	idle	10	118.83	2.46	115.54	122.51
trot	eg	pre	10	98.34	1.69	95.55	99.96
trot	eg	idle	10	98.34	1.24	96.43	100.01
trot	nos	pre	10	84.30	1.81	81.80	87.50
trot	nos	idle	10	79.70	0.75	78.88	80.84
bound	eb	pre	9	119.86	2.23	116.62	123.20
bound	eb	idle	10	113.74	1.13	111.93	115.76
bound	nos	pre	9	144.93	4.00	139.24	150.67
bound	nos	idle	5	158.19	2.49	155.18	161.58
free	eb	pre	8	81.18	1.48	78.61	83.59
free	eb	idle	12	83.08	1.37	81.01	85.59
free	eg	pre	10	90.46	2.18	86.95	95.57
free	eg	idle	10	89.72	1.02	88.36	91.23
free	nos	pre	10	61.24	0.51	60.42	62.09
free	nos	idle	10	59.81	0.59	58.69	60.58

Table A.4.: Complete statistics of power usage, standard deviation (Std) for each gait, style, and setting (Set)

Gait	Style	Set	Count	Mean force (N)	Std
default	eb	pre	56866	101.69	37.92
default	eb	idle	73225	102.60	32.25
default	eg	pre	53293	100.77	42.56
default	eg	idle	46100	102.91	36.49
default	nos	pre	84410	120.37	52.55
default	nos	idle	84636	117.49	38.43
trot	eb	pre	83817	124.52	75.22
trot	eb	idle	82886	116.59	65.87
trot	eg	pre	87251	126.03	61.29
trot	eg	idle	90771	129.09	60.88
trot	nos	pre	87826	123.73	55.99
trot	nos	idle	89483	126.13	46.80
bound	eb	pre	68816	116.88	63.24
bound	eb	idle	77213	112.69	55.63
bound	nos	pre	69202	126.53	77.43
bound	nos	idle	33533	118.19	67.12
free	eb	pre	61147	118.94	53.15
free	eb	idle	92633	119.42	42.91
free	eg	pre	77700	101.33	38.08
free	eg	idle	77519	97.44	32.61
free	nos	pre	86040	124.58	52.88
free	nos	idle	87544	125.08	45.78

Table A.5.: Complete statistics of average combined foot force, standard deviation (Std) for each gait, style, and setting (Set)

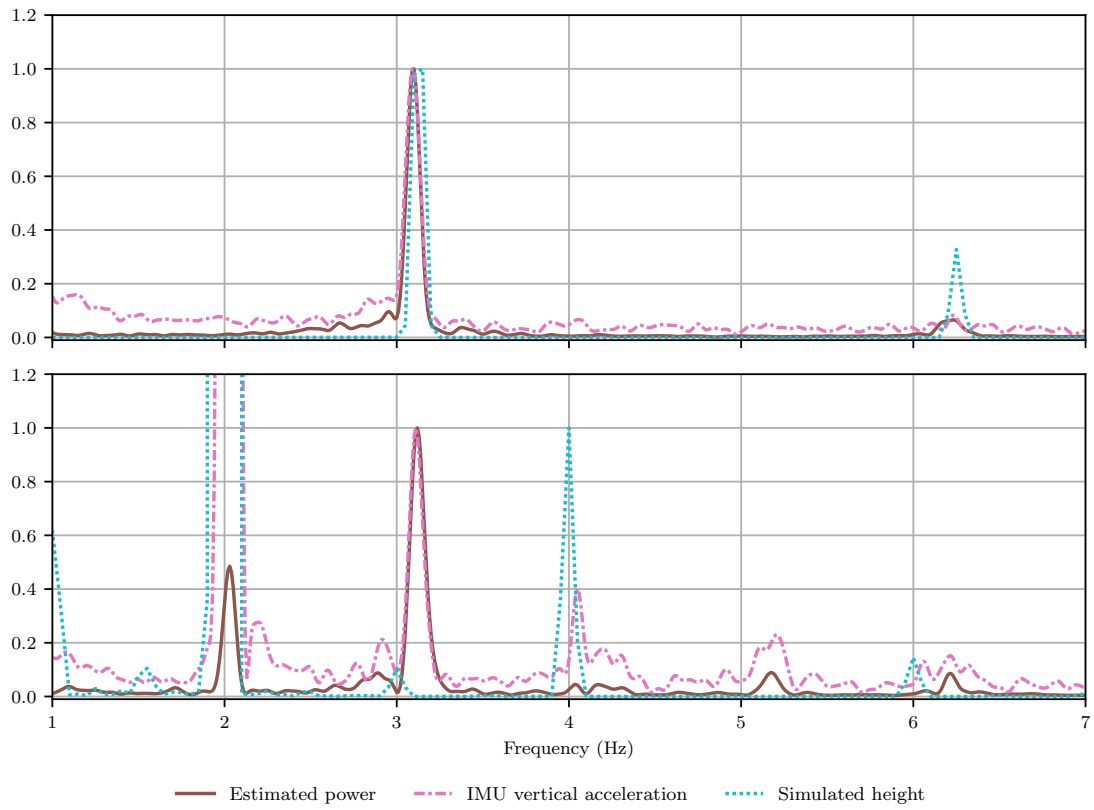


Figure A.3.: PSD of the *default eb* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

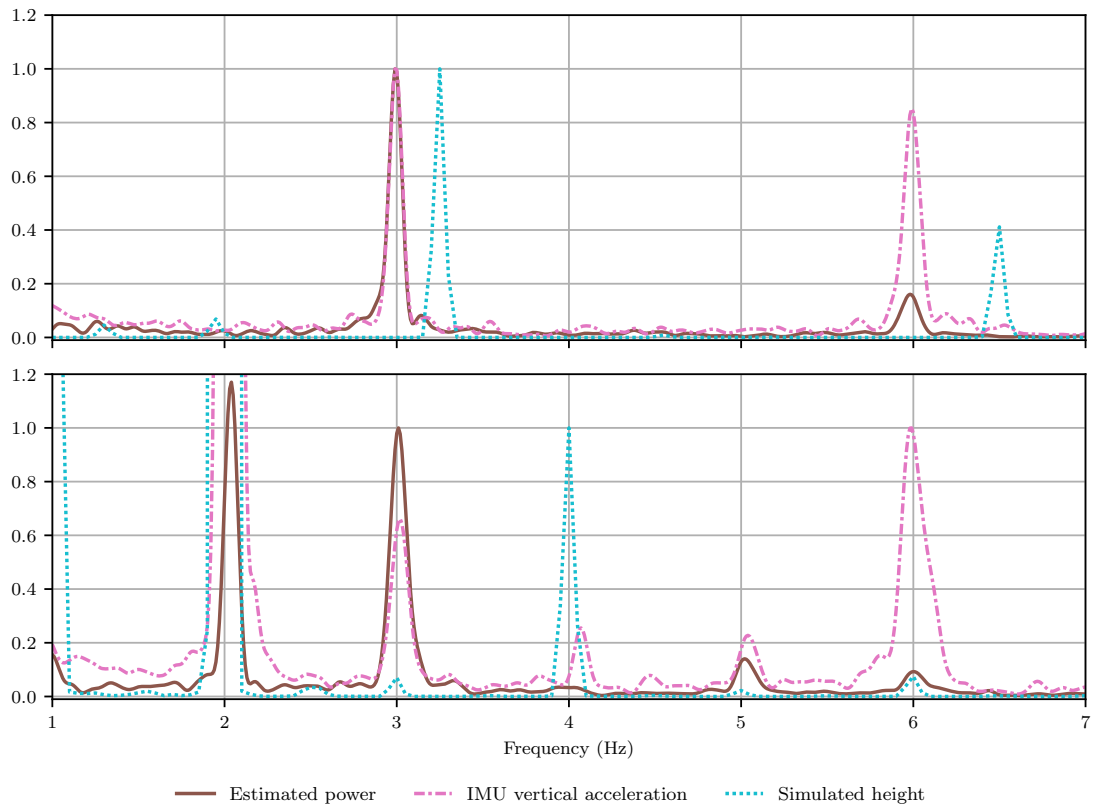


Figure A.4.: PSD of the *default eg* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

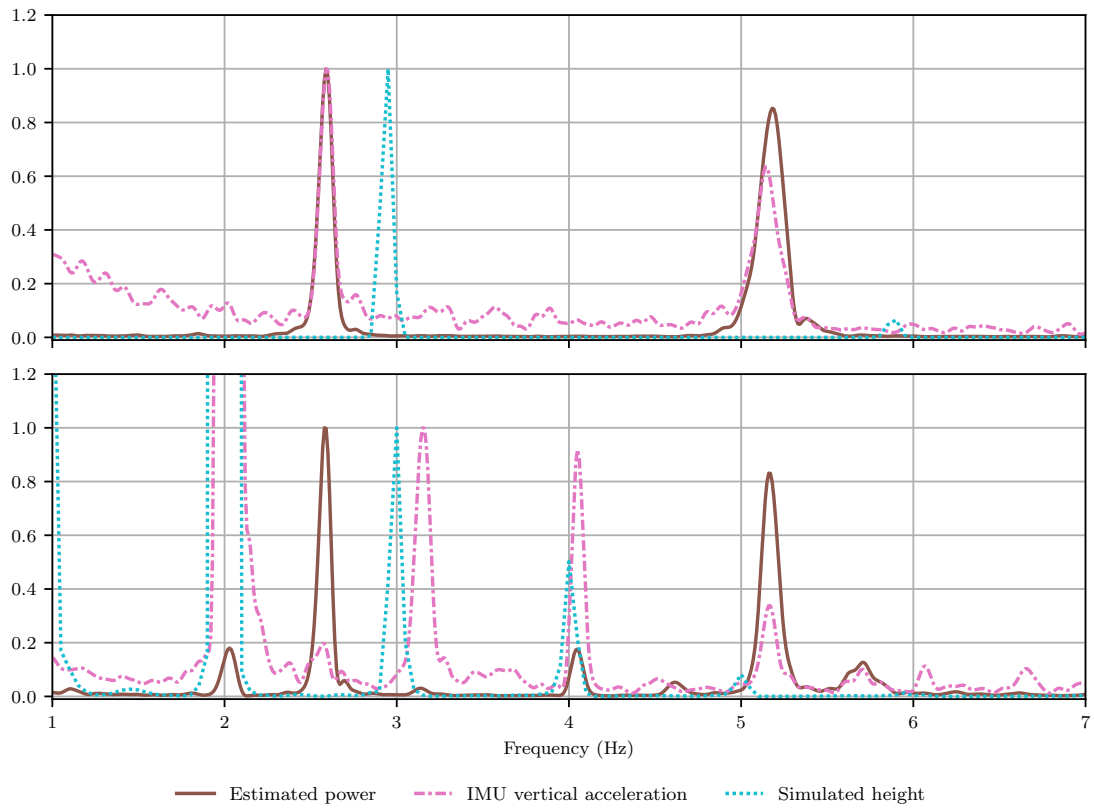


Figure A.5.: PSD of the *default nos* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

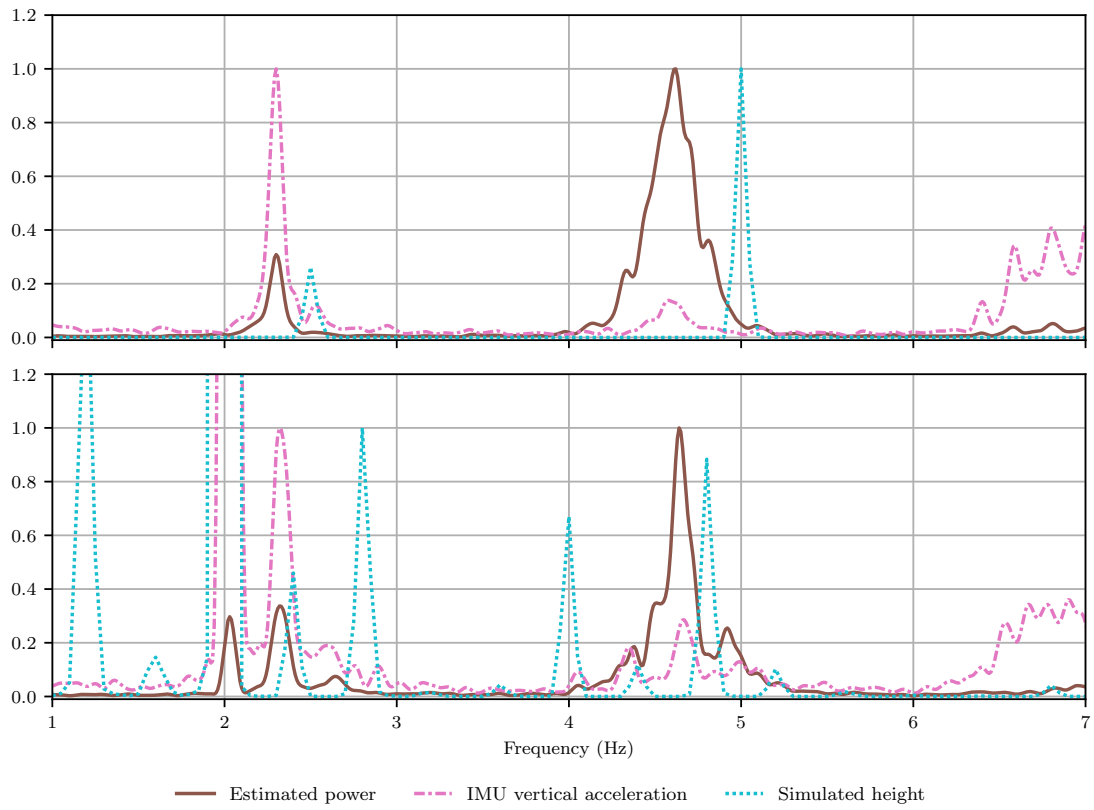


Figure A.6.: PSD of the *trot eb* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

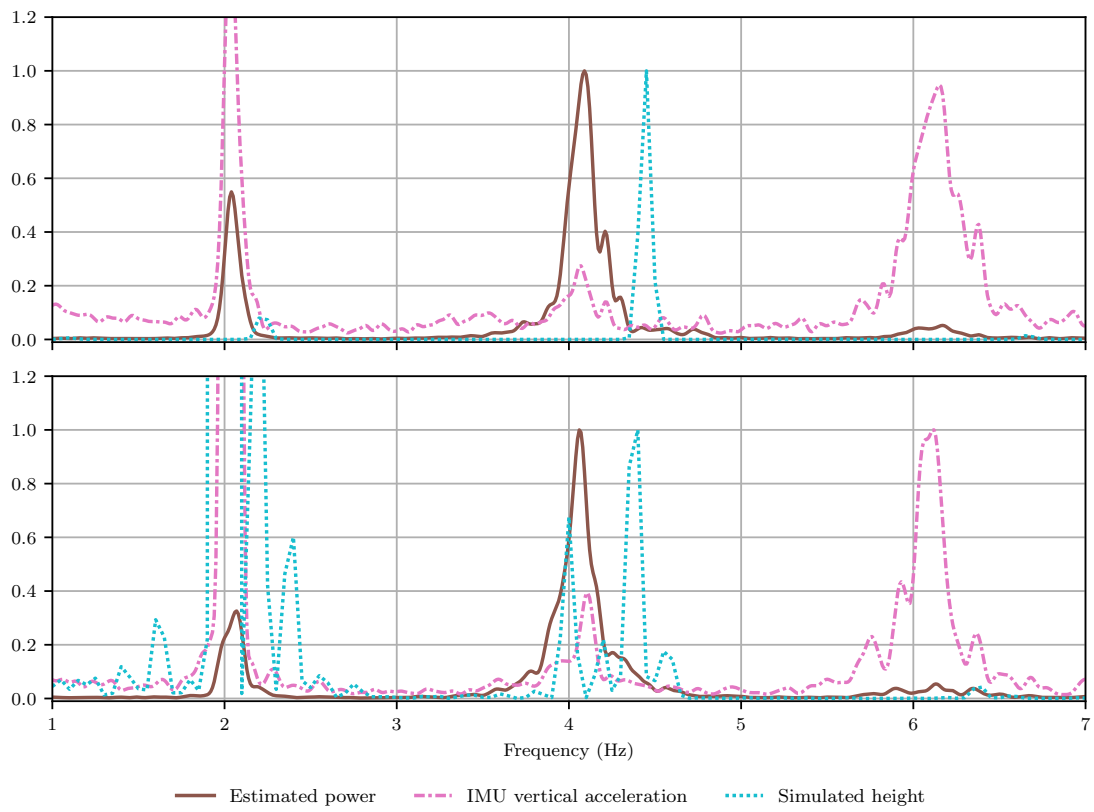


Figure A.7.: PSD of the *trot eg* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

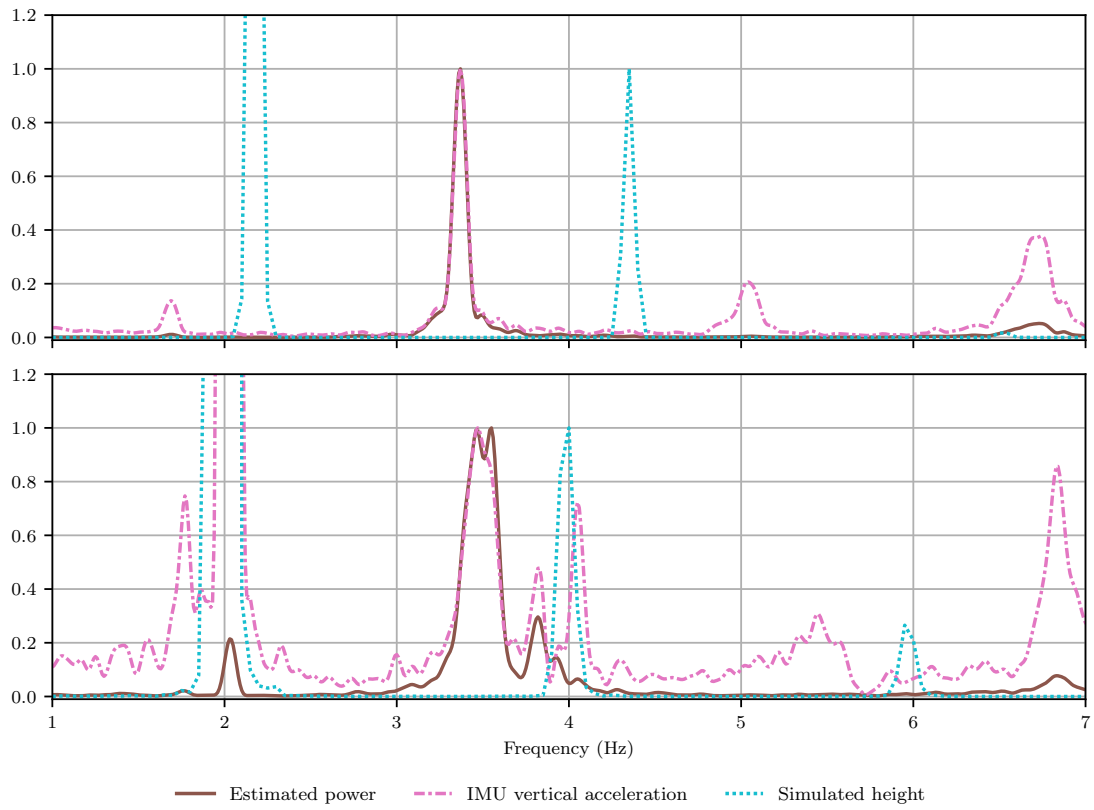


Figure A.8.: PSD of the *trot nos* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

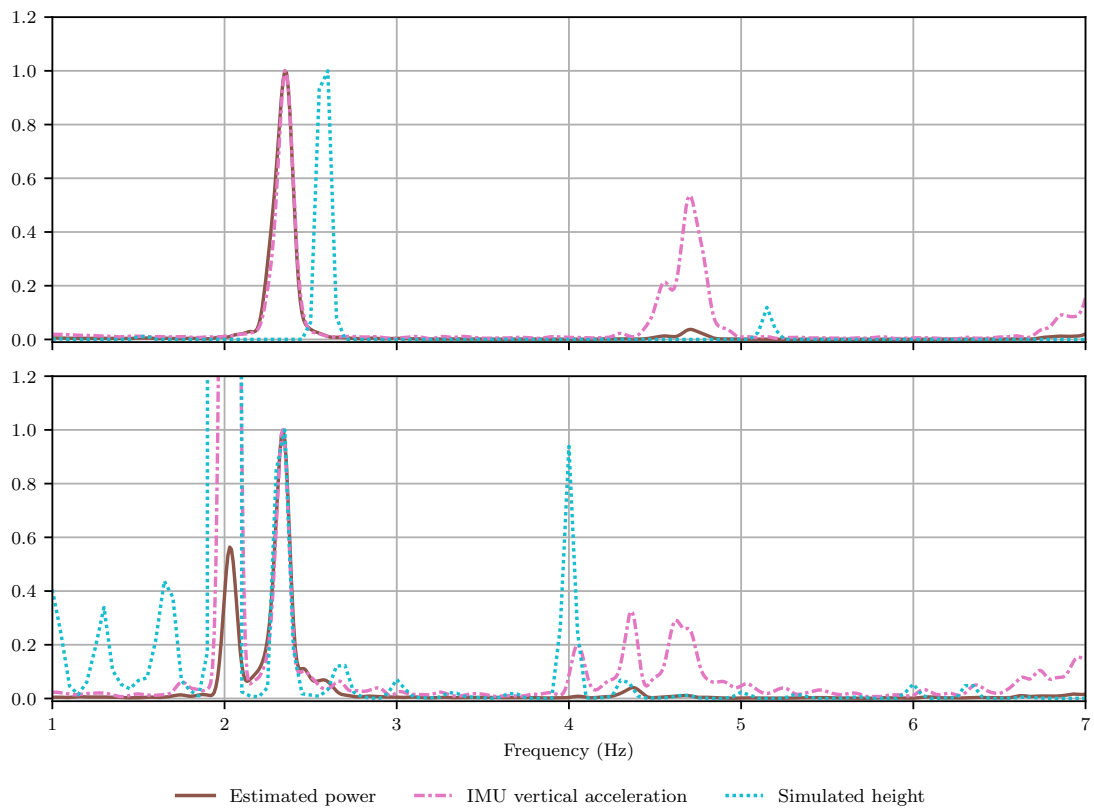


Figure A.9.: PSD of the *bound eb* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

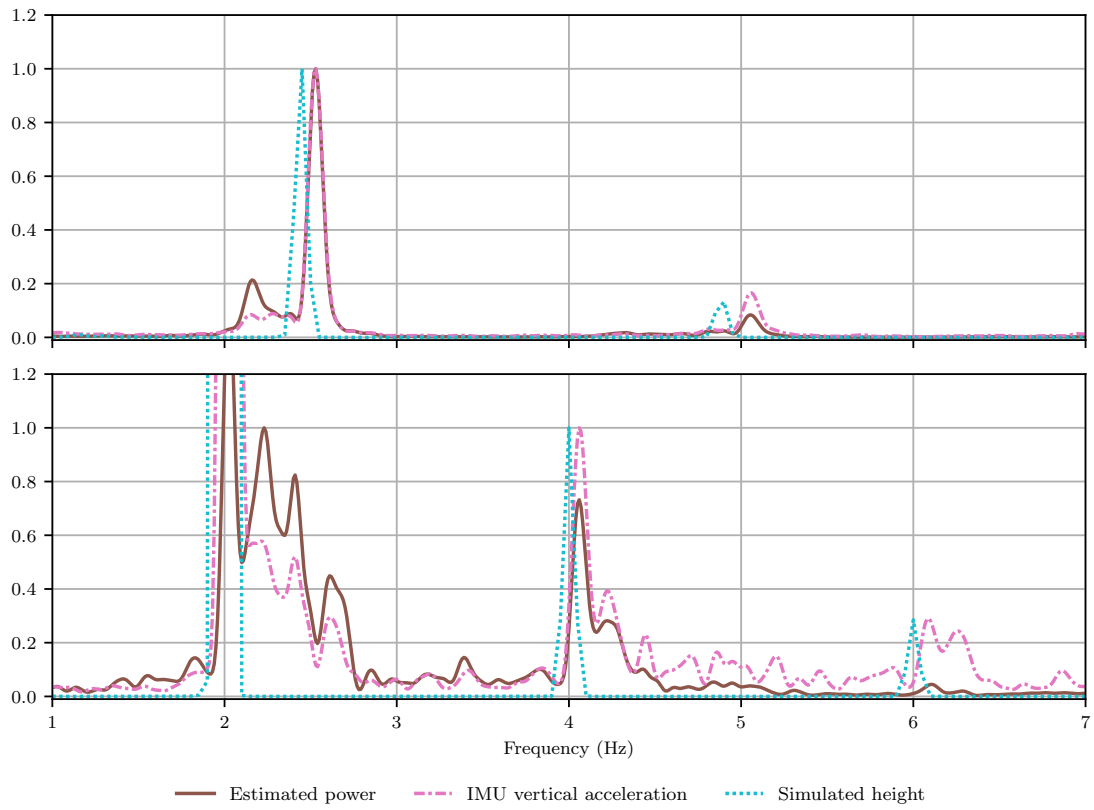


Figure A.10.: PSD of the *bound nos* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

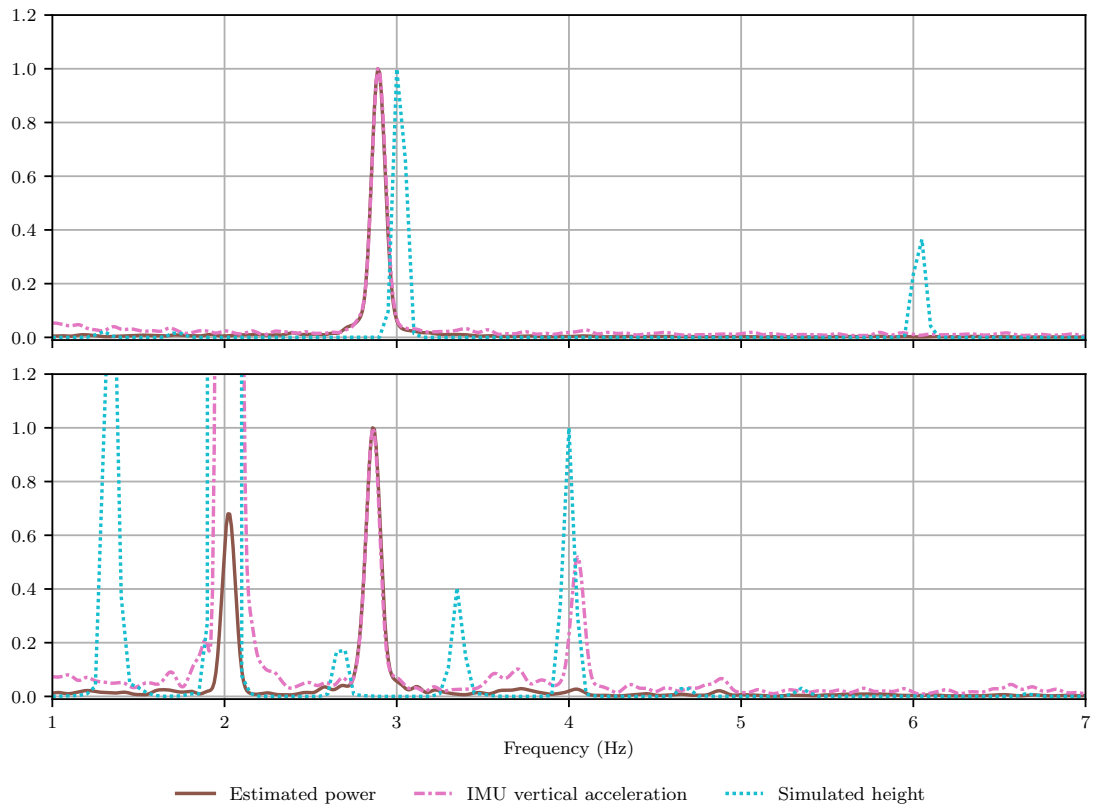


Figure A.11.: PSD of the *free eb* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

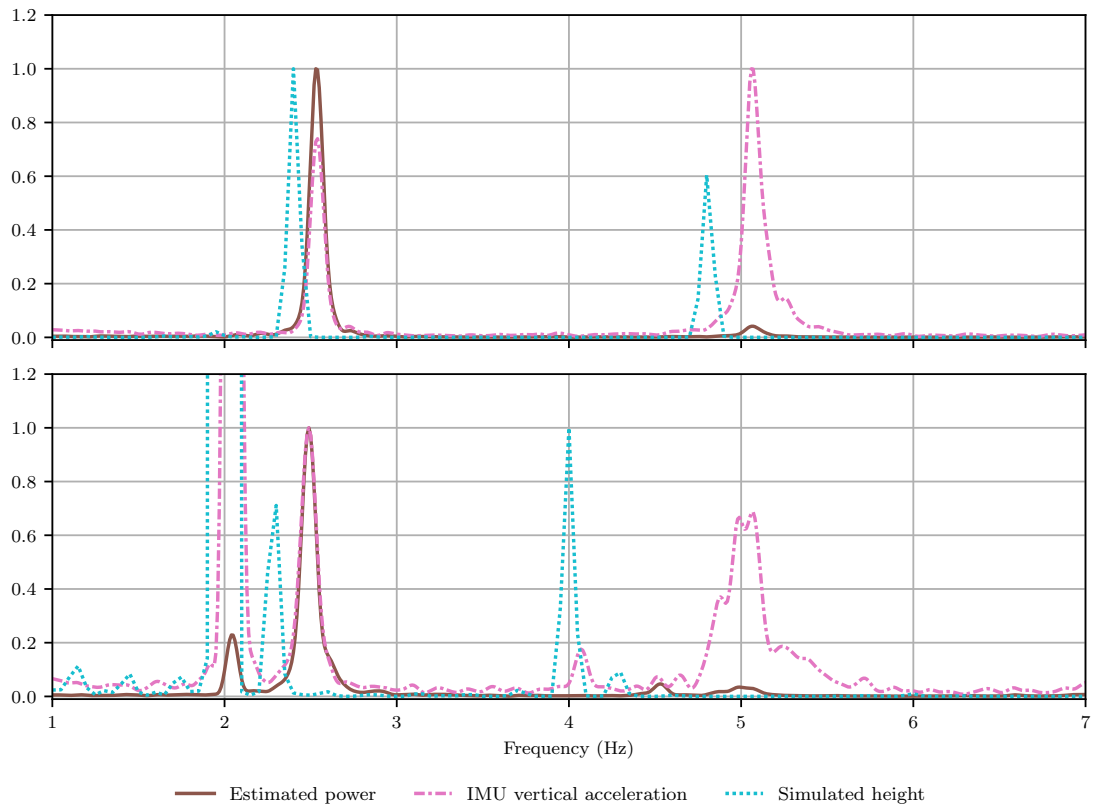


Figure A.12.: PSD of the *free eg* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.

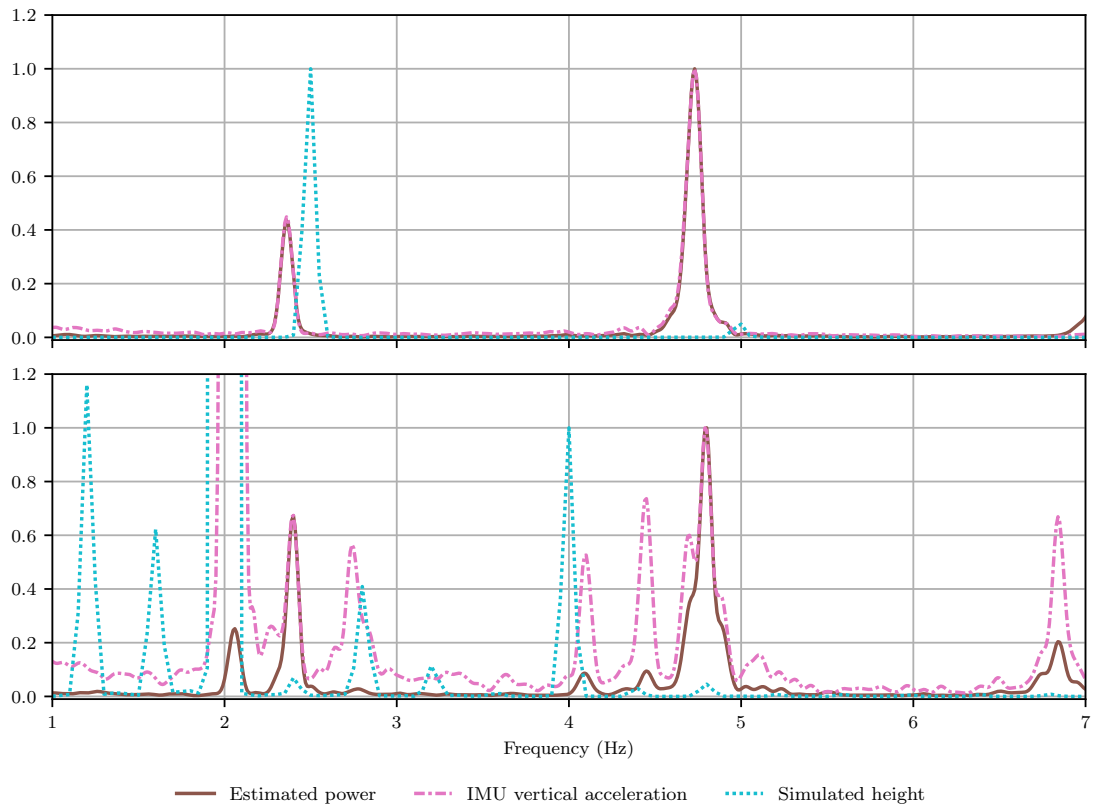


Figure A.13.: PSD of the *free nos* policy on the idle bridge (top) and the oscillating bridge (bottom). The power spectra are normalized to the highest peak above 2.2 Hz, comparing the estimated power, z-axis acceleration and the simulated height.