Towards Natural Versatility in Legged Robots through Bio-inspired and Learning-based Frameworks

Joseph Elliot Humphreys

University of Leeds

School of Mechanical Engineering

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

December, 2025

Intellectual Property and Publication Statement

I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution and the other authors to this work has been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Papers contributing to this thesis:

1. Humphreys, J., Peers, C., Wan, Y., Richardson, R., Zhou, C., 2022. **Teleoperation of a Legged Manipulator for Item Disposal**. In UKRAS22 Conference "Robotics for Unconstrained Environments" Proceedings (pp. 44-45). EPSRC UK-Robotics and Autonomous Systems (UK-RAS) Network.
    - Used in Chapter 3.
    - Humphreys, J. (myself): conceptualisation, formal analysis, algorithm design, control framework design, software design, simulation design, hardware experiment design, data collection and analysis, figure creation, wrote the manuscript.
    - Peers, C.: review and editing.
    - Wan, Y.: review and editing.
    - Richardson, R.: review and editing, supervision, funding acquisition.
    - Zhou, C.: review and editing, supervision, funding acquisition.
2. Humphreys, J., Peers, C., Li, J., Wan, Y., Sun, J; Richardson, R., Zhou, C., 2022. **Teleoperating a Legged Manipulator Through Whole-body Control**. In Annual Conference Towards Autonomous Robotic Systems (pp 63-77). Springer International Publishing.

- Used in Chapter 3.
- Humphreys, J. (myself): conceptualisation, formal analysis, algorithm design, control framework design, software design, simulation design, hardware experiment design, data collection and analysis, figure creation, wrote the manuscript.
- Peers, C.: review and editing.
- Wan, Y.: review and editing.
- Sun, J.: review and editing.
- Richardson, R.: review and editing, supervision, funding acquisition.
- Zhou, C.: review and editing, supervision, funding acquisition.

3. Humphreys, J., Peers, C., Li, J., Wan, Y. and Zhou, C., 2023. **High utility teleoperation framework for legged manipulators through leveraging whole-body control**. Journal of Intelligent & robotic systems, 108(3), p.57.
   - Used in Chapter 3.
   - Humphreys, J. (myself): conceptualisation, formal analysis, algorithm design, control framework design, software design, simulation design, hardware experiment design, data collection and analysis, figure creation, wrote the manuscript.
   - Peers, C.: review and editing.
   - Li, J.: review and editing.
   - Wan, Y.: review and editing.
   - Zhou, C.: review and editing, supervision, funding acquisition.
4. Humphreys, J., Li, J., Wan, Y., Gao, H. and Zhou, C., 2023. **Bio-inspired Gait Transitions for Quadruped Locomotion**. IEEE Robotics and Automation Letters, 8(10), pp.6131-6138.
   - Used in Chapter 4.
   - Humphreys, J. (myself): conceptualisation, formal analysis, algorithm design, control framework design, software design, simulation design, hardware experiment design, data collection and analysis, figure creation, wrote the manuscript.

- Li, J.: prior control framework development that this work uses to test the new proposed method, review and editing.
- Wan, Y.: review and editing.
- Gao, H.: review and editing, supervision, funding acquisition.
- Zhou, C.: review and editing, supervision, funding acquisition.

5. Humphreys, J. and Zhou, C., 2025. **Learning to Adapt through Bio-inspired Gait Strategies for Versatile Quadruped Locomotion**. Nature Machine Intelligence, 7, pp.1141-1153.
    - Used in Chapter 5.
    - Humphreys, J. (myself): conceptualisation, formal analysis, algorithm design, control framework design, software design, simulation design, hardware experiment design, data collection and analysis, figure creation, wrote the manuscript.
    - Zhou, C.: review and editing, supervision, funding acquisition.

Joseph Elliot Humphreys

December, 2025

Acknowledgements

First, I would like to thank my supervisor Chengxu Zhou for his unwavering support throughout this project. Whether it was discussing ambitious experiments or last-minute conference submissions, your insight and guidance have been invaluable. You have inspired me to reach greater heights than I ever thought possible, and for that I am truly grateful.

I would also like to thank Robert Richardson, George Jackson-Mills, and Andy Barber for helping me with all the requests, questions and challenges I had during the project. From sourcing equipment to providing feedback on experiments, you have all made the lab an amazing place to work.

To my friends and colleagues who have helped me with experiments, long days of filming demos, and fixing robots after the inevitable crash from pushing the robot just a little too far, I am sincerely grateful. Thank you for all your time and effort you have expended on my behalf, with a special mention to Jun Li, Yuhui Wan, Christopher Peers, and Jingcheng Sun.

I would also like to thank my Mum, Dad, Anna, Theo, and Johnny for their support and enthusiasm in my work, even when the main result was the robot taking a few steps before falling over.

To my partner Holly, who has essentially lived through this project just as much as I have. From supporting me in the toughest of times, to celebrating even the smallest steps forward, your belief in me means more than you know.

Finally, to Clarence, the best robot dog. You have put up with more trials, tribulations, and failed high-speed sprint attempts than the best of us. Without you, this project would have been very literally impossible.

Abstract

Quadrupedal legged manipulators (QLMs) exhibit great potential in replacing or aiding humans in hazardous and laborious jobs to mitigate risk to human health. However, these systems are incredibly complex due to their high number of degrees of freedom (DoF) and end-effectors. Hence, for successful deployment in these applications, where the environment is often complex and dynamic, a highly versatile control framework is required. For successful deployment, a framework must generate adaptable and optimal locomotion while undertaking complex manipulation tasks. Animals in the wild demonstrate highly proficient locomotion while interacting with their environment. Yet, existing frameworks fail to achieve both these aspects and struggle to generalise to the chaotic nature of the real-world.

This thesis presents a novel solution to this problem through the development of a bio-inspired hierarchical control framework, working towards achieving the same level of versatility that animals exhibit within QLMs.

This was achieved through instilling the abstracted proficiencies of animal locomotion within various levels of the hierarchical architecture. Animal gait selection strategies were instilled within a DRL gait selection policy, gait memory was embedded within a bio-inspired gait scheduler (BGS), and adaptive motions were realised by a DRL locomotion policy. To facilitate manipulation, whole-body motions, and refine the output of the upstream DRL policies, a whole-body controller (WBC) was developed as the final layer of the hierarchy. Initially, the control modules were tested and validated separately; the locomotion modules achieved zero-shot deployment on challenging terrain, while the WBC completed whole-body manipulation tasks. Once validated on hardware, all control modules were augmented, retrained and unified for loco-manipulation tasks using a QLM, while preserving their respective proficiencies. The resultant framework completed a range of dynamic loco-manipulation tasks in simulation, validating optimal gait selection, whole-body manipulation, and adaptive locomotion simultaneously across a variety of challenging terrains.

Contents

List of Figures

List of Tables

Abbreviations

BGS	Bio-inspired Gait Scheduler
CoM	Center of Mass
CoP	Center of Pressure
CoT	Cost of Transport
CPG	Centeral Pattern Generator
CV	Coefficient of Variation
DoF	Degrees of Freedom
DRL	Deep Reinforcement Learning
EKF	Extended Kalman Filter
GRF	Ground Reaction Force
ID	Inverse Dynamics
IK	Inverse Kinematics
IMU	Inertial Measurement Unit
LIP	Linear Inverted Pendulum
LTA	Learning to Adapt
LTA-M	Learning to Adapt with Manipulation
MDP	Markov Decision Process
MLR	Mesencephalic Locomotor Region
MPC	Model Predictive Control
OTP	Optimal Transition Point
PPO	Proximal Policy Optimisation
QLM	Quadrupedal Legged Manipulator
QP	Quadratic Programming
SE	State Estimator
WBC	Whole-body Controller

Chapter 1

Introduction

Since the birth of robotics, it has long been promised that the research of today will enable the robots of tomorrow to replace humans in strenuous jobs, mitigating risk to human health. Modern robots have the potential to be capable of such feats as they are well-equipped to traverse and interact with their surroundings. However, for proficient implementation, robots must be able to do this within both natural environments and a world built for humans, which are unpredictable, dynamic and contact-rich. To operationalise the challenges that robots must overcome to achieve this goal, the key criteria for effective implementation can be summarised as follows:

1. Exhibit optimal locomotion in terms of efficiency and stability through seamlessly adapting to the immediate environment.

2. Able to interact with the environment to complete complex manipulation tasks, facilitated through effective deployment of the system's utility.

3. Successfully undertake hybrid loco-manipulation tasks while exhibiting the same proficiencies detailed above.

However, robots are highly complex systems, often featuring a large number of degrees of freedom (DoF), making control considerably challenging. One robot type that features a reduced number of DoF is wheeled robots, which have been the subject of extensive research [1] and deployment in industry [2]. However, the inherent limitations of their form prevent them from tackling unstructured terrain such as large debris and

step-like obstacles[1], meaning they can never satisfy all three criteria. On the contrary, legged robots present the ideal systems for meeting the defined criteria, as they are able to navigate uneven and complex environments [3] while also able to complete manipulation tasks if outfitted with a manipulator arm [4]. However, despite significant progress being made in legged robot control, deployment and use of these robots outside structured environments remains sparse. Specifically, deep reinforcement learning (DRL) policies trained on the equivalent of years of data, and optimal control schemes meticulously crafted to account for highly dynamic control, still often fail to generalise past the environments they were developed in, and even in cases where generalisation is possible, their operational scope remains limited, leaving optimal locomotion unattainable. In turn, this prevents criteria points 1 and 3 from being realised.

However, when observing the legged robot's natural counterpart, legged terrestrial animals exhibit no such limitations. They traverse the natural world with ease, despite its inherently unpredictable and dynamic nature. The methods which animals utilise to achieve such proficiency are a well-studied field within biomechanics, with the most substantial and impactful findings coming from the study of terrestrial quadruped mammals, which focuses on the strategies they utilise to achieve efficient and adaptable locomotion [5, 6, 7]. Considering that legged robots and animals share many morphological characteristics, there is great potential to leverage the intrinsic principles of animal locomotion within advanced control frameworks for legged robots, enabling highly adaptable and optimal locomotion in unstructured natural environments, which would in turn resolve criterion 1.

Assuming the successful development of such a locomotion framework, this still leaves the challenge of manipulation unanswered; for successful implementation of robots within jobs designed for humans, they need to be able to interact with their environment to complete manipulation tasks in dynamic scenarios. To facilitate dynamic manipulation tasks within a legged robotic system, manipulators are often mounted on the base of the robot to form a legged manipulator [8] and can realise criteria 2. However, the locomotion aspect of such systems is often limited, being confined to use only within structured and static environments, which could be due to the inherent limitations of the leading control methods of the field, namely optimal control [9]

[1] While specialist wheel designs could overcome such limitations, this is impossible without conceding generality and efficiency on all terrains, while legged systems exhibit unrivalled adaptability

and DRL [10]. Optimal control methods are fully defined and interpretable, allowing precise control over system behaviour [9], but this comes at the cost of high computational demands. On the other hand, DRL enables highly efficient policy execution while considering full robot dynamics [11]. However, it sacrifices transparency and controllability, resulting in less predictable and potentially unsafe behaviour, in turn presenting a significant challenge during safety-critical manipulation tasks. Hence, for a robot to successfully complete challenging manipulation tasks while also exhibiting adaptable and optimal locomotion, in order to meet all criteria simultaneously, these advanced control methods need to be leveraged to capitalise on their proficiencies but augmented to offset their limitations.

It is clear that legged robots present great potential in completing highly dynamic and complex tasks, yet without further development, the proficiency of these systems will never attain that of humans in the same task. Therefore, to address this challenge, this project aims to investigate and develop methods of augmenting legged robot locomotion and manipulation to work towards the formulation of a loco-manipulation control framework capable of highly adaptable and optimal locomotion coupled with safe and versatile manipulation. To refine this aforementioned general project scope, the project research questions, aims, and objectives are as defined in the following sections.

1.1 Research Questions

1. How can current state-of-the-art control methods be improved and leveraged to facilitate highly adaptive locomotion and versatile manipulation?

2. How can advanced control methods be augmented to realise safe, dexterous and versatile manipulation capabilities in legged robot systems?

3. Can advanced control methods be integrated with the intrinsic principles of animal locomotion to develop locomotion control methods for legged robots, capable of realising highly adaptable and optimal locomotion in unstructured natural environments?

4. Can these methods be unified within a loco-manipulation framework to achieve simultaneous realisation of adaptable locomotion and versatile manipulation?

1.2 Project Aim

Investigate how characteristics of animal locomotion proficiencies can be instilled within modern control techniques in the development of a quadrupedal legged manipulator control framework to achieve adaptable, versatile and efficient dynamic loco-manipulation tasks on challenging terrain.

1.3 Research Objectives

To achieve the outlined project aims, the following objectives must be met:

1. Complete an extensive literature review on current state-of-the-art legged robot control frameworks, techniques and relevant technologies, along with the relevant findings within biomechanics that uncover the intrinsics of terrestrial animal locomotion strategies and behaviour.
   (a) Review literature surrounding optimisation-based techniques in legged robot control frameworks, including whole-body controllers (WBC) and model predictive control (MPC).
   (b) Review literature surrounding DRL-based methods in legged robot control frameworks.
   (c) Review literature surrounding other critical framework modules, including gait schedulers and state estimators.
   (d) Review literature within biomechanics that develops metrics and models that characterise animal locomotion.
2. Develop a WBC for a legged manipulator with an attached arm to facilitate safe and dexterous manipulation.
   (a) Formulate a WBC that is generic enough to accommodate a range of versatile and high-utility tasks and constraints to enable it to be utilised in manipulation tasks, but also to be integrated within a loco-manipulation framework.
   (b) Develop a teleoperation control framework to test and evaluate the performance of the WBC; a WBC cannot operate alone, as it requires reference trajectories.

(c) Test and evaluate the performance in a set of complex manipulation and basic locomotion tasks in simulation and on hardware, involving all key frames of the robot.

3. Develop a bio-inspired locomotion controller for optimal and adaptable locomotion.

   (a) Augment models and metrics used within biomechanics for use in locomotion control modules to instil core animal locomotion proficiencies and behaviours.

   (b) Create locomotion control modules using these augmented bio-inspired metrics.

   (c) Develop a locomotion framework to test and evaluate the developed bio-inspired control modules.

   (d) Deploy the framework in simulation and on hardware, in a range of complex and natural environments.

   (e) Investigate if animal locomotion proficiency attributes and behaviours have indeed been instilled within the control modules, and the extent to which optimal and adaptable locomotion is achieved.

4. Develop a loco-manipulation framework that unifies both WBC and bio-inspired locomotion controllers to simultaneously exhibit the proficiencies of all controllers.

   (a) Formulate the final loco-manipulation control framework.

   (b) Complete any augmentation to the controllers that are required to accommodate their integration into the loco-manipulation framework.

   (c) Deploy the loco-manipulation in simulation to complete dynamic loco-manipulation tasks on different types of terrain.

   (d) Evaluate the performance of the framework and assess if the locomotion and manipulation proficiencies are upheld.

1.4 Summary of Contributions

The contributions of this project to the field can be summarised as follows:

- A highly generic WBC with dynamic tasks and constraints for legged manipulators for effective DoF exploitation, in turn providing high utility, adaptable, and safe control for these systems.
- (Minor Contribution) An intuitive teleoperation framework for controlling complex robotic systems.
- A bio-inspired gait scheduler able to generate contact and Cartesian references of a diverse set of gaits online, with stable transitions between them, through leveraging animal gait strategies.
- A DRL locomotion controller for quadruped robots that is capable of zero-shot adaption to real-world environments without exteroceptive data.
- A DRL gait selection module for optimal gait selection, in terms of stability and efficiency, that applies metrics developed within biomechanics to instil animal gait strategies within it.
- A loco-manipulation control framework that unifies DRL, optimal control, and bio-inspired control modules for high adaptability, stability and efficiency in loco-manipulation tasks in challenging environments.

As a testament to this project's contribution to the field, throughout the course of the project, journals and conference papers were produced and published from its findings and results. Through publication, this ensures that the contributions of this project to the field are effectively disseminated and aid in pushing the boundaries of the current state-of-the-art. The papers produced from this project are as follows:

1. Humphreys, J., Peers, C., Wan, Y., Richardson, R., Zhou, C., 2022. **Teleoperation of a Legged Manipulator for Item Disposal**. In UKRAS22 Conference "Robotics for Unconstrained Environments" Proceedings (pp. 44-45). EPSRC UK-Robotics and Autonomous Systems (UK-RAS) Network.
2. Humphreys, J., Peers, C., Li, J., Wan, Y., Sun, J; Richardson, R., Zhou, C., 2022. **Teleoperating a Legged Manipulator Through Whole-body Control**. In

Annual Conference Towards Autonomous Robotic Systems (pp 63-77). Springer International Publishing.

3. Humphreys, J., Peers, C., Li, J., Wan, Y. and Zhou, C., 2023. **High utility teleoperation framework for legged manipulators through leveraging whole-body control**. Journal of Intelligent & robotic systems, 108(3), p.57.

4. Humphreys, J., Li, J., Wan, Y., Gao, H. and Zhou, C., 2023. **Bio-inspired Gait Transitions for Quadruped Locomotion**. IEEE Robotics and Automation Letters, 8(10), pp.6131-6138.

5. Humphreys, J. and Zhou, C., 2025. **Learning to Adapt through Bio-inspired Gait Strategies for Versatile Quadruped Locomotion**. Nature Machine Intelligence, 7, pp.1141-1153.

Papers 1-3 contribute to Chapter 3, paper 4 contributes to Chapter 4, and paper 5 contributes to Chapter 5. At the time of this thesis's submission, despite its clear contribution towards adaptable and efficient loco-manipulation in legged robots, the work in Chapter 6 is yet to be used in the production of a paper due to hardware testing becoming infeasible due to project time and resource constraints. However, following the submission of this thesis, work on these hardware experiments will commence in an effort to produce a paper from this chapter's contributions, subject to the success of these hardware tests.

1.5 Thesis Overview

The thesis takes the following structure:

- **Chapter 2: Literature Review**
  - Establishes the current state of the field and identifies critical gaps in current research, which leads to the development of the high-level targeted control framework architecture that the following chapters will work towards developing. Answers Research Question and Objective 1.
- **Chapter 3: Development of a Whole-body Controller for High-utility Manipulation**

  - Details the formulation of the whole-body controller, the creation of a teleoperation framework to enable the rigorous testing of the controller, its deployment in simulation and on hardware, and analysis of its performance. Answers Research Question and Objective 2.

- **Chapter 4: Leveraging Biomechanics for the Development of a Gait Scheduler**
  - Details the formulation of the bio-inspired gait scheduler, simulation and hardware deployment, analysis of its generality, a comparison to the resultant performance to animal behaviour, and a comparative study between this novel formulation and existing gait schedulers. Works towards answering Research Question and Objective 3.

- **Chapter 5: Instilling Animal Gait Strategies within a DRL Locomotion Framework**
  - Details the development of the DRL training environments that enable the instillation of animal gait strategies within DRL control policies for adaptable and optimal locomotion, further development of the bio-inspired gait scheduler, and deployment in simulation and on hardware. This leads to a study into the performance gain through instilling animal gait strategies, a study into the emergence of animal gait strategies within the robot performance in comparison to real animal data, and deployment of the control policies in natural environments and in critical stability loss scenarios to assess the realised adaptable and optimal locomotion behaviour. Conclusively answers Research Question and Objective 3.

- **Chapter 6: Unifying Bio-inspired Locomotion Control with Whole-body Manipulation Control to Achieve Dynamic Loco-manipulation Tasks**
  - Details method of unifying all developed control modules to form a loco-manipulation framework, and the development of the DRL learning environment to train the loco-manipulation policy. This leads to the deployment of the framework in simulation only and performance analysis in a set of loco-manipulation tasks to assess if the previously seen proficiencies of

the manipulation and locomotion control modules are upheld and expanded upon to realise dynamic manipulation tasks. Answers Research Question and Objective 4.

- **Chapter 7: Closing Discussion and Conclusions**
  - A review of the project to determine if all research questions have been thoroughly addressed and answered, a discussion into the successes and limitations of the project, and a proposition of potential future work directions.

Chapter 2

Literature Review

To refine the focus of this project, an extensive literature review has been completed to investigate the limitations of current control frameworks, identify potential methods to mitigate these limitations, and uncover the findings within biomechanics capable of augmenting legged robot control. The critical foundation for any control system is its architecture, as this not only defines the overall control philosophy and scope, but also the control modules required to be developed. As such, this literature review will first focus on establishing the current state-of-the-art legged robots and loco-manipulation control architectures. From this initial investigation, a legged robot morphology and control architecture will be selected, based on their potential to realise the success of the project as defined in Chapter 1, from which the following sections will focus on the control modules featured in the selected architecture. The literature review then moves on to investigating the findings within biomechanics that could be leveraged to realise optimal and adaptable locomotion within the control framework. Closing out, the literature review finished with a summary of its findings and the proposed method of overcoming the uncovered limitations within the field.

2.1 Legged Robots

From the rise in demand for legged robotic solutions and research, a wide array of legged robots of different morphologies have been developed, as presented in Fig. 2.1. By far the most popular type of legged robot within industry and the research field is the quadruped robot [12]. This robot strikes an effective balance between control capabilities and complexities with 12 DoF, while also mirroring the morphology of a

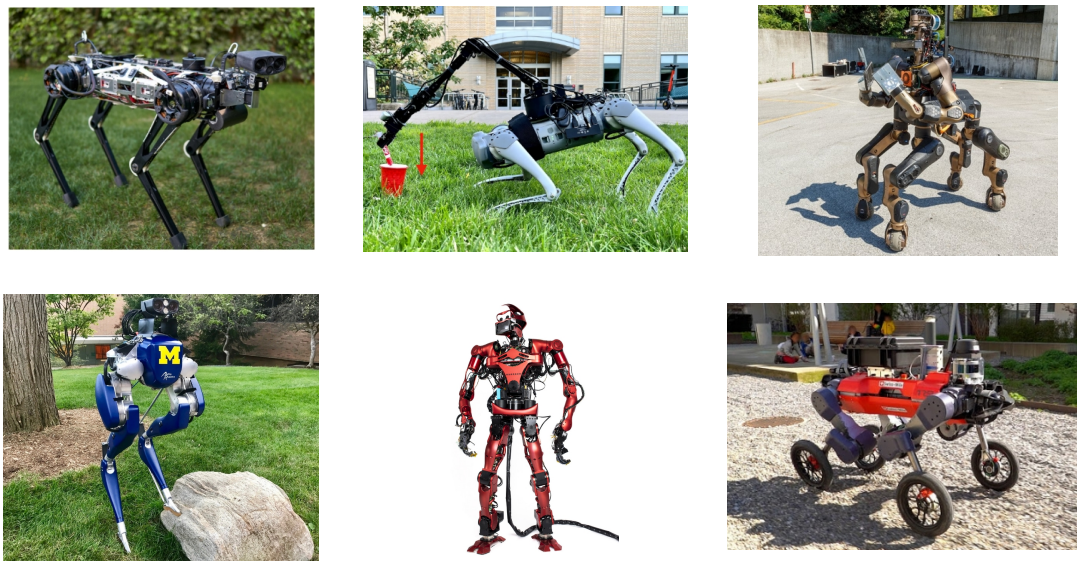


Figure 2.1: Examples of different legged robot morphologies: (from left to right) quadruped [12], quadrupedal legged manipulator [13], wheeled quadrupedal humanoid [14], bipedal [15], humanoid [16], and wheeled quadruped [17]

quadrupedal animal, allowing for investigations that bridge the gap between robotics and biomechanics. In turn, this has resulted in it being the subject of many applications and studies. Its morphology also offers inherent kinematic stability, a feature not present in bipedal morphologies [15], which aids in traversing over uneven and complex terrain. However, in this configuration, its manipulation capabilities are limited to tasks that only feature pushing via either its base or point feet. A popular method of overcoming this limitation, at the expense of increasing control complexity, is through attaching a manipulator arm to the quadruped's base [13]. This widens the scope of tasks which this robot can complete considerably, with it also now being capable of refined manipulation tasks (such as pick-and-place tasks or opening doors). Other variants aim to instil bimanual capabilities through taking on a partial [14] or full humanoid [16] form. Although this does successfully provide bimanual functionality, this dramatically increases control complexity and, in turn, computational demand due to the exceedingly high DoF these systems require [18]. A more radical approach to legged robot design is attaching wheels to their legs for simplified locomotion over flat terrain [17]. However, wheels are far from a natural phenomenon, in turn preventing learnings from the study of animals from being applied to augment the control of these

systems. With a core aim of this project focusing on leveraging the findings within biomechanics to improve legged robot locomotion, these wheeled systems lie out of scope. Considering these findings, the system with the best potential for achieving the objectives of this project is the quadrupedal legged manipulator (QLM), as not only can this system accommodate proficient locomotion and manipulation capabilities, with it sharing the same morphology as quadruped animals (excluding the manipulator), it enables almost direct application of biomechanics findings. Therefore, within this project, the robotic system of choice will be a QLM, with the only exception being the removal of the arm during work that focuses only on locomotion.

2.2 Loco-manipulation Control Framework Architectures for Legged Robots

Solving the complete rigid body dynamics and kinematics problems that govern legged robots involves computing huge equations involving thousands of variables. Although solving these equations would result in almost ideal motion generation for these robots, this is simply not feasible to realise in a real-time control application. To address this challenge, several different control architectures have been developed, namely hierarchical optimal control, end-to-end DRL control, and hybrid hierarchical DRL-optimal control. Within this section, these frameworks will be reviewed in an effort to identify which architecture should be applied within this project.

2.2.1 DRL Preliminaries

The training of DRL policies, whether within end-to-end or hierarchical frameworks, can be classified into two distinct methods: model-based and model-free. Model-based frameworks are much more sample efficient and have the potential to offer much better performance than model-free methods [19]. However, these methods are subject to easily inheriting bias within the trained model if there are any inaccuracies within the training environment compared to the real-world. To overcome this limitation, an option is to complete training on the real robot, which has seen some promising initial results [20, 21], yet the presented risk to the robot and the safeguards put in place to reduce this risk make this method impractical and results in limited performance. On the contrary, although model-free offers poor sample efficiency, the resultant policies

are much more resilient to training inaccuracies, easier to implement and much more stable. Model-free methods themselves can be further broken down into two different groups: Q-learning and policy optimisation. Yet again, although the main advantage of Q-learning is its enhanced sample efficiency [22], policy optimisation is preferred in the application of robotics due to its ability to directly target the desired performance. In turn, this provides improved stability and sim-to-real transfer [23]. More specifically, the variant of policy optimisation that has produced the best results in state-of-the-art control frameworks is proximal policy optimisation (PPO) [24] due to its ability to reduce undesirable characteristics being produced by the trained policy by limiting the changes within the policy between policy updates, a critical function in this application due to the complexity of legged robot systems where it is often seen to out perform its predecessors: deep deterministic policy gradient [25] and trust region policy optimization [26]. Other hybrid algorithms and methods have aimed to combine Q-learning with policy optimisation (DDPG), namely soft actor-critic [27] and twin delayed DDPG [28], to offset each method's limitations. However, their implementation is much more complex and unstable, making these methods much less popular than PPO.

2.2.2 End-to-end DRL Control

DRL for legged robot control has had a dramatic impact on the field, driving a shift from optimal control to a data-driven approach. The key to the success of DRL for this application is its ability to train control policies on the full dynamics of the robot without conceding computational cost during inference and deployment. In turn, this has led to the development of the end-to-end DRL control framework, where one DRL policy is trained on the robot's full dynamics to realise all targeted functionality through generating joint commands from high-level user commands. Such frameworks have yielded highly impressive locomotion capabilities, where they are demonstrated to be capable of high-speed locomotion [29], traversing complex terrain [30], climbing over obstacles [31], generalising to different platforms [32], and jumping across gaps [33]. Application of end-to-end DRL framework has also been expanded to loco-manipulation frameworks for QLMs to enable dynamic manipulation tasks such as object disposal [13] and opening doors [34].

However, due to the single trained policy within end-to-end frameworks being required to handle all major functionalities, from stability to reference tracking, careful

curation of their learning environments is required, resulting in highly complex reward functions and learning techniques to avoid local minima during the training process. In turn, without a direct method of guiding policy training towards the desired performance, this makes safe, adaptable and efficient deployment exceedingly challenging to achieve. Additionally, in the case of the application for QLMs, these frameworks still only offer simple locomotion and limited manipulator end-effector tracking performance. Specifically, a single DRL policy would struggle to encapsulate all functionalities that are required to realise dexterous manipulation and optimal locomotion. This is insufficient for realising this project's objectives, let alone implementation for jobs in hazardous environments.

2.2.3 Hierarchical Optimal Control

As opposed to DRL, optimal control offers a predictable and explicitly controllable method of solving rigid body dynamics without requiring large amounts of data. However, the main caveat of this method is its computational efficiency when it comes to real-time applications; solving the full dynamics of a legged robot system in real-time is simply not possible with the amount of computational capacity available on a deployed robot. To overcome this limitation, the highly popular hierarchical optimal control framework was produced; through dividing control functionalities into different optimal control modules, real-time control becomes possible. From high to low level control, these control modules split the control problem into reference trajectory generation, trajectory optimisation and joint optimisation. These take the form of gait pattern and manipulation trajectory generators, whole-body planners, and WBCs, respectively. A prime example of this is presented in [35], where such a framework is utilised to produce motion that mimics that of animal gaits for quadruped robots. This is achieved through using re-targeted animal data for reference trajectory generation, an MPC for trajectory optimisation, and an inverse dynamics (ID) based WBC for joint optimisation, as illustrated in Fig. 2.2. This results in dynamic and fluid motions while exhibiting seamless changes in gait phase. This general architecture of hierarchical reference trajectory generation, trajectory optimisation and joint optimisation control modules has seen wide use within the field. The most famous implementation of this architecture was produced from a line of research by ETH. First presented in [36], this framework features a gait pattern generator for reference generation, an MPC for

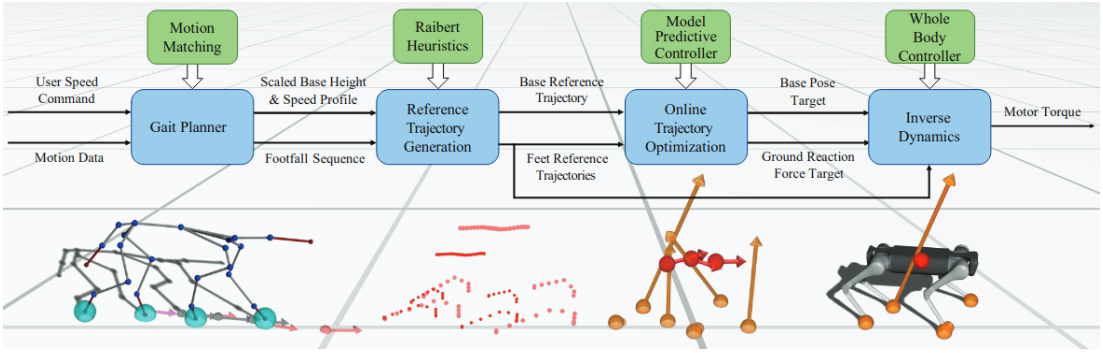


Figure 2.2: The hierarchical optimal control framework developed in [35].

trajectory optimisation, and an ID-based WBC for trajectory realisation. The strength of this framework lies in its formulation of an advanced and modular MPC that features centre of mass (CoM) and foothold optimisation for accurate modelling of the robot dynamics, which in turn produces optimal motion planning feasible for execution by the WBC. This framework is adapted for wheeled legged robots [37, 38], and legged manipulators [39] through further expanding the MPC formulation to account for the dynamics of wheels and manipulators. This results in a framework capable of generating dynamic motion quadruped-based robots, with a range of robotic augments, capable of navigating a wide variety of obstacles present in real-world applications, including stairs and heavy doors. This framework has also seen use to realise a wider variety of motions, such as different gaits [12] and jumping [40], in turn broadening the potential of these frameworks for adaptable control.

With optimal control offering predictable and direct control, this enables safe and accurate manipulation capabilities within legged robots. In [41], these features are expertly leveraged to achieve dexterous tasks, such as wiping a table while trotting, through defining tasks and constraints within its whole-body optimal control problem centred around manipulability to effectively capture this quality for enhanced manipulation with QLMs. Even more impressive manipulation capabilities are presented in [42], where trajectory optimisation is coupled with informed graph searching to enable the QLM to complete complex tasks such as opening spring-loaded doors and opening a dishwasher door initially with the manipulator arm and then closing it with the feet.

However, none of these aforementioned frameworks are able to exhibit this proficiency on natural, unpredictable and highly complex terrain. This is due to the simplifications made at the trajectory optimisation level. While DRL is able to train policies

on the full dynamics of the robot, the dynamics model used within MPCs is simplified to enable real-time control [43, 44]. This results in impressive simulation results, but significant stability issues when transferring to hardware in highly dynamic scenarios. There have been attempts at increasing the complexity of the dynamics model to use a centroidal dynamics model, such as in [45], but due to its non-linearity, computational speed was limited to an average of 60ms, making real-time control impossible.

2.2.4 Hybrid Hierarchal DRL-optimal Control

In the effort of trying to capture the strengths of DRL and optimal control methods, while offsetting each other's weaknesses, several works have developed hybrid frameworks featuring both classes of methods. Building around the standard hierarchical optimal control framework, [46] implements a suite of DRL policies to offset the model inaccuracies and limitations of the MPC, including a domain adaptive tracker and a recovery policy. However, this framework's training method and architecture are incredibly complex, as shown in Fig. 2.3, and still only allows for one gait to be deployed and only basic uneven terrain to be traversed. Similarly, the work in [47] also builds upon the existing hierarchical optimal control framework with a trained DRL residual controller applying adjustments to the output of the optimal control modules. Although this method is able to expand the robot's functional scope to include manipulation tasks using its feet, no deployment outside of flat terrain is exhibited, nor are multiple gaits achieved. In turn, this suggests that just building around an optimal control framework may not be sufficient in realising adaptable and efficient operation.

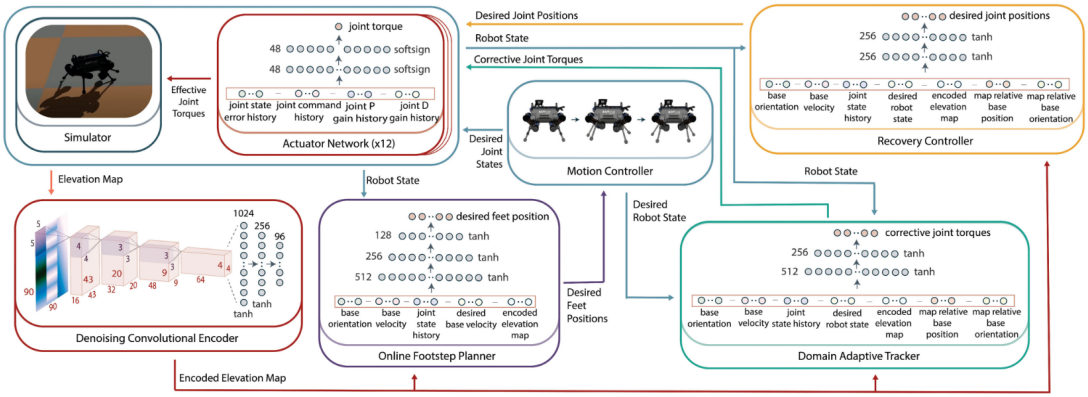


Figure 2.3: The highly complex hybrid control framework developed in [46].

Taking a more direct approach, the work in [48] completely replaces the WBC of the hierarchical optimal control framework with a DRL policy to generate joint actions from the input of an MPC and gait reference planner. This results in the policy being able to realise a wide set of gaits, presenting the potential to exhibit high adaptability despite results only showing deployment on flat grassy terrain. Although not being able to realise multiple gaits, the work in [49] also replaces the WBC with a DRL policy, but with the end goal of realising highly stable locomotion on complex terrain. However, both [48] and [49] have not been formulated to account for the addition of a manipulator upon the robot base; hence, with both being reliant on an MPC whose computational time scales with robot DoF, this presents a significant risk of being unable to generalise for use in QLMs.

A hybrid framework that is able to accommodate manipulation capabilities is presented in [51], where control is segmented into manipulation and locomotion; a DRL policy solely responsible for locomotion through generating joint commands for the quadruped base, while an MPC controls the manipulator. Although this framework results in impressive wrench control at the manipulator end-effector and realises stable locomotion while experiencing perturbations, the framework does not allow for the realisation of end-effector orientation trajectories; it is capable of position-only tasks, hence this prevents this framework from completing tasks that require high dexterity.

Taking a different approach, the methods in [50] and [52] instead replace the MPC with a DRL policy while keeping the WBC, as presented in Fig. 2.4. In both works, this results in stable locomotion in comparison to the standard hierarchical optimal control framework. While both works are relatively undeveloped, being the first step into exploring this type of framework, the reasoning behind the formulation appears to be strong. The bottleneck in the hierarchical optimal control framework is the MPC,

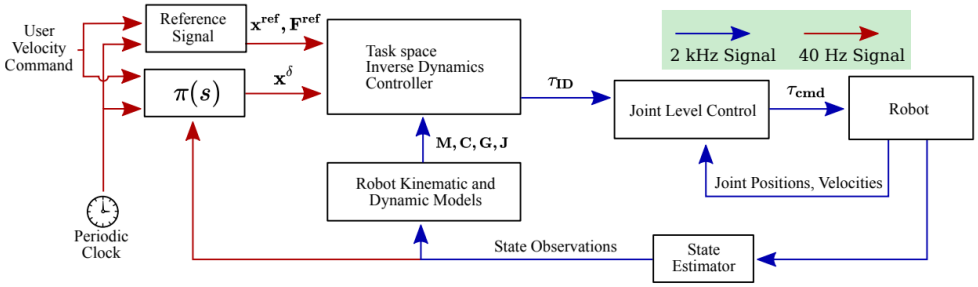


Figure 2.4: An example of a DRL-WBC framework developed in [50].

as it houses the reduced dynamics model; hence, if it is replaced via a DRL policy that can be trained on the full robot dynamics and still operate in real time, this presents great potential in realising adaptable and optimal locomotion while still preserving the safety and accuracy that the WBC provides during manipulation.

2.2.5 Architecture Selection

Overall, from this investigation into existing legged robot control framework architecture, it is clear that DRL methods far exceed that of optimal control for locomotion, due to their ability to capture the robot's full dynamics without sacrificing significant computation, while optimal control is much better when it comes to safe and dexterous manipulation tasks. In turn, this would manifest in a hybrid hierarchical DRL-WBC framework, similar to that in [50] and [52]. This would enable the DRL policy to generate references that consider the full dynamics of the robot and remove the limitations that an MPC would otherwise inflict, while the WBC would act as a safety and predictability layer for both locomotion and manipulation. Additionally, the WBC would be able to offset some functional scope from the DRL policy, removing the limitation that an end-to-end DRL framework would result in. Further on this point, the hierarchical architecture enables further functional scope reduction of the DRL policy through not requiring it to operate without reference motions, which should lead to more refined and diverse motions, similar to those found in [48]. As such, the first draft of the general framework architecture to be developed in this work is presented in Fig. 2.5. The following literature review sections will focus on uncovering how each module within Fig. 2.5 could be developed and potential methods of instilling animal locomotion proficiencies within it.

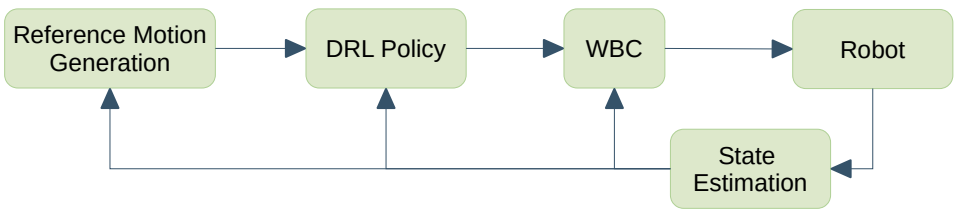


Figure 2.5: The first draft of the framework architecture to be used in this project.

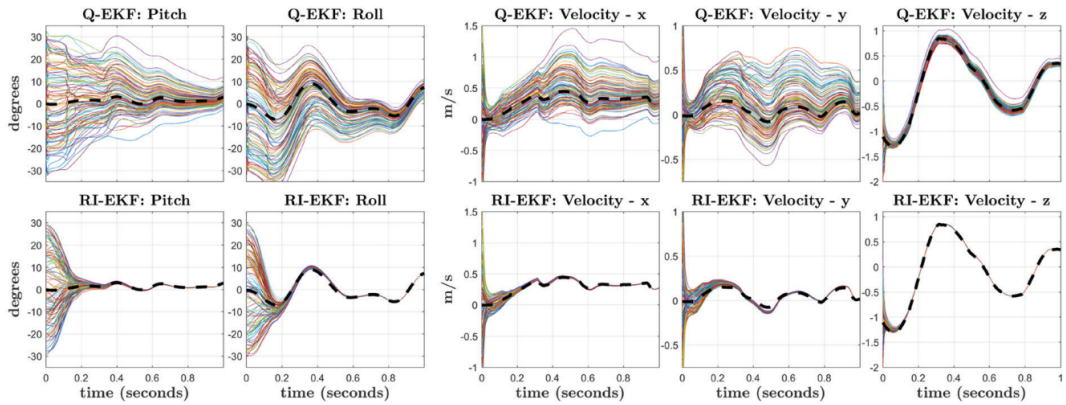


Figure 2.6: A comparison between a quarternion based state estimator and the invariant extended Kalman filter convergence rates developed in [53].

2.3 State Estimators

In simplistic terms, legged robots are floating bases on support legs, occasionally featuring extra limbs for manipulation purposes; hence, knowing the state of the floating base is paramount in producing an accurate dynamic and kinematic model of the robot. An accurate estimate of the floating base is impossible with pure sensor feedback. Consequently, a core component of a robust control framework is an accurate and fast state estimator. MIT implemented a decoupled state estimator in their Cheetah 3 robot, combining sensor feedback and an extended Kalman Filter (EKF), to estimate its state during various dynamic gaits [12]. However, no data has been provided to explicitly demonstrate its performance and accuracy. More rigorous research in state estimators has been completed by Fink and Semini, where, through combining an attitude observer, leg odometry, and a sensor fusion algorithm to create a state estimator able to initialise with an initial estimate containing large errors, omitting the divergence that would be seen with estimators built around an EKF [54]. However, over the course of 300 seconds, a significant drift between the state estimator's z position and the ground truth z is observed. A more advanced state estimator is developed by Hartely et al., who present an invariant EKF which utilises the contact state of the robot (in this case the Cassie bipedal robot) acquired from the foot contact sensors [53]. This work demonstrates convergence rates much better than other existing state estimators and has enabled the Cassie robot to complete prolonged mapping missions, as demonstrated in Fig. 2.6. However, no part of the state estimator accounts for the

variation in ground level present in uneven terrain. Pseudo measurements acquired via perception techniques could be implemented to solve this issue. Overall, highly accurate, fast and robust state estimators have been developed in research. However, they are a component often overlooked in the control frameworks in Section 2.2 despite playing a crucial role in prolonged legged robot operation. This presents an opportunity to further improve upon existing frameworks by adopting these aforementioned state estimators.

2.4 Gait and Manipulation Reference Generation

Throughout the frameworks discussed in Subsection 2.2.4, gait and manipulation references are vital in defining the immediate task and guiding the characteristics of task execution. Specifically, it is a popular approach to provide gait pattern references to define the gait style used by the robot, and Cartesian trajectories are provided for the completion of manipulation tasks. In this section, with the exception of manual reference design[1], the most popular and successful methods of doing so will be explored, namely gait schedulers for gait references, teleoperation for Cartesian trajectories for manipulation references, and offline motion capture for a general approach.

2.4.1 Offline Motion Capture

One approach for generating reference trajectories for both locomotion and manipulation is through collecting key frame and joint data from motion capture suites, videos, and manual human demonstrations to compile an offline expert motion library. For quadruped locomotion, current data collection methods include collecting this data from existing animation [55], motion capture suites placed on dogs [56], and video data from quadruped animal locomotion [57]. This data requires the use of motion re-targeting algorithms, but once complete, these generated motions can then be used to train DRL policies [57] or provide the input reference trajectories to MPCs [56, 55]. In turn, this offline motion library enables the framework to demonstrate impressive dynamic motions and a variety of gaits. However, the significant caveat of this method is that the proficiency of the resultant motion is entirely dependent on the quality of the

[1]This is not covered due to its simplicity, yet it is still implemented in this project due to its ease of use and enabling efficient experiment completion.

motion data and if the required motion is far from the reference motions of the library, this results in poor performance or even failure. With real-world environments often causing out-of-distribution states, this would exemplify the limitations of this method; such brittleness makes this method infeasible in dynamic environments.

For use in manipulation tasks, the most popular method of collecting motion data is through manually positioning the robot arm to complete the task while collecting state data, which removes the requirement of re-targeting methods and teleoperation systems. This data has then been used to train policies through imitation learning and DRL to the end of completing stacking tasks [58], vegetable peeling [59], and even playing hockey [60]. Although these tasks are impressive, in the application of using a QLM for dynamic loco-manipulation, this method becomes infeasible as it assumes a static base state and environment; neither of these assumptions is maintained during locomotion, which only further breaks down in real-world environments. Overall, although using offline motion capture to generate motion libraries can produce impressive results, this method is simply too brittle and restrictive for loco-manipulation tasks in complex environments where adaptability is critical. Therefore, this method will not be adopted for use in this project.

2.4.2 Gait Schedulers

Gait schedulers typically output a foot contact schedule for the robot to carry out in order to realise different gait styles, often designed based on real animal gait patterns. An example of this base-level gait scheduler is found in [12], where a phase variable and offset are used to provide an updated online contact schedule every timestep. However, in the scenario of a gait transition, a critical aspect of adaptable locomotion, the scheduler in [12] is only capable of instantaneous gait transitions with no method of generating transition references; the preservation of stability is left entirely to the low-level controllers. In highly dynamic or stability-critical scenarios, this method will surely fail without an appropriate transition design.

A popular method of generating stable gait transitions is through the utilisation of a central pattern generator (CPG) [61, 62, 63]. These works develop a simplified neural network to generate a cyclic gait pattern dependent on the robot's base velocity, where natural gait transitions emerge as this velocity increases. However, with CPGs being entirely dependent on the periodic cycles of the gait pattern being preserved,

this method is prone to failure in contact-rich environments, which could introduce asynchronicity into the system with irregular contact events. Another gait scheduler that features a transition method is presented in [64], where gait utility is leveraged to determine the output contact schedule. However, this method fails to capture efficiency as a determinant for gait transition, which is a vital characteristic in realising adaptable and efficient locomotion.

Directly leveraging locomotion efficiency to drive and govern gait transitions, the works in [65] and [66] train DRL policies to act as a gait scheduler and gait selection module (coupled with a standard gait scheduler), respectively. In both works, the efficiency is achieved via utilising the locomotion metric cost of transport (CoT), a standard metric used within robotics and biomechanics, which is further discussed in Section 2.7.2, formulated within the reward function of the trained policies. Although this does lead to efficient gait transitions, not only is stability not accounted for (potentially resulting in failures on rough terrain), but their generality is also heavily limited due to utilising a DRL approach and hence dependent on the training environment; motion references generated at this high level within a hierarchical framework should be as general as possible to not inflict any training bias or model limitations on the lower level control modules and hence creating a bottle neck.

Overall, for a gait scheduler to effectively facilitate a diverse set of gaits within a hierarchical control framework, while accounting for efficiency and stability, it needs to be generic, independent and simple enough to be suitable for a wide variety of environments and applications while also not bottlenecking performance at the high levels of the framework. None of the aforementioned methods is capable of achieving this, yet the method in [67] comes the closest to doing so through being independent from the framework and featuring a generic transition method. However, this framework does not account for stability or efficiency when generating these transitions, hence presenting the opportunity to augment this method to develop a gait scheduler capable of facilitating the realisation of this project's objectives.

2.4.3 Teleoperation

Although much progress has been made in autonomously generating target reference trajectories for complex task completion [68, 69], they are not proficient enough to handle sensitive tasks with high-severity risks or those that require specialist expertise.

An elegant solution to this challenge is to integrate a human-in-the-loop via a teleoperation system, which also removes the requirement of manual or complex Cartesian reference trajectory generation, although the best method of doing so remains a topic for debate. The most common method is through the teleoperator using a joystick, as seen in [70] for aerial manipulation and in [71] for quadrupedal manipulation. However, completing dexterous tasks using a joystick can prove challenging due to either limited controllable DoF [70] or the increased complexity caused by adding more controllable DoF [71]. In an effort to make locomotion control more intuitive, the work in [72] utilises the whole teleoperator's body to generate commands for the robot via a force plate that they stand on while measuring body tilt using a specialised suite to control a wheeled biped. However, this method is heavily restrictive on the teleoperator and does not facilitate manipulation. Coupling manipulation and locomotion within one teleoperation method, the work in [73] utilises 2 and 3 DoF controllers for locomotion and manipulation, respectively. However, this system still does not allow for dexterous manipulation tasks, as a minimum of 6 DoF trajectories are required for these tasks.

To overcome the inherent limitations of teleoperation via joystick, motion capture suits are a popular system to facilitate the generation of 6 DoF trajectories, seeing use to control humanoid robots [74], quadrupeds [75], wheeled manipulators [76], and manipulators [77]. As the trajectory generated by the teleoperator using this method is typically feasible and utilises human-in-the-loop feedback, these teleoperation systems are often coupled with WBCs to enable the robot to realise the input trajectory at the kinematic level while utilising its redundancy. This method has produced a range of impressive results, including tracking motion and muscle activity for control of a wheeled manipulator [78], retargeting motion capture suit data for generic robot control [79], direct kinematic tracking for control of a humanoid robot [80], and coupling motion capture suit data and joystick data to control a humanoid robot for low and high level control [81]. However, these methods are only able to map data between two morphologically similar systems–humans and humanoid robots, as shown in Fig. 2.7. In turn, this prevents effective teleoperation control of QLMs that provide enhanced stability, adaptability, and real-time control. The work in [75] is able to overcome this limitation through toggling between base and manipulator end-effector control, but in doing so, it has reverted back to using a complex joystick and only being able to control end-effector position and not orientation. Overall, the current methods of teleoperation

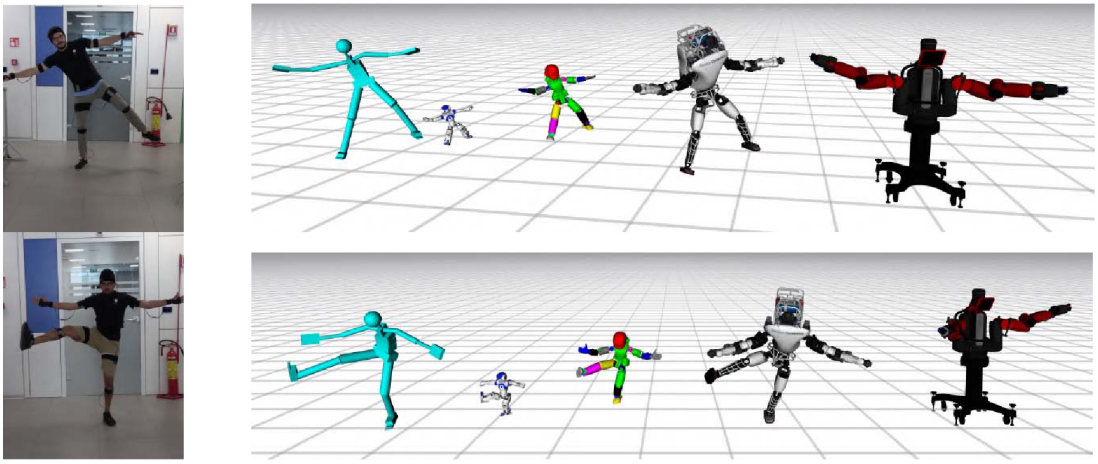


Figure 2.7: Although generic to the humanoid morphology, most motion capture suit teleoperation methods can't generalise outside this morphology [79].

for generating 6 DoF trajectories for dexterous manipulation tasks remain considerably limited. In turn, this presents the opportunity to further develop such systems for the complex manipulation tasks that this project's framework is aiming to achieve.

2.5 Whole-body Controllers

WBCs excel at utilising DoF redundancy to realise task-space operations in high DoF systems, such as QLMs. WBCs will typically consist of a quadratic programming (QP) problem, where the tasks and constraints dictate the output of the controller. These QP problems will be formulated around either a dynamics or a kinematics model, in turn constructing the two most popular forms of WBCs, ID or inverse-kinematics (IK) based WBCs. The most popular WBC type within the field of legged robots is the ID-WBC which optimise for joint torques and contact forces, allowing them to maintain stability and reject perturbations[1] hence finding use in QLMs [82], wheeled bipedal robots [83], and humanoid robots [84, 85]. The first meaningful development of an ID-WBC for quadruped robots was presented in [86], where stability is maintained through the formulation of a set of tasks and constraints. These stability-oriented constraints include contact force limits, joint limits, and contact constraints. As a result, this WBC enables a quadruped robot to traverse simple, unknown environments. To overcome

[1]Before the rise of DRL these WBCs provided much needed stability in these systems at the dynamics level.

this framework's primary limitation, in that it can not execute dynamic gaits, [87] presents a method of optimising for contact forces and joint accelerations rather than just joint torques. In turn, this improves stability to the point of enabling trotting, walking and pacing gaits. To enable the WBC to facilitate manipulation tasks, an ID-WBC is augmented to account for Cartesian tasks and constraints within [88] and [42] to realise position and force references at the manipulator end-effector while also completing a locomotive gait, as shown in Fig. 2.8. However, these controllers need manual tuning of many tasks and constraints, and solve at the dynamics level, which can result in a high computational cost. With the overarching goal of integrating a WBC with a DRL policy, as shown in Fig. 2.5, this mitigates the requirement for the WBC to account for the robot's dynamics, as the DRL policy will be trained to learn the full dynamics. In turn, this presents a strong case to develop an IK-WBC in this work instead, as not only will its simplicity (through only solving at the kinematic level) result in lower computational cost, as supported by the investigation in [89], but it also will not require the same magnitude of manual parameter tuning, hence reducing development time.

Figure 2.8: An ID-WBC being used to enable a QLM to undertake a loco-manipulation task [88].

Despite not accounting for dynamics, IK-WBCs still offer impressive performance, capable of rejecting high velocity disturbances in highly redundant systems [90], obstacle and singularity avoidance during loco-manipulation tasks [91], bi-manual manipulation tasks [92], and climbing stairs [93]. This is achieved through very similar methods and techniques to those used in ID-WBCs, but just at the kinematic level and solving for joint velocities rather than joint torques. However, the real strength of IK-WBCs lies in their ability to be generic and computationally efficient. This proficiency is highlighted within [94], where a generic formulation of the IK-WBC is presented, where tasks and constraints can easily be adjusted online for any open-chain system to account for a wide variety of tasks, constraints and behaviours. Although this work was exclusively

on humanoid robots, this IK-WBC appears to be the ideal controller to augment for use with QLMs for loco-manipulation tasks; not only does the inherent generality aid in developing a highly versatile framework, but it also presents the opportunity to investigate how effective IK-WBCs are when coupled with DRL for loco-manipulation tasks in QLMs, a pairing that hasn't previously been investigated. The hypothesised result is that the IK-WBC is able to reduce the operational scope of the DRL policy, reducing training time and refining performance, while the DRL can provide whole-body dynamics aware references to the IK-WBC, in turn offsetting its limitation of only considering kinematics. The logic of offsetting dynamic awareness from an IK-WBC to other control modules has seen previous success in [95] and [96] to improve stability, in turn providing confidence that this approach will be successful.

2.6 DRL Locomotion Controllers

The first significant breakthrough in DRL locomotion controllers was presented in [97], where for the first time such a controller exhibited stable locomotion and even recovered from a fall. This was achieved through leveraging a separately trained actuator network that was deployed only during training to improve simulation accuracy and consequently achieved successful sim-to-real transfer. This method was then further developed through applying a temporal convolutional network rather than a multi-layer perceptron, and utilising exteroceptive and proprioceptive perception methods during training and operation [98] to enable the robot to traverse natural terrain and complete a long-distance hike [99]. To expand this framework to realise locomotion over extreme obstacles, the work in [100] trains individual policies for select skills, such as climbing, jumping and walking, resulting in the framework being able to traverse large gaps and tall platforms. However, throughout these works, only one ground-level locomotion skill is deployable, in turn limiting its ability to realise optimal locomotion while also exhibiting adaptability; the framework only accounts for stability with no solution for optimally efficient locomotion. Additionally, in [100], a new policy has to be trained for each new skill, making the framework highly laborious to develop and brittle to further developments, as with the addition of a new skill, the high-level planning policies need to be retrained.

To improve the adaptability and optimal locomotion potential of a DRL controller, enabling it to deploy a diverse set of gaits is paramount; as identified in biomechanics, no

single gait is universally optimal across all scenarios [5, 6, 7] (this concept will be covered in detail in Section 2.7). To achieve this, one approach is to train individual policies for each gait and then unify them under one connecting policy [101] or use them to train a student policy [66]. However, similar to the limitations of [100], these development pipelines are laborious and don't account for optimal stability and efficiency to adapt to the immediate scenario ([66] does account for efficiency, but can only deploy three different gaits and only accounts for efficiency). The alternate approach is to train a single policy with gait references provided within the observation vector, resulting in a much more generic and adaptable DRL locomotion policy. This approach has been taken within [102, 103, 104, 105] in which a phase variable oscillates the desired state of each foot to be in either stance or swing while the offset between each leg's phase variable constructs the gait pattern, as described in Section 2.4.2. This method effectively enables the deployment of a wide variety of gaits with these policies on basic real-world terrain, as presented in Fig. 2.9.

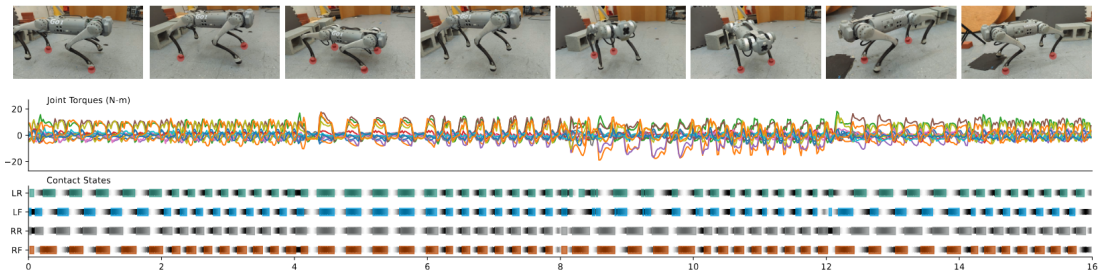


Figure 2.9: Deployment of the DRL policy developed in [104] to realise various gaits, where the top subplot shows the joint torques while the bottom plot shows the contact state of the robot's feet.

However, these existing frameworks exhibit several limitations, limiting their adaptability and constraining them to these basic terrains. The first is that the provided gait references are either only during training and in deployment, only the phase variable is provided [102, 103], or just contact-related references provided online [104, 105]. If the terrain varies significantly and causes earlier or later contact timings than defined by the gait references or experienced during training, this will likely cause a severe loss in stability. Secondly, none of these works feature a gait transition and selection strategy that accounts for stability and efficiency, as they solely rely on manual gait selection and linear interpolation of gait parameters to switch gaits, which will likely

fail in more complex environments. In turn, this prevents these existing frameworks from realising optimal and adaptable locomotion. Nonetheless, this presents the opportunity to instil animal locomotion behaviour and strategies within these types of DRL locomotion controllers to achieve highly adaptable and optimal locomotion; initial attempts of utilising gait references for multi-gait single policies appear effective, yet significant augmentation is required for realising the locomotion skills required to meet this project's objectives.

2.7 Animal Locomotion and Bio-inspired Control

Analysis in this section is published in Nature Machine Intelligence:

Humphreys, J. and Zhou, C., 2025. **Learning to Adapt through Bio-inspired Gait Strategies for Versatile Quadruped Locomotion**. Nature Machine Intelligence, 7, pp.1141-1153.

It has been postulated that findings from biomechanics could aid in the development of proficient locomotion in robotic systems. With the field of biomechanics being so vast, a scrutinous investigation must be carried out to identify the exact metrics and methods that can indeed improve legged robotic control.

Quadruped animals exhibit remarkable locomotion adaptability, shaped via innate and environmentally induced factors [106, 107, 108], through being able to deploy a diverse and rich set of skills. To refine the set of terrestrial quadruped animal locomotion skills and behaviours to be targeted within this project to achieve its objectives, the following locomotion classification is proposed based on if it can be characterised by the Froude number [109] (detailed in Section 2.7.1) or not:

- Froude-characterised locomotion
  - This includes walking, running, hopping, ambling, pronking, pacing, bounding, limping, cantering, galloping and other rhythmic gaits.
  - Can account for $\approx 70 - 90\%$ of the daily locomotion utilised by terrestrial animals [110, 111].
  - Gait transitions need to occur at both low and high velocity magnitudes, hence requiring the ability to rapidly change gait under dynamic conditions.

  - Features a desired velocity that has no normal component to the ground plane.

- Froude-free locomotion
  - This includes climbing, jumping, crouching and other non-rhythmic gaits or motion skills.
  - Can account for $\approx 10 - 30\%$ of the daily locomotion utilised by terrestrial animals [110, 111].
  - Gait and skill transitions occur at low velocity magnitudes due to the extreme differences in motion.
  - Features a desired velocity that has normal components to the ground plane.
  - Breaks the assumptions that are used in the formulation of the Froude number due to large variations in the height of the base relative to the ground.

With Froude-characterised locomotion being the primary locomotion utilised by animals to traverse terrain in the natural world, this highlights the significance of robots mastering this skill for optimal performance in the majority of scenarios. Furthermore, optimal deployment of Froude-characterised locomotion is more than capable of realising this project's objectives; hence, it is this highly adaptive gait-based locomotion and transition strategies seen in animals that we aim to capture within this project. Froude-free locomotion could help to further expand the framework's operational scope; however, to also achieve this within the project's time frame is simply infeasible and hence out of scope; achieving adaptable and optimal Froude-characterised locomotion is more than sufficient.

The rest of this section will investigate the specific metrics and methods used within biomechanics that can be leveraged to realise adaptable and optimal Froude-characterised locomotion and existing works that have also aimed to achieve this.

2.7.1 Froude Number

Based on the dynamic similarity hypothesis [112], derived from the linear inverted pendulum (LIP) model, the Froude number can be used to characterise gait selection and transition behaviour between animals of the same morphology, irrespective of their height and mass, based on their velocity. More specifically, it has been found that

animals exhibit gait transitions at the same Froude number across a wide variety of different animals of diverse heights and masses [109]. This proves to be a powerful tool in providing a generic method of predicting locomotion behaviour, and in turn, this presents the Froude number as being an ideal metric to govern gait transitions; as per the dynamic similarity hypothesis, this metric should easily transfer to quadruped systems, as supported by the Froude number's application in [66]. However, the Froude number studies have so far only focused on nominal locomotion skills, such as trotting and running, whereas animals are capable of a much larger repertoire of gaits and skills, in turn suggesting that more than just the Froude number needs to be investigated and leveraged for proficient robot locomotion.

2.7.2 Animal Gait Selection Strategies

The leading hypothesis for the driver behind animal gait transitions, and the consequential emergence of the relationship described by the Froude number, is that they are trying to preserve their efficiency through switching to a more efficient gait for their current speed [113]. However, these studies only focus on animals locomoting on flat terrain with only forward linear velocity while using standard gaits. While in the wild, animals adapt their locomotion by employing nominal gaits such as ambling, trotting, and running [114], while switching to specialised gaits like hopping, pronking, and bounding for off-nominal tasks such as predator evasion or obstacle navigation [115]. In turn, this suggests that many other factors other than just efficiency are at play in more dynamic and off-nominal scenarios. In the following sections, the main drivers for animal locomotion strategies discovered within biomechanics are investigated to identify their associated characteristic metrics.

Cost of Transport

The cost of transport (CoT) metric is a dimensionless quantity that characterises how much energy has been expended during locomotion normalised by the system's mass and speed:

$$\mathrm{CoT} = \frac{P}{mgv}, \tag{2.1}$$

where P is the power consumed, m is the system's mass, g is the gravity constant, and v is the system's velocity. In the case of animal locomotion, P is found through

measuring how much CO_2 is generated and O_2 is consumed, and v is assumed to be the speed of the treadmill the animal is running on [116, 117, 109]. Within these studies, a strong relationship between CoT and switching gaits is uncovered; CoT increases with speed, but after an apparent threshold is breached, a transition to a more efficient gait is initiated, resulting in a reduction of CoT as exhibited in Fig. 2.10. Additionally, due to the dimensionless nature of the CoT metric, data can be gathered across animals of different sizes and weights to further support this discovery, in turn demonstrating its generality. However, although it appears that in nominal locomotion, animals will often choose the most efficient gait, in many other scenarios, other factors could become a higher priority, such as stability when traversing rough terrain. Following this logic, it would be reasonable to assume that CoT alone is not sufficient for characterising animal locomotion strategies.

Musculoskeletal Forces

In several studies, experimental results suggest a link between musculoskeletal forces and the initiation of a gait transition to reduce the stress within the body, measured through foot ground reaction forces [118, 119, 120]. This factor seems to act as a hard trigger, with the gait transition being initiated once a foot reaction force threshold is reached due to the animal increasing its locomotion speed, a trend observed across horses of different sizes and weights. These results were strengthened in [118], where some of the horses completed the experiment carrying and not carrying a payload, with the horse initiating a gait transition at the same ground reaction force in both experiments. However, this method seems flawed in that it would increase the instability of the horse due to a shift in their centre of mass, which is not addressed in [118]. Consequently, although this metric seems useful for preserving the safety of a system, further investigation of the link between this metric and system stability might further uncover the causes of this observation.

Mechanical Work Efficiency

Some studies within biomechanics have discovered a link between preserving mechanical work efficiency and animals transitioning gaits [122, 123, 124]. Mechanical work losses arise due to the variation of the animal's base height during a gait cycle; if they are to have perfect mechanical work efficiency, there would be a net zero change in

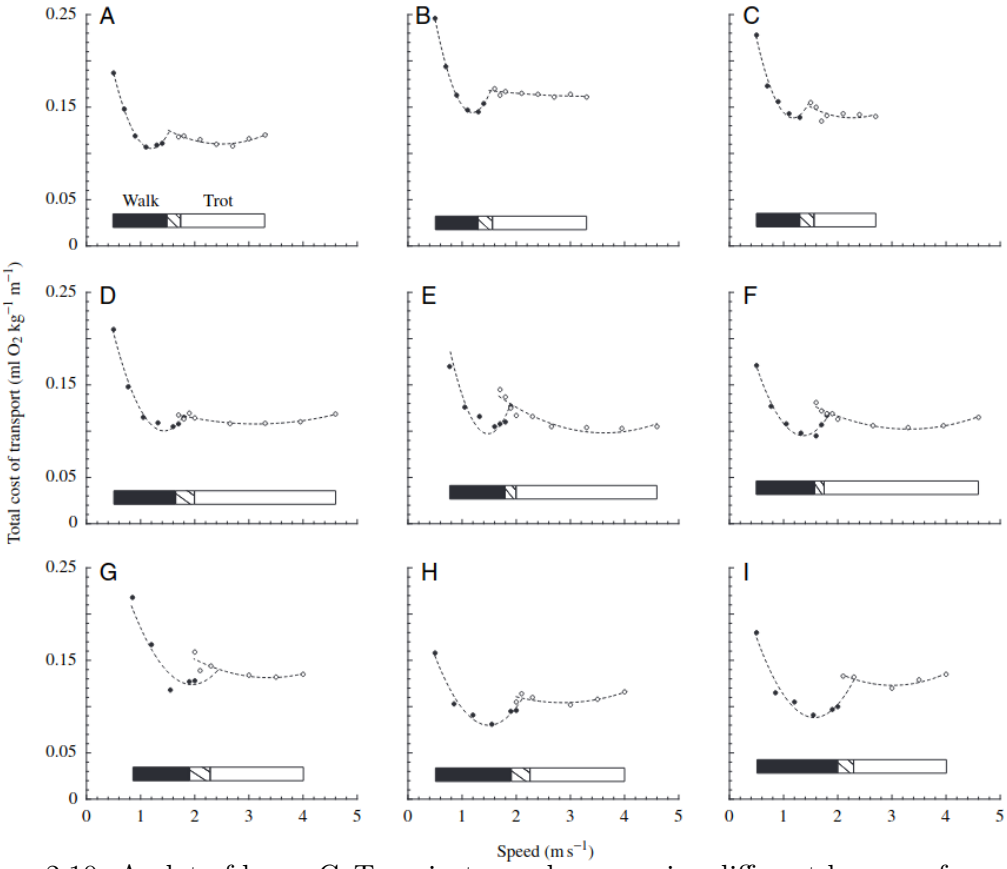


Figure 2.10: A plot of horse CoT against speed across nine different horses, of varying physical statures, to investigate the relationship between CoT preservation and gait transitions [113].

external work over the duration of a gait cycle, as there would be perfect exchange between kinetic and potential energy. Hence, mechanical work efficiency in animals is characterised by the sum of the change in kinetic and potential energy [122] or the sum of external work of the animal over the duration of a gait cycle [123], calculated via base height variations and O_2 uptake. Switching to a gait that has a higher frequency or prolonged aerial phases can reduce this variation and improve locomotion efficiency. However, large variations in base height might also have strong links with stability, particularly when traversing uneven terrain. Hence, similar to the trend related to musculoskeletal forces and gait transitions, this metric would also benefit from further investigation.

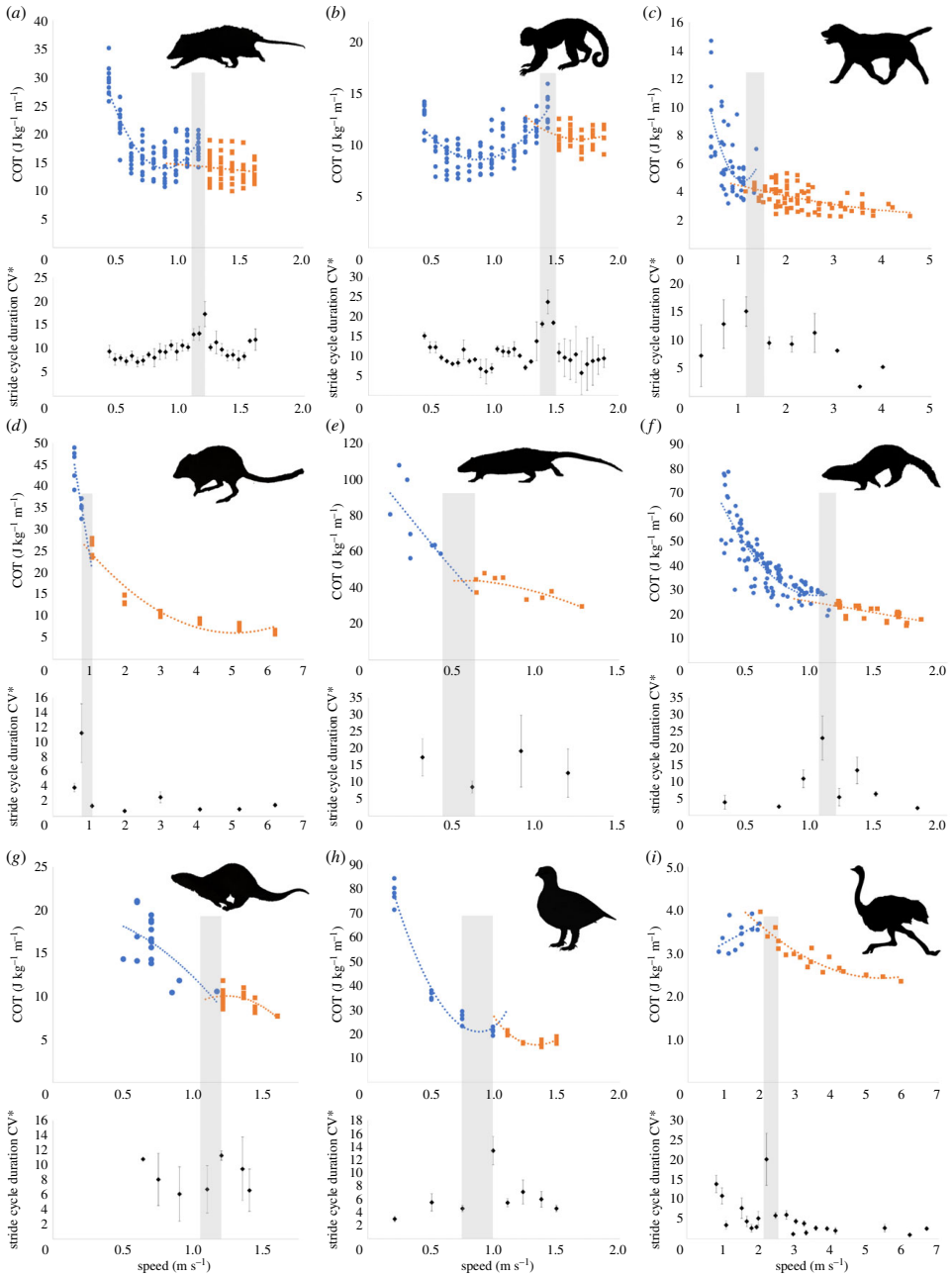


Figure 2.11: Inter-species results of CoT and stride CV against speed, exhibiting how gait transitions are able to reduce CoT and stride CV to improve efficiency and stability [121].

Stability

It has been discovered that stability also drives changes in gaits across different species of animals, spanning significantly different sizes and weights [121, 125]. Within [121], stability is characterised through the coefficient of variation (CV) of gait periodicity over a test period (typically a set of gait cycles). Explicitly put, as animal gaits are highly rhythmic, if there are large variations in foot contact timings between gait cycles, this indicates increased disturbance and instability within the system. As such, the data within [121] shows a strong correlation between an animal changing gait, a reduction in CV and hence an increase in stability as the animal increases its speed across a diverse set of quadruped animals, as presented in Fig. 2.11. This metric is particularly useful as not only can it characterise stability during nominal locomotion but also during stability critical scenarios. In turn, this metric should prove highly valuable for application in robotic locomotion control frameworks, but also investigate the relationship between the aforementioned metrics and stability.

Overall, it is clear that although the focus of individual studies is on these metrics, they are all interlinked in some aspect, as inefficiency could be the product of instability and vice versa. As such, there is strong logic to support the notion that rather than a single metric being responsible for governing animal locomotion strategies, it is instead the case that all metrics could play a role and could uncover the intricacies of animal transition strategies through their unification, a hypothesis also arrived upon in [121, 126, 127]. Furthermore, as quadruped robots also utilise gaits originally inspired by quadruped animals [66, 67] and share the same morphology, this presents a strong case for it to be possible to apply these metrics to instil animal locomotion proficiency within a locomotion control framework.

2.7.3 Bio-inspired Locomotion Control

Although sparse, there have been several works that have recognised the potential of leveraging findings from biomechanics for enhanced locomotion. The most popular of these existing methods utilises central pattern generators (CPG) for gait motion generation. CPGs, modelled around the mesencephalic locomotor region (MLR) of animals [128, 129] in an attempt to mimic their neural mechanisms for locomotion, utilise coupled oscillators that communicate with each other to determine a rhythmic

contact schedule and foot trajectories, while also considering foot contact forces [130]. Through the use of CPGs, quadruped robots have been shown to realise spontaneous gait transitions [61] and mimic certain animal behaviours [131]. However, their real-world performance is rather limited, often only capable of unidirectional motion and counterintuitively unnatural behaviour. This limited performance arises due to the contract-rich environment the natural world presents, which often causes irregular contact states. This results in instability within the coupled oscillators of CPGs. An alternate approach for harnessing animal locomotion behaviours is through utilising animal motion capture data and remapping this data onto quadruped robots [132]. The most advanced version of this method is presented in [133], where dog motion capture data is used to train several DRL policies to learn animal-like motions (including walking, running and jumping) at the low-level, and learn how to play at the high-level (although this playing behaviour was not inherited through utilising animal data). However, as this method focuses on directly realising low-level animal motions, there is no guarantee that the resultant locomotion is optimal; it is unreasonable to assume that mechanically different systems would perform efficiently with the same low-level behaviour. CPG and motion capture methods both attempt to replicate animal locomotion mechanisms precisely, yet both exhibit prevalent limitations. In turn, these limitations suggest that instead locomotion frameworks should be augmented with high-level attributes derived from animal locomotion to instil the proficiency observed in nature. Furthermore, there has been no attempt within the field to couple bio-inspired locomotion with manipulation capabilities, which could explain both why current loco-manipulation frameworks exhibit limited functionality, but also present a significant opportunity to develop and investigate this approach in this project.

2.8 Summary of Literature Review Findings

Through investigating the different current control framework architectures for loco-manipulation in the field, it became clear that only through taking on the form of a hybrid hierarchical DRL-optimal control framework would it be possible to maximise the potential of achieving the project's objectives. However, with many different potential methods and techniques that could be utilised for each control module of the framework, a thorough investigation of each component in the proposed framework in Fig. 2.5 was completed. From this investigation, Fig. 2.5 is redrafted as Fig. 2.12 with

the specific control modules and methods selected from the findings. For reference motion generation, a gait scheduler and motion capture teleoperation will be used for locomotion and manipulation, respectively, with teleoperation being interchangeable with manual design when suitable. This is due to their highly generic yet proficient capabilities for enabling adaptable locomotion and dexterous manipulation agnostic to the formulation of the more complex control modules of the framework. This also decreases project risk as the final formulation of the WBC and DRL locomotion policy cannot be predicted in the early stages of the project. For state estimation, as this does not form a significant aspect of this project's contribution, the SE used in [134], which draws upon the works explored in Section 2.3, will be implemented as its proficiency is sufficient for this project.

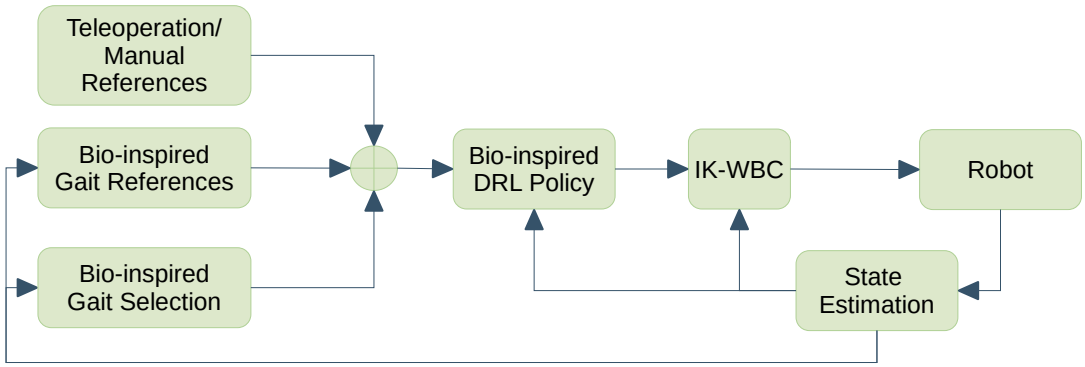


Figure 2.12: A revised draft of the high-level loco-manipulation framework architecture to be targeted based off the findings of the literature review.

For achieving dexterous and high utility manipulation, it has been opted to utilise an IK-WBC, as although it may not account for system dynamics, the computational efficiency this method achieves presents it as the perfect controller to provide generality, predictability and safety while integrated with a DRL policy that does account for the full robot dynamics, offsetting the limitations of DRL methods. Furthermore, coupling an IK-WBC with a DRL policy for loco-manipulation tasks has not previously been attempted, presenting the opportunity to investigate the validity and proficiency of this proposed framework architecture. For the development of the DRL controller, this will be trained using a reference tracking approach for gait and manipulation realisation, as this will enable multiple gaits and skills to be contained within a DRL policy. However, as identified, simply passing gait references to the policy during

training results in significant inherent limitations to arise in terms of optimal, stable and adaptable locomotion. As such, implementing findings from biomechanics that study animals that exhibit no such limitation presents the ideal opportunity to overcome these aforementioned challenges. From reviewing the relevant biomechanics metrics and models, along with current attempts at integrating biomechanics with legged robot control, it appears that taking a high-level approach to instilling these behaviours would yield the best results; animals and robots are fundamentally different systems, so it would be unreasonable that the same low-level methods would be just as effective across them. This approach will be applied to gait reference generation, the DRL locomotion policy, and gait selection in the effort of achieving the same locomotion proficiencies seen in animals. This approach will be used the most extensively in the development of the gait selection module; animal gait strategies primarily revolve around optimal gait transitions. Therefore, effective implementation of the previously discussed animal gait selection strategies within this module will be critical in instilling the same behaviour within the framework in 2.12. The following chapters will work towards developing its core control modules of Fig. 2.12.

Chapter 3

Development of a Whole-body Controller for High-utility Manipulation

The first control module to be developed of the framework presented in Fig. 2.12 is the inverse kinematics whole-body controller (IK-WBC), as this will form the foundation of the framework by ensuring joint-level commands sent to the robot are feasible for hardware deployment. In this section, a teleoperation framework is also developed as although this does not form a main contribution of this work, it is a necessary tool to develop in order to test the IK-WBC as not only is manual trajectory design impractical and unrealistic for real-world deployment, but it also enables real time stress testing of the IK-WBC as this presents a very challenging use case of all developed functionalities. Work contributing to this chapter has been published in the following journal and conferences:

- Humphreys, J., Peers, C., Wan, Y., Richardson, R., Zhou, C., 2022. **Teleoperation of a Legged Manipulator for Item Disposal**. In UKRAS22 Conference "Robotics for Unconstrained Environments" Proceedings (pp. 44-45). EPSRC UK-Robotics and Autonomous Systems (UK-RAS) Network.
- Humphreys, J., Peers, C., Li, J., Wan, Y., Sun, J; Richardson, R., Zhou, C., 2022. **Teleoperating a Legged Manipulator Through Whole-body Control**. In Annual Conference Towards Autonomous Robotic Systems (pp 63-77). Springer International Publishing.
- Humphreys, J., Peers, C., Li, J., Wan, Y. and Zhou, C., 2023. **High utility teleoperation framework for legged manipulators through leveraging**

whole-body control. Journal of Intelligent & robotic systems, 108(3), p.57.

3.1 Introduction

As identified in Chapter 2, to enable dexterous manipulation capabilities in systems of high DoF redundancy that are aware of hardware limits and kinematic feasibility, an IK-WBC is to be developed for use within Fig. 2.12. However, the existing IK-WBCs reviewed in Chapter 2 provide limited functionality, in turn reducing the variety and complexity of the tasks they are capable of completing. The breadth of functionality a framework provides can be classified by the amount of utility they have; these frameworks of limited functionality have a low level of utility. This low level of utility seen in existing IK-WBCs dramatically reduces their potential for successful deployment in challenging environments and tasks. This apparent limitation is exacerbated when considering that real-world tasks could require utilising several frames of the robot and a fine level of control. Therefore, to develop a framework that can handle tasks in challenging environments, they must have high utility to tackle the large variety and high complexity these tasks present.

Considering this, the work in this chapter aims to not only develop a high utility IK-based WBC but also a teleoperation framework that facilitates this utility and the analysis of the WBC. In essence, this is realised through leveraging a WBC within a teleoperation framework to achieve high utility. In turn, this would allow a teleoperator to complete complex tasks efficiently via controlling the main frames of the robot, along with providing useful functionalities. This utility is heightened when considering that if the foot frames of the legged manipulator can be controlled for simple manipulation tasks, this mitigates the requirement of additional arms being added to the system, which also preserves payload capacity, and in turn would rigorously test the performance of the WBC. However, it should be noted that foot manipulation capabilities will be limited to this chapter; this functionality is only meant to demonstrate the IK-WBCs' generality and versatility while becoming redundant during the loco-manipulation tasks of Chapter 6.

As discussed in Chapter 2, there is a range of hardware solutions for generating the teleoperation commands to be passed to the IK-WBC. Traditionally, joystick controllers are used to generate teleoperation commands; however, not only can these controllers be unintuitive to use, but they typically require multiple controllers if trajectories of more

than 3 DoF are required. Surpassing this limitation, only a whole-body motion capture suite provides the required expressiveness while remaining intuitive to use. Explicitly, a whole-body motion capture suit enables the teleoperator to generate 6-DoF trajectories via the pose of their hands relative to their torso, while selecting operational modes using open-closed states of their hands [135]. As such, a Noitom Perception Neuron motion capture suit is selected to build the teleoperation framework around due to its ability to record complete skeletal data, including hand states.

The utility of the framework stems initially from the generality of the set of tasks and constraints that formulate the IK WBC. Then, through leveraging these tasks and constraints, a set of teleoperation strategies are developed that each, in turn, deliver unique functionality and control features, providing the utility required to complete complex tasks; specifically, these strategies provide the teleoperator with control of all main frames of the robot, providing functionality to select which DoF to control frames in and a static gait. This is all packaged into the complete framework, detailed in Fig. 3.1, and as of writing, this is the only teleoperation framework that provides this level of utility while requiring no additional hardware to be added to the robot and being manageable by a single teleoperator.

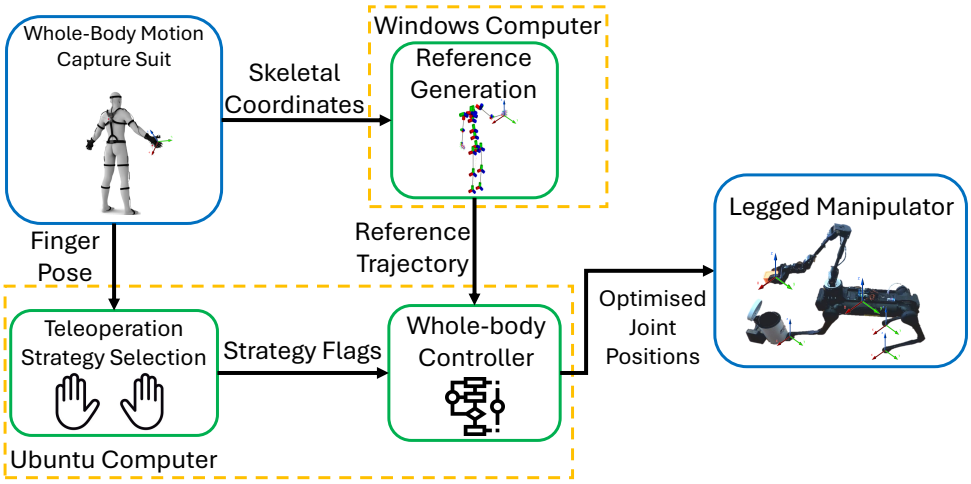


Figure 3.1: System overview of the teleoperation framework to test the IK-WBC.

3.2 Methods

3.2.1 Formulation of WBC Optimisation Problem

The optimisation problem within the WBC is to be formulated as a quadratic programming (QP) problem based on the IK cost function, where it will optimise for joint velocities $\dot{\boldsymbol{q}}$ (as represented by the inverse kinematics block in Fig. 3.2). Along with a set of constraints, the overall QP is formulated as

$$\min_{\dot{\boldsymbol{q}}} \quad \frac{1}{2}\dot{\boldsymbol{q}}^T \boldsymbol{A}^T \boldsymbol{A}\dot{\boldsymbol{q}} - \boldsymbol{b}^T \boldsymbol{A}\dot{\boldsymbol{q}} \tag{3.1}$$

$$\text{s.t.} \quad \boldsymbol{C}_{\text{lb}} \leq \boldsymbol{J}_{\text{CoM}}\dot{\boldsymbol{q}} \leq \boldsymbol{C}_{\text{ub}}\,, \tag{3.2}$$

$$\boldsymbol{J}_{\text{halt}}\dot{\boldsymbol{q}} = 0\,, \tag{3.3}$$

$$\dot{\boldsymbol{q}}_{\text{lb}} \leq \dot{\boldsymbol{q}} \leq \dot{\boldsymbol{q}}_{\text{ub}}\,, \tag{3.4}$$

$$\boldsymbol{q}_{\text{lb}} \leq \boldsymbol{q} \leq \boldsymbol{q}_{\text{ub}}\,, \tag{3.5}$$

where, $\boldsymbol{A}$ and $\boldsymbol{b}$ are used to stack all Cartesian, joint and CoM tasks so that the QP problem can realise several tasks simultaneously:

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} w_{\text{Cart}}\boldsymbol{A}_{\text{Cart}} \\ w_{\text{CoM}}\boldsymbol{A}_{\text{CoM}} \\ w_{\text{Jnt}}\boldsymbol{A}_{\text{Jnt}} \end{bmatrix} \in \mathbb{R}^{(6\text{m}+3+\text{n})\times\text{n}}, \\ \boldsymbol{b} &= \begin{bmatrix} w_{\text{Cart}}\boldsymbol{b}_{\text{Cart}} \\ w_{\text{CoM}}\boldsymbol{b}_{\text{CoM}} \\ w_{\text{Jnt}}\boldsymbol{b}_{\text{Jnt}} \end{bmatrix} \in \mathbb{R}^{6\text{m}+3+\text{n}}, \end{aligned} \tag{3.6}$$

specifically, the CoM task has been added to provide improved stability control to aid during teleoperation. $n = 12 + 5$ is the number of DoF of the legged manipulator (12 leg DoFs and 5 arm DoFs), $m = 4 + 1 + 1$ is the number of Cartesian tasks (four feet, one base and one end-effector), and w_{Cart}, w_{Jnt}, and w_{CoM} are the tasks weights used to prioritise certain tasks over others based on the control strategy. This task prioritisation method has been utilised within this work because it reduces the computational demand of the WBC as only one QP problem needs to be solved for each time step, as opposed to a hierarchical prioritisation method that needs to solve several sequential QP problems.

Further, to enhance this QP for use in teleoperation, (3.1) is subject to i) a CoM stability constraint (3.2) which takes the full form of

$$\frac{(\boldsymbol{x}_{\text{lb}} - \boldsymbol{x}^{\text{fk}}_{\text{CoM}})}{\delta t} \leq \boldsymbol{J}_{CoM}\dot{\boldsymbol{q}} \leq \frac{(\boldsymbol{x}_{\text{ub}} - \boldsymbol{x}^{\text{fk}}_{\text{CoM}})}{\delta t}, \tag{3.7}$$

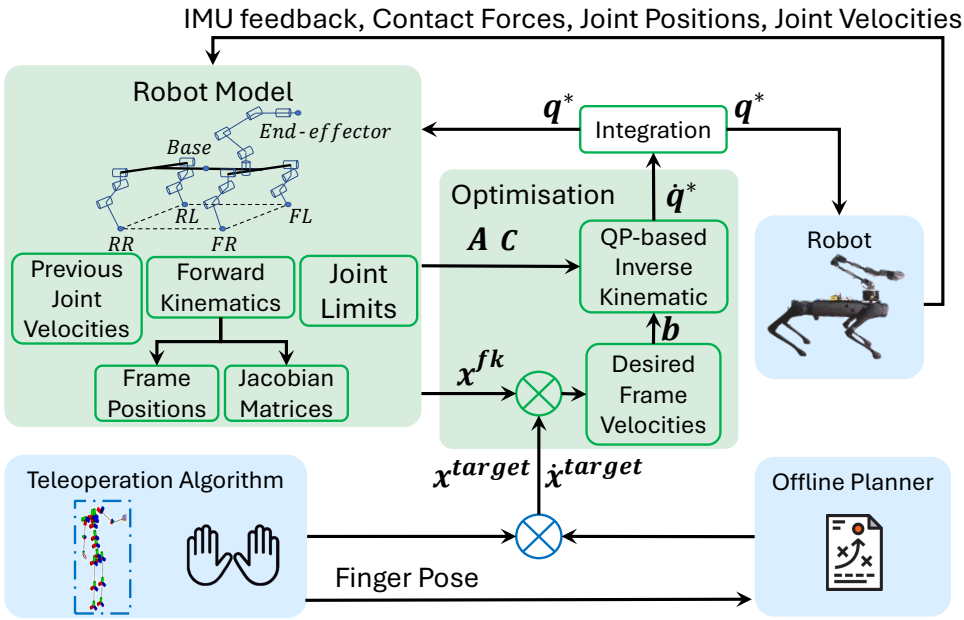


Figure 3.2: Block diagram of the IK-WBC, built around the QP problem.

ii) a halt constraint (3.3), along with iii) joint velocity (3.4) and position (3.5) constraints (these joint constraints were adopted from [94]). It should be noted that within (3.4) a safety factor of 2 is added to protect against aggressive motion caused by latency. The specifics of the CoM stability constraint (3.7), halt constraint (3.3) and all tasks are detailed later in this section.

The Cartesian tasks that are used within the WBC reduce the residual between a target Cartesian position and the current Cartesian position of a frame by enforcing a velocity upon it. This is achieved by utilising the Jacobian matrix of a frame which can map the current joint velocities to a Cartesian velocity, which can be formulated into $\boldsymbol{A}$ and $\boldsymbol{b}$ as

$$\boldsymbol{A}_{\text{Cart}} = \boldsymbol{J} \in \mathbb{R}^{6\text{m}\times\text{n}}, \quad \boldsymbol{b}_{\text{Cart}} = \left[\dot{\boldsymbol{x}}_1^T \cdots \dot{\boldsymbol{x}}_m^T\right] \in \mathbb{R}^{6\text{m}}, \tag{3.8}$$

where $\boldsymbol{J} = \left[\boldsymbol{J}_1^T \cdots \boldsymbol{J}_m^T\right]^T$, $\boldsymbol{J}_i \in \mathbb{R}^{6\times\text{n}}$ is the combined Jacobian matrix of all frames of interest and the target Cartesian task frame velocity

$$\dot{\boldsymbol{x}}_i = \dot{\boldsymbol{x}}_i^{\text{target}} + \boldsymbol{K}_{\text{Cart}}(\boldsymbol{x}_i^{\text{target}} - \boldsymbol{x}_i^{\text{fk}}) \tag{3.9}$$

describes the required velocity for a frame of interest of the robot to execute a desired trajectory using the velocity control law presented in [136], where $\dot{\boldsymbol{x}}_i^{\text{target}}$ is the target

frame velocity, $\boldsymbol{x}_i^{\text{target}}$ is the target frame configuration, $\boldsymbol{K}_{\text{Cart}}$ is the task gain, and $\boldsymbol{x}_i^{\text{fk}}$ is the current frame configuration calculated through forward kinematics. For the work in this project, the tracking of six frames of a legged manipulator will be used. This being, in reference to Fig. 3.2, all end-effector frames of the robot (four feet and one end-effector) with the addition of the base frame. These frames have been selected to achieve a strong level of control.

CoM trajectory planners see frequent use for quadrupedal robot locomotion, such as in [137, 134]. Therefore, in order to allow this WBC to be compatible with these planners, a CoM tracking task is included. A Jacobian for CoM can be found based on the robot configuration. Consequently, a CoM tracking task can be easily added to the QP problem, in the same manner as the Cartesian task, through defining $\boldsymbol{A}_{\text{CoM}}$ and $\boldsymbol{b}_{\text{CoM}}$ to be placed within (3.6) as

$$\boldsymbol{A}_{CoM} = \boldsymbol{J}_{\text{CoM}} \in \mathbb{R}^{3\times \text{n}}, \quad \boldsymbol{b}_{\text{CoM}} = \dot{\boldsymbol{x}}_{\text{CoM}} \in \mathbb{R}^3, \tag{3.10}$$

in which $\boldsymbol{J}_{\text{CoM}}$ is the Jacobian of the CoM and $\dot{\boldsymbol{x}}_{\text{CoM}}$ is the target CoM velocity. $\dot{\boldsymbol{x}}_{\text{CoM}}$ is derived by applying the same velocity control law as seen in (3.9).

Within the WBC, several joint tasks are utilised to enforce specific behaviours on each joint. Two different joint tasks are imposed on the joints within this study, the manipulability gradient and damping joint tasks. The manipulability gradient task aims to reduce the likelihood of a singularity being produced, while the joint damping tasks aim to reduce high-frequency oscillations of the joints. Aligning with (3.6), these joint tasks take the following form,

$$\begin{aligned} \boldsymbol{A}_{\text{Jnt}} &= w_{\text{mnp}}\boldsymbol{S}_{\text{mnp}} + w_{\text{prev}}\boldsymbol{S}_{\text{prev}}, \\ \boldsymbol{b}_{\text{Jnt}} &= w_{\text{mnp}}\nabla f(\boldsymbol{q}) + w_{\text{prev}}\dot{\boldsymbol{q}}^*_{\text{i prev}}, \end{aligned} \tag{3.11}$$

where w_{mnp}, $\boldsymbol{S}_{\text{mnp}} \in \mathbb{R}^{\text{n}\times\text{n}}$, and $\nabla f(\boldsymbol{q}) \in \mathbb{R}^{\text{n}}$ are the manipulability gradient joint task weight, selection matrix, and manipulability gradient, $f(\boldsymbol{q}) = \sum_{i=1}^{m}\sqrt{\det(\boldsymbol{J}_i\boldsymbol{J}_i^T)} \in \mathbb{R}^n$ is the sum of all limbs' manipulabilities, w_{prev}, $\boldsymbol{S}_{\text{prev}} \in \mathbb{R}^{\text{n}\times\text{n}}$, and $\dot{\boldsymbol{q}}^*_{\text{i prev}} \in \mathbb{R}^{\text{n}}$ are the damping joint task weight, selection matrix, and optimised joint velocities for the last time step respectively. It should be noted that $w_{\text{mnp}} + w_{\text{prev}} = 1$. Further joint tasks could be used within (3.11) such as a Tikhonov Regularization task, see [94] for details.

To improve the solution space of the QP problem, several constraints are added to it. To promote the stability of the robot, a CoM stability constraint (3.7) has been defined. Through utilising the mapping between joint space and Cartesian space, the

constraint restricts the position of the CoM to always lie within the bounds of the support polygon, while there are three or more points of contact. For (3.7), $\boldsymbol{x}_{\text{lb}}$ and $\boldsymbol{x}_{\text{ub}}$ are the lower and upper bounds of the support polygon respectively, $\boldsymbol{x}^{\text{fk}}_{\text{CoM}}$ is the estimated position of the CoM, $\boldsymbol{J}_{\text{CoM}}$ is the CoM Jacobian, and δt is the time step. Another constraint that has been applied to the QP problem is the halt constraint (3.3), in which $\boldsymbol{J}_{\text{halt}} = \left[\boldsymbol{J}_1^T \cdots \boldsymbol{J}_c^T\right]^T \in \mathbb{R}^{6\text{c}\times\text{n}}$, where $\boldsymbol{J}_i \in \mathbb{R}^{6\times\text{n}}$ is the Jacobian of the frame to be constrained, and c is the number of these constraints. This constraint enforces zero velocity at a frame and can be applied to the frames in contact with the ground to enforce a non-slip condition. This constraint can also be adapted to halt a frame locally to another frame in position,

$$(\boldsymbol{J}_{\text{halt}} - \boldsymbol{J}_{\text{local}})\dot{\boldsymbol{q}} = 0, \tag{3.12}$$

where $\boldsymbol{J}_{\text{local}}$ is the Jacobian of the frame that the halted frame has zero velocity in respect to.

3.2.2 WBC Applications for Teleoperation

Through this formulation of the WBC, several tasks and constraints can be combined to improve the utility of the teleoperation framework. For the teleoperator to fully control the camera view of the robot, the orientation of the frame of the robot that the camera module is placed upon should be controlled by generating a rotational trajectory that matches the orientation of the teleoperator's head, and realised through (3.9). This will enable the teleoperator to change their field of view with ease, improving operability. This task would have a relatively low weight as it should not hinder more critical tasks from being met.

Quadrupedal legged manipulators will often complete manipulation tasks as instructed by the teleoperator. Once the manipulator has picked up an object, it could be desirable to keep that object stationary in space. An example of this would be when the robot has picked up an object but must move the quadruped section of the robot out of the way of hazards. This would involve enforcing zero velocity in selected DoFs of the end-effector frame by applying a halt constraint on it, as detailed in (3.3) where any component within $\boldsymbol{J}_{\text{halt}}$ related to a DoF not to be constrained is set to zero. This implementation is included in Section 3.2.3 across various teleoperation strategies and is achievable as (3.3) can be added and removed dynamically.

The static gait allows for the controlled locomotion of quadruped-like robots, making it ideal for manipulation tasks. As such, this gait should be included in this framework. A simple static gait can be realised by assigning a Cartesian task a suitable trajectory for each foot frame and a CoM task, with the CoM task being controlled using either an offline planner or automatically due to the teleoperator only using high-level control. In the case of using an offline planner, a trajectory would be supplied to the CoM task to ensure that it moves to a position within the dynamic support polygon. For the WBC to control the CoM automatically, the target CoM position is set as the centre of the support polygon, as illustrated in Fig. 3.3.

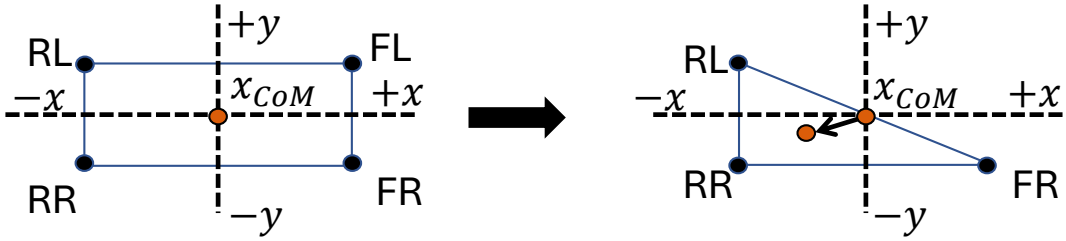


Figure 3.3: Instantaneous translation of the CoM when a foot is no longer in contact with the ground, the orange circles represent the target position and the black circles represent the feet in contact with the ground.

During operation, the teleoperator may not consider or have a full understanding of how their commands may violate the WBC constraints, such as trying to move the base frame to a position that would result in the CoM leaving the area bound by the CoM constraint; without intervention, the input command would not be realised. To mitigate this limitation, the framework tracks the residual, $\boldsymbol{D}$, between the target CoM position and the current CoM position,

$$\boldsymbol{D} = \boldsymbol{J}_{\text{CoM}}\dot{\boldsymbol{q}} - \dot{\boldsymbol{x}}_{\text{CoM}} \quad \in \mathbb{R}^3. \tag{3.13}$$

If a component of $\boldsymbol{D}$ surpasses a defined limit, a command would be sent to a walking pattern generator [134]. Furthermore, the direction of this gait can be determined by the component of $\boldsymbol{D}$ that breached the limit. Note that the vertical DoF is not considered here. Overall, this results in moving the bounds of the support polygon and workspace in the direction of the reference trajectory.

During manipulation tasks, often the manipulator will be grasping an object making it unavailable for further manipulation tasks. In humanoid robots, this issue is easily mitigated by utilising the other arm of the robot, however, legged manipulators typically do not feature an additional manipulator due to the additional payload limiting performance. Consequently, using the feet frames of the robot to complete simple manipulation tasks, such as pushing obstructing obstacles, is the favored option over adding an additional manipulator to the system. This is achieved through utilising a Cartesian task for any of the feet frames.

3.2.3 Teleoperation Strategies

To enable the teleoperator to utilise all main frames of the robot to complete tasks despite the kinematic dissimilarities between themselves and the legged manipulator, a set of teleoperation strategies is developed. Each strategy is selected based on the hand posture of the teleoperator, as detailed in Table 3.1 along with their functionalities.

Once a strategy is entered, the motion capture suit provides the current pose of the right hand in the hip frame. This is set as the start of the reference trajectory, with any subsequent movement from this point being sent as a reference trajectory to the WBC at 500 Hz, as illustrated in Fig. 3.2. To map human motions to the robot, a scaled relative pose relationship at each time step t is derived as,

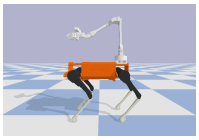
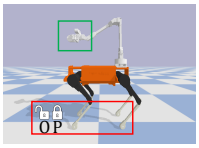
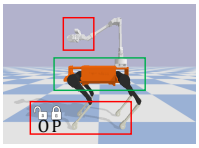
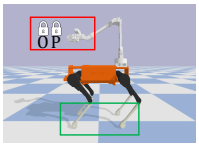
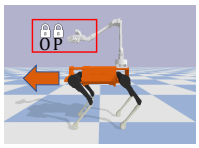
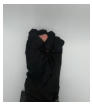
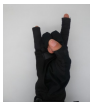
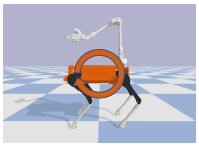
$$\boldsymbol{x}_t^{target} = \boldsymbol{x}_0^{target} + \boldsymbol{\mu}(\boldsymbol{x}_t^m - \boldsymbol{x}_0^m), \tag{3.14}$$

where $\boldsymbol{x} = [p_x; p_y; P_z; \theta_r; \theta_p; \theta_y]$ is a vector of displacements in position and rotation, in which p_x, p_y, p_z, θ_r, θ_p and θ_y correspond to x, y, z, roll, pitch and yaw respectively, and $\boldsymbol{\mu}$ is the scaling factor for the mapped motion. Superscripts m and $target$ refer to master, and slave target trajectory respectively. Subscript 0 depicts the initial state once a strategy is entered. This enables the teleoperator to exit a strategy, using TS0, and then re-enter with all subsequent movements starting from that point, allowing the teleoperator to either readjust their posture when they are at the limits of their movement or pause operation safely.

3.3 Results and Discussion

To test and validate the effectiveness of the teleoperation framework for use with hardware in real-time, two experiments were completed. One was a general demonstration

Table 3.1: Teleoperation Strategies: in the 'Left Hand' and 'Right Hand' columns 0 depicts a digit is open and 1 is closed. Some pictures of the teleoperation strategy hand poses are provided to illustrate how the pose array represents a real hand pose.

Name	Reference Diagram	Frame Controlled		Features	Left Hand	Right Hand
TS0		n/a		No reference sent to the robot.	0,0,0,0,0 	0,0,0,0,0 
TS1		TS1a	End-effector control	• Feet locked in position and orientation. • End-effector can be opened and closed.	0,0,0,0,0 	1,1,1,1,1 
		TS1b	End-effector orientation		0,0,0,0,0	0,0,1,1,1
		TS1c	End-effector position		0,0,0,0,0	0,1,1,1,1
TS2		TS2a	Base	• End-effector locked in position.	1,1,1,1,1	0,1,1,1,1
		TS2b		• End-effector locked in position and orientation. • Automatic stepping.	1,1,1,1,1 	0,0,1,1,1 
TS3		TS3a	FR foot	• End-effector is locked in position and orientation. • Automatic CoM adjustments and foot lifting.	0,1,1,1,1 	0,0,1,1,1 
		TS3b	FL foot		0,0,1,1,1	0,0,1,1,1
		TS3c	RR foot		0,0,0,1,1	0,0,1,1,1
		TS3d	RL foot		0,0,0,0,1	0,0,1,1,1
TS4		n/a		• Robot completes static gait. • End-effector is locked locally in position.	1,1,1,1,1 	1,0,1,1,0 
TS5		n/a		• Resets all frames to a home position.	1,1,1,1,1 	1,1,1,1,1 

of the majority of the teleoperation strategies, and the other was a real-world task of disposing of an object, representing a hazardous substance, in a push pedal bin. To analyse the performance of the framework across all experiments, the joint positions of the robot were recorded and were then passed to a simulated robot in PyBullet [138], from which the positions and orientations of all frames of interest were extracted. A supplementary video of these experiments can be found at [139][1]. However, to first test the generality of the framework and tune the controller gains, several initial simulation experiments were completed first.

3.3.1 Initial Simulation Experiments and Generality Validation

Before the teleoperation framework was tested on hardware, a set of simulations were completed in order to tune the WBC task weights and gains, test the framework's modularity, and ensure it was safe to deploy on the real robot. These simulations consisted of a Pybullet physics base simulation where three different quadruped robots (A1, Laikago and Aliengo all from Unitree) paired with the ViperX 300 arm were tested using pre-recorded motion capture suit data that involved utilising all teleoperation strategies.

Tuning the task weights and gains involved a simple trial and error method, where an initial estimate was made, and then they were either increased or decreased by a factor of 10 or 2 for the weights and gains respectively. As a result, similar whole-body motions were observed across all three robots to realise the reference trajectory, as illustrated in Fig. 3.4. Furthermore, as illustrated in Fig. 3.4, an increase in z displacement and *pitch* rotation is observed before any x displacement. This is a direct result of the CoM stability constraint (3.7) causing the WBC to prioritise trajectories in DoF that do not cause a shift in the CoM. The effectiveness of the WBC being able to realise reference trajectories that would otherwise be impossible is highlighted in Fig. 3.5, where without the WBC activated, the input of the reference trajectory results in a singularity being reached. Additionally, in Fig. 3.5 the WBC's ability to preserve stability using a combination of (3.7) is highlighted, where in the case where a foot is lifted from the ground, the WBC causes the CoM to shift within the new support polygon; with the WBC inactive, the robot is observed to fall.

[1] https://youtu.be/iL_K_CVA0pU

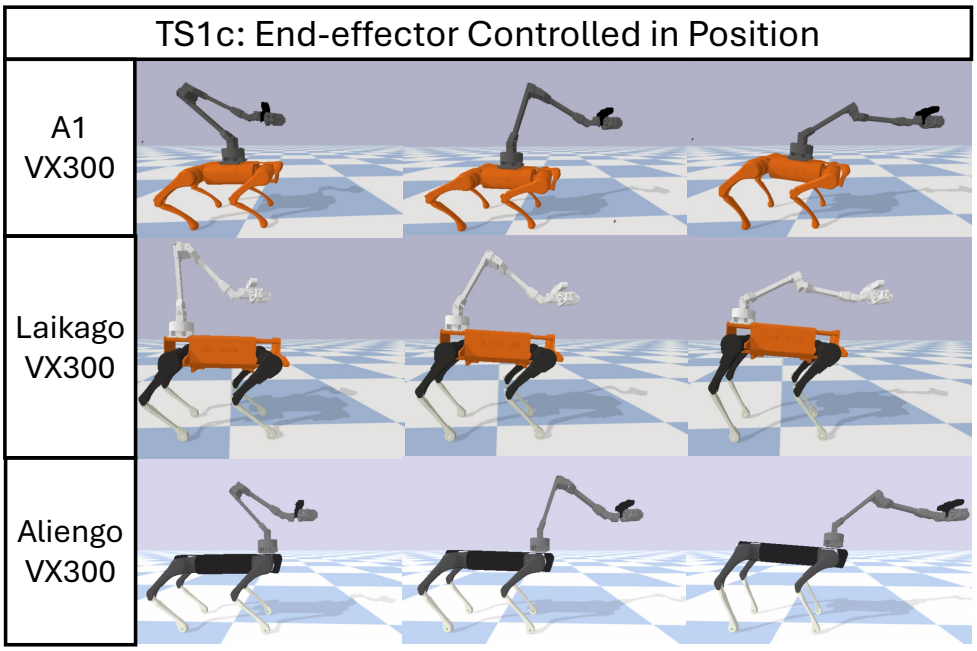


Figure 3.4: Snapshots of TS1c being used across a range of different robots to test framework modularity.

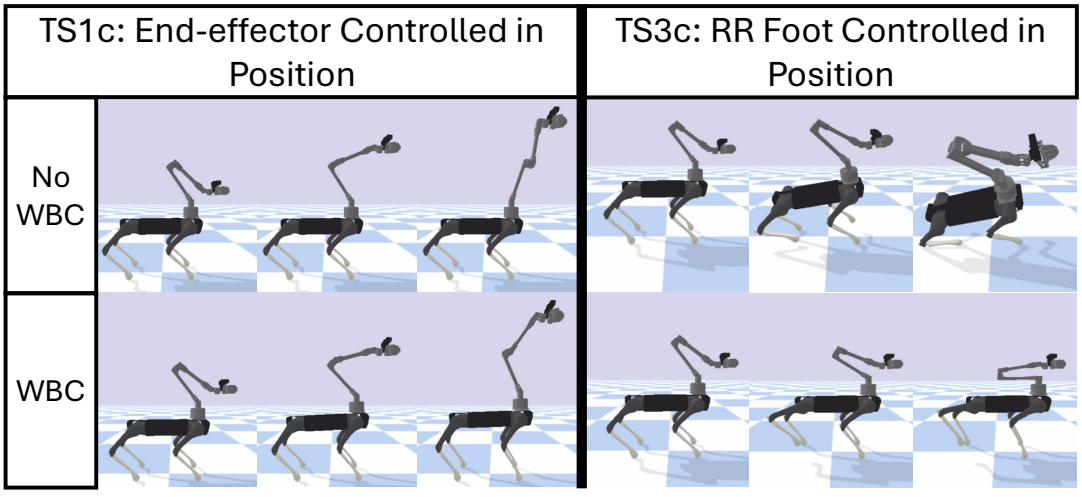


Figure 3.5: Comparing the performance with and without the WBC enabled.

3.3.2 General Hardware Demonstration

In this experiment, one of three different movement types was used to provide a reference trajectory to the WBC, using different strategies to evaluate their performance. This being in x, y, z, or yaw, which will be referred to as $M1$, $M2$, $M3$ and $M4$. It should be noted that with the base being aligned with the world frame, $+x$ was in the direction the base is facing, $+y$ was to the left of the base, and $+z$ was vertically upwards. The strategies used and analysed in this test are TS0, TS1, TS2 and TS3.

Initially, TS1b was entered and an $M4$ was generated using the motion capture suit to be realised by the end-effector frame. As such, the end-effector frame rotates in θ_y while the base frame was observed to adjust its pose to aid in the realisation of this trajectory, as seen in the data in Fig. 3.6 and demonstrated in Fig. 3.7. However, a small displacement of the end-effector was seen in p_y despite the use of a halt constraint applied to its position. This was due to the WBC not accounting for the imperfect[1] joint motors of the robot hence not minimising error when a significant shift in the CoM tending towards an individual leg occurs, which was further exacerbated by the weight of the arm. This is highlighted in Fig. 3.8 where the y component of the centre of pressure (CoP) approaches the boundary. This issue could potentially be resolved by building the WBC around ID instead of IK or having the IK-WBC paired with a DRL policy that can consider the full dynamics, as not only would it better consider the weight of the arm, but even though the Aliengo motors can be position controlled they operate through torque control. Switching to TS1c, a $M1$ and a subsequent $M3$ trajectory were generated and executed by the end-effector frame. Again for both trajectories whole-body motion was observed in the robot to help the end-effector realise the reference trajectory, as presented in Fig. 3.6 and Fig. 3.7, while also demonstrating accurate realisation of the WBC output. Furthermore, the orientation halt constraint proves to be effective in TS1b, as minimal rotation of the end-effector frame was observed.

[1] The joint position commands sent to the robot's motors will not be realised perfectly, a factor not currently considered by the IK-WBC.

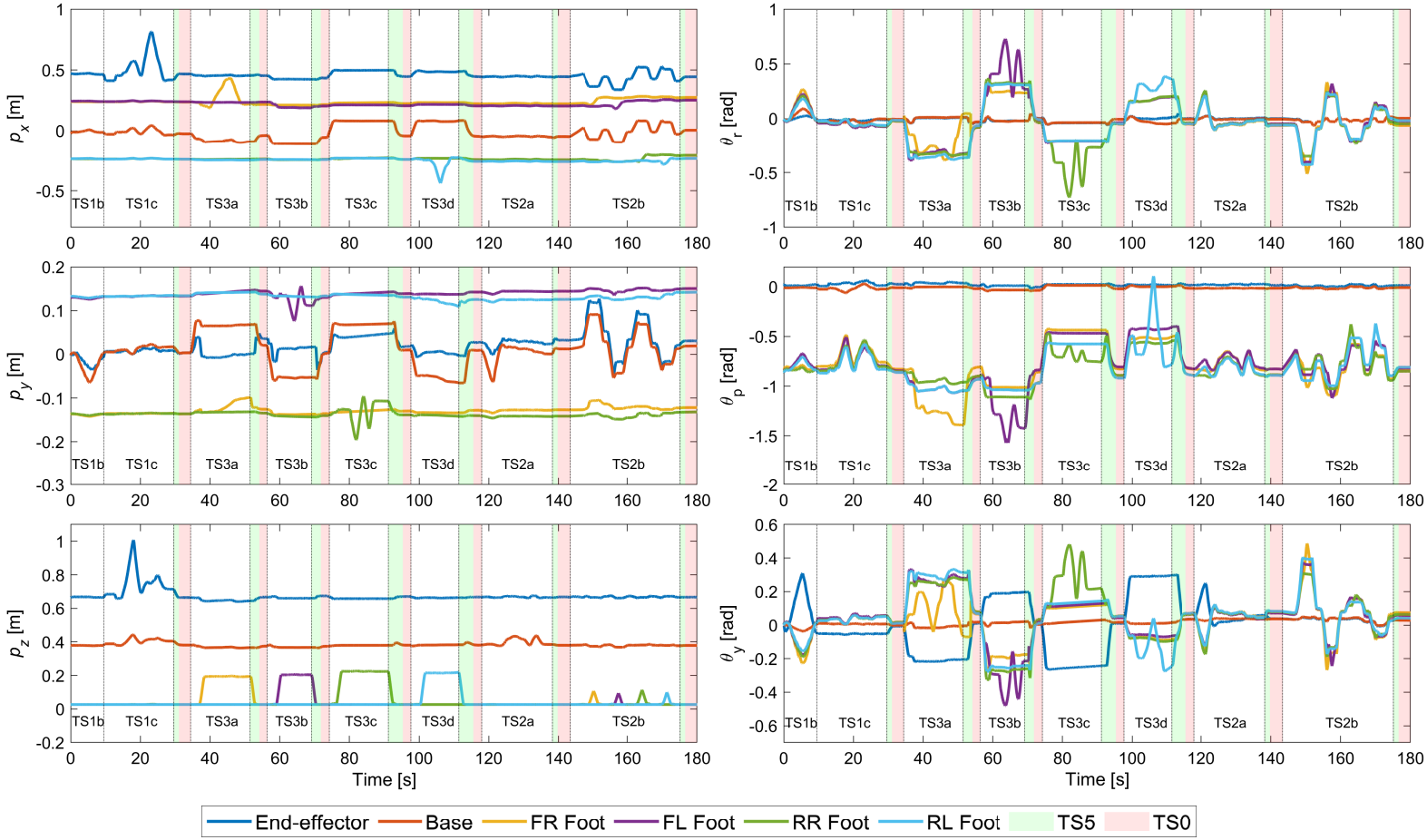


Figure 3.6: The position (left) and orientation (right) WBC and real output of all main frames of the robot during the general demonstration.

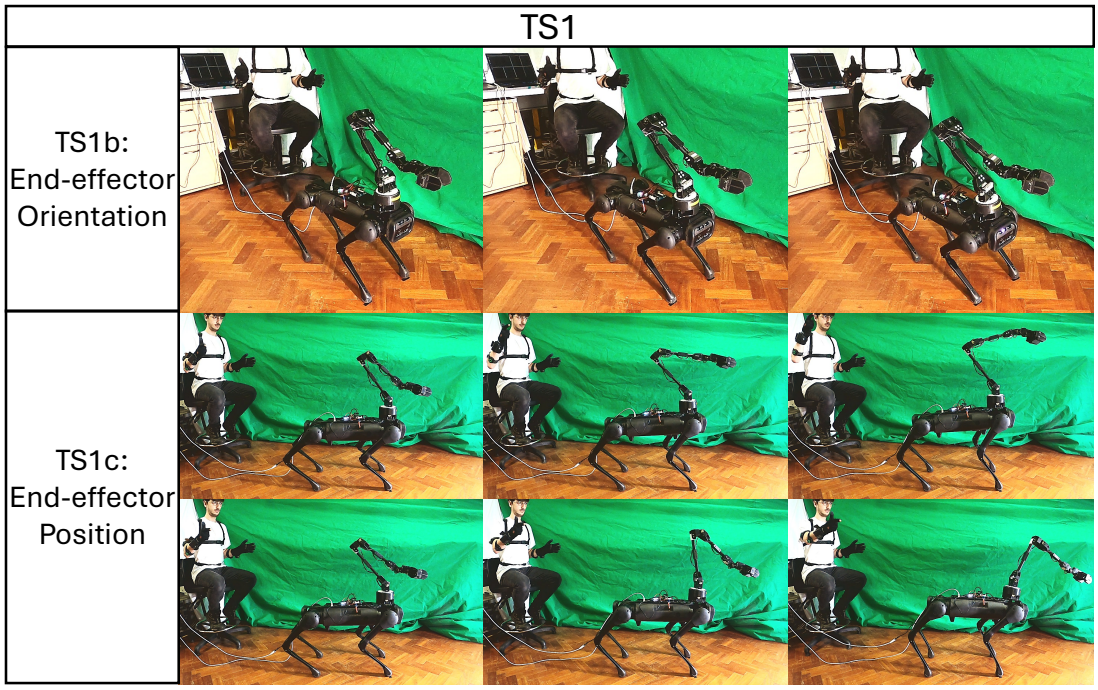


Figure 3.7: Snapshots of teleoperation using TS1.

For all strategies of TS3 (TS3a, TS3b, TS3c and TS3d), before each foot was lifted the offline planner provided a reference trajectory for a CoM Cartesian task to ensure the CoM lies within the new support polygon to preserve stability. This stability is supported by Fig. 3.8 where even with a foot lifted from the ground the CoP always lies within the stability bounds of the support polygon. Foot frames FR and RL completed an $M1$ trajectory, while frames FL and RR completed an $M2$ trajectory. Pose adjustments were observed by the base to assist the foot frames to achieve their reference trajectories, however, these are only minor adjustments, due to the CoM constraint, so stability was maintained while there is a reduced support polygon, as shown in 3.6 and Fig. 3.9. Furthermore, the end-effector frame was observed to stay almost completely fixed in position while the foot frames completed their tasks, demonstrating the effectiveness of the halt constraint, with the only caveat of the imperfect joint motors which caused the end-effector frame to displace slightly in p_y.

In TS2a, a reference trajectory of $M2$ and then $M3$ were generated by the motion capture suit and sent to the WBC to realise these trajectories using a base Cartesian task. As there was a halt constraint in place for the end-effector's position, as the base

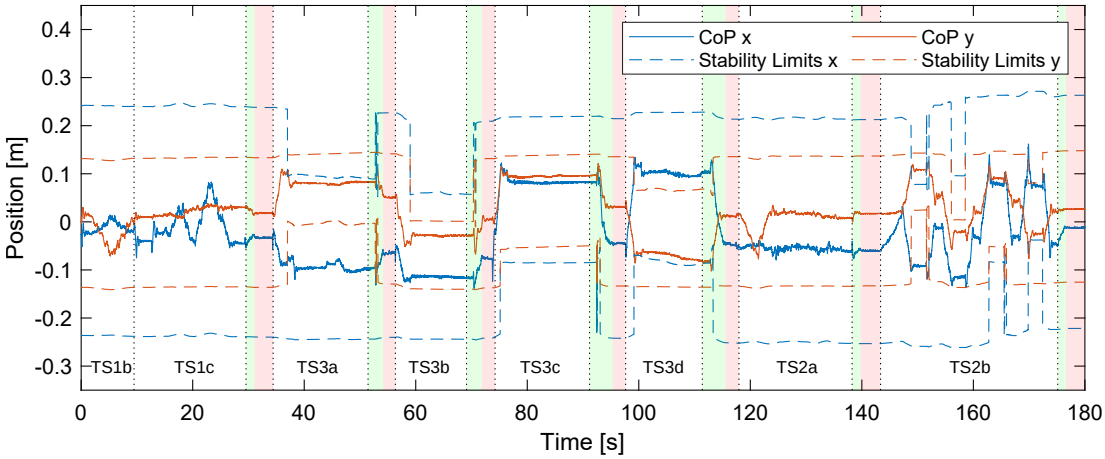


Figure 3.8: x and y components of the robot's CoP during the general demonstration.

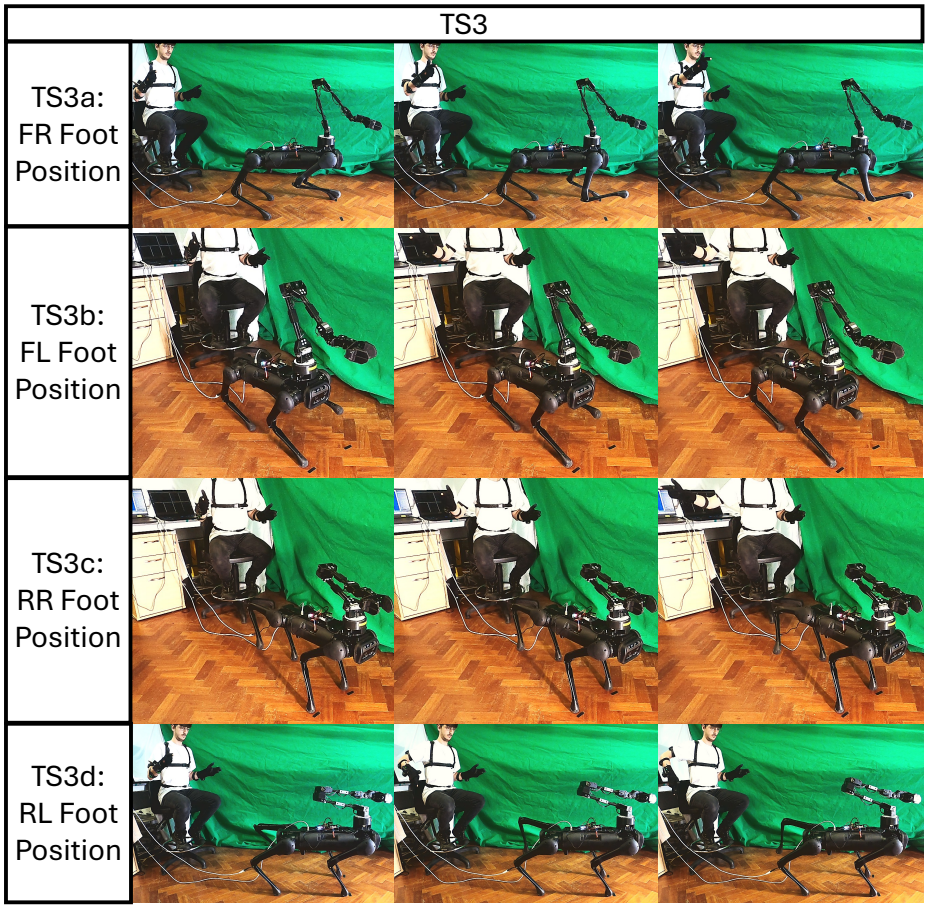


Figure 3.9: Snapshots of using TS3 during the hardware general experiment.

completed both reference trajectories, in the WBC output the end-effector remained locked in position, as presented in Fig. 3.6. However, minor displacement was observed in the real robot in p_y due to a combination of imperfect motors unable to perfectly compensate for the movement of the base and the IK-based WBC not considering the dynamics of the robot. This displacement could be reduced through the implementation of a state estimator, such as [15], which considers the dynamics and contact state of the robot to provide an accurate estimation of the base's position and orientation. After completion of these trajectories, TS2b was entered and an $M1$ trajectory was realised by the base while the end-effector frame was locked in position and orientation, using a halt constraint, as shown in the data presented in Fig. 3.6. This significantly shifted the CoM and, which breached the limit described in Section 3.2.2. This caused the offline planner to automatically generate the reference trajectories for a static stepping gait. During this gait, the was end-effector locked locally to the base, using a local halt constraint, while its orientation was still locked in the world frame. The gait was observed to complete successfully; overall forward displacement can be observed in Fig. 3.6 and its stability is observed in Fig. 3.8 where the x and y components of the CoP never leave the stability limits, demonstrating the effectiveness of the CoM task (3.10).

3.3.3 Object Disposal

The objective of this test was to dispose of a small box in a push pedal bin. This task requires the robot to simultaneously hold open the bin with the pedal while dropping the box in the bin, for which a range of frames need to be controlled in specific DoF at a time (for dexterity) all while stability is maintained. Therefore, utilising the high utility of the teleoperation framework is critical to the success of this test. The box had a width of 65 mm while the end-effector had a maximum gripping width of 76 mm; the completion of this task also demonstrates the high level of dexterity achieved by this framework. Initially, TS1a was used to utilise both position and orientation trajectories, generated by the motion capture suit, to position the end-effector frame close to the box and roughly align with its orientation, all the while the WBC produced whole-body motion to aid in the realisation of this task. Next, TS1c was used to for finer movement to position the end-effector ready to pick up the box, while it is locked in orientation, as shown in Fig. 3.10.

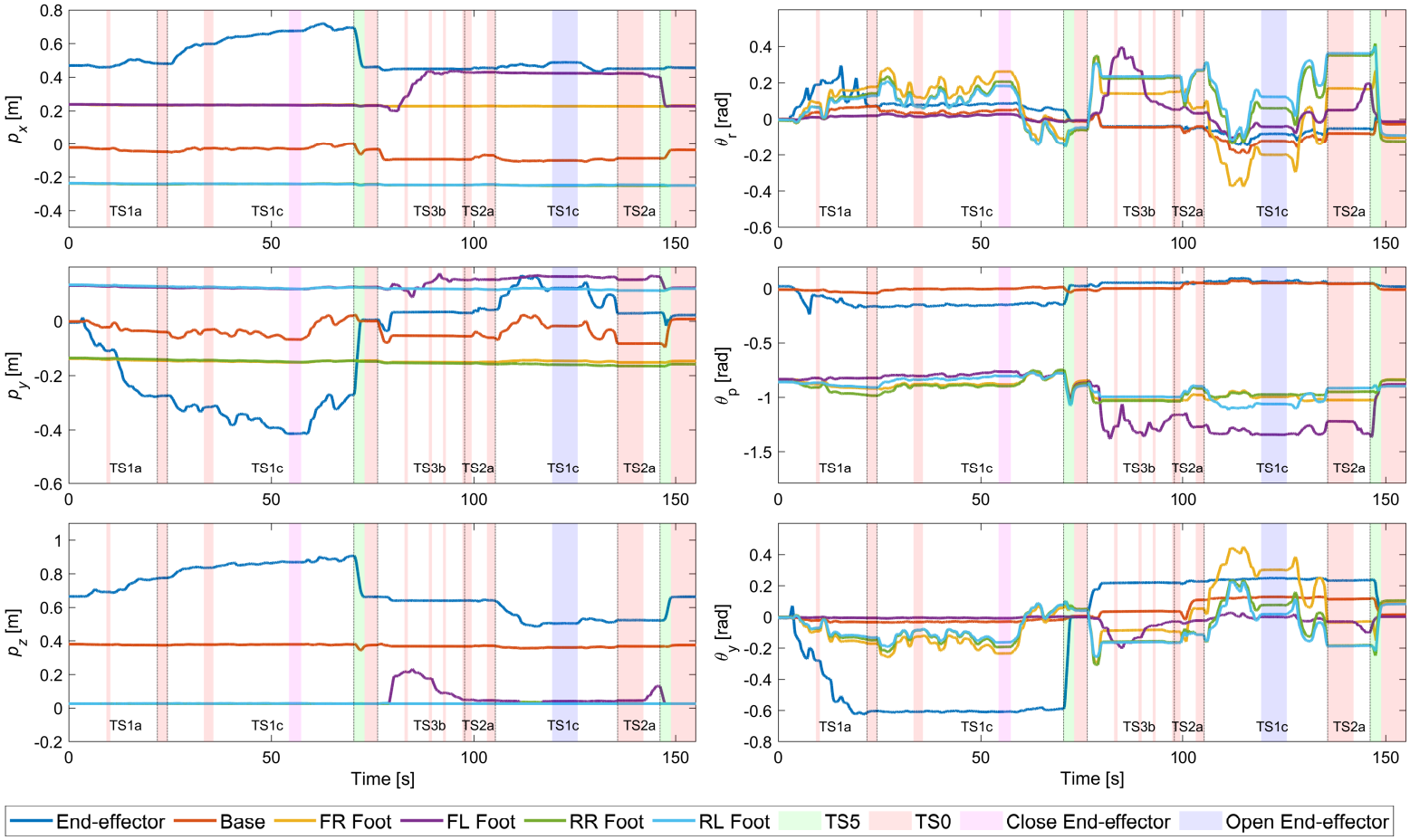


Figure 3.10: The position (left) and orientation (right) WBC and real output of all main frames of the robot during the object disposal experiment.

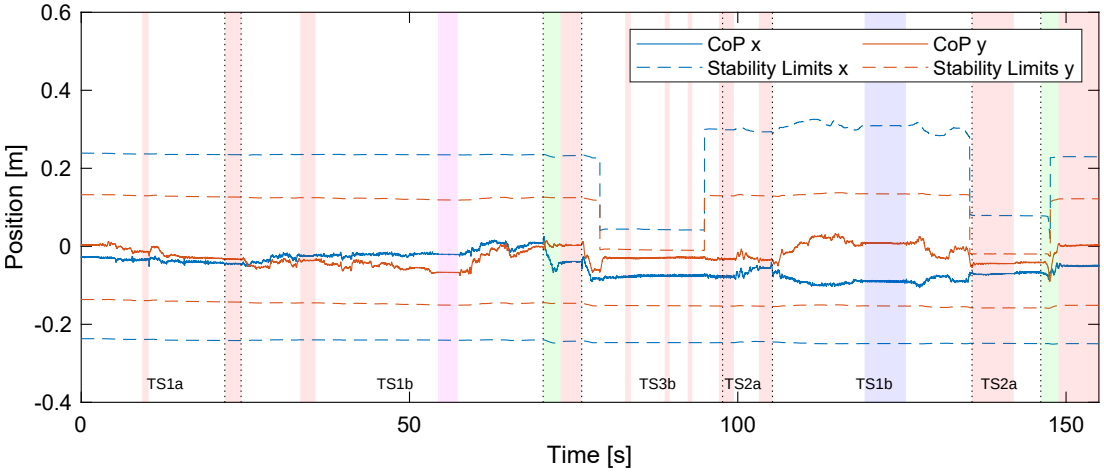


Figure 3.11: The x and y components of the robot's CoP during the object disposal experiment.

The end-effector then closed and picks up the box using TS1c, after which TS5 was used to bring the robot back into the home configuration for quick repositioning of the end-effector and box. TS3b was then used to pick up the FL foot and push the pedal on the bin, lifting the lid. Due to the imperfect motors, the base rotates in θ_r when the FL foot lifts off the ground. This in turn causes the real robot to experience error from the WBC output, in the end-effector and FL frames, as detailed in Fig. 3.10. However, through observation, the teleoperator compensated for this, which lead to the successful opening of the bin.

While the end-effector and foot were locked in position using a halt constraint, ensuring the box was stable, and the foot remained on the pedal, TS2a was used to orientate the base to look into the bin to simulate the scenario outlined in Section 3.2.2 and the teleoperator only having vision from the front quadruped cameras. Minor displacement is observed in the end-effector frame, although this could be resolved through the implementation of a state estimator. Utilising TS1c, the box was positioned over the bin and then dropped into it. During both the TS2a and TS1c phases, negligible displacement of the FL foot is observed, shown in Fig. 3.10, even while it is on the unstable surface of the bin pedal and while the base and end-effector frames complete trajectories. This demonstrates the effectiveness of the halt constraint (3.3) and the stability of the framework enforced by the CoM constraint (3.7), which is further supported by Fig. 3.11 where the CoP is observed to stay within the bounds of the stability

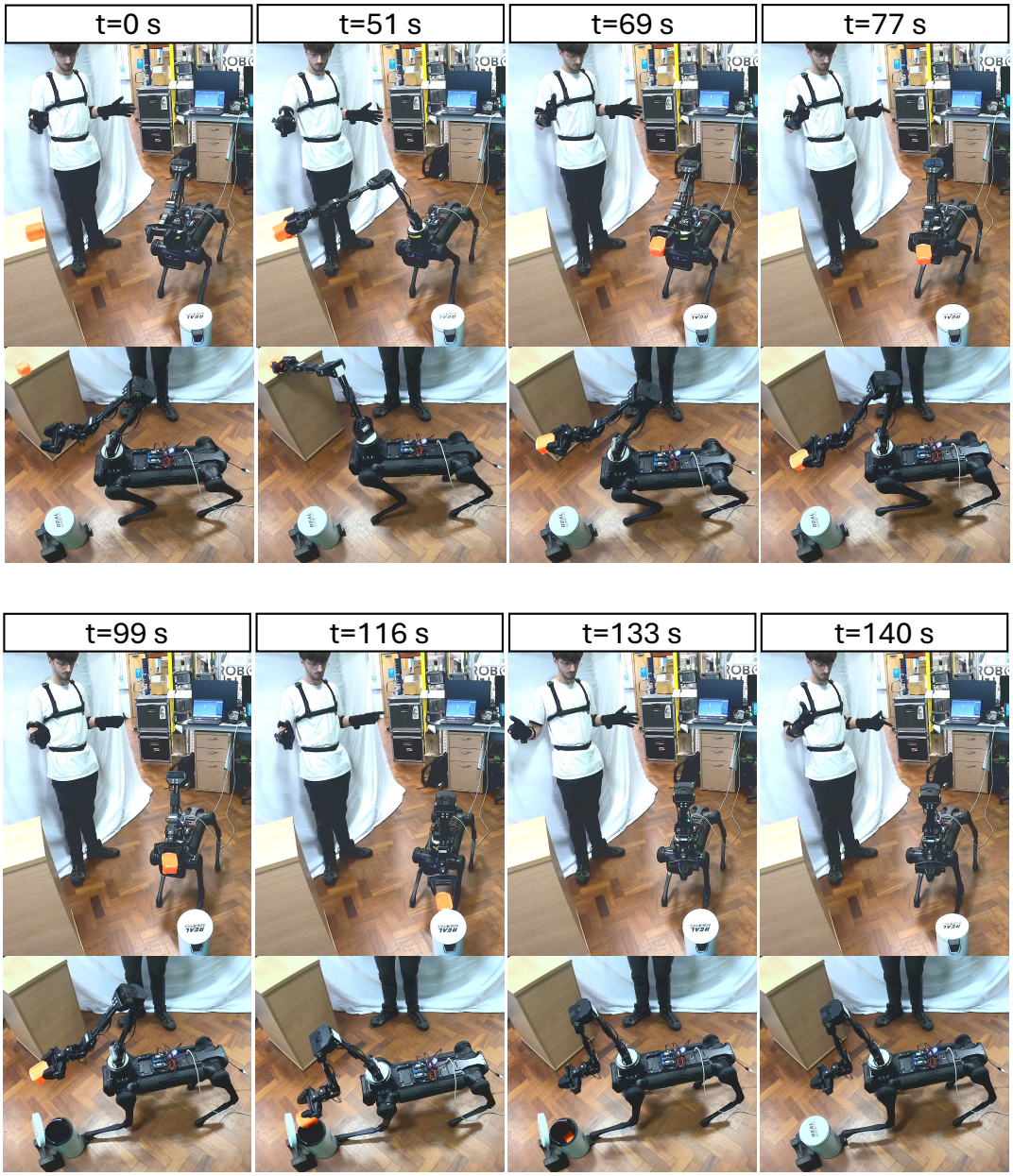


Figure 3.12: Snapshots of the teleoperator controlling the robot during the object disposal experiment.

limits. TS3b was then re-entered to lift the FL foot, releasing the pedal. Finally, TS5 returned the robot to the home position.

This experiment is presented in Fig. 3.12, where it can be observed that not only does the framework enable the successful execution of a complex experiment, where three frames of the robot are used in conjunction to complete tasks, but also the robot remained stable even when one foot was lifted and the base was completing a trajectory (further evidence is provided in the supplementary video). In turn, this demonstrates the effectiveness of this framework as without the ability to control a range of frames of the robot this task would prove to be impossible without adding another manipulator to the robot, although doing so would only be a detriment to its performance and payload capacity.

Despite latency being a common issue in teleoperation, in terms of both time delay and data loss which can both result in robots exhibiting dangerous behaviour, no latency issues were encountered during both this experiment and the general demonstrations. This was primarily due to the joint velocity dampening tasks (Tikhonov Regularization and joint dampening) that penalize large changes in joint velocities and the CoM constraint (3.7) which together both ensured that any delay or intermittent references passed to the framework due to latency would neither cause unstable or dangerous behaviour in the robot. Additionally, as mentioned in Section 3.2.1, to add further protection against aggressive motions the joint velocity limits have a safety factor of 2 applied to them.

The only limitation experienced through deploying the framework to complete this task was the high mental load on the teleoperator, where completing the task required the use of many of the control options, which could have potentially led to increased time in completing this complex task. One method of alleviating this mental load on the teleoperator would be to add further automation to shift more of the workload onto the framework, such as automating which DoF a frame is being controlled in through using computer vision and machine learning techniques to infer the task requirements.

3.4 Chapter Conclusion

In this chapter, a teleoperation framework of high utility, which utilises a tailor-made WBC, a range of teleoperation strategies, and a motion capture suit to control legged manipulators has been developed. Through implementing a set of tasks and constraints

specialised for teleoperation with a QP optimisation problem, forming the core of the WBC, the teleoperation framework can take input reference trajectories generated by a motion capture suit to control a range of functionalities and all main frames of the legged manipulator. This work has demonstrated the effectiveness of each teleoperation strategy through a general demonstration of the strategies and through completing an object disposal experiment that would be impossible to complete without the use of these strategies and no modifications to the robot's hardware.

Overall, with the rigorous testing featured within this chapter it has been demonstrated that the developed WBC has achieved project Objective 2 and in turn is primed for use within the proposed loco-manipulation framework of this project. The highly generic and versatile tasks and constraints formulated in Section 3.2.1 have been proven to be able to realise complex multi-frame manipulation tasks, further demonstrating its proficiency, which in turn will later be used to realise dynamic loco-manipulation tasks. That being said, in its exact formulation presented in Section 3.2.1 it may not be perfect for integration with a DRL policy that provides the IK-WBC reference trajectories, however due to its ability to dynamically adjust the active tasks and constraints further adaption of its formulation should not prove to be an issue.

Chapter 4

Leveraging Biomechanics for the Development of a Gait Scheduler

In order to develop the DRL locomotion policy in Fig. 2.12, first the gait scheduler and automatic gait selection control methods must first be produced, as theses gait references are integral to the successful training of this control module. As the literature revealed, gait and gait transition generation is a natural ability animals possess, yet for quadruped robots this presents a significant challenge when striving to achieve optimal and stable locomotion. As such, to develop the gait scheduler, featuring an automatic gait switching method, findings from biomechanics will be leveraged to realise optimal locomotion with stable and seamless transitions in the effort to harness the same proficiency that we see in animals in the wild. Work contributing to this chapter has been published in the following journal:

- Humphreys, J., Li, J., Wan, Y., Gao, H. and Zhou, C., 2023. **Bio-inspired Gait Transitions for Quadruped Locomotion**. IEEE Robotics and Automation Letters, 8(10), pp.6131-6138.

A video summary, featuring the simulation and hardware experiments, can be found at [140][1].

[1] https://www.youtube.com/watch?v=vhA0tiHbSSk

4.1 Introduction

Efficient and stable locomotion is a vital capability for quadruped robots to carry out tasks effectively in real-world applications. Generallhttps://myukke.work/y, this encompasses moving at various speeds depending on the situation, similar to how animal quadrupeds display a spectrum of gaits subject to the task. As a result, it is standard procedure to base robotic quadruped gaits on categorised real animal gaits through manual design [141] or model matching techniques [35]. These gaits can subsequently be stored and regulated by a gait scheduler for implementation on quadruped robots. The stored gaits can vary from slow to highly dynamic, thereby necessitating a robust control framework for their stable execution in quadruped robots. As detailed in Chapter 2, legged locomotion frameworks typically fall into two groups: optimal control, which utilises an MPC-WBC combination [142], and DRL for end-to-end control [29]. Despite their respective limitations, both are able to exhibit dynamic locomotion.

In spite of the competence of these control frameworks, a single gait type is not suitable for all speeds; much like how animals transition between gaits to optimise energy efficiency [143], quadruped robots can adopt this strategy to attain comparable outcomes [66]. In the wild, this manifests as animals transitioning to preserve their CoT; when energy consumption surpasses natural limits, quadrupeds transition to gaits that are more efficient at higher speeds, and vice versa [118]. Advantageously, CoT is a metrics that can also be applied to quadruped robots, in turn presenting the theory of governing gait transitions based on CoT an attractive concept. However, to leverage this strategy and the efficacy of existing and future control frameworks, any proposed gait transition method must satisfy three primary prerequisites: 1) it must be independent from the control framework (it necessitates no additional effort or alteration to the existing control framework for its implementation), 2) it must be general enough to be applicable to quadrupeds of any size or weight, and 3) it must be resource efficient, not depending on a significant amount of time for implementation nor relying on data that may be inaccessible for all quadrupeds. Specific to this project, these prerequisites are essential for ensuring that the developed gait scheduler is compatible with any locomotion controller architecture that will arise in the next chapter. In addition to these primary prerequisites, of course, any gait transition method must maintain stability and efficiency. For instance, the conventional method of transitioning between gaits is simply to trust the stability that the control framework offers to mitigate the need

for designing gait transitions, defaulting to abrupt gait switching [12]. Nonetheless, switching gaits during highly dynamic locomotion can engender serious stability issues.

In terms of meeting all three key prerequisites, the closest existing approach is linear interpolation of gait parameters [67]. This method is independent of the control framework and has potential for generality. Albeit in its current state, it is inefficient and does not account for without automatic parameter adjustment. Albeit, in its current state it does not account for stability nor efficiency; this transition method is naive and similarly to hard switching primarily relies on the inherent proficiency of the wider control framework. Without informed automatic adjustments of gait parameters during transition, this method remains inefficient. Evidently, no existing method has achieved all prerequisites, unveiling the core limitations with these methods. The solution developed in this chapter, rooted in biomechanics, utilises the Froude number, which links gait selection with the linear velocity of various-sized quadrupeds [113, 109]. Thus, a Froude number-based bio-inspired gait transition method offers universality in gait selection and transition design (Fig. 4.1). Being dimensionless and unaffected by physical parameters, it allows for dynamic transition modifications, paving the way for an independent, general, and efficient method. This chapter aims to achieve the chapter objectives through the following actions, which double up as contributions to the field:

1. Analysis of CoT's variation across gaits with increasing Froude number to discern the ideal shift point for efficiency.

2. Leveraging the Froude number to devise governing equations defining gait transition traits.

3. Assessment of the bio-inspired transition's generality and proficiency through tests on four quadruped robots, and comparison with two other methods' CoT efficiency.

4. Stability, efficiency, and generality validation of the gait transitions via hardware tests involving an A1 robot demonstrating gait switches under nominal and enhanced mass and height configurations.

5. Comparative analysis of the proposed method's stability and CoT proficiency against two other methods in hardware tests.

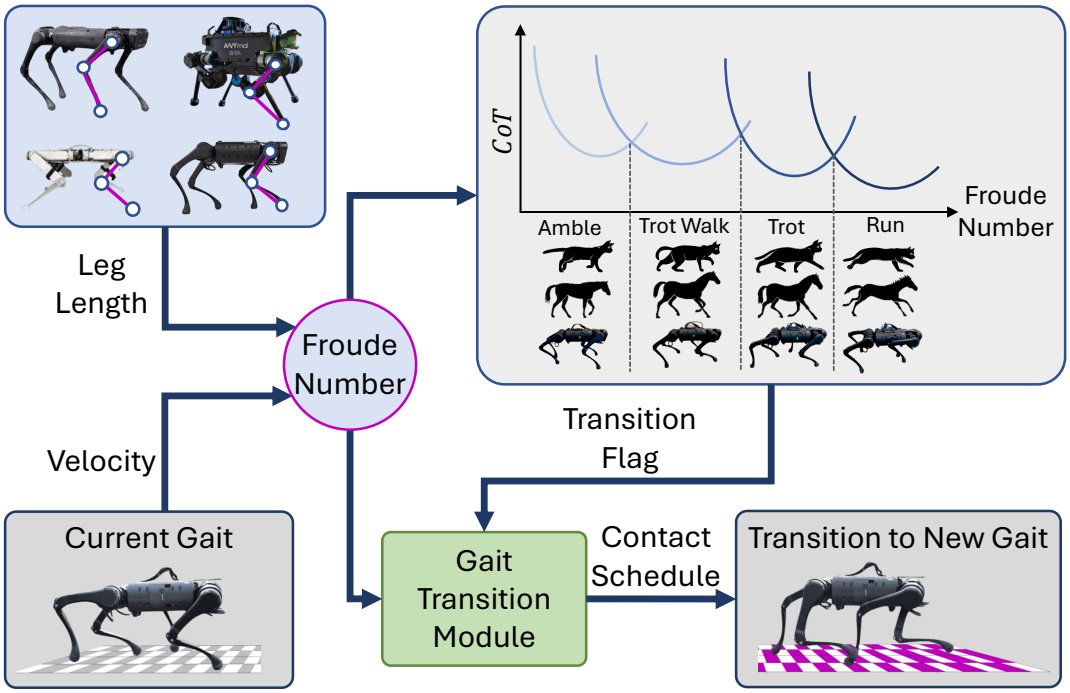


Figure 4.1: An overview of how the Froude number can be leveraged to govern and trigger gait transitions.

The remainder of this chapter is organised as follows: Section 4.2 introduces the control framework for the gait scheduler and expounds the gait scheduler and bio-inspired transitions. Section 4.3 presents the simulations, analyses shaping the transitions, and validates these transitions via hardware experiment. Lastly, Section 4.4 concludes and suggests potential research trajectories.

4.2 Methods

4.2.1 Control Framework

To test the developed gait transition and selection method a locomotion control framework is required. With the DRL locomotion policy yet to be developed, a standard optimal control framework is used and the developed bio-inspired gait scheduler (BGS) is simply embedded within it, as show in Fig. 4.2. The optimal control framework contains four main units, namely a gait generation module, a motion planning module, a whole-body control module, and a state estimation module. The gait generation

module is responsible for generating the contact schedule and contact timings. The motion planning module computes the desired body pose and ground reaction forces.

Gait Generator

The gait generator is designed based on a bio-inspired method detailed in Section 4.2.2 to transition between gaits and schedule contact states and timings for all feet according input velocity references. The desired contact states and contact timings will be used in the motion planning module.

Motion Planner

The motion planner receives input velocity commands and contact timings of all feet from the gait scheduler and computes body motion references, foot trajectories and ground reaction forces with a 3-D single rigid body model. This simplified model promotes computational efficiency and enables continuous online re-planning of the motion references, resulting in reactive behaviours of the robot. The desired foot location is based on the Raibert heuristic with a capture-point-based feedback term [144], while the reaction force is obtained by solving a QP minimising the body state error while respecting body dynamics [142].

Inverse Dynamics Whole-Body Controller

With the reference motions and forces from the motion planner a whole-body hierarchical inverse dynamics controller [134], accounting for system dynamics, is needed to compute joint torque commands. It formulates all robot tasks as affine functions of generalised accelerations and external generalised forces, and defines strict priorities and the importance of weighting of each task within a hierarchy. After solving a sequence of QPs to get the optimal generalised accelerations and external forces, joint torques are calculated from them to track as closely as possible the desired task-space motion and force references.

State Estimator

The state estimator provides a high-rate, low-latency estimate of the full state of the robot for the motion planner and whole-body controller at every control loop. A proprioceptive state estimator [12] is used in the framework, fusing inertial measurement unit

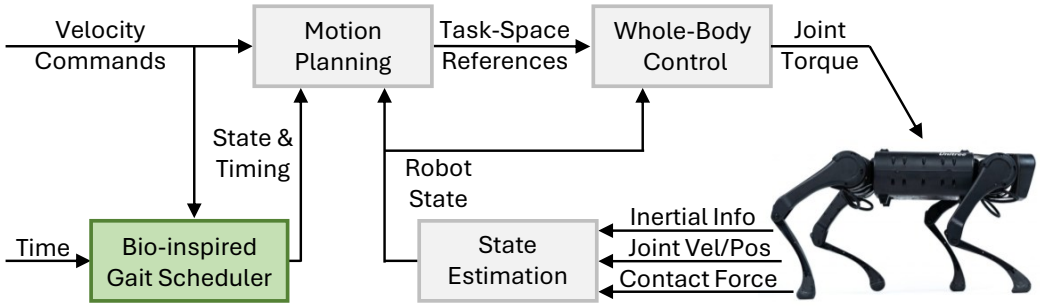


Figure 4.2: The control framework used within this chapter to analyse and validate the proposed BGS.

(IMU) information, force data, and kinematic measurements. Unlike other estimation approaches that solve a fully coupled inference problem to deduce position and orientation pose estimates, the state estimator decouples orientation estimation from the base velocity and position estimation. With the IMU gyroscope and accelerometer readings, an orientation filter is employed to estimate the body orientation expressed as a rotation matrix. A linear Kalman filter is used to estimate the base velocity and position with the orientation result and leg kinematics measurements.

4.2.2 Bio-inspired Gait Scheduler

Within the gait generation framework (Fig. 4.2), the BGS module is responsible for generating the contact schedule and timings. In this chapter, the standard gait scheduler is further developed, though taking the findings of bio-mechanics studies related to the Froude number [113, 109] and applying it quadruped robots, to implement automatic gait switching and stable transitions between a set of designed gaits based on real animal locomotion.

Baseline Gait Scheduler

The gait scheduler determines the time and duration that each leg is in contact with the ground, based on a set of gait parameters, to generate periodic phase-based gaits. The scheduler stores a set of phase variables, $\phi_i \in [0, 1]$, for each leg $\phi_1, \ldots, \phi_4$, where the phase depicts the state of each leg in the gait cycle. At the start of the gait cycle, each leg starts in stance at a phase value of $\phi_i = 0$. As ϕ_i increases, and passes the threshold of $\phi_i > \phi_{stance}$, the leg switches to a swing state. Once it further increases

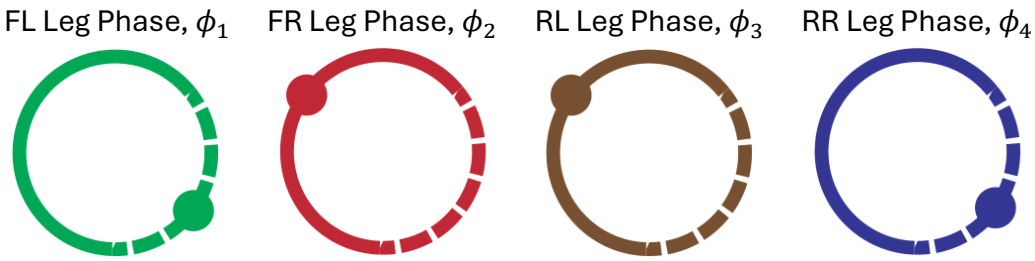


Figure 4.3: The large hollow cycle includes stance (solid) and swing (dashed) states. The small solid cycle represents the phase offset.

to a maximum value of 1, the phase wraps around to zero, starting the cycle again.

To design gaits to be used in this scheduler, three parameters must be defined: gait period, duty factor and phase offset for each leg. Fig. 4.3 shows the typical mapping of a trot gait. The gait period is used to define the stepping frequency; at higher velocities, a smaller gait period is more suitable. During locomotion where a gait transition is not occurring, the phase can be calculated by

$$\phi_i = \frac{t - t_{i,0}}{T}, \tag{4.1}$$

where t is the current time, $t_{i,0}$ is the start time of the current gait period of the i-th leg, and T is the gait period. The duty factor, $d_i \in [0, 1]$, determines the proportion of the gait cycle during which a leg is in stance by switching the leg to swing when $\phi_i = d_i$. With a constant gait period, a smaller d_i would result in each leg having an increased aerial time and create a more dynamic gait. The phase offset, $\theta_i \in [0, 1]$, serves to coordinate the legs to produce different gait patterns based on the leading leg through the relationship $\phi_i = \phi_1 + \theta_i$, adjusting the phase of leg i. These phase offsets can be selected based on studying animal gaits in the effort of recreating them.

Designed Gaits

For the design of the gaits, direct inspiration is taken from real animal gaits to determine the gait parameters of an amble and trot [114]. The trot gait is further broken down into walking trot, trot and run gaits for improved locomotion at different speeds; in accordance with Section 4.2.2, the walking trot has an increased T and d_i while the run has a decreased T and d_i. The static parameters for these gaits are defined in Table 4.1.

Table 4.1: Gait design parameters.

Name	Period	Duty Factor	Phase Offset
Amble	1.00	0.80	0.00, 0.75, 0.50, 0.25
Walking Trot	0.50	0.60	0.00, 0.50, 0.50, 0.00
Trot	0.40	0.50	0.00, 0.50, 0.50, 0.00
Run	0.30	0.40	0.00, 0.50, 0.50, 0.00

Governing Gait Transitions with Froude Number

As discussed in Section 4.1, the Froude number is a dimensionless number used to characterise gait transition speeds of different-sized quadrupeds,

$$\Omega = v^2/gh, \tag{4.2}$$

where Ω is the Froude number, v is the linear velocity of the base (in this case, the robot's base), g is the acceleration due to gravity, and h is the maximum leg length. This formulation has been built around the dynamic similarity hypothesis, which states that quadrupeds of different sizes will have the same gait characteristics at a given Ω independent of their size [112]. As such, h can be any parameter as long as it is a geometric dimension that can be multiplied by a factor to get the same dimension of another. In this work, h is chosen as the maximum leg length rather than the hip height, as unlike animals that have a constant hip height at stance, robot quadrupeds have a configurable stance height.

As it is known that animals transition between gaits when energy consumption becomes inefficient at the current gait, which tends to occur at similar Ω values, the same principles can be applied to quadruped robots. In this case, energy consumption can be quantified by CoT [145], which is formulated in this work as P,

$$P = \frac{\sum_{i=1}^{k} \tau_i \dot{\theta}_i}{mgv}, \tag{4.3}$$

where τ_i is the joint torque, $\dot{\theta}_i$ is the joint velocity, k is the total number of actuated joints, m is the mass of the robot, and v_{cmd} is the velocity command sent to the robot. By comparing Ω against P as velocity increases for each gait outlined in Table 4.1, the optimal Ω ranges for each gait can be selected. It should be noted that during the

calculation of Ω and P, v is estimated through the use of the state estimator outlined in Section 4.2.1. Data for this comparison is collected in simulation; hence, details on how the Ω ranges detailed in Table 4.2 are selected are presented in Section 4.3.1.

This formulation of CoT has been used in this work to find these ranges, as it is a standard metric used in studying legged locomotion efficiency. However, any metric or extension of this P term that describes legged locomotion efficiency, such as accounting for the efficiency of motors, could be used as Ω itself is not dependent on efficiency. For example, if two identical quadrupeds that only differ in their motor efficiency travelled at the same velocity, even though the less efficient quadruped would expend more power, both quadrupeds would still have the same Ω. This line of thought can be extended to where one of the quadrupeds is walking on flat ground and the other is on rough terrain. As long as they are travelling at the same speed, they will still have the same Ω regardless of the extra energy the quadruped on rough terrain has to expend to keep up. Combining this with the ability for Ω to ignore the difference in quadruped size, it can be stated that this method to find the optimal Ω ranges for each gait across different velocities is sufficient to enable generic optimal gait selection in quadruped robots.

Gait Blending for Transitions

To ensure that the transition between gaits is stable, the gait parameters of the current and new gait are blended together. Taking inspiration from [146], weights are used to transition between the gait parameters based on transition duration. These weights are defined as

$$w_1(n_{\text{trans}}) = 1 - (n_{\text{trans}}/D_{\text{trans}}), \quad w_2(n_{\text{trans}}) = n_{\text{trans}}/D_{\text{trans}}, \tag{4.4}$$

where $w_1(n_{\text{trans}})$ and $w_2(n_{\text{trans}})$ are the weights used for the current and new gait parameters respectively, n_{trans} is the n-th time step of the transition and D_{trans} is the number of time steps in the transition. These weights are applied to all parameters detailed in Table 4.1 for a stable and smooth transition. The only parameter not blended using weights is the phase offset, which will be covered later in this section. The transition duration, in terms of D_{trans}, is defined as

$$D_{\text{trans}} = \left(\frac{D_{\text{current}} + D_{\text{new}}}{2}\right) C, \tag{4.5}$$

where D_{current} and D_{new} are the number of time steps in the current and new gait respectively, and C is the number of cycles the transition should occur over. Through modulating C, the duration of the transition can be controlled. With faster transitions at higher velocities becoming more stable, the following definition of C is formulated to create an inverse relationship between Ω and transition duration,

$$C = \left(1 - \frac{\Omega^*}{\Omega_{\text{crit}}}\right) C_{\text{max}}, \tag{4.6}$$

where Ω^* is the Froude number at which the transition begins, Ω_{crit} is the maximum stable value of Ω for a given gait, and C_{max} is the maximum cycles any transition should occur over. Ω_{crit} is conservatively selected to have a value of 1 as this the limit of a stable walking gait [147]; the closer the Froude number is to this limit, not only is a fast transition more stable but is also required to preserve stability and helps improve efficiency by changing to a more efficient gait quicker. Furthermore, C_{max} is selected as 2 as this is the longest number of cycles quadrupeds take to transition between gaits [143] as although faster transitions are more desirable, if they are too quick this can cause instabilities due to the resultant reduced resolution of the transition, hence C_{max} is selected as this conservative value.

To ensure that this gait transition method can handle rapid accelerations, a technique for decreasing the transition duration is devised. In a scenario where the velocity increases so rapidly that it causes another transition to be initiated before the current transition is complete, this may result in a significant loss of stability and efficiency. To address this issue, the value of n during transition becomes a function of Ω according to the following relationship:

$$n_{\text{trans}}(\Omega) = \begin{cases} 1 + \left(\frac{\Omega - \Omega^{\text{lb}}_{\text{up}}}{\Omega^{\text{ub}}_{\text{up}} - \Omega^{\text{lb}}_{\text{up}}}\right) D_{\text{trans}} & \text{if } \Omega \geq \Omega^{\text{ub}} \\ 1 + \left(\frac{\Omega^{\text{ub}}_{\text{down}} - \Omega}{\Omega^{\text{ub}}_{\text{down}} - \Omega^{\text{lb}}_{\text{down}}}\right) D_{\text{trans}} & \text{if } \Omega \leq \Omega^{\text{lb}} \end{cases} \tag{4.7}$$

where $\Omega^{\text{lb}}_{\text{up}}$ and $\Omega^{\text{ub}}_{\text{up}}$ are the lower and upper Ω bounds of the shift up gait, $\Omega^{\text{lb}}_{\text{down}}$ and $\Omega^{\text{ub}}_{\text{down}}$ are the lower and upper Ω bounds of the shift down gait, and Ω^{lb} and Ω^{ub} are the lower and upper Ω bounds of the current gait. With this new definition of $n_{\text{trans}}(\Omega)$, the transition duration has an inverse quadratic relationship with velocity, as the aforementioned blending weight is dependent on this parameter.

The only other parameter left to blend during transitions is the phase offset. Rather than blending the two offsets, it is adjusted during transition so that it phases into the

offset of the gait that is being transitioned to,

$$\theta_{j\text{trans}} = \left(\frac{\theta_{j\text{new}} - \theta_j}{D_{\text{trans}}} \right) n_{\text{trans}}(\Omega), \tag{4.8}$$

where $\theta_{j\text{trans}}$ is the phase offset during transition, and $\theta_{j\text{new}}$ and θ_j are the new and current gait phase offsets respectively. This enables smooth transition between gaits of different phase offsets, while also adjusting the phase offset at the rate defined by (4.7). Overall, with gait transitions being governed by the quadruped's Ω, as visualised by the flowchart of be BGS in Fig. 4.4, transitions occur automatically with velocity and adjust accordingly based on its magnitude to preserve stability and efficiency. Furthermore, this leveraging of Ω enables its use with a variety of quadruped robots, independent of their size and motor dynamics.

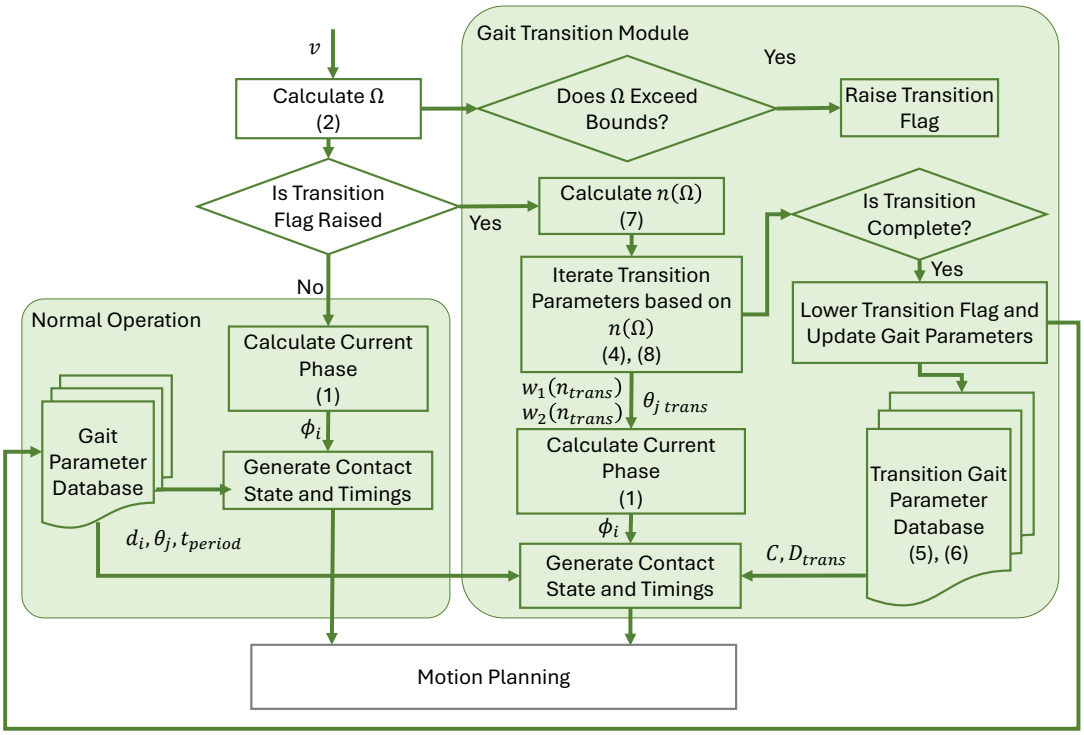


Figure 4.4: Flowchart of the BGS.

4.3 Results and Discussion

To realise the proposed Ω-governed method, first the Ω bounds have to be attained through simulation across a diverse set of quadruped robots. This is to test if the dynamic similarity hypothesis that Ω is generated from holds true in quadruped robots and to promote the generality of the method. Once complete, these Ω bounds are implemented within the BGS and a set of simulation and hardware experiments are completed to analyse its performance.

4.3.1 Gait Froude Number and CoT Analysis

To analyse transitions and identify the Ω bounds outlined in Section 4.2.2, a set of simulations was completed. For these simulations, the physics-based simulator PyBullet [138] was used, and the control frequency was set to 500 Hz. For identifying the Ω bounds, all gaits in Table 4.1 were used with a linear velocity profile, starting from 0 m/s up to 1.5 m/s and at an acceleration of 0.1 m/s^2. The simulation was completed with the A1, AlienGo, ANYmal B and Solo12 quadrupeds to validate that the Ω value can be used to select points of gait transitions to improve efficiency through selecting the gait with the minimum CoT at a given velocity.

From analysing Fig. 4.5, it can be observed that the Ω ranges at which transitioning between gaits to minimise CoT are very similar across all quadrupeds. In turn, this validates the theory that Ω can be used to select the optimal point of gait transition for quadrupeds, regardless of their size. However, without a method of selecting the best Ω value to transition, selecting this value is ambiguous. To this end, a weighted average is used to select the Ω values by utilising a dimensionless metric of the need for a legged system to remain stable, $E = g/hf^2$ where f is the average of the stride frequency of the gaits in question [109], as the weight,

$$\Omega^* = \frac{\sum_{i=1}^{r} \Omega_i^{\text{mean}} E_i}{\sum_{i=1}^{r} E_i}, \tag{4.9}$$

where Ω_i^{mean} is the mean of the Ω intersection values and r notates the quadrupeds used in this investigation. Through using E as the weight, it creates a bias towards the Ω^{mean} of the smaller quadrupeds as due to their size they have a greater requirement to maintain equilibrium [109], which overall improves the transition conditions for these quadrupeds.

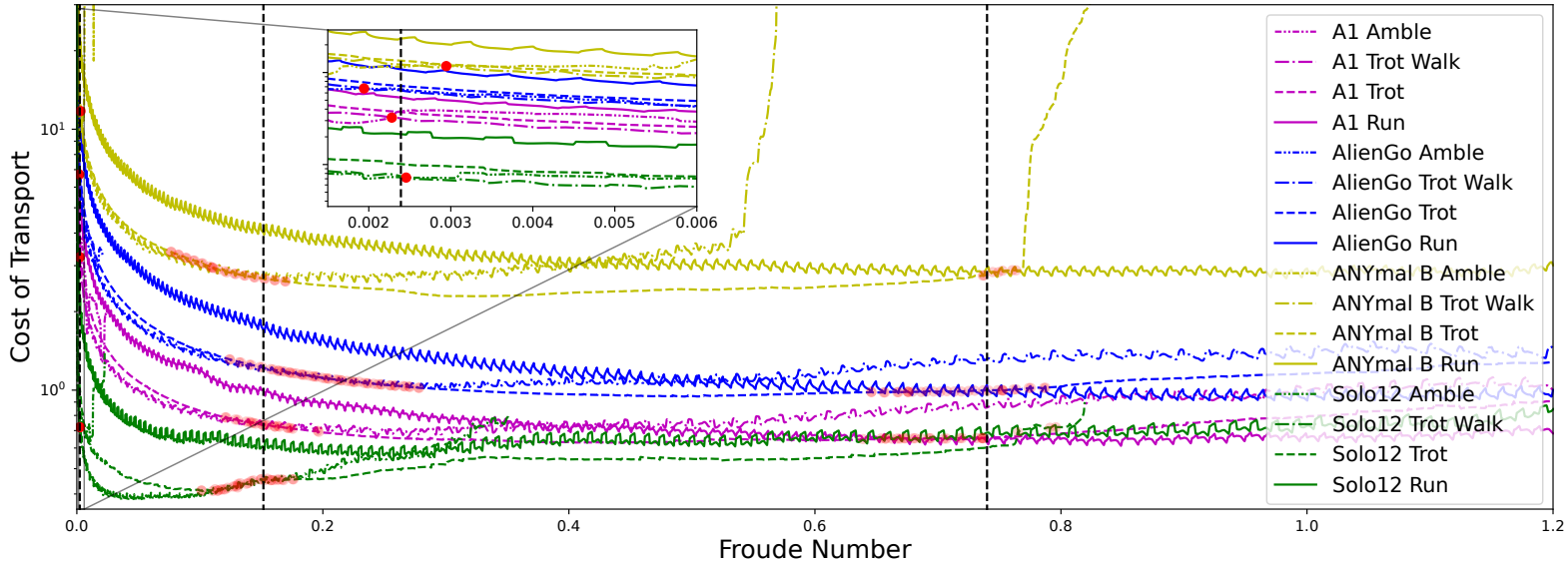


Figure 4.5: To find the optimal points to transition between gaits, a set of gaits are run on the A1, AlienGo, ANYmal B and Solo12 to select and verify the Ω bounds. This is done through taking a weighted average of points of intersection between subsequent gaits indicated by the red dot markers. Through defining these bounds, as indicated by the black dashed lines, the BGS enables the robot to track the minimal CoT possible with the available gaits as velocity increases.

It took around 2 minutes for all final Ω points to be selected, which are presented in Table 4.2, which are then implemented within the BGS and tested on the A1, AlienGo, ANYmal B and Solo12 quadrupeds, as illustrated in Fig. 4.6.

Table 4.2: Gait Froude number bounds.

Name	Amble	Walking Trot	Trot	Run
Ω^{lb}	0.0009	0.0024	0.7400	0.7400
Ω^{ub}	0.0024	0.1517	0.7400	1.0000

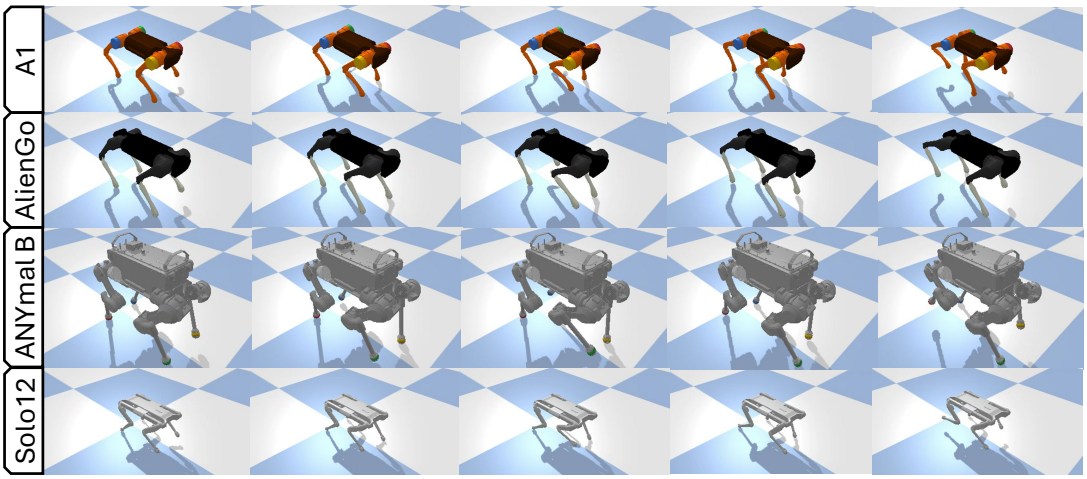


Figure 4.6: The proposed bio-inspired gait transitions has been tested on the A1, AlienGo, ANYmal B and Solo12 Quadrupeds. This figure highlights the transition between amble and walking trot.

4.3.2 Comparing Gait Transition Methods

To evaluate the bio-inspired gait transition method of the BGS against other existing methods, namely hard gait switching and linear interpolation, simulations were completed with the four quadrupeds for all three transition methods. These other transition methods were selected for this study as they are the only other methods that offer the same level of generality while being independent from the rest of the control framework. In each simulation an increasing velocity command, accelerating at a rate of 0.4 m/s^2 up to a maximum of 1.5 m/s, was sent to the control framework. Once this maximum value was reached, the command velocity decelerated back to 0 m/s at the same rate.

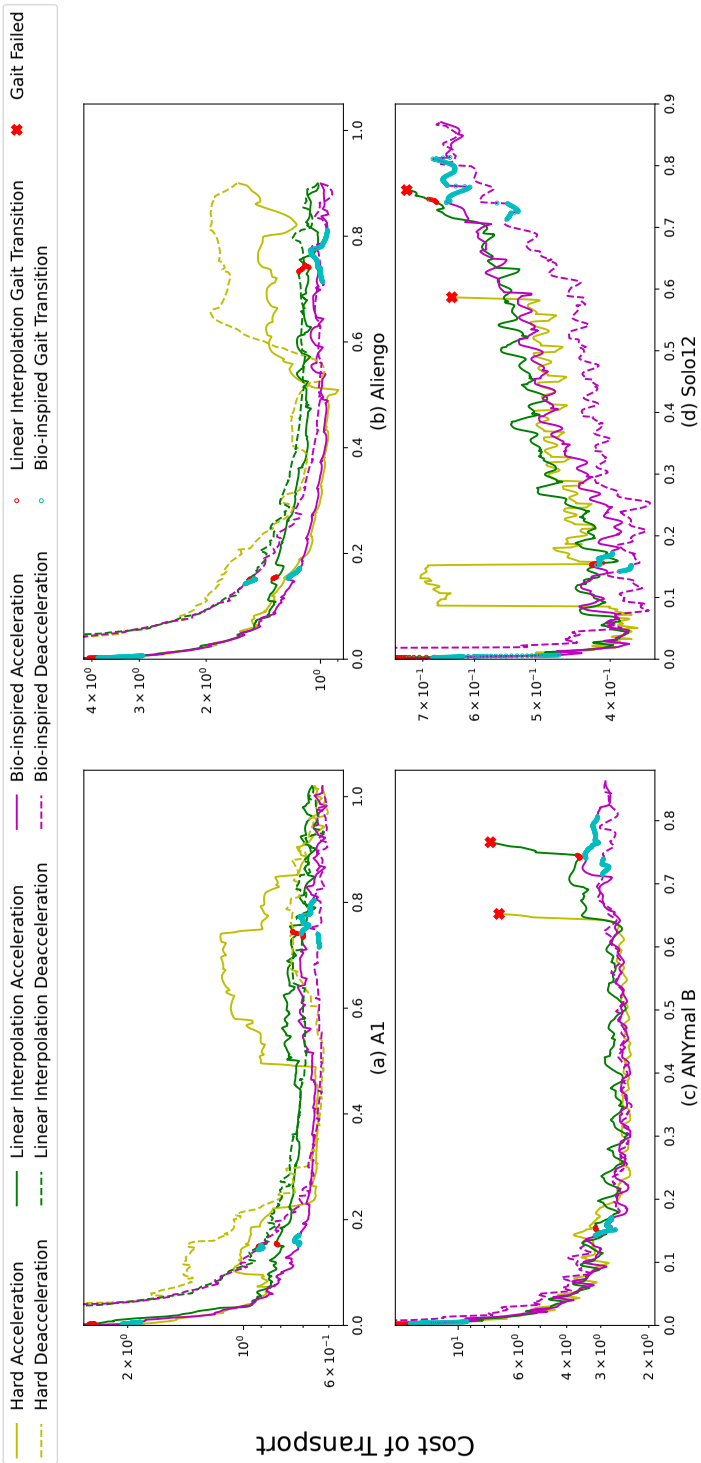


Figure 4.7: Comparing the CoT, the y-axis values, across different gait transition methods with varying Ω, the x-axis values.

From the results presented in Fig. 4.7, although similar performance is observed at lower Ω values, at higher velocities the bio-inspired gait transitions significantly outperformed the other methods in terms of CoT, and in the cases of ANYmal B and Solo12 it prevents the gait from failing. This is primarily due to the hard switching and linear interpolation methods having fixed and shorter (in the case of hard switching, it is instantaneous) transition durations. Although this could reduce CoT, the reduced transition resolution produces instabilities. It could be argued that simply increasing the transition duration could resolve this issue. However, this requires meticulous tuning of the transition parameters which may not be suitable for various accelerations and will definitely not be generic enough for the same parameter to be used across different quadrupeds. In contrast, in the case of the bio-inspired gait transitions, the transition duration is adaptable based on the acceleration to provide stability, using (4.6) and (4.7), and generality is provided through governing these equations with Ω. This improved stability at high speeds and accelerations is further highlighted in Fig. 4.8, where both hard switching and linear interpolation transition methods cause the gait to fail during the transition between trot to run, while accelerating at 0.4 m/s^2 at a velocity of approximately 1.3 m/s, whereas the bio-inspired method preserved stability during transition.

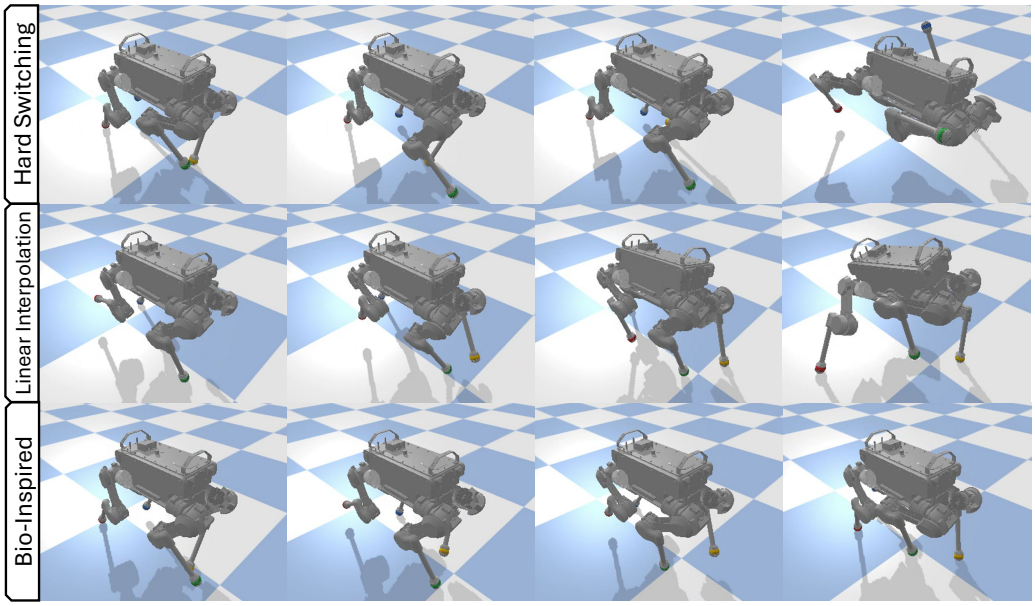


Figure 4.8: A comparison between gait transition methods when transitioning from trot to run at 1.3 m/s and 0.4 m/s^2.

4.3.3 Hardware Deployment and Validation

To validate the bio-inspired gait transitions and compare its performance to other transition methods, a set of experiments were completed using A1. During the experiments, the velocity command followed a profile that switched between accelerating at 1 m/s^2 and maintaining constant velocity, up to a maximum of 1.2 m/s from rest. Once this maximum value was reached, a similar velocity profile was used to reduce the velocity back to 0 m/s. To complete a comparative study, the bio-inspired, linear interpolation, and hard switching gait transition methods were all run ten times using this profile, and points of transition were determined by the Ω bounds in Table 4.2, with the snapshots of the bio-inspired method's transitions presented in Fig. 4.9, and the corresponding data from these experiments being presented in Fig. 4.10. To validate the method's generality, this experiment was again run ten times but with the A1 modified to mimic a taller and heavier robot. To control the A1 robot, the control framework was run on an Ubuntu computer at 500 Hz and sent torque commands via an Ethernet connection.

Figure 4.9: Snapshots of the gait transitions achieved through the bio-inspired gait transition method's hardware deployment.

Transition Stability

After tracking the success rate of each transition method across their ten respective test runs, as summarised in Table 4.3, it was found that the bio-inspired transition method significantly outperformed the other methods. Most failures occurred at high velocities, and this vast difference in success rate is deduced to result from the bio-inspired

Table 4.3: The success rates of each transition method across ten runs of the experiment.

Method	Successes	Failures	Success Rate
Hard Switching	0	10	0%
Linear Interpolation	2	8	20%
Bio-inspired: Nominal	9	1	90%
Bio-inspired: Mass and Height ↑	9	1	90%

method's ability to adjust transition resolution to preserve stability. Combining the fact that hard switching and linear interpolation methods lack this functionality and that real-world operation introduces disturbances not modelled in simulation, the instabilities observed in Fig. 4.7 are further exacerbated, causing repeated failures. In turn, this demonstrates that the bio-inspired method can reliably outperform these existing methods in terms of stability. Moreover, throughout the successful implementations of the bio-inspired approach, a maximum roll of 0.12 rad was discerned, as depicted within Fig. 4.11, further corroborating the maintained stability. It warrants mention that while stability on flat terrain is evidenced, the preservation of such stability during gait transition on uneven terrain may be compromised. This owes to the fact that our proposed technique does not incorporate additional dynamics considerations, thereby revealing its principal constraint.

Cost of Transport

Across all experiments, the CoT is observed to decrease as velocity increases and the gait is transitioned to one more suitable for faster velocity commands, as detailed in Fig. 4.10. Similar CoT trends are observed compared to the simulation results presented in Fig. 4.7, demonstrating that tracking of the lowest CoT across gaits for the corresponding Ω values transfers successfully to hardware. Similar to the trends seen in Fig. 4.7, the bio-inspired method maintains the lowest CoT across the different methods for the majority of the experiment due to the stability during transition that it provides. This is particularly evident at higher velocities, where it can be seen that the transition duration is adjusted so that the transition has a higher resolution for stability, to account for this highly dynamic scenario, while remaining short enough to preserve CoT.

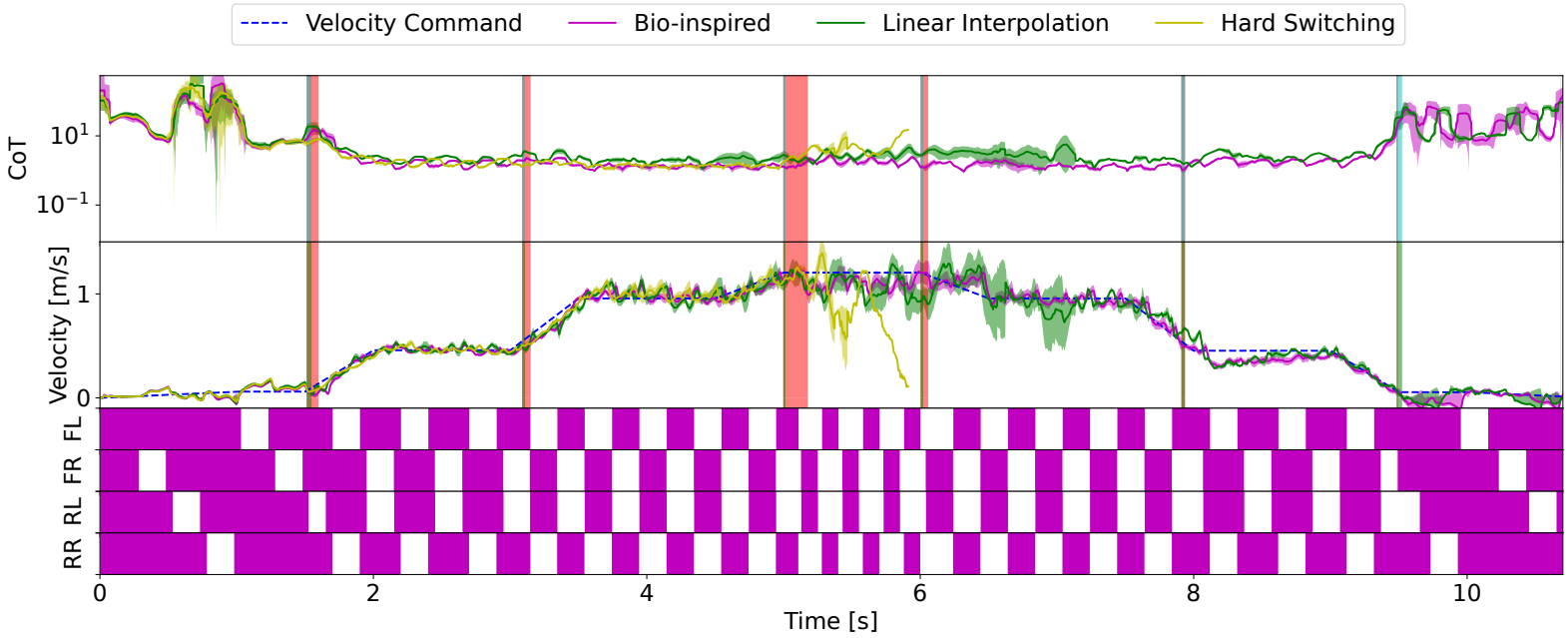


Figure 4.10: Graph showing the CoT, command velocity, estimated velocity, and contact states of the real robot during the experiment (from top to bottom). In the CoT and velocity plots, data from all ten experiments, for each transition method, are presented as average trend lines and variance displaced by the shaded regions. In the bottom four plots, a shaded region shows that a foot is in contact with the ground, while the white spaces show that it is in swing. The bio-inspired gait transition phases are highlighted in red, while the linear interpolation transition phases are in blue.

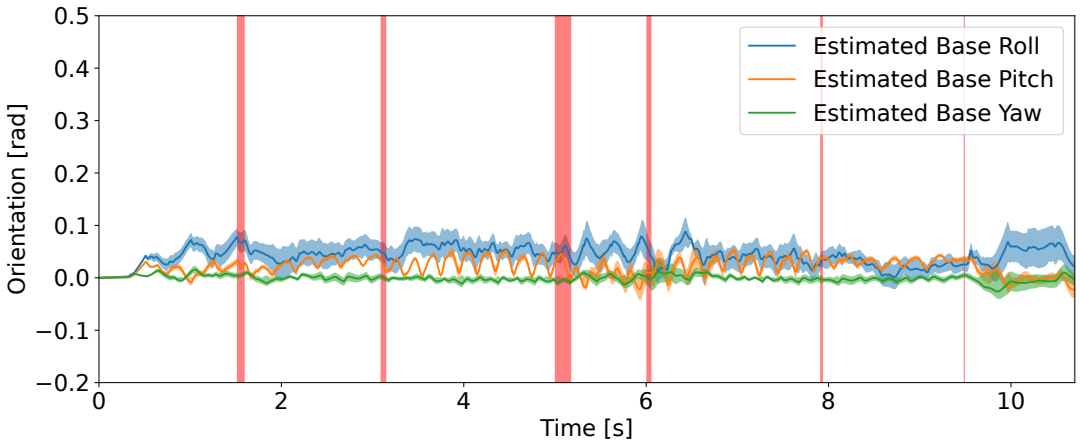


Figure 4.11: Plot of the A1's base roll, pitch, and yaw for all 10 hardware experiments using the developed bio-inspired transition method corresponding with Fig. 4.10.

Whereas, in the cases of linear interpolation and hard switching, the instability caused by their inability to adjust transition parameters results in increased CoT at higher velocities; this effect can be seen to have a lasting effect even after a transition is complete.

Velocity Command Tracking

By comparing the plot of the velocity command to the estimated linear velocity during the bio-inspired method experiments (estimated by the state estimator) in Fig. 4.10, it is observed that during gait transition tracking of the velocity command is virtually unaffected. This validates that the bio-inspired method successfully transitions between gaits while preserving stability and normal function of the robot, as further supported by Fig. 4.11 where the low variations in robot base orientation is a strong indication of stability. However, this is not the case for both linear interpolation and hard switching methods at high velocities, where the previously discussed loss instability at this high velocities causes poor command tracking. This further demonstrates that the bio-inspired method is able to outperform these other existing methods.

Contact State During Transitions

The contact states during this experiment are presented in Fig. 4.10, which details how the contact states are adjusted during transitions for one run of the bio-inspired transition method. During these transitions, the duration and frequency of these states can be seen to blend between the current and new gaits. This demonstrates how the gait blending process detailed in Section 4.2.2 facilitates smooth transitions, aiding in preserving stability. This is further emphasised by there being no point during the experiment when only one leg in stance or any harsh changes in contact states occur. Furthermore, it can be observed that the duration of transitions while the robot is decelerating is significantly shorter compared to when accelerating. This is due to the gradient of the velocity profile being steeper, which, as outlined by (4.7), causes faster transitions to preserve stability.

Transition Generality

The high success rate achieved by the bio-inspired transition method without necessitating any modifications during deployment, substantiates that our proposed method boasts considerable generality. To further scrutinise this generality and fortify our assertion, the A1 was configured to transport a payload equating to 26% of its mass and reach a target height that is 15% higher than its customary height, thereby simulating a larger and heavier robot, as depicted in Fig. 4.12. This adjusted configuration was tested ten times and also demonstrated a high success rate of 90%. These results clearly attest to the generality of the proposed method.

Figure 4.12: Snapshots of the nominal (top) and modified (bottom) A1 during a trot to run transition.

4.4 Chapter Conclusion

In this chapter, a bio-inspired gait transition approach has been developed and integrated into a gait generator to form the BGS, wherein the Froude number is employed to establish several correlations governing automatic gait transitions, transition duration, and gait blending. As the Ω value at a specific speed is independent of the robot's size and motor dynamics, this method proves to be compatible with various quadruped robots, as confirmed through simulation. Furthermore, the transition approach has been validated on physical hardware, demonstrating its generality and proficiency in producing stable transitions that preserve the cost of transport while surpassing other transition methods in these metrics during a comparative analysis.

However, despite the BGS's proficiency in this efficiency focused approach, there remains several key limitations:

- The BGS currently doesn't account for terrain variations and has only been tested on flat terrain.
- The BGS currently does not account for off nominal scenarios, such as a critical loss of stability, and has not been tested for such scenarios.
- If the BGS is to be required to switch to a gait that can handle off nominal scenarios, such as bounding or pronking, this is not possible with the current formulation as non-specific gait-to-gait transitions are not possible while Ω bounds are used to initiate gait transitions.

As such, along with the development of the DRL locomotion policy, the next chapter will also augment the BGS for use with a wider set of gaits with gait-to-gait agnostic transitions and a DRL gait selection policy that utilises animal gait transition strategies to inform optimal gait selection. In turn this will fully realise project —Objective 3.

Chapter 5

Instilling Animal Gait Strategies within a DRL Locomotion Framework

With the first iteration of the BGS complete, and therefore generation of gait references possible, this primes the development and training of the DRL locomotion control modules outlined in Fig. 2.12. However, as discussed in the Section 4.4, the BGS first needs to be augmented to enable the framework to also use off nominal gaits, such as pronking and limping, to provide enhanced adaptability and utilise a similar range of gaits that animals use in their day-to-day lives; not only should the framework optimise for efficiency but also stability to realise the highly adaptable and versatile locomotion this project aims to produce. Once this next iteration of the BGS is complete, foundational animal locomotion metrics and models will be leveraged to instill the intrinsics of animal locomotion proficiency within a DRL locomotion and gait selection policies to achieve zero-shot deployment on a wide range of natural real-world terrains and stability critical scenarios. Work contributing to this chapter has been published in the following journal:

- Humphreys, J. and Zhou, C., 2025. **Learning to Adapt through Bio-inspired Gait Strategies for Versatile Quadruped Locomotion**. Nature Machine Intelligence, 7, pp.1141-1153.

A video summary of this chapter, featuring the hardware experiments, can be found at [148][1].

[1] https://www.youtube.com/watch?v=ecplBINQ3Tg

5.1 Introduction

Originally inspired by the remarkable adaptability of quadruped mammal locomotion, an ability shaped by innate and environmentally induced factors [106, 107, 108], the field of quadruped robotics has invested substantial resources into developing equally proficient locomotion frameworks. At present, as discussed in Chapter 2, the most advanced systems rely on end-to-end deep reinforcement learning (DRL), which involves training a multilayer perceptron [149] capable of navigating diverse environments.

These frameworks demonstrate impressive realisation of terrestrial locomotion skills, which can be classified into two groups (previously detailed in Section 2.7); Froude-characterised locomotion (no desired velocity normal to the ground plane and hence upholds the assumptions of the Froude number [109]) such as walking or running in real-world [99, 30], urban [150, 97] and deformable[30] terrains, and Froude-free locomotion (features a desired velocity normal to the ground plane) such as jumping between platforms [33], climbing over obstacles [100], and sure-footedness [49]. Despite these achievements, when it comes to Froude-characterised locomotion, which can account for about $\approx 70 - 90\%$ daily animal locomotion [110, 111], the adaptability of these frameworks remains constrained, as most systems are limited to deploying a single targeted gait or locomotion strategy.

In contrast, biomechanics research has shown that no single gait is universally optimal across all scenarios within Froude-characterised locomotion [5, 6, 7]. Animals adapt their locomotion by employing nominal gaits such as ambling, trotting, and running [114], while switching to specialised gaits such as hopping, pronking and bounding for off-nominal tasks such as predator evasion or obstacle navigation [115]. Current DRL frameworks fall short of replicating this level of Froude-characterised locomotion versatility. To address this limitation, some approaches have focused on training DRL policies to learn multiple gaits by providing reference motions during training [102, 48, 103, 104], or by learning from policies that specialise in specific gaits [66]. However, these methods remain insufficient when compared to the extensive capabilities observed in animal locomotion, which include:

- Adaptation of gait style for optimal performance in response to challenging terrains and perturbations.
  - Enabled by advanced gait selection strategies.

- Rapid deployment of a diverse set of task- and state-specific gaits.
  - Attributable to gait procedural memory.
- Seamless deviation from nominal gait motions to address off-nominal contact states.
  - Achieved through precise motion adjustments tailored to the environment.

Although existing DRL frameworks have demonstrated progress in implementing learned gaits, none successfully integrate all three attributes simultaneously. This gap highlights the considerable potential of biomechanics-inspired approaches to advance robotic locomotion. Animal gait transition strategies, which contribute to optimal performance and enable navigation of challenging environments, are believed to emerge from the minimisation of metrics related to energy consumption [116, 117, 109], mechanical work [122, 123, 124], instability [121, 125], and musculoskeletal forces [118, 119, 120]. However, no singular metric has been definitively identified as the sole driver of these transitions. Instead, it is hypothesised that a combination of these factors influences gait transition strategies [121, 126, 127].The concept of gait procedural memory, which facilitates the rapid deployment of a range of gaits, is thought to reside within the cerebellum of the animal brain. This region governs the coordination of limb movements for each gait learned by the animal [151, 152]. Similarly, adaptive motion adjustments, crucial for seamless adaptation to off-nominal contact states, are achieved through coordination between the mesencephalic locomotor region (MLR), which oversees locomotion execution [153], and the cerebellum. These adjustments rely on sensory feedback to modify limb movements in response to the animal's current state [154].

Despite these insights, there has been no prior attempt to simultaneously integrate all these attributes, gait transition strategies, procedural memory, and adaptive motion adjustments, within existing DRL locomotion frameworks. To address these current limitations and achieve this chapter's objectives, the DRL locomotion framework presented in Fig. 5.1 is designed to incorporate these key animal locomotion attributes. The framework, which hence forth with be referred to the Learning to Adapt (LTA) framework, demonstrates exceptional adaptability through zero-shot deployment in complex, real-world environments, relying solely on interoceptive sensors.

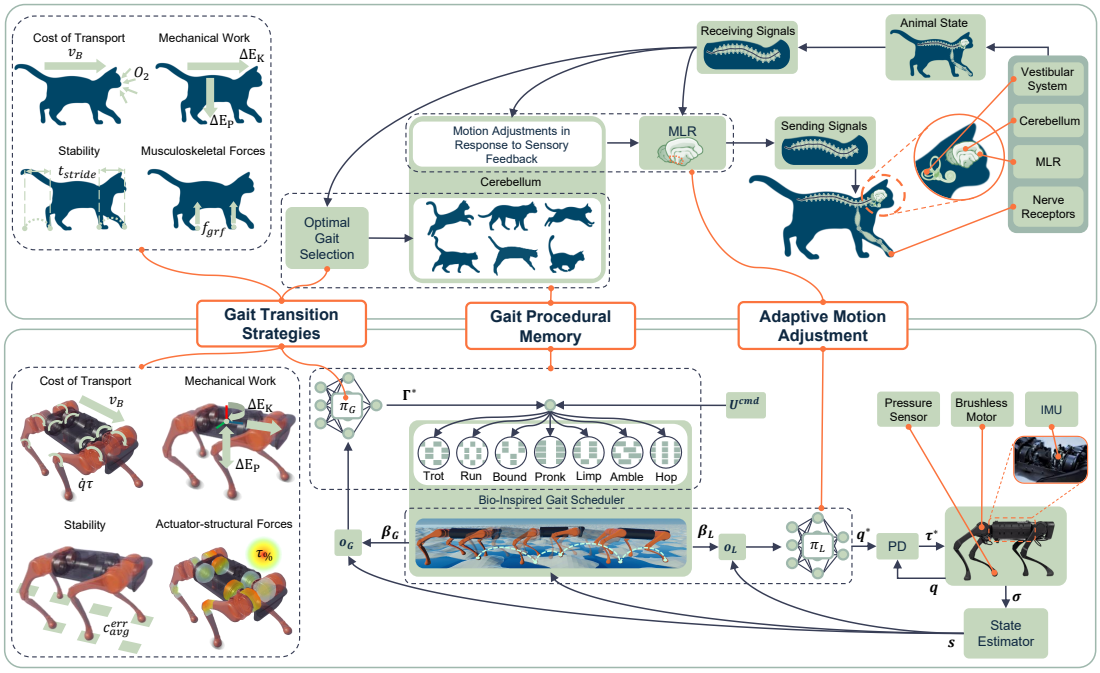


Figure 5.1: The LTA Framework. From taking an abstracted view of animal locomotion to determine the attributes of proficient locomotion, animal gait transition strategies are instilled within the gait selection policy, π_G, gait procedural memory is embedded within the bio-inspired gait scheduler, and adaptive motion adjustments are realised by the locomotion policy, π_L. π_G has been trained to minimise the animal gait transition metrics applied to the quadruped robot based on the current robot state, $\boldsymbol{s}$, and relevant bio-inspired gait scheduler output, $\boldsymbol{\beta}_G$, to select the optimal gait, Γ^*. The bio-inspired gait scheduler then generates the gait references informed by $\boldsymbol{s}$ from encoded high-level gait parameters. The gait references, $\boldsymbol{\beta}_L$, are then passed to π_L to inform of any adjustments to the nominal gait motions. The control framework is constant between training and deployment to improve sim-to-real transfer. Regarding acronyms, MLR is, mesencephalic locomotor region, PD is proportional derivative controller, and IMU is inertial measurement unit.

5.2 Methods

Within the LTA framework, presented in Fig. 5.1, a gait selection policy, π_G, is trained for optimal gait selection through minimising gait transition metrics adopted from biomechanics to generate the output $\Gamma^* \in [0, 7]$ which maps to a specific gait within [*stand*, *trot*, *run*, *bound*, *pronk*, *limp*, *amble*, *hop*]. This selected gait, coupled with the velocity command within $\boldsymbol{U}^{\text{cmd}} = [v_x^{\text{cmd}}, v_y^{\text{cmd}}, \omega_z^{\text{cmd}}, \Gamma^*] \in \mathbb{R}^4$, where v_x^{cmd}, v_y^{cmd} and ω_z^{cmd} are base velocities in x, y and yaw respectively, is passed to the bio-inspired gait scheduler (BGS) to generate gait references (along with transition references between any gait pair), as detailed in Section 5.2.2. These gait references are contained within the BGS outputs $\boldsymbol{\beta}_L$ and $\boldsymbol{\beta}_G$ for the locomotion policy, π_L, and π_G respectively through inclusion within their observation vectors $\boldsymbol{o}_L$ and $\boldsymbol{o}_G$. In this respect, the BGS acts as pseudo gait procedural memory. The gait references are adjusted based on the robot's state, $\boldsymbol{s}$, generated by the SE from the sensor feedback vector, $\boldsymbol{\sigma}$, which in turn reflects the relationship between the cerebellum and MLR. To realise the output joint positions of π_L, $\boldsymbol{q}^*$, they are passed through a proportional derivative controller to generate joint torque commands $\boldsymbol{\tau}^*$.

5.2.1 Control Framework Overview

At the core of this work, the Unitree A1 quadruped robot used in all experiments features 12 degrees of freedom, n, which are all modelled as revolute joints, with their angular positions denoted as $\boldsymbol{q} \in \mathbb{R}^n$ and its base orientation represented as a rotation matrix $\boldsymbol{R}_B \in SO(3)$. As outlined in Fig. 5.1, both π_L and π_G are integrated within a control framework and supported by the SE and BGS for generation of the robot's state data and gait references respectively. The final output of π_L is target joint positions, $\boldsymbol{q}^*$, which are converted into joint torques, $\boldsymbol{\tau}^*$, through the following proportional derivative controller that get sent to the motors,

$$\boldsymbol{\tau}^* = K_p(\boldsymbol{q}^* - \boldsymbol{q}) - K_d\dot{\boldsymbol{q}}, \tag{5.1}$$

where K_p and K_p are the proportional and derivative gains respectively. Throughout this work, a constant $K_p = 25$ N/m and $K_d = 1$ Ns/m are used while running at 1000 Hz, while π_L and π_G are run at 500 Hz and 100 Hz respectively.

5.2.2 Bio-inspired Gait Scheduler

As much more diverse behaviour is being targeted by the LTA, the BGS originally developed in Chapter 4 needs to have its functional scope expanded significantly. This takes the form of facilitating additional gaits, generating swing foot trajectories, and enabling gait agnostic transitions. Furthermore, as identified in Chapter 2 and Section 5.1, the LTA needs to preserve stability and efficiency to effectively achieve the same locomotion proficiencies that animals exhibit. With the previous method of Chapter 4 centering gait transition events around CoT, and in turn only considering efficiency, this method is replaced by the gait selection policy, π_G, to enable the instillation of all animal gait transition strategies. Consequently, the method of utilising Froude number bounds to trigger gait transitions is discarded in this revised BGS. Accommodating this expanded functional scope requires significant adjustments the formulation of the BGS. However, credited to its highly generic formulation, such adjustments to the BGS can be implemented with minimal effort and the core functionality of gait generation and proficient gait transitions are maintained.

The overhauled BGS primary output, $\boldsymbol{\beta}_L = [\boldsymbol{c}^{\text{ref}}, \boldsymbol{p}_x^{\text{ref}}, \boldsymbol{p}_y^{\text{ref}}, \boldsymbol{p}_z^{\text{ref}}] \in \mathbb{R}^{16}$, now defines the reference contact state of each foot, $\boldsymbol{c}^{\text{ref}} \in \mathbb{B}^4$, and their reference Cartesian position in the world frame x-axis, $\boldsymbol{p}_x^{\text{ref}} \in \mathbb{R}^4$, y-axis, $\boldsymbol{p}_y^{\text{ref}} \in \mathbb{R}^4$, and z-axis, $\boldsymbol{p}_z^{\text{ref}} \in \mathbb{R}^4$ which are calculated online using the Raibert heuristic [155] to account for the current state of the robot. Throughout this chapter, the limits enforced on the generation of $\boldsymbol{p}_x^{\text{ref}}$, $\boldsymbol{p}_y^{\text{ref}}$ and $\boldsymbol{p}_z^{\text{ref}}$ are 0.3 m, 0.2 m and 0.1 m respectively from the nominal local foot position. An adjusted version of the BGS output, $\boldsymbol{\beta}_G$, is used for π_G as not all the information in $\boldsymbol{\beta}_L$ is required. This has the form of $\boldsymbol{\beta}_G = [\boldsymbol{c}^{\text{ref}}, \boldsymbol{p}_z^{\text{ref}}, \Omega_{\text{stab}}, \kappa] \in \mathbb{R}^{10}$ where $\Omega_{\text{stab}} \in \mathbb{R}$ characterises the inherent stability of a gait [109], and $\kappa \in \mathbb{B}$ is a logical flag to indicate a state of gait transition.

Originally in Chapter 4, the Froude number [109], Ω, is used to trigger gait transition based exclusively on CoT which results in a set order of transitions. However, this method is not entirely suitable as now multiple biomechanics metrics and a set of auxiliary gaits need to be considered. One issue is that $\Omega > 1$ values are not compatible when calculating how many gait cycles, C, should a transition occur over. This has been resolved through calculating C through

$$C = e^{-2\Omega} \tag{5.2}$$

This relationship ensures an almost instantaneous transition at $\Omega \geq 2$, which is the typical value that quadruped animals transition to a run [109] instantaneously. Another limitation is that the calculation of the transition resolution, δ, (how quickly a transition should be progressed each time step) only enables the transition between set gait pairs; this was not an issue in Chapter 4 as CoT efficiency was the only metric considered. As π_G requires any gait transition pair to be possible, $\Omega_{\text{stab}} = g/hf^2$ [109] is utilised, where g is gravitational field strength, h is hip height and f is gait frequency. Through the use of Ω_{stab}, it is now possible to determine an indication of the inherent stability of any gait, hence a transition between a higher Ω_{stab} gait to a lower one should have smaller values of δ to increase the smoothness of the transition to promote stability. In the reverse scenario, a more harsh transition is more feasible hence larger values of δ should be produced for rapid transition. As such, δ is now calculated by

$$\delta = 1 + \frac{\Omega_{\text{stab}}}{\Omega_{\text{stab}}^{\text{next}}} \tag{5.3}$$

where $\Omega_{\text{stab}}^{\text{next}}$ is the Ω_{stab} of the gait that's being transitioned to. In essence, f of the current and next gait dictates the harshness of the transition. This behaviour is also reflected in animal gait transitions, where transitioning from running (higher f) to trotting (lower f) the transition is slower compared to the opposite scenario [156]. Overall, this augmented version of the BGS can achieve transition between any designed gait, while considering the inherent stability of the transition. Complete details of how $\boldsymbol{c}^{\text{ref}}$ is generated for each gait can be found back in Chapter 4 as the remainder of the process is kept the same. As a wider range of gaits are to be utilised in this locomotion framework than in Chapter 4 (to accommodate improved stability and adaptability) a wider range of gaits and their design parameters have been designed and are detailed in Table 5.1.

5.2.3 Policy Training

To simplify the training process, for both the locomotion policy, π_L, and gait selection policy, π_G, the training method, environment and network architecture are kept constant. Both policies are modelled as an MLP with hidden layer sizes $[512, 256, 128]$ and LeakyReLU activations. Subscripts L and G represent the specific parameters for the locomotion policy and gait selection policy respectively. The model-free DRL training problem for the policies is represented as a sequential Markov decision process (MDP),

Table 5.1: Expanded gait design parameters.

Name	Period	Duty Factor	Phase Offset
Trot	0.40	0.50	0.00, 0.50, 0.50, 0.00
Run	0.30	0.40	0.00, 0.50, 0.50, 0.00
Bound	0.40	0.40	0.00, 0.00, 0.50, 0.50
Pronk	0.50	0.50	0.00, 0.00, 0.00, 0.00
Amble	0.50	0.55	0.00, 0.50, 0.25, 0.75
Unnatural	0.40	0.50	0.05, 0.50, 0.50, 0.00
Hop	0.30	0.50	0.00, 0.00, 0.00, 0.00

which aims to produce a policy that maximises the expected return of the policy π,

$$J(\pi) = \mathbb{E}_{\xi \sim p(\xi|\pi)} \left[\sum_{t=0}^{N-1} \gamma^t r \right], \tag{5.4}$$

in which $\gamma \in [0, 1)$ is the discount factor, ξ is a finite-horizon trajectory dependent on π with length N, $p(\xi|\pi)$ is the likelihood of ξ, and r is the reward function. The PPO algorithm [24] is used to train all policies and the hyperparameters used are detailed in Table 5.2, which were selecting through the standard method of parameter tuning.

An estimate the state of the robot during training through using an SE. Hence, in terms of applying state feedback noise for domain randomisation to improved sim-to-real transfer, this only needs to be implemented on the input sensor data vector of the SE, $\boldsymbol{\sigma} = [\boldsymbol{\omega_B}, \dot{\boldsymbol{v}}_B, \boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{\tau}, \boldsymbol{f}_{\text{grf}}]$. This vector includes base angular velocity, $\boldsymbol{\omega_B} \in \mathbb{R}^3$, base linear acceleration, $\dot{\boldsymbol{v}}_B \in \mathbb{R}^3$, joint positions, $\boldsymbol{q}$, joint velocities, $\dot{\boldsymbol{q}}$, joint torques, $\boldsymbol{\tau}$, and foot ground reaction forces, $\boldsymbol{f}_{\text{grf}} \in \mathbb{R}^4$. As the initial state of the robot and its performance can never be guaranteed during real-world deployment, the initial configuration of the robot in each episode, $\boldsymbol{q}_{\text{init}}$, the mass of the robot's base, m_B, the fiction coefficient between the robot's foot and the ground, μ_{fric}, K_p and K_p are all randomised as well. Additionally, to ensure that a rich variation of $\boldsymbol{U}^{\text{cmd}}$ is experienced during training, randomly sampled gaits, velocity commands and velocity change durations (to achieve random acceleration), t_{acc}, are implemented during training. These are within defined maximum, $t_{\text{acc}}^{\text{max}} = 0.5$s, and minimum, $t_{\text{acc}}^{\text{min}} = 0$s, acceleration durations. All random sampeling was attained from either uniform or normal distributions, as detailed

Table 5.2: PPO Hyperparameters.

Parameter	Value
Number of Environments	240
Clip Range	0.2
Max Steps per Batch	400
GAE λ	0.95
Learning Epochs per Batch	4
Learning Rate	5e-4
Number of Mini-batches	4
Minimum Policy std	0.2
Reward Discount Factor	0.99
Optimizer	Adam

in Table 5.3.

The environment itself is constructed using RaiSim [157], as the vectorized environment setup allows for efficient training of policies. Additionally, the observation normalisation functionality offered by RaiSim is also used for improved training. During the training of π_L only flat terrain is present within the environment to isolate and highlight the effect of implementing $\boldsymbol{\beta}_L$. A core point of investigation of this chapter is that the implementation of $\boldsymbol{\beta}_L$ aims to impart gait procedural memory within π_L^{bio} hence if rough terrain was observed during training it will become ambiguous if the improved performance is a direct result of implementing $\boldsymbol{\beta}_L$. However, for training π_G, flat to very rough terrain is implemented using fractal noise, enabling the policy to learn to employ the use of each gait minimising biomechanics metrics on a variety of terrains. All variations of π_L and π_G were trained for 20k iterations, taking 6 and 9 hours respectively, on a standard desktop computer with one Nvidia RTX3090 GPU with a training frequency of 100Hz. It is also important to note that the training of all π_G policies only utilise the final proposed bio-inspired locomotion framework π_L^{bio}.

Table 5.3: Noise and Sampling Distributions.

Parameter	Distribution
$\boldsymbol{\omega}_B$, $\dot{\boldsymbol{v}}_B$	$0.015\mathcal{N}(0,1)$
$\boldsymbol{q}$	$0.005\mathcal{N}(0,1)$
$\dot{\boldsymbol{q}}$	$0.15\mathcal{N}(0,1)$
$\boldsymbol{\tau}$, $\boldsymbol{f}_{\text{frc}}$	$\mathcal{N}(0,1)$
μ_{fric}	$\max(0.4, \min(0.6 + 0.5\mathcal{N}(0,1), 1))$
m_B	$\max(-1, \min(\mathcal{N}(0,1), 3))$
K_p	$K_p \max(0.9, \min(1 + 0.05\mathcal{N}(0,1), 1.1))$
K_d	$K_d \max(0.9, \min(1 + 0.05\mathcal{N}(0,1), 1.1))$
v_x^{cmd}	$\mathcal{U}(0, 1.5)$
ω_z^{cmd}	$\mathcal{U}(-1, 1)$
Γ	$\mathcal{U}(0, 6)$
t_{acc}	$\mathcal{U}(t_{\text{acc}}^{\min}, t_{\text{acc}}^{\max})$

5.2.4 Locomotion Policy

The goal of the locomotion policy π_L is to realise the input $\boldsymbol{U}^{\text{cmd}}$ while exhibiting stable and versatile behaviour. As such, π_L is trained to generate the action, $\boldsymbol{q}^*$, from an input observation, $\boldsymbol{o}_L = [\boldsymbol{\beta}_L, \boldsymbol{s}, \boldsymbol{v}_B^{\text{cmd}}] \in \mathbb{R}^{69}$, where $\boldsymbol{v}_B^{\text{cmd}} = [v_x^{\text{cmd}}, v_y^{\text{cmd}}, \omega_z^{\text{cmd}}] \in \mathbb{R}^3$ is the high-level velocity command of the robot's base within $\boldsymbol{U}^{\text{cmd}}$, as outlined in Fig. 5.1. $\boldsymbol{s}$ is generated from the output of the SE and is defined as $\boldsymbol{s} = [\boldsymbol{\alpha}\boldsymbol{R}_B^T, \boldsymbol{q}, \boldsymbol{\omega}_B, \dot{\boldsymbol{q}}, \boldsymbol{v}_B, z_B, \boldsymbol{\tau}, \boldsymbol{c}] \in \mathbb{R}^{50}$, where $\boldsymbol{\alpha} = [0, 0, 1]^T$ is used to select the vertical z-axis, $\boldsymbol{\omega}_B \in \mathbb{R}^3$ is the base angular velocity, $\boldsymbol{v}_B \in \mathbb{R}^3$ is the base linear velocity, z_B is the current base height, and $\boldsymbol{c} \in \mathbb{B}^4$ is the contact state of the feet. The locomotion reward function, r_L, is formulated so that the the output of the policy can realise the reference gait patterns and velocity commands stably, smoothly and accurately,

$$r_L = \mathrm{w}_\eta r_\eta + \mathrm{w}_{v^{\text{cmd}}} r_{v^{\text{cmd}}} + \mathrm{w}_f r_f + \mathrm{w}_{\text{stab}} r_{\text{stab}}, \tag{5.5}$$

where r_η, $r_{v^{\text{cmd}}}$, r_f and r_{stab} are the grouped reward terms focusing on efficiency, velocity command tracking, gait reference tracking and stability respectively. w_η, $\mathrm{w}_{v^{\text{cmd}}}$, w_f and w_{stab} are the weights of each reward and are valued at -1.5, 15, -10, and -5

respectively. r_η aims to minimise joint jerk, $\dddot{\boldsymbol{q}}$, joint torque, and the difference between $\boldsymbol{q}^*$ and the previous action, $\boldsymbol{q}^*_{t-1}$,

$$r_\eta = \| \dddot{\boldsymbol{q}} \|^2 + \|\boldsymbol{\tau}\|^2 + \|\boldsymbol{q}^* - \boldsymbol{q}^*_{t-1}\| \tag{5.6}$$

$r_{v^{\text{cmd}}}$ minimises the difference between the commanded and current base velocity,

$$r_{v^{\text{cmd}}} = \psi \left(\|\boldsymbol{v}_B - \boldsymbol{v}_B^{\text{cmd}}\|^2 \right), \tag{5.7}$$

in which the function $\psi : x \to 1 - \tanh\left(x^2\right)$ is used to normalise the reward term so that their maximum value is 1 to prevent bias towards individual rewards, $\boldsymbol{v}_B = [v_x, v_y, \omega_z] \in \mathbb{R}^3$ is the current base x, y and yaw velocities. r_f ensures the robot realises the commanded gait references within $\boldsymbol{\beta}_L$,

$$r_f = |\boldsymbol{c}^{\text{err}}| + \sum_{i=1}^{4} \|\boldsymbol{p}_i - \boldsymbol{p}_i^{\text{ref}}\|^2, \tag{5.8}$$

where $\boldsymbol{c}^{\text{err}} \in \mathbb{B}^4$ defines the feet that violate the desired contact state, with $\boldsymbol{p}_i \in \mathbb{R}^3$ and $\boldsymbol{p}_i^{\text{ref}} \in \mathbb{R}^3$ being the current and reference Cartesian positions of the i-th foot. r_{stab} aims to prevent contact foot slip, large hip joint motions and undesirable base orientations,

$$\begin{aligned} r_{\text{stab}} = \sum_{i=1}^{F} \|\dot{\boldsymbol{p}}_i\|^2 + \|\boldsymbol{\omega}_B\|^2 + \psi \left(\|\boldsymbol{\alpha}\boldsymbol{R}_B - \boldsymbol{\alpha}\boldsymbol{R}_B^{\text{des}}\|^2 \right) \\ -\psi \left((z_B - z_B^{\text{nom}})^2 \right) + \|\boldsymbol{q}_{\text{hip}}\|^2, \end{aligned} \tag{5.9}$$

where $\boldsymbol{p}_i \in \mathbb{R}^3$ is the velocity of the i-th foot scheduled to be in stance, F is the number of stance feet, $\boldsymbol{\omega}_B = [\omega_x, \omega_y] \in \mathbb{R}^2$ where ω_x and ω_y are roll and pitch base velocities respectively, $\boldsymbol{R}_B^{\text{des}} \in SO(3)$ is the desired base orientation, z_B^{nom} is the nominal base heights respectively, and $\boldsymbol{q}_{\text{hip}} \in \mathbb{R}^4$ is the hip angular joint positions.

5.2.5 Biomechanics Gait Transition Metrics

Although the set of biomechanics metrics applied in this work were designed to accommodate different animals of the same morphology, even when animal body size and weight vary considerably, the fact still remains that they were designed for the analysis of animal locomotion. Hence, several adjustments to how they are calculated needs to be implemented; for example, energy consumption in animals is often measured through the rate of consumption of O_2, hence unsuitable for the application of robotics. Additionally, as robots provide a wide array of feedback data some of the metrics

have also been augmented to better reflect the characteristic that these biomechanics metrics are attempting to characterise.

Energy Efficiency

The calculation of CoT takes the general form of

$$\text{CoT} = \frac{P}{mgv}, \tag{5.10}$$

where P is power consumed and m is the system's mass. When studying animal locomotion, P is found through measuring how much CO_2 is generated and O_2 is consumed and v is assumed to be the speed of the treadmill the animal is running on [116, 117, 109]. For the case of the robot, P is calculated from $\boldsymbol{\tau}$ and $\dot{\boldsymbol{q}}$ with an adjustment term, adopted from [65], and v is assumed to be the magnitude of the robot base velocity command to take a similar approach to animal studies and for consistent metric use between simulation and real-world deployment; completely accurate measurement of the robot's linear base velocity is impossible during real-world deployment due to the accumulation of error within the SE. As such, calculation of the robot's CoT is formulated as

$$\text{CoT} = \sum_{i=1}^{n} \frac{\max(\tau_i \dot{q}_i + 0.3\tau_i^2, 0)}{mg|\boldsymbol{v}_B|}, \tag{5.11}$$

where m is the robot's mass and g is gravity. It should be noted that CoT is only calculated and applicable when $|\boldsymbol{v}_B^{\text{cmd}}| > 0$.

Actuator-structural Forces

As gaining an exact understanding of the actuator-structural forces within animals is infeasible, researchers have opted instead to measure the peak ground reaction forces of the animal's stance feet during locomotion using force plates [118]. Other methods include adding strain gauges to the bones of the animals [119]. However, in the case of robots we have the privilege of having access to joint state feedback while also knowing the exact limitations of the hardware. Therefore, when considering the biomechanics hypothesis that animals aim to minimise actuator-structural forces to prevent injury and that torque is proportional to strain and force, in the case of the robot actuator-structural forces are characterised through joint torque saturation, $\tau_{\%}$,

which is calculated by

$$\tau_{\%} = \left| \frac{\boldsymbol{\tau}}{\boldsymbol{\tau}_{\text{lim}}} \right| \frac{1}{n}, \tag{5.12}$$

where $\boldsymbol{\tau}_{\text{lim}} \in \mathbb{R}^n$ is the joint torque limits (assumed based on manufacturers specification), which proves particularly usefully when considering that the hip joints of most quadruped robots, including the A1, are often more sensitive to forces at the foot due to their distance from the point of ground impact and the only motor of the leg set in this plane; this would not be considered if just ground reaction force was used to characterise actuator-structural forces.

Mechanical Work Efficiency

During animal locomotion, if they are to have perfect mechanical work efficiency there would be a net zero change in external work over the duration of a gait cycle as there would be perfect exchange between kinetic and potential energy [122]. As expected, perfect mechanical work is never seen in nature, hence mechanical work efficiency in animals is characterised by the sum of the change in kinetic and potential energy [122] or the sum of the external work of the animal [123] over the duration of a gait cycle. As this is typically calculated through measuring the O_2 uptake, for the case of robots the calculation of the external work, W_{ext}, is formulated through

$$W_{\text{ext}} = \sum_{i=0}^{t_{\text{gait}}} (\Delta E_{\text{k},i} - \Delta E_{\text{p},i}) \tag{5.13}$$

where t_{gait} is the duration of the current gait cycle, and $\Delta E_{\text{k},i}$ and $\Delta E_{\text{p},i}$ are the changes in kinetic and potential energy over a control time step respectively. The primary difference between the metrics seen in biomechancis and this formulation of W_{ext} is that $\Delta E_{\text{k},i}$ accounts for not only forward linear velocity but also lateral and angular velocity whereas originally only forward linear velocity is considered.

Stability

The best indication of stability in animals is their stride duration coefficient of variation (CV). This metric characterises periodicity, which is a primary indication of stable locomotion [121]. However, to accurately calculate this, the mean and standard deviation of the stride duration needs to be taken over an extended period of time for appropriate

data generation. This is sufficient for undertaking analysis similar to that presented in [121], but this presents an issue when it comes to analysing the performance of the proposed control framework; it is common for multiple speed commands being used within the duration of one stride. To overcome this limitation $c^{\text{err}}_{\text{avg}} = |\boldsymbol{c}^{\text{err}}|/4$ is formulated, which can be measured every time step rather than just at each foot touchdown event; the gait references generated by the BGS have a constant and periodic stride duration, therefore an accurate tracking of this reference would in turn indicate high periodicity.

5.2.6 Gait Selection Policy

To achieve optimal gait selection for a given state, the biomechanics metrics are leveraged within the reward function of π_G, r_G. For the different variations of π_G trained within this chapter, each policy's reward function only features the metric that its focusing on within r_G but π^{uni}_G unifies all metrics hence uses the full form of r_G with all metrics. In addition, as the biomechanics metrics all describe characteristics that animals try to minimise through changing gaits, they can be directly applied within r_G with some normalisation where appropriate. The full form of r_G is

$$r_G = \mathrm{w_u} r_{\mathrm{u}} + \psi(\mathrm{CoT}) + \psi(\tau_{\%}) + \psi(c^{\text{err}}_{\text{avg}}) + \psi(W_{\text{ext}}), \tag{5.14}$$

where r_{u} is the utility reward term which all π_G use and $\mathrm{w_u}$ is its weight with a value of 0.4. r_{u} ensures the smoothness of the output Γ^*, the standing gait is only used when appropriate, and any select gait is able to follow $\boldsymbol{v}^{\text{cmd}}_B$. To achieve this, r_{u} has the form

$$r_{\mathrm{u}} = r_{v^{\text{cmd}}} + r_{\text{stand}} + r_{\text{smooth}}, \tag{5.15}$$

in which $r_{v^{\text{cmd}}}$ is taken from (5.7), and r_{stand} is set to -10 if a stand gait is used when $|\boldsymbol{v}^{\text{cmd}}_B| > 0$ or not used when $|\boldsymbol{v}^{\text{cmd}}_B| = 0$. For r_{smooth}, the reward aims to penalise unnecessary changes in Γ^* to remove rapid gait changes when two gaits could achieve similar metric minimisation for a given task and state. As such, if there is a gait change between time steps it is calculated as $r_{\text{smooth}} = -\psi(\mathrm{CoT} + \tau_{\%} + c^{\text{err}}_{\text{avg}} + W_{\text{ext}})$ otherwise it is set to 0. To generate Γ^*, π_G takes in input observation vector $\boldsymbol{o}_G = [\boldsymbol{s}, \boldsymbol{\beta}_G, \boldsymbol{v}^{\text{cmd}}_B, \dot{\boldsymbol{v}}^{\text{cmd}}_B, \Gamma^*_{t-1}] \in \mathbb{R}^{66}$ in which Γ^*_{t-1} is the previous output action to aid in action smoothing. Appropriate selection of the data provided to π_G is critical in order to achieve targeted minimisation of the biomechanics metrics. As such, the inclusion of $\boldsymbol{s}$ coupled with $\boldsymbol{c}^{\text{ref}}$, $\boldsymbol{p}^{\text{ref}}_z$, $\boldsymbol{v}^{\text{cmd}}_B$, $\dot{\boldsymbol{v}}^{\text{cmd}}_B$ and Ω_{stab} informs the policy of its current and

demanded stability, while the terms $\boldsymbol{\tau}$ and $\dot{\boldsymbol{q}}$ within $\boldsymbol{s}$ capture the power consumption of the robot and the forces to which it is subjected.

5.3 Results

To evaluate the developed framework a set of studies are completed, which in turn demonstrate how it outperforms others by exhibiting quadruped animal locomotion strategies, and validating this gained proficiency on real-world terrain, as presented in Fig. 5.2 and the video found at [158][1].

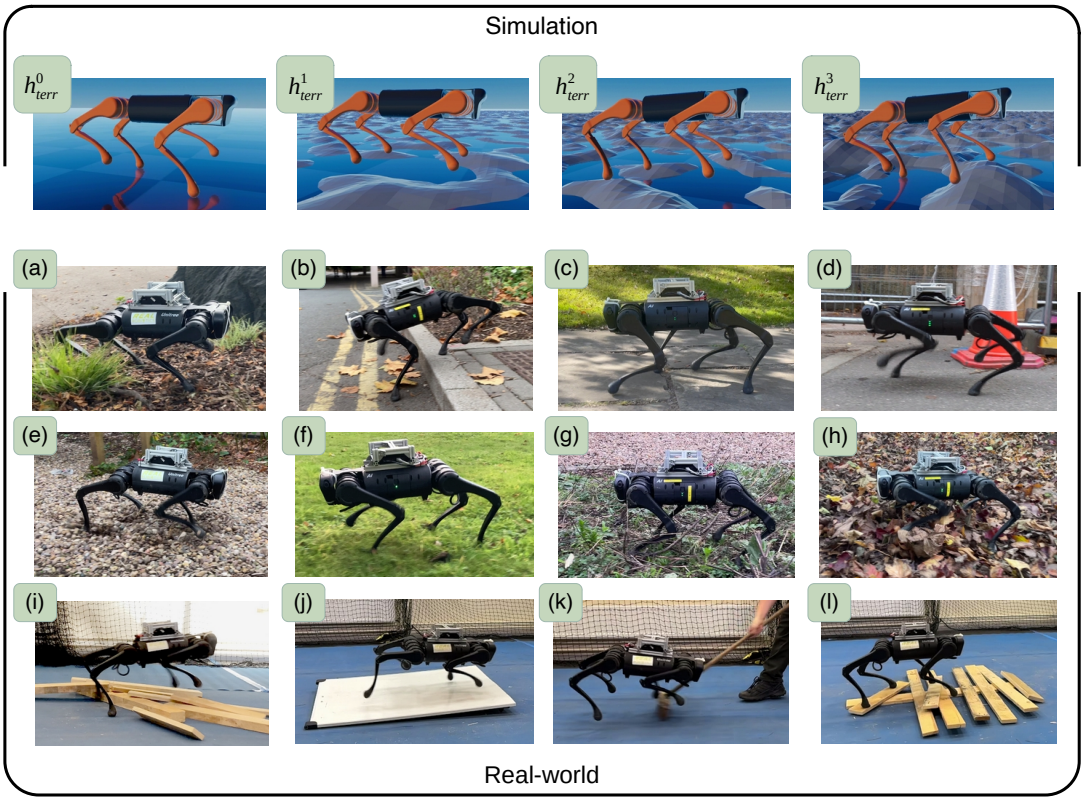


Figure 5.2: Simulated terrains h^0_{terr} through h^3_{terr} are generated with fractal noise with maximum heights of 0 m, 0.06 m, 0.13 m and 0.2 m. The LTA framework has been deployed (a) wood-chip, (b) a large step, (c) concrete slabs with large cracks, (d) tarmac, (e) deep rocks, (f) grassy terrain, (g) overgrown roots, (h) fallen leaves, (i) loose timber, (j) low-friction ramp, (k) flat terrain with perturbations, and (i) balanced timber.

[1] https://youtu.be/NwHoB7pErYQ

5.3.1 Achieving Adaptive Motion Adjustment with a Diverse Set of Gaits

To evaluate the method of instilling adaptive motion adjustment and procedural memory for diverse gait deployment, a comparison study is completed between the bio-inspired locomotion policy π_L^{bio}, a standard multi-gait locomotion policy with no pseudo procedural memory $\pi_L^{\text{no}\boldsymbol{\beta}_L}$, and a policy trained which also uses $\boldsymbol{\beta}_L$ within $\boldsymbol{o}_L$ but implements the standard approach of extracting the observations from the simulator, π_L^{noSE}, as shown in the video found at [159][1]. From the results of this study, shown in Fig. 5.3, Fig. 5.4, and Fig. 5.5, the proficiency of π_L^{bio} over $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE} in terms of velocity tracking error v_B^{err}, contact schedule tracking error $c_{\text{avg}}^{\text{err}}$, and base stability (magnitude of undesirable base angular velocities) ω_B^{err}, is stark.

As shown in Fig. 5.3, on flat terrain, π_L^{bio} achieves lower v_B^{err}, $c_{\text{avg}}^{\text{err}}$, and ω_B^{err} than $\pi_L^{\text{no}\boldsymbol{\beta}L}$ and π_L^{noSE}, averaging 15%, 21%, and 10% lower respectively (excluding failures), with errors increasing at higher velocity commands. This performance gap widens on rough terrain, where $\pi_L^{\text{no}\boldsymbol{\beta}L}$ and π_L^{noSE} fail frequently, especially at higher speeds and rougher surfaces, while π_L^{bio} succeeds in all trials despite being trained only on flat terrain. This highlights its adaptability to unseen environments (see Section 5.2.3 for rationale behind omitting rough terrain in π_L training).

This incompetence of $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE} is caused by a lack of adaptive swing foot motion adjustments captured within $\boldsymbol{\beta}_L$ based on $\boldsymbol{s}$ and the accumulation of error within the SE respectively. Both of these factors are substantially affected by the instabilities rough terrain inflicts upon the robot. Considering that the nominal swing foot peak height is defined as 25% of the nominal base height and for h_{terr}^2 and h_{terr}^3 (as defined in Fig. 5.2) the peak terrain height is 44% and 67% of the base height, having no data or strategy to account for this harsh terrain results in the rapid deterioration of proficiency.

For $\pi_L^{\text{no}\boldsymbol{\beta}L}$, Fig. 5.4 shows large spikes in $c_{\text{avg}}^{\text{err}}$ and ω_B^{err} when encountering h_{terr}^3, highlighting its inability to adapt swing foot trajectories and overcome an unrefined solution space. Fig. 5.5 further shows sustained instabilities in base height and orientation after contact with steps at 17% of the nominal base height. For π_L^{noSE}, Figs. 5.4 and 5.5 show that poor reference tracking and stability worsen with velocity command magnitude and time, as it lacks strategies to mitigate error build-up in the SE.

[1] https://youtu.be/-DfkDFA3KkI

Figure 5.3: Each policy follows a set of command velocities in x and yaw between 0 to 1 m/s and -1 to 1 rad/s respectively. During each velocity pair, the commanded gait is cycled through all gaits, switching every 1 s. This repeated 5 times over flat, h^0_{terr}, to very rough terrain, h^3_{terr} (shown in Fig. 5.2) with the average performance being plotted. A number rather than a magnitude indicates the count of experiments that the policy failed.

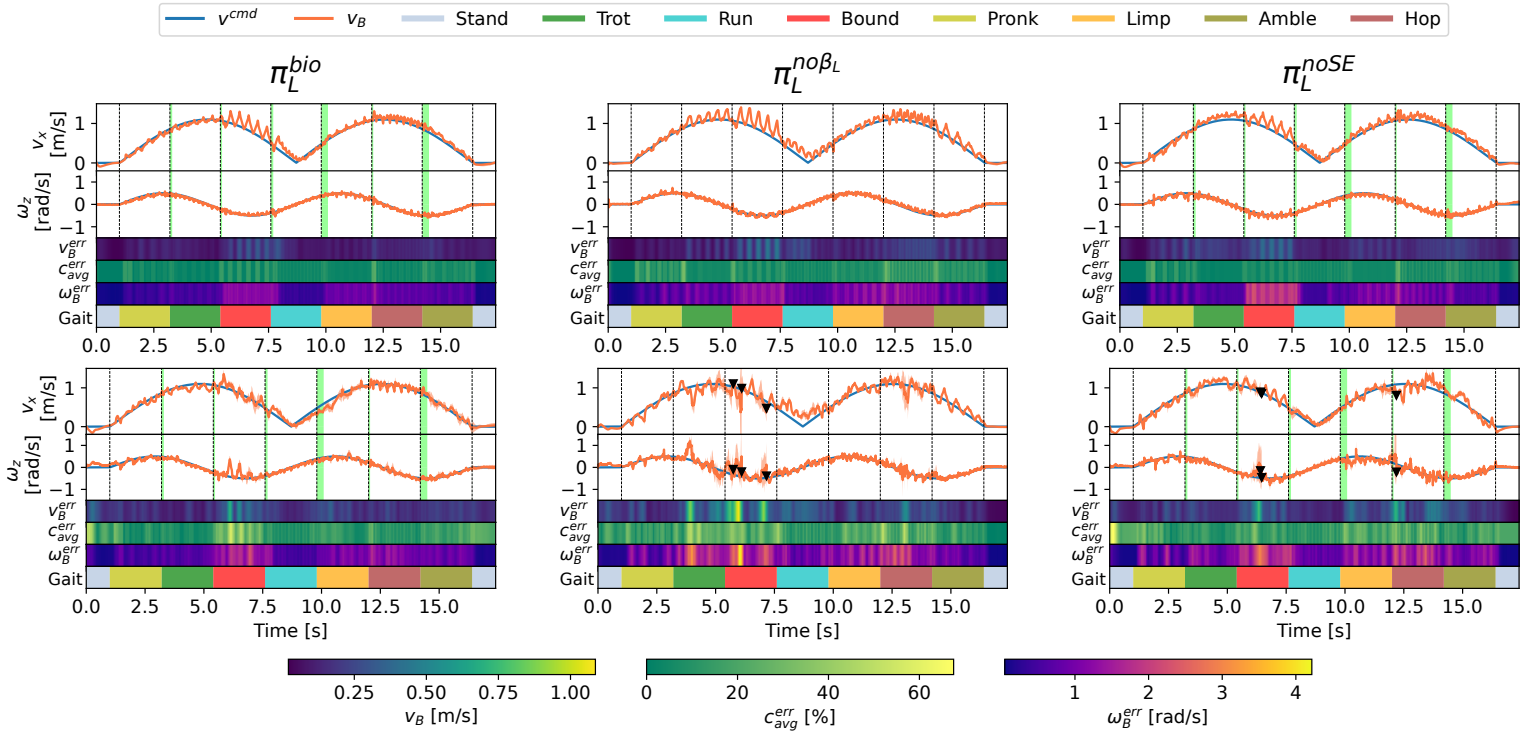


Figure 5.4: Each policy follows a sinusoidal trajectory in x, v_x, and in yaw, ω_z, while switching gaits every 2 s, which is repeated 5 times over h^0_{terr} and h^3_{terr} terrain. Green highlighted areas indicating transition phases, for which the function and formulation are detailed in Section 5.2.2, and black triangles representing points of failure.

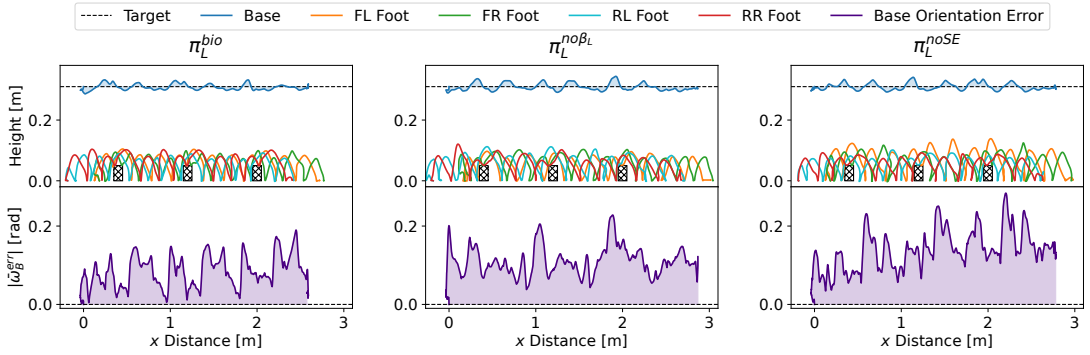


Figure 5.5: Policies follow a command of just $v_x = 0.5$ m/s while using a constant trot gait while encountering rectangular steps with a height of 0.05 m, where $|\bar{\omega}_B^{\text{err}}|$ is the magnitude of the desired orientation error. FL is front left, FR is front right, RL is rear left, and RR is rear right in reference to the robot's feet.

Since π_L^{bio} does not experience any of these limitations, this explicitly demonstrates the effectiveness of implementing $\boldsymbol{\beta}_L$ and $\boldsymbol{s}$ within $\boldsymbol{o}_L$; π_L^{bio} successfully generalises across gaits and terrain, demonstrating successful instillation of adaptive motion adjustment and gait procedural memory.

5.3.2 Applying Biomechanics Metrics For Optimal Gait Selection

Directly applying biomechanics metrics to instil animal gait transition strategies is unsuitable due to differences between animals and robots, as well as π_G training requirements. Instead, CoT, torque saturation ($\tau_{\%}$), external work (W_{ext}), and foot contact tracking error ($c_{\text{avg}}^{\text{err}}$) are used to minimise energy use, actuator-structural forces, mechanical work, and instability, respectively. Details and justification of these metrics are in Section 5.2.5. In accordance with Section 5.2.6, these metrics are unified within the training of the gait selection policy π_G^{uni} to instil the strategies animals use for optimal gait selection to achieve exemplary adaptability. To investigate whether π_G^{uni} effectively minimises these metrics through gait selection, the results of competing the highly demanding velocity command trajectory presented in Fig. 5.6 and within a video found at [160][1] for π_G^{uni} paired with π_L^{bio} are collected, along with that for all individual gaits deployable by π_L^{bio}, for flat terrain and terrain h_{terr}^2. Within Fig. 5.6, the contact state of the feet and the gait that π_G^{uni} realises is also shown beneath the time series plots

[1] https://youtu.be/y4KnzMEdf78

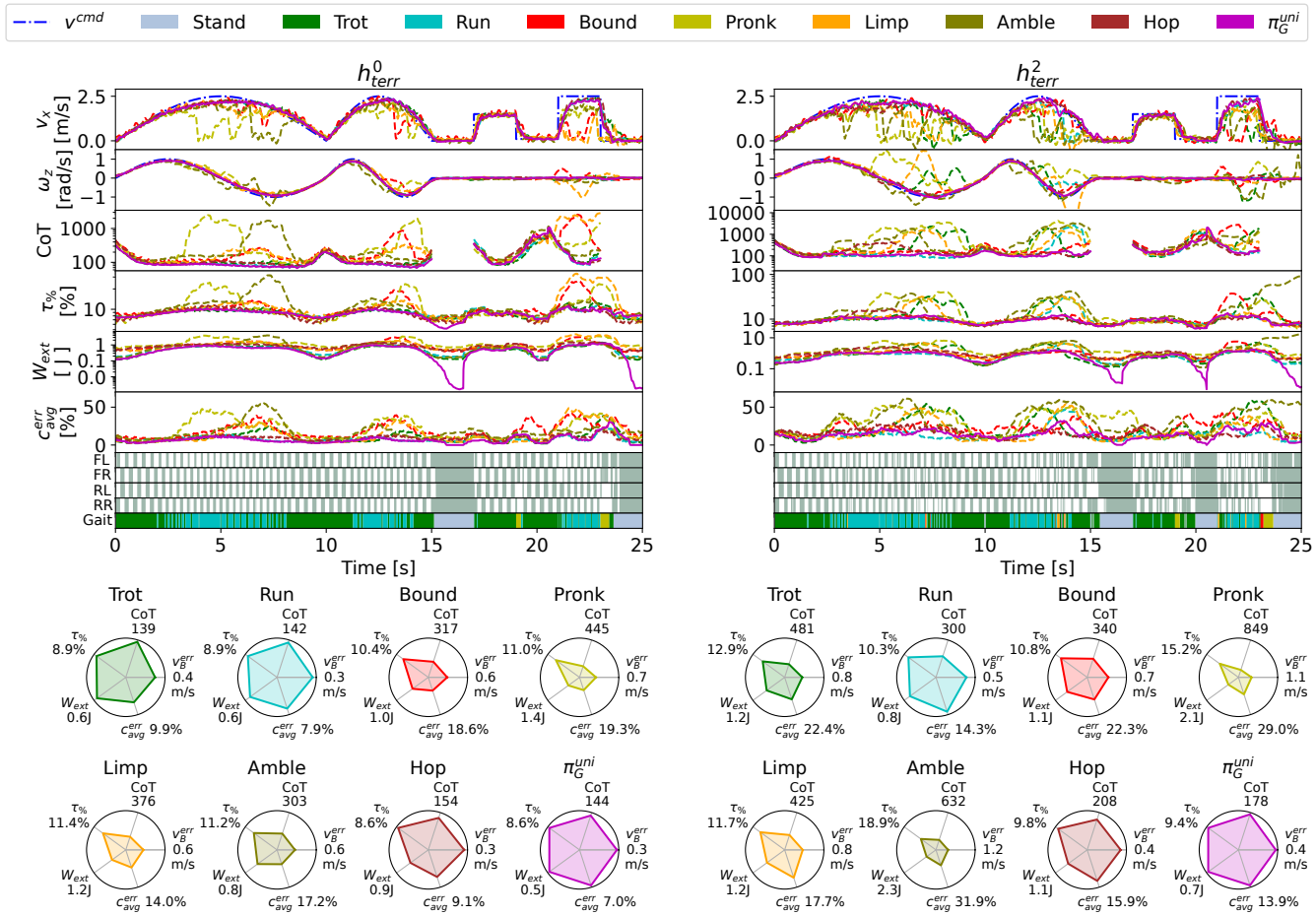


Figure 5.6: For terrains h^0_{terr} and h^2_{terr}, each isolated gait and π_G^{uni} are given a velocity command to follow to assess their performance in terms of CoT, $\tau_{\%}$, W_{ext}, and $c^{\text{err}}_{\text{avg}}$; for each individual gait π_L^{bio} just run with the gait statically selected and for π_G^{uni} it is coupled with π_L^{bio} for autonomous optimal gait selection.

Fig. 5.6 shows that π_G^{uni} exclusively uses trotting at low speeds and running at high speeds. When accelerating, π_G^{uni} oscillates between trotting and running, with an increasing bias towards running, to increase the stride frequency as visualised by the foot contact data in Fig. 5.6. Consequently, π_G^{uni} not only reliably tracks the optimal gait but outperforms individual gaits, which is only aided by its transition to standing during $v_x^{\text{cmd}} = 0$ $\omega_z^{\text{cmd}} = 0$ events for minimal $\tau_{\%}$, W_{ext} and $c_{\text{avg}}^{\text{err}}$. This behaviour, although never targeted, is reflected in animal locomotion strategies, where gait stride frequency increases with speed, and transitional phases blend gaits to minimise energy use [123]. Under rough terrain and high acceleration, π_G^{uni} employs additional gaits to manage instability. This leads to a gait classification: trot and run serve as nominal gaits for low and high speeds, while bound, pronk, limp, amble, and hop act as auxiliary gaits for off-nominal conditions like stability recovery.

When inspecting the radar charts in Fig. 5.6, which depict relative performance in terms of the metrics[1], the origin of this emerged gait selection strategy becomes clear. Across the gaits, on flat terrain trot and run gaits achieve the best relative performance. However, when it comes to overcoming rough terrain bound, hop and limp gaits all gain relative performance in $\tau_{\%}$, W_{ext} and $c_{\text{avg}}^{\text{err}}$, while trot and run gaits exhibit a reduced dominance in relative proficiency. In addition to this observation providing an insight as to why π_G^{uni} chooses to utilise these auxiliary gaits it also provides evidence to suggest that $\tau_{\%}$, W_{ext} and $c_{\text{avg}}^{\text{err}}$ can effectively characterise stability. This observation is further investigated and discussed in the following sections. Overall, π_G^{uni} outperforms all individual gaits across all metrics, with the exception of CoT for trot and run gaits where the difference is negligible, demonstrating the successful minimisation of the metrics and successful instillation of gait procedural memory of how to utilise each gait given the robot's state and task.

5.3.3 Comparison Between Robot and Animal Gait Selection

When developing metrics to characterise gait transitions in animals, data is collected over intervals of increasing forward velocity on flat terrain [113, 118, 123, 121]. Hence, within Fig. 5.7 the same approach is taken. This experiment is also repeated with h_{terr}^3

[1]The average relative performance in terms of these metrics is displayed in the radar plots, where each gait's performance is normalised to that of the best performer for each metric; the higher the value within the radar plot, the more effectively that metric has been minimised.

to investigate correlations between the metrics and the effects of introducing rough terrain, as presented in Fig. 5.8. Additionally, four further π_G policies are trained that individually minimise energy consumption π_G^{CoT}, actuator-structural forces $\pi_G^{\tau_\%}$, mechanical work $\pi_G^{W_{\text{ext}}}$ and stability $\pi_G^{c^{\text{err}}}$, in accordance with Section 5.2.5, to compare their performance with π_G^{uni}. One unanimous observation across Fig. 5.7 is that animals experience a gait transition phase over a range of velocities [121, 113, 161]. This behaviour is reflected in π_G^{uni}, where a transition phase is classed as where no individual gait occupies more than 75% of the gaits used at a specific speed.

Energy Expenditure – Cost of Transport

An animal's gait transition aligns with the CoT optimal transition point (OTP), λ^{CoT}, to minimise energy cost [109, 123], as shown in Fig. 5.7. π_G^{uni} mirrors this, tracking the lowest CoT gait with transitions centred on λ^{CoT}. In contrast, π_G^{CoT} lacks a defined transition phase and switches earlier, due to training on rough terrain where hopping improves CoT (Fig. 5.6). Fig. 5.8 further shows h^3_{terr} induces similar CoT distributions but with greater gait variance, as auxiliary gaits become more effective. This highlights that CoT-only training reduces generality and diverges from natural gait selection.

Actuator-structural Forces – Foot Contact Forces

Animals are observed to change gait to minimise actuator-structural forces (i.e. musculoskeletal forces) [118], which in biomechanics is measured through foot ground reaction force, f_{grf}. Similarly, π_G^{uni} and $\pi_G^{\tau_\%}$ reduce f_{grf} through selecting the optimal gait for minimising $\tau_\%$. This supports that $\tau_\%$ is a suitable alternative to f_{grf}, which is further validated by a strong correlation between them within Fig. 5.8. However, not only does $\pi_G^{\tau_\%}$ maintain a trotting gait past optimal f_{grf}, but also the transition itself is instantaneous. In turn, $\pi_G^{\tau_\%}$ better reflects the animal data from [118] than π_G^{uni}. This could be explained by the metric being misinterpreted in [118]; with a high correlation between $\tau_\%$ and f_{grf} with stability metric $c^{\text{err}}_{\text{avg}}$, it suggests instability causes transition, which is rare on flat terrain, as addressed in Section 5.4.

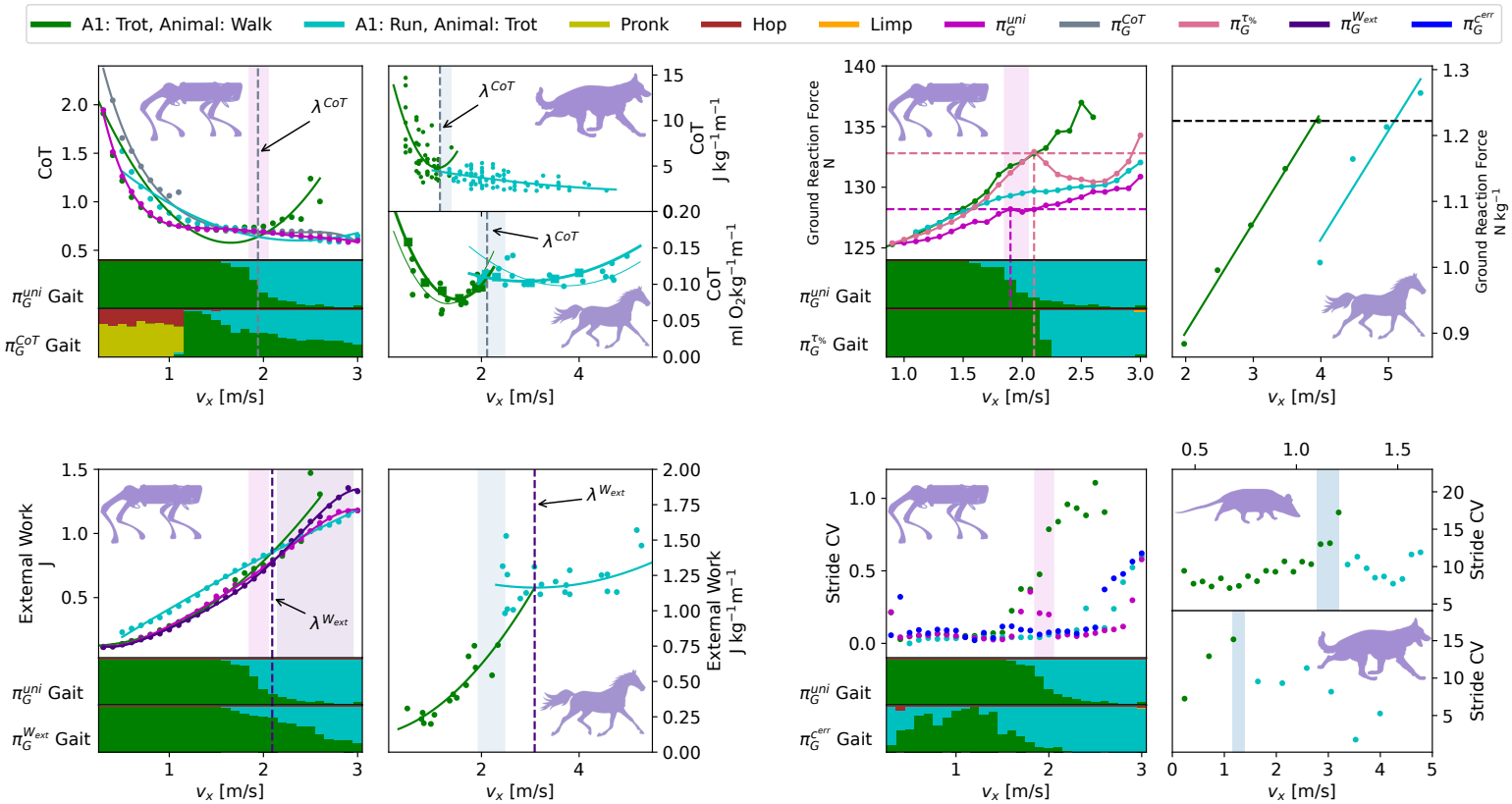


Figure 5.7: The bottom two plots of all robot data indicate the percentage of each gait utilised at that velocity. Magenta, purple and blue shaded regions indicate the transition phases of π_G^{uni}, $\pi_G^{W_{\text{ext}}}$ and animals respectively. This study compares transition strategies of (top left) π_G^{uni} and π_G^{CoT} to data collected from dogs [121] and horses [113, 123] in terms of CoT, (top right) π_G^{uni} and $\pi_G^{\tau_{\%}}$ to data collected from horses [118] in terms of foot ground reaction forces, (bottom left) π_G^{uni} and $\pi_G^{W_{\text{ext}}}$ to horses [123] in terms of external work, and (bottom right) π_G^{uni} and $\pi_G^{c^{\text{err}}}$ to opossums and dogs [121].

Figure 5.8: Mapping the correlation between metrics and the average gait ID selected across velocities for terrains h^0_{terr} and h^3_{terr}, with the dataset of each terrain consisting of 31 data points averaged over 1000 samples.

Mechanical Work – External Work

Animals transition to preserve external mechanical work, W_{ext}, but do so before $\lambda^{W_{\text{ext}}}$, indicating relaxed minimisation (Fig. 5.7). π_G^{uni} reflects this behaviour; transition occurs before the mechanical work OTP, $\lambda^{W_{\text{ext}}}$, yet minimal W_{ext} is preserved. While $\pi_G^{W_{\text{ext}}}$ can reduce W_{ext}, its transition phase is extended over a larger range of velocities compared to animals, occurring just after $\lambda^{W_{\text{ext}}}$. In turn, this suggests that switching gaits between trotting and running offers minimal reductions in W_{ext} resulting in a less definitive transition. However, W_{ext} also seems to capture stability due to its high correlation with the stability metrics on h^3_{terr} in Fig. 5.8, providing insight into its reduced role in Fig. 5.7 where only flat terrain is present.

Stability – Stride Duration Coefficient of Variation

Within [121] animals are shown to reduce their stride duration CV to preserve stability as high gait periodicity indicates stability. This behaviour is presented in Fig. 5.7, where animals are seen to initiate a transition phase when there is a considerable increase in stride CV, consequently leading to a decrease in CV and an increase in stability. π_G^{uni} inherits the same strategy as only when a spike in stride CV is experienced does a transition phase begin resulting in improved stability through lowering stride CV. However, this is not the case with $\pi_G^{c^{\text{err}}}$ as it acts to reduce stride CV much more aggressively by mixing both slow, fast and auxiliary gaits which results in no clear transition phase being produced.

Only π_G^{uni} consistently reflects all animal data sets and demonstrates the successful instillation of animal gait transition strategies. This also supports the notion that no singular metric can characterise animal gait selection and only through unification can similar behaviour in robots arise; the minimisation of the metrics is expected and is seen across all π_G policies, but the intricacies of animal gait transition strategies are only seen in π_G^{uni}. Additionally, this behaviour is verified to transfer to real-world deployment in the following section.

5.3.4 Preservation of Biomechanics Metrics on Grass

When deploying π_L^{bio} with π_G^{uni} on a smooth low-friction floor and uneven grassy terrain π_G^{uni} can utilise the same behaviour animals demonstrate in minimising CoT; comparing the performance of maintaining a static gait of both trot and run gaits within Fig. 5.9, π_G^{uni} can select the most energy efficient gait and even prevent failure as the trot gait cannot maintain stability for the entire experiment. Additionally, the distribution of gait usage between trot and run gaits varies between the two terrains. This is due to the uneven grassy terrain causing reduced stability of the robot when trotting, as indicated by $c_{\text{avg}}^{\text{err}}$ in Fig. 5.9, which in response π_G^{uni} exhibits increased modulation between trotting and running to preserve stability. This preservation of stability while minimising CoT is achieved through the resultant modulation of the stride frequency as gaits with higher stride frequency offer improved stability [121] and efficiency on rough terrain [123]. Overall, on smooth low-friction floor π_G^{uni} is able to achieve a 15% and 24% reduction in CoT compared to trotting and running gaits and a reduction of 18% and 30% on uneven grass.

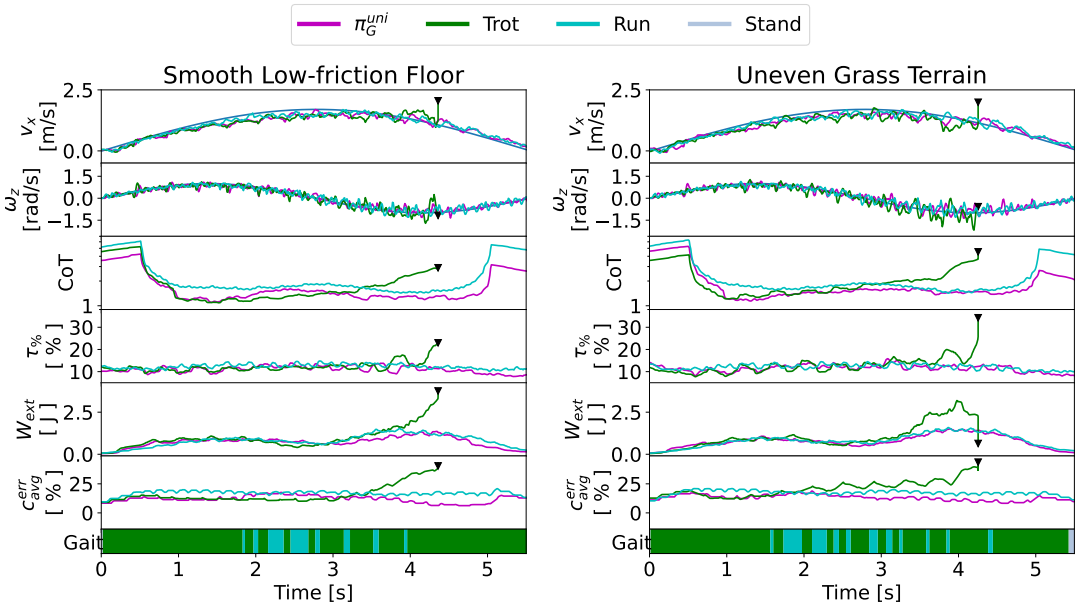


Figure 5.9: Deployment of π_L^{bio} and π_G^{uni} on smooth low-friction floor and uneven grassy terrain, with the black triangles indicating a point of failure.

Figure 5.10: The deployment of π_L^{bio} on uneven grassy terrain with manual gait selection to cycle through all gaits with a $v_{\text{x}}^{\text{cmd}}$ of 0.5 m/s where red shaded regions indicate a transition period.

5.3.5 Adaption to Real World Terrain

Although animal gait transition strategies, gait procedural memory, and adaptive motion have been instilled into the LTA framework, its real-world proficiency without hardware deployment is uncertain. Grassy terrain may trap swing feet, and the ground often features irregularities. However, despite π_L^{bio} only observing flat terrain during training, during hardware experiments it successfully realises all seven gaits on this terrain, as illustrated in Fig. 5.10, demonstrating that gait procedural memory and adaptive motion adjustment successfully transfer to real-world environments, providing a high level of adaptability, as shown in the video at [162][1].

Terrain that causes states of instability presents a substantial risk to the robot, hence the limits of the LTA framework were tested through deployment on loose timber, muddy grass, and a low-friction board, as presented in Fig. 5.11a, Fig. 5.11b and Fig. 5.11c respectively and consolidated in the video found at [163][2]. Each of the presented experiments showcases an off-nominal stability recovery event; however in the nominal scenario, the LTA framework can maintain stability without changing gaits. In the case of loose timber, critical instability is caused when one rear foot slips on a plank, causing it to collide with another. In response, π_G^{uni} utilises auxiliary gaits pronk and bound to recover, as depicted in Fig. 5.11a. This strategy of utilising the auxiliary gaits for stability recovery is seen across all experiments and reflected in animals as highlighted in Fig. 5.12 where a horse is observed to utilise bounding and limping gaits to traverse down complex rock formations.

In all experiments presented, a considerable increase in a combination of W_{ext}, $\tau_{\%}$ and $c_{\text{avg}}^{\text{err}}$ is experienced before a gait transition, while a weaker correlation is observed with CoT; spikes in W_{ext}, $\tau_{\%}$ and $c_{\text{avg}}^{\text{err}}$ directly coincide or even preempt gait changes while CoT peaks lag. This is expected with $c_{\text{avg}}^{\text{err}}$ and W_{ext} as they capture periodicity and base height respectively. However, this is less expected for $\tau_{\%}$ to correlate with stability; this was never a factor investigated within [118], however, the correlation observed in these experiments and in Fig. 5.11a provides strong evidence that this is the case.

[1] https://youtu.be/I02DQ1RGdyw
[2] https://youtu.be/f6CqJ7gb3ZM

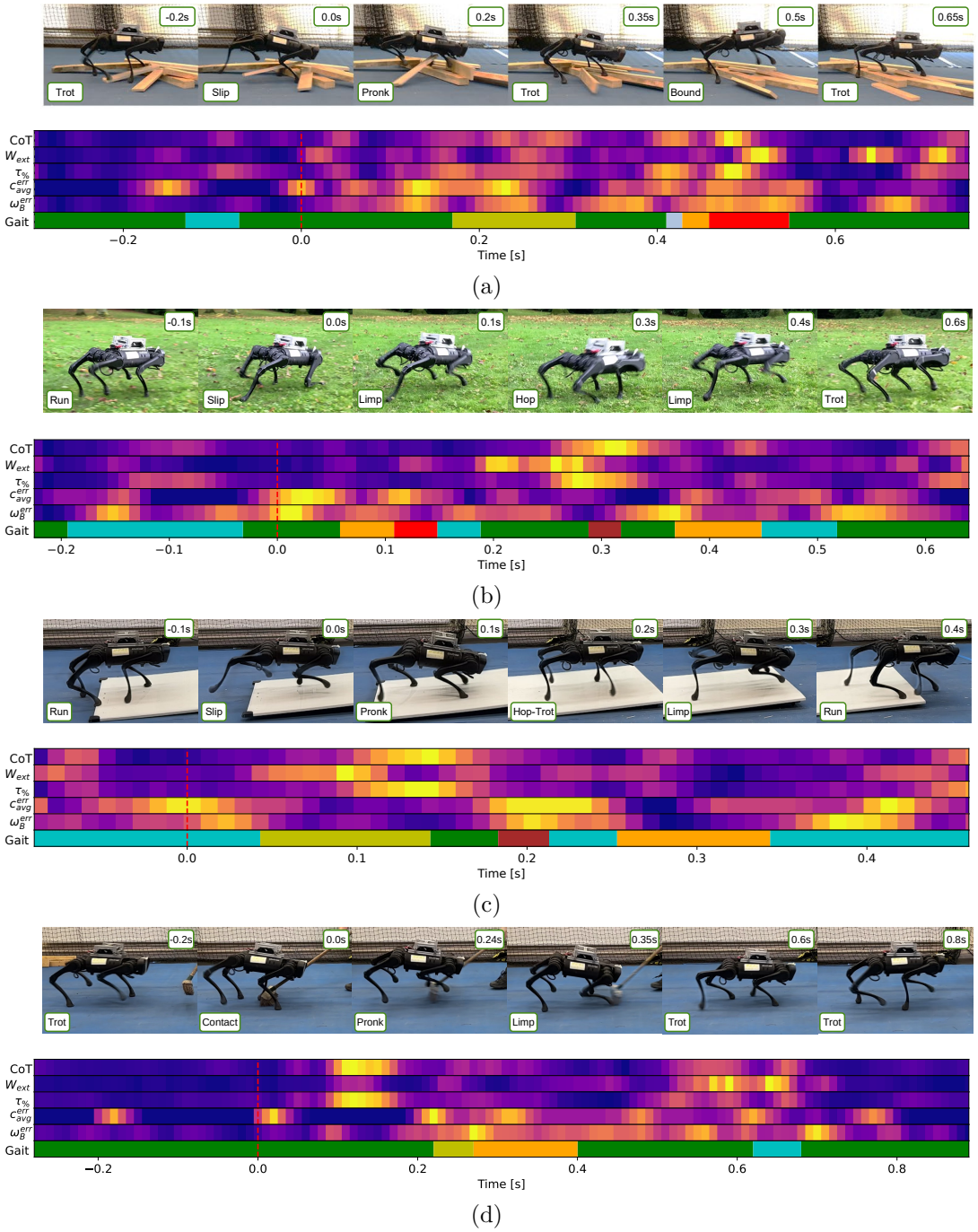


Figure 5.11: The LTA framework is deployed on loose timber, muddy grass, a low-friction board, and subject to pertubations, with maximum velocity magnitudes of 1 m/s, 2 m/s, 1.7 m/s, and 1 m/s respectively, where critical loss of stability is indicated by the red dashed line, the bottom subplot uses the same gait colour code as Fig. 5.6, and the heatmaps show the magnitudes of each metric increasing from purple to yellow.

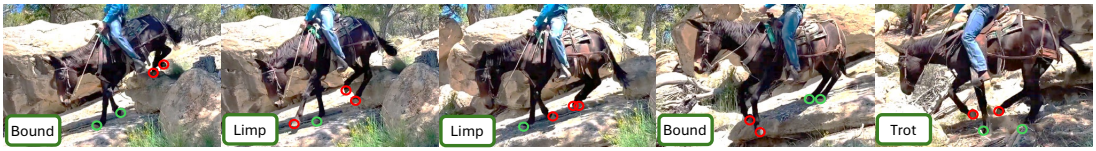


Figure 5.12: Snapshots showing how animals also use a mixture of auxiliary gaits to overcome challenging terrain [164], where red circles are swing feet and green circles are stance feet.

5.4 Discussion

5.4.1 Framework Performance

From taking inspiration from animal locomotion proficiency attributes, a locomotion framework has been developed, capable of traversing complex and high-risk terrain despite the robot not utilising exteroceptive sensors nor experiencing any rough terrain during the training of π_L^{bio}. For π_L^{bio}, this is achieved through including the BGS output $\boldsymbol{\beta}_L$ (which encodes state dependent pseudo gait procedural memory and adaptive motion adjustments) within the observation space $\boldsymbol{o}_L$. This proves to be effective at overcoming rough terrain within Fig. 5.3, Fig. 5.4, and Fig. 5.5 as without the presence of $\boldsymbol{\beta}_L$ within $\boldsymbol{o}_L$, increased failure and instability are observed. This is the equivalent of removing an animal's cerebellum functionality (resulting in reduced limb coordination and stability [154]) in turn supporting the claim that $\boldsymbol{\beta}_L$ effectively encodes pseudo gait procedural memory.

π_G^{uni} for optimal gait selection greatly expands adaptability through instilling gait selection strategies used by animals. As demonstrated in Fig. 5.11, π_G^{uni} can maintain stability in the event of the terrain undergoing radical structural or friction coefficient adjustments. These scenarios pose risks to robots with vision systems, as they typically cannot detect ground friction or terrain changes beyond their front legs. Through the use of π_G^{uni}, this limitation is mitigated without implementing resource-heavy exteroceptive sensors. Comparing Fig. 5.11 to Fig. 5.12 showcases that animals and π_G^{uni} utilise multiple auxiliary gaits to prevent failure. This behavior, untargeted during training, suggests that unifying these metrics encodes the intricacies of animal gait transitions.

One provoking observation is that actuator-structural forces appear to characterise instability. Considering that [118] validates by applying increased loads that could cause

instability, it suggests this metric was initially misunderstood. Additionally, despite the employed biomechanics metrics only being tested on animals completing a linear forward trajectory on flat terrain, π_G^{uni} upholds animal gait transition strategies across a wide range of terrains and base velocity commands. This supports that these metrics successfully characterise gait transitions and the notion that robots can indeed test biomechanics hypotheses, avoiding the resource, compatibility, and ethical challenges of animal testing.

5.4.2 Comparison to Other Froude-characterised Locomotion Frameworks

To benchmark the performance of the LTA framework, all available performance metrics and characteristics of other similar frameworks have been compiled in Table 5.4, with the most relevant ones being:

- DeepTransition [131]
- WalkTheseWays [104]
- RL+MPC [48]
- Phase-guided [103]
- EnergyLocomotion [66]

Comparing the LTA framework to the others that take a bio-inspired approach, DeepTransition and EnergyLocomotion, it has been deployed in a much wider variety of terrains. In respect of DeepTransition, this disparity could be due to their limited number of deployable gaits and their method of utilising a bio-inspired approach at the low-level, which as discussed in the main text, would limit its performance even though both this framework's methods and theirs take inspartation from stability and efficiency animal gait strategies. In the case of EnergyLocomotion, as only one energy efficiency animal gait strategy was implemented, only three gaits emerged which in turn results in limited adaptability in comparison to the LTA framework.

Additionally, both these frameworks cannot achieve arbitrary gait transitions as their gait selection strategies are embedded within their locomotion method, while the LTA framework is able to achieve this due to π_L^{bio} and π_G^{uni} being independent policies

Table 5.4: Comparison between Froude-characterised Locomotion Frameworks

	Phase-guided	WalkTheseWays	RL+MPC	DeepTransition	EnergyLocomotion	**LTA**
Framework Type	CPG-RL	RL	RL-MPC	CPG-RL-IK	RL	RL
Gaits	Amble, Pace, Trot, Bound	Pronk, Pace, Trot, Bound	Trot, Pace Pronk, Bound Gallop	Amble, Trot, Pronk Bound	Pronk Amble, Trot,	Amble, Run Pronk, Bound Limp, Amble, Hop
Level of Bio-inspired Implementation	✗	✗	✗	Low-Level Stability and Efficiency	High-level for Efficiency	High-level for Stability and Efficiency
Optimal Gait Selection	✗	✗	✗	✓	✓	✓
Arbitrary Gait Transitions	✓	✓	✓	✗	✗	✓
Hardware	Black Panther	Go1	Go1	A1	A1	A1
Terrains Traversed	Grass	Sand, Steps, Grass, Wood-chip	Grass,Tarmac, Wooden Boards	Flat Terrain with Large Gaps	Cardboard Sheets Grass, Rocks, Bush, Wooden Planks	Wood-chip, Deep Rocks, Large Step, Concrete Slabs, Tarmac, Loose Timber, Overgrown Roots, Leaves, Grass, Balanced Timber, Low-friction Ramp

which in turn promotes its suitability to be combined with Froude-free locomotion frameworks. In the comparison between the LTA framework and those who haven't implemented neither optimal gait selection nor implemented any animal gait strategies (WalkTheseWays, RL+MPC and Phase-guided), with the LTA framework being able to successfully traversing a much more diverse set of terrains and realise more types of gaits, this supports the notion that the method provides considerable improvements in adaptability and proficiency in Froude-characterised locomotion.

5.4.3 Comparison Between Froude-characterised to Froude-free Focused Frameworks

Comparison between Froude-characterised focused and Froude-free focused locomotion frameworks is somewhat impractical. Primarily, Froude-free locomotion frameworks measure success through failure rate over challenging terrain and stability, with no concern or analysis on efficiency. Additionally, these frameworks are only able to realise walking or running gaits, with other Froude-characterised gaits and optimal gait

selection being outside their operational scope [33, 100, 49]. On the other hand, the LTA framework, which focuses on Froude-characterised locomotion, determines its success based on command tracking error, rapid but stable gait transitions, efficiency and adaptability. Consequently, direct comparison between these two locomotion groups is not constructive as they are inherently different in terms of the challenges they are designed to overcome. However, an important and valuable direction for further research would be investigating how to effectively combine both framework types.

For example, although the Froude-free locomotion could potentially overcome the terrains featured in Fig. 5.11, they would do so without minimising energy consumption, with increased energy expenditure likely to be caused by excessive motion; as it has been proven that the terrains in Fig. 5.11 are traversable through just switching Froude-characterised gaits, use of Froude-free skills like climbing, jumping and surefootedness would incur unreasonable energy costs. However, as the LTA framework explicitly focuses on optimal Froude-characterised locomotion, through instilling animal gait strategies, it is not only able to traverse these terrains while preserving efficiency through optimal gait selection, but also recover from critically unstable states without requiring the use of inefficient Froude-free locomotion through rapid gait transitions to auxiliary gaits and then back to a nominal gait in a matter of milliseconds, as exhibited in Fig. 5.11. Therefore, to achieve both optimal Froude-characterised locomotion while also having proficient Froude-free locomotion capabilities, the LTA framework could be integrated with the works presented in [100, 49, 133]. Taking a specific example, the work in [100] uses a single walking gait policy for all Froude-characterised locomotion tasks, but if this policy was directly replaced with the LTA framework not only could optimal Froude-characterised locomotion be achieved but obstacles that require Froude-free locomotion are also traversable. However, such combination of these frameworks lie outside the scope of this project and hence are deferred to future work.

5.5 Chapter Conclusion

Overall, within this chapter a highly adaptable locomotion framework with optimal gait selection, realised through instilling animal locomotion gait strategies and proficiency attributes, has been developed and validated through its deployment in a diverse set of natural and challenging terrains, all of which the policies had never previously observed. Such highly adaptable and optimal Froude-characterised locomotion has never previ-

ously been achieved within the field of robotic quadruped legged locomotion, which not only demonstrate how powerful bio-inspired techniques can be but also lay a robust foundation for the locomotion aspect of the final targeted framework in Fig. 2.12. As such, through the amalgamation of the work within this Chapter and Chapter 4, it can confidently be stated that project objectives 3a through 3e have been completed and consequently the second project aim has been met.

However, with the environment developed to train the locomotion and gait selection policies, its requires significant augmentation for training loco-manipulation control policies and host legged manipulator agents. These augmentations required to develop and train the final control modules of Fig. 2.12, and consequently meeting the final project objectives, are completed in the next chapter.

Chapter 6

Unifying Bio-inspired Locomotion Control with Whole-body Manipulation Control to Achieve Dynamic Loco-manipulation Tasks

With all core locomotion and manipulation control modules and methods that construct the framework presented in Fig. 2.12 developed, only unification remains. However, before the final deployment of this framework can be realised, a range of augmentations to the inverse kinematics-based whole-body controller (IK-WBC) and training environments developed in Chapters 3 and 5, respectively, need to be completed to enable this unification. As such, this chapter will first detail these augmentations, outline the resultant training and deployment architectures, and analyse the performance of the framework in a set of loco-manipulation tasks in simulation. It should be noted that only simulation results are featured in this chapter due to the time constraints of the project.

6.1 Introduction

As uncovered in Chapter 2, a hierarchical framework provides the best architecture for generating adaptive, stable and efficient performance in legged loco-manipulation tasks in complex environments. Through splitting the core aspects of loco-manipulation between control modules, hence refining the functional scope of each module, this en-

ables the framework to:

- Ensure all module proficiencies are preserved simultaneously. This being the realisation of a diverse set of gaits while also completing challenging loco-manipulation tasks that require whole-body motions.
- Control modules are able to offset the limitations of others. In this framework, the IK-WBC constrains the deep reinforcement learning (DRL) policy output within the physical limits of the robot while completing whole-body motions during locomotion.
- Inject bio-inspired methods at select hierarchical levels to instil targeted animal locomotion proficiencies for stability, efficiency, and adaptability across different terrains.

However, so far this is only true in concept, with all control modules detailed in Fig. 2.12 being tested external to the other modules or only tested for a quadruped platform. As such, this chapter first focuses on unifying all control modules to formulate the final control framework, defining the architecture of how each control module interacts with each other. The IK-WBC is overhauled to realise loco-manipulation tasks from input joint and end-effector commands generated from the upstream DRL loco-manipulation controller. The DRL locomotion policy of Chapter 5 is augmented to not only generate adaptive locomotion for a quadrupedal legged manipulator (QLM), but also generate an adjustment term to aid in manipulation accuracy to offset the model inaccuracies of the IK-WBC. The DRL gait selection module is also retrained for deployment on a QLM. With all control modules augmented and unified to form the loco-manipulation framework, it is extensively tested on a variety of terrains. This is to investigate if the proficiencies of each respective control module have been preserved post-unification, and to evaluate its performance in terms of stability, efficiency, adaptability, and manipulation accuracy. If successful, this not only achieves project Objective 4 but also presents significant progress in working towards real-world deployment of QLM for loco-manipulation tasks. As the loco-manipulation framework is essentially an extension of the learning to adapt (LTA) framework developed in Chapter 5, the loco-manipulation framework will henceforth be referred to as the learning to adapt with manipulation (LTA-M) framework.

6.2 Methods

With the addition of the Z1 robot arm upon the B1 quadruped as depicted in Fig. 6.1, the robot is now modelled to have $n = 18$ degrees of freedom, all modelled as revolute joints, with their angular positions denoted as $\boldsymbol{q} \in \mathbb{R}^n$ and its base orientation represented as a rotation matrix $\boldsymbol{R}_B \in SO(3)$. This methods section covers the final control framework architecture in detail, the re-configuring and training setup of the new learning environment, and the revised IK-WBC formulation for loco-manipulation tasks. As commands and references in this Chapter are provided in various frames of the robot, it is important to note that a subscript B refers to the robot's base frame, subscript EE refers to the arm end-effector frame, and subscript W refers to the world frame.

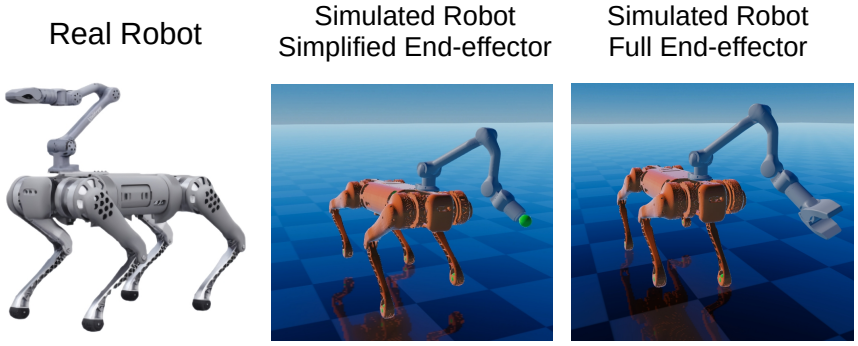


Figure 6.1: Snapshots showing the Unitree Z1 arm mounted on the B1 quadruped, forming the legged manipulator to be used in this Chapter. During training the end-effector was simplified, indicated by the green sphere in the middle snapshot, to reduce training time.

6.2.1 LTA-M Framework Architecture

As detailed in Fig. 6.2, the key changes between Fig. 5.1 and Fig. 6.2 is switching out the locomotion policy, π_L, for the loco-manipulation policy π_{LM} and the integration of the IK-WBC that takes input references from π_{LM}. Additionally, the arm end-effector command is added as

$$\boldsymbol{M}^{\text{cmd}} = \left[\boldsymbol{p}^{\text{cmd}}, \boldsymbol{\theta}^{\text{cmd}}\right] \in \mathbb{R}^6, \tag{6.1}$$

$$\text{where} \quad \boldsymbol{p}^{\text{cmd}} = \left[p^{\text{cmd}}_{x,\text{B}\{x,y,yaw\}}, p^{\text{cmd}}_{y,\text{B}\{x,y,yaw\}}, p^{\text{cmd}}_{z,\text{W}}\right] \in \mathbb{R}^3, \tag{6.2}$$

$$\boldsymbol{\theta}^{\text{cmd}} = \left[\theta^{\text{cmd}}_{x,\text{B}\{yaw\}}, \theta^{\text{cmd}}_{y,\text{B}\{yaw\}}, \theta^{\text{cmd}}_{z,\text{B}\{yaw\}}\right] \in \mathbb{R}^3 \tag{6.3}$$

which includes commands in the x direction, $p^{\text{cmd}}_{x,\text{B}\{x,y,yaw\}}$, y direction, $p^{\text{cmd}}_{y,\text{B}\{x,y,yaw\}}$, z direction, $p^{\text{cmd}}_{x,\text{W}}$, *roll* orientation, $\theta^{\text{cmd}}_{x,\text{B}\{yaw\}}$, *pitch* orientation, $\theta^{\text{cmd}}_{y,\text{B}\{yaw\}}$, and *yaw* orientation, $\theta^{\text{cmd}}_{z,\text{B}\{yaw\}}$. Components $p^{\text{cmd}}_{x,\text{B}\{x,y,yaw\}}$ and $p^{\text{cmd}}_{y,\text{B}\{x,y,yaw\}}$ are expressed in the base frame in only x, y, and *yaw* as this ensures that x and y move with the robot base while orientated in the *yaw* direction the robot is facing to keep control intuitive, while being decoupled in z, *roll*, and *pitch* to ensure the whole-body motions of the robot do not effect this command. For example, if the goal is to move the arm end-effector to pick up an object off the ground, it is desirable for the robot to adjust its base's z and *pitch* to aid in this task, but it would be undesirable for this motion to feed back into the input command.

Additionally, to simplify control and promote command kinematic feasibility, $\theta^{\text{cmd}}_{x,\text{B}\{yaw\}}$, $\theta^{\text{cmd}}_{y,\text{B}\{yaw\}}$, and $\theta^{\text{cmd}}_{z,\text{B}\{yaw\}}$ are set to always ensure that the end-effector is pointing at the target Cartesian position, while aligned with the robot base's *yaw*. This is achieved through the following process:

$$\hat{\mathbf{x}} = \frac{\boldsymbol{p}^{\text{cmd}}}{\|\boldsymbol{p}^{\text{cmd}}\|}, \tag{6.4}$$

$$\hat{\mathbf{z}} = \frac{\boldsymbol{\alpha}^T - \left(\boldsymbol{\alpha}^T \cdot \hat{\mathbf{x}}\right)\hat{\mathbf{x}}}{\|\boldsymbol{\alpha}^T - \left(\boldsymbol{\alpha}^T \cdot \hat{\mathbf{x}}\right)\hat{\mathbf{x}}\|}, \tag{6.5}$$

$$\hat{\mathbf{y}} = \frac{\hat{\mathbf{z}} \times \hat{\mathbf{x}}}{\|\hat{\mathbf{z}} \times \hat{\mathbf{x}}\|}, \tag{6.6}$$

$$R = \left[\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\right], \tag{6.7}$$

$$\text{RPY}(R) = \left[\theta^{\text{cmd}}_{x,\text{B}\{yaw\}}, \theta^{\text{cmd}}_{y,\text{B}\{yaw\}}, \theta^{\text{cmd}}_{z,\text{B}\{yaw\}}\right], \tag{6.8}$$

where $\boldsymbol{\alpha} = [0, 0, 1]$ and $\text{RPY}(\cdot)$ is a function to convert a rotation matrix to Euler angles. Additionally, as the *roll* of the end-effector may need to be adjusted based on the object and its orientation, that is being manipulated. Hence, an adjustment in *roll*, $\Delta\theta^{\text{cmd}}_{x,EE}$, can be achieved without effecting the direction the manipulator is facing simply through adding the desired roll in the end-effector frame to $\theta^{\text{cmd}}_{x,\text{B}\{yaw\}}$.

To account for error build-up between the robot model in the IK-WBC, originating from the input data generated by the state estimator (SE), π_{LM} not only outputs joint position references, $\boldsymbol{q}$, but also an adjustment term, $\Delta\boldsymbol{p}^{\text{ref}}$, to be added to $\boldsymbol{p}^{\text{cmd}}$. However, to better condition this output and remove high-frequency noise, $\Delta\boldsymbol{p}^{\text{ref}} + \boldsymbol{p}^{\text{cmd}}$ is passed through a Kalman filter, based off this implementation [165], to generate output $\boldsymbol{p}^{\text{ref}}_{\text{filt}}$, resulting in a smooth end-effector reference trajectory to be passed to the

IK-WBC to be tracked through a Cartesian task.

The functionality of π_G^{uni}, bio-inspired gait scheduler (BGS), SE, proportional derivative (PD) controller, and base velocity command, $\boldsymbol{U}^{\text{cmd}}$, are all the same as Fig. 5.1. The PD controller has the same formulation as in (5.1), but now $K_p = 300$ N/m and $K_d = 8$ Ns/m for the B1 quadruped, $K_p = 600$ N/m and $K_d = 12$ Ns/m for the Z1 arm, running at 1000 Hz. Additionally, all control modules run at 500 Hz with the exception of π_G^{uni} which runs at 100 Hz.

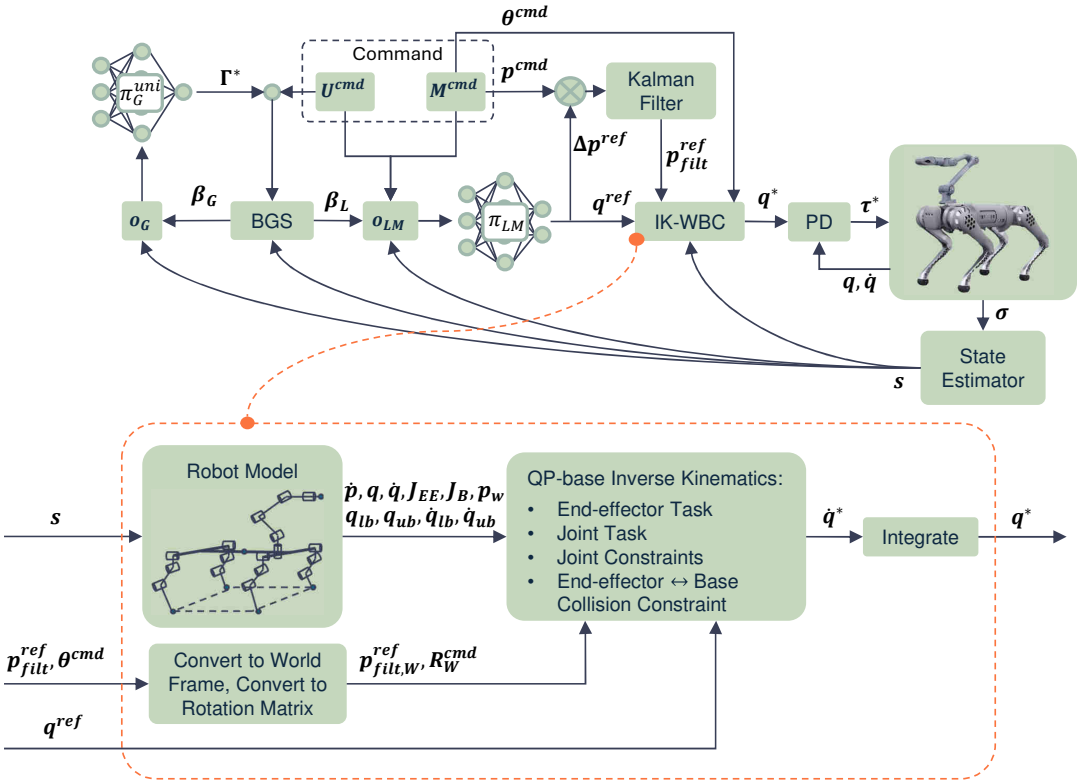


Figure 6.2: The final architecture of the LTA-M framework, featuring the BGS from Chapter 5, a retrained π_G^{uni} for B1Z1, π_{LM} the loco-manipulation policy to be trained in this Chapter, and the IK-WBC from Chapter 3 reconfigured for loco-manipulation tasks. Unlike π_L from Chapter 5, π_{LM} outputs not only joint references, $\boldsymbol{q}^{\text{ref}}$, but also a Cartesian command correction, $\Delta\boldsymbol{p}^{\text{ref}}$, to offset errors caused by SE drift and disturbances. A Kalman filter has also been added for smoothing of $\Delta\boldsymbol{p}^{\text{ref}}$.

6.2.2 Reconfiguring the IK-WBC for Loco-manipulation

Although the IK-WBC developed in Chapter 3 excels at manipulation tasks, in its current form, it is not completely suitable for loco-manipulation tasks within this framework. Instead of the quadruped joints being controlled via the Cartesian tasks, detailed in (3.9), it should be controlled through a joint position tracking task. Highly adaptable locomotion has been achieved in Chapter 5 with π_L^{bio} using an action space of joint positions, hence the best way to harness the proficiency of such a policy's output is to track its output joint positions, which also removes the requirements for the halt constraint (3.3). Furthermore, with more dynamic contact states being executed by the robot and even occasionally exhibiting aerial phase (such as during a running gait), the CoM task (3.10) and the CoM constraint (3.7) are also now not required. Instead, a collision constraint is added to prevent collisions occurring between the robot's base and the arm end-effector; more dynamic motions increase the risk of such collisions. As such, the new QP is formulated as

$$\min_{\dot{\boldsymbol{q}}} \quad \frac{1}{2}\dot{\boldsymbol{q}}^T \boldsymbol{A}^T \boldsymbol{A}\dot{\boldsymbol{q}} - \boldsymbol{b}^T \boldsymbol{A}\dot{\boldsymbol{q}} \tag{6.9}$$

$$\text{s.t.} \quad \boldsymbol{C}_{\text{col}}\dot{\boldsymbol{q}} \geq c_{\text{col}} \,, \tag{6.10}$$

$$\dot{\boldsymbol{q}}_{\text{lb}} \leq \dot{\boldsymbol{q}} \leq \dot{\boldsymbol{q}}_{\text{ub}} \,, \tag{6.11}$$

$$\boldsymbol{q}_{\text{lb}} \leq \boldsymbol{q} \leq \boldsymbol{q}_{\text{ub}} \,, \tag{6.12}$$

where $\boldsymbol{A}$ and $\boldsymbol{b}$ are now formulated to contain a Cartesian task for the end-effector and joint tasks for all quadruped joints:

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} \boldsymbol{A}_{\text{Cart}} \\ \boldsymbol{A}_{\text{Jnt}} \end{bmatrix} \in \mathbb{R}^{(6+\text{n})\times \text{n}}, \\ \boldsymbol{b} &= \begin{bmatrix} \boldsymbol{b}_{\text{Cart}} \\ \boldsymbol{b}_{\text{Jnt}} \end{bmatrix} \in \mathbb{R}^{6+\text{n}}, \end{aligned} \tag{6.13}$$

for which it should also be noted that no task weights are used this time as both tasks share the same priority. $\boldsymbol{A}_{\text{Cart}}$ and $\boldsymbol{b}_{\text{Cart}}$ have a very similar formulation to that in (3.8) but is updated notation to reflect that only one Cartesian task is being used and the use of the Kalman filter:

$$\begin{aligned} \boldsymbol{A}_{\text{Cart}} &= \boldsymbol{J}^{EE} \in \mathbb{R}^{6\times n}, \\ \boldsymbol{b}_{\text{Cart}} &= \begin{bmatrix} K_p^{\text{Cart}} \left(\boldsymbol{p}_{\text{filt,W}}^{\text{ref}} - \boldsymbol{p}_{\text{W}}^{EE} \right) + K_d^{\text{Cart}} \frac{\left(\boldsymbol{p}_{\text{filt,W}}^{\text{ref}} - \boldsymbol{p}_{\text{W}}^{EE} \right)}{\delta t} \\ K_p^{\text{Cart}} \boldsymbol{R}_{\text{W}}^{EE} \left(\log \left(\left(\boldsymbol{R}_{\text{W}}^{EE} \right)^T \boldsymbol{R}_{\text{W}}^{\text{cmd}} \right) \right)^{\vee} \end{bmatrix} \in \mathbb{R}^6, \end{aligned} \tag{6.14}$$

where $\boldsymbol{J}^{EE}$ is the Jacobian matrix of the end-effector, $K_p^{\text{Cart}} = 500$ is the Cartesian task proportional gain, $\boldsymbol{p}_{\text{filt,W}}^{\text{ref}} \in \mathbb{R}^3$ is $\boldsymbol{p}_{\text{filt}}^{\text{ref}}$ converted to the world frame, $\boldsymbol{p}_{\text{W}}^{EE} \in \mathbb{R}^3$ is the current end-effector position in the world frame, $\delta t = 0.002$ s is the timestep of the IK-WBC, $\boldsymbol{R}_{\text{W}}^{EE}$ is the current rotation matrix of the end-effector, $\boldsymbol{R}_{\text{W}}^{\text{cmd}}$ is the rotation matrix of $\boldsymbol{\theta}^{\text{cmd}}$ converted to the world frame, and $\vee : \mathfrak{so}(3) \mapsto \mathbb{R}^3$ transforms a 3×3 skew-symmetric matrix to a $\mathbb{R}^3$ vector [166]. The most significant change in this new formulation of (3.8) as (6.14) is a new method of calculating the task orientation residual. With $\boldsymbol{U}^{\text{cmd}}$ containing a *yaw* component, the robot model will experience significant changes in its orientation, resulting in the previous method being at significant risk of not calculating a suitable desired orientation trajectory. However, with this revised formulation, it is guaranteed to produce the shortest orientation trajectory between the current end-effector orientation and the commanded orientation [166]. As a result, this significantly improves orientation control during dynamic locomotion.

To realise $\boldsymbol{q}^{\text{ref}}$, $\boldsymbol{A}_{\text{Jnt}}$ and $\boldsymbol{b}_{\text{Jnt}}$ are now formulated as a joint position tracking task:

$$\boldsymbol{A}_{\text{Jnt}} = \boldsymbol{S} \in \mathbb{R}^{n \times n}, \quad \boldsymbol{b}_{\text{Jnt}} = K_p^{\text{Jnt}} \left(\frac{(\boldsymbol{q}^{\text{ref}} - \boldsymbol{q})}{\delta t} \right) + K_d^{\text{Jnt}}(-\dot{\boldsymbol{q}}) \in \mathbb{R}^n, \tag{6.15}$$

where $\boldsymbol{S}$ is a selection matrix to select all quadruped joints (ignores all arm joints through setting all corresponding arm elements to 0 and all quadruped elements to 1), $K_p^{\text{Jnt}} = 0.45$ is the joint proportional gain, and $K_d^{\text{Jnt}} = 0.001$ is the joint derivative gain. To formulate a collision constraint between the end-effector and robot base, inspired by the work in [94], the following is used:

$$\text{for } d < d_i, \quad \boldsymbol{C}_{\text{col}} = \left(\frac{\boldsymbol{p}^B - \boldsymbol{p}^{EE}}{d} \right)^T \left(\boldsymbol{J}^B - \boldsymbol{J}^{EE} \right) \in \mathbb{R}^n, \quad c_{\text{col}} = -\xi \frac{d - d_s}{d_i - d_s}, \tag{6.16}$$

where $d = \|\boldsymbol{p}^B - \boldsymbol{p}^{EE}\|$ is the distance between the position of the quadruped base, $\boldsymbol{p}^B \in \mathbb{R}^3$, and end-effector, $\boldsymbol{p}^{EE} \in \mathbb{R}^3$, $\boldsymbol{J}^B \in \mathbb{R}^3$ is the position Jacobian matrix of the quadruped base, $d_i = 0.3$ m is the influence distance, and $d_s = 0.2$ m is the security distance. In essence, this works to increasingly decrease the velocity of the end-effector as it comes closer to the base until it reaches the security distance, at which point no further velocity can be produced that would reduce the distance between the end-effector and the base. It should be noted that if $d < d_i$, this constraint is not active. Overall, in this reduced form, not only is the IK-WBC in the ideal form to realise both input joint reference and end-effector Cartesian task during locomotion, but through removing several now redundant tasks and constraints, this considerably improved the

computation time of the QP; this serves to improve both dynamic performance and training time of π_{LM}. This relatively straightforward process of switching out tasks and constraints exemplifies the generic capabilities of the IK-WBC.

6.2.3 Augmenting the Learning Environments for Loco-manipulation

From the success of the work in Chapter 5, much of the policy training setup is kept constant; all details set out in Section 5.2.3 are kept constant for the training of π_{LM} and π_G^{uni}. The core changes are instead related to enabling π_{LM} to generate loco-manipulation reference motions in joint space for the quadruped and task space for the arm end-effector for the IK-WBC to realise. As such, the action vector of π_{LM} is fomulated as $\boldsymbol{a}_{LM} = \left[\boldsymbol{q}^{\text{ref}}, \Delta\boldsymbol{p}^{\text{ref}}\right] \in \mathbb{R}^{15}$. $\Delta\boldsymbol{p}^{\text{ref}}$, an adjustment term to add to $\boldsymbol{p}^{\text{cmd}}$, rather than providing the a full stand-alone Cartesian reference as not only does this addition of $\Delta\boldsymbol{p}^{\text{ref}}$ and $\boldsymbol{p}^{\text{cmd}}$ help guide and better condition the training of π_{LM} (reduces the scope of what π_{LM} has to learn), but it also ensures a suitable mean for the policy output as the mean output is always the value of $\boldsymbol{p}^{\text{cmd}}$ which ensures that even in the early stages of training π_{LM} can quickly learn to refine the solution space.

In terms of the noise and sampling used within the training environment, all parameters within Table 5.3 are also used in this learning environment, with the exception of the distributions for selected gait ID, Γ, and with the addition of the sampling distributions for $\boldsymbol{p}^{\text{cmd}}$. As loco-manipulation tasks require accurate tracking of the end-effector command trajectory to ensure successful manipulation, gaits that rely on large base pose changes are no longer feasible. As such, bounding and pronking gaits are removed from the gait pool, while keeping trotting and running for efficient and smooth nominal locomotion and ambling, limping, and hopping as auxiliary gaits to aid in stability recovery. Hence, the distribution of Γ becomes $\mathcal{U}(0, 4)$, to select a gait from the gait pool $[stand, trot, run, limp, amble, hop]$. To sample $\boldsymbol{p}^{\text{cmd}}$, the following distribution and constraints are used to ensure that the command remains within the sphere of arm reachability and doesn't intersect with the quadruped base,

$$\boldsymbol{p}^{\text{cmd}} \circ \boldsymbol{p}^{\text{cmd}} \leq \mathcal{U}(0, \vartheta)^2, \tag{6.17}$$

$$\boldsymbol{p}_B^{\text{box},lb} \leq \boldsymbol{p}_B^{\text{cmd}} \leq \boldsymbol{p}_B^{\text{box},ub}, \tag{6.18}$$

where $\boldsymbol{p}_B^{\text{cmd}}$ is the end-effector command in the world frame, $\vartheta = 0.8$ m is the radius of the sphere in which the arm can reach, while $\boldsymbol{p}_B^{\text{box},lb} = [-0.8, -0.4, -0.8]$ and $\boldsymbol{p}_B^{\text{box},y} =$

$[0.3, 0.4, 0.8]$ are the lower and upper bounds of a blocking box that the end-effector cannot intersect with.

To ensure all key state and command data is observed by π_{LM}, creating a rich observation space, the observation vector is defined as $\boldsymbol{o}_{LM} = \left[\beta_L, \boldsymbol{s}, \boldsymbol{v}_B^{\text{cmd}}, \boldsymbol{p}_B^{EE}, \dot{\boldsymbol{p}}_B^{EE}\right] \in \mathbb{R}^{100}$, where $\boldsymbol{p}_B^{EE}$ and $\dot{\boldsymbol{p}}_B^{EE}$ are the end-effector Cartesian position and velocity in the base frame, while β_L, $\boldsymbol{s}$, and $\boldsymbol{v}_B^{\text{cmd}}$ are the same as defined in Chapter 5. The loco-manipulation reward function, r_{LM}, is constructed to ensure both manipulation and multi-gait locomotion can be realised simultaneously while remaining stable, efficient, and accurate,

$$r_{LM} = \mathrm{w}_\eta r_\eta + \mathrm{w}_{v^{\text{cmd}}} r_{v^{\text{cmd}}} + \mathrm{w}_f r_f + \mathrm{w}_{\text{stab}} r_{\text{stab}} + \mathrm{w}_{\text{e}} r_{\text{e}}, \tag{6.19}$$

where efficiency, base velocity command tracking, gait reference tracking, and stability reward terms r_η, $r_{v^{\text{cmd}}}$, r_f and r_{stab} are inherited from (5.5) with the same values for the reward weights. The addition of the end-effector task reward, r_e, with its task weight, $\mathrm{w}_e = 15$, ensures that π_{LM} learns to accurately track the manipulation command $\boldsymbol{M}^{\text{cmd}}$ in both position and orientation, taking the form of

$$r_{\text{e}} = \psi\left(\left\|\boldsymbol{p}_W^{EE} - \boldsymbol{p}_W^{\text{cmd}}\right\|^2\right) + \psi\left(\left\|\boldsymbol{\theta}_W^{EE} - \boldsymbol{\theta}_W^{\text{cmd}}\right\|^2\right), \tag{6.20}$$

where $\psi : x \to 1 - \tanh\left(x^2\right)$ is used to normalise the reward term, $\boldsymbol{p}_W^{\text{cmd}} \in \mathbb{R}^3$ the end-effector position command in the world frame, and $\boldsymbol{\theta}_W^{\text{cmd}} \in \mathbb{R}^3$ the end-effector orientation command in the world frame. Additionally, as π_{LM} outputs $\Delta\boldsymbol{p}^{\text{ref}}$, smoothing of this output needs to be added to r_η, taking the updated form of

$$r_\eta = \|\ddot{\boldsymbol{q}}\|^2 + \|\boldsymbol{\tau}\|^2 + \|\boldsymbol{q}^* - \boldsymbol{q}_{t-1}^*\| + \|\Delta\boldsymbol{p}^{\text{ref}} - \Delta\boldsymbol{p}_{t-1}^{\text{ref}}\| \tag{6.21}$$

The training and reward function of π_G^{uni} remains that same as in Chapter 5 with the exception of adding the IK-WBC, $\boldsymbol{p}^{\text{cmd}}$ generation, its action becoming $\Gamma^* \in [0, 5]$ to account for the reduced gait pool, and with the additional joints of the arm the observation vector becomes $\boldsymbol{o}_G \in \mathbb{R}^{84}$. Tracking of $\boldsymbol{p}^{\text{cmd}}$ is omitted from the reward function r_G as π_G^{uni} should only be concerned about the efficiency and stability of the quadruped base, with the motions of the arm being interpreted by π_G^{uni} as just a perturbation the base is subject to. For example, the CoT reward should only consider the efficiency of the quadruped as only these joints are used in realising locomotion, while the efficiency of the manipulation should be independent from CoT calculations.

For the training process itself, π_{LM} and π_G^{uni} were trained for 20k iterations in RaiSim [157], taking 18 and 24 hours respectively (the increased training time being a result of adding the IK-WBC to the environment), on a standard desktop computer with one Nvidia RTX3090 with a training frequency of 100 Hz.

6.3 Results

To validate and analyse the proficiency of the developed framework, a set of experiments are devised to assess its ability to deploy a diverse set of gaits during end-effector control, evaluate if the animal gait strategies leveraged in Chapter 5 have also been successfully instilled within this framework, and set of pick-and-place tasks during locomotion of different terrain to assess its loco-manipulation capabilities; hence determining if the developed framework has successfully achieved the objectives of this project. Testing this final framework on hardware during real-world experiments would provide further validation of this framework; however, due to the time constraints of this project, this will form part of the proposed future work.

6.3.1 Multi-gait Deployment During Manipulator Control

To assess the framework's ability to not only deploy all targeted gaits but also complete end-effector trajectories while switching between gaits, two sets of experiments were completed. The first consisted of a constant base velocity command of $\boldsymbol{U}^{\text{cmd}} = [0.75, 0.0, 0.0]$ m/s, switching gaits every 2 s, while completing a sinusoidal $\boldsymbol{p}^{\text{cmd}}$ trajectory with x, y, and z maximum amplitudes of 0.3 m, 0.3 m, and 0.6 m respectively, relative to the nominal end-effector position. It should be noted that as part of this experiment's objective is to assess the deployment of each gait within the gait pool, and the ability to switch between them, each gait is manually cycled through to ensure each gait is deployed. Hence π_G^{uni} is not used in this experiment. This experiment was completed on both flat and rough terrain, for which data is presented in Fig. 6.3 and Fig. 6.4, respectively. The rough terrain for this experiment is generated through using fractal noise, with a maximum height of 54% of the nominal robot base height. Across both flat and rough terrain, π_{LM} in combination with the IK-WBC successfully realise all targeted gaits and end-effector commands $\boldsymbol{p}^{\text{cmd}}$ and $\boldsymbol{\theta}^{\text{cmd}}$ with considerable accuracy, as reflected by the relatively low average tracking errors detailed in Table 6.1.

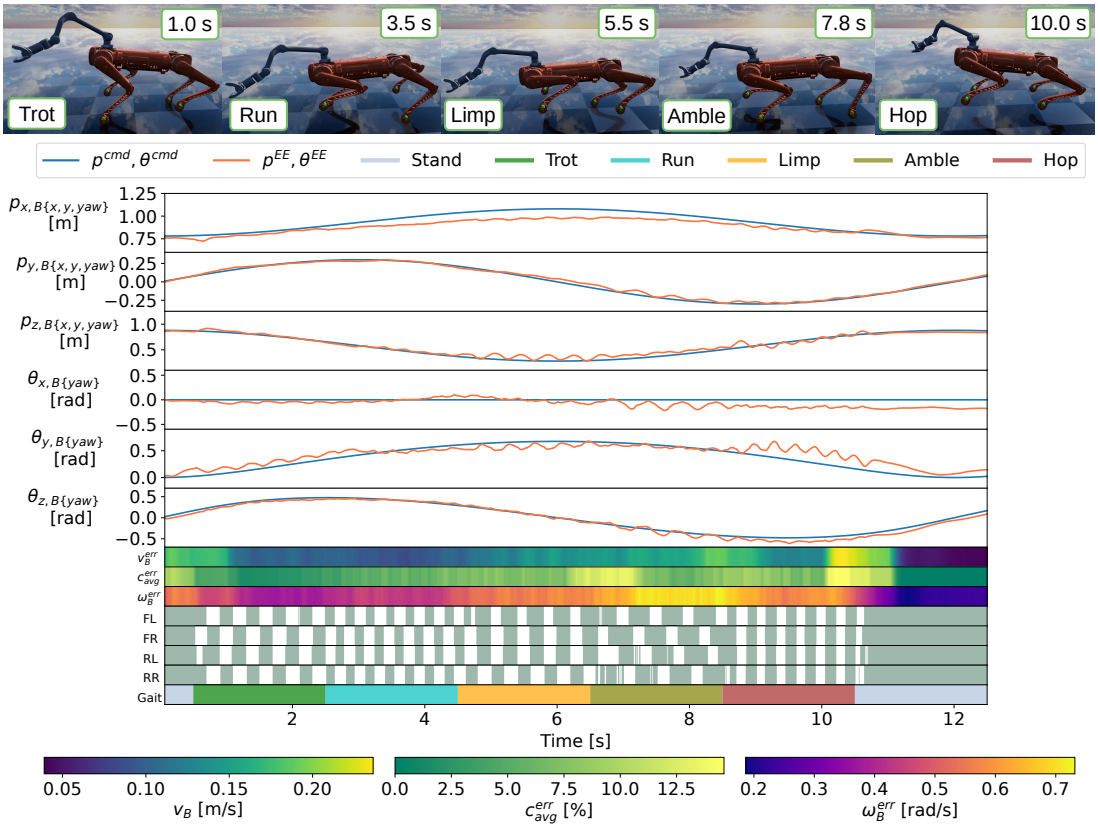


Figure 6.3: Plots of the data collected for the gait switching experiment, with constant base velocity, as depicted in the snapshots at the top of the figure. The top six subplots depict the sinusoidal command $\boldsymbol{p}^{\text{cmd}}$, and resultant $\boldsymbol{\theta}^{\text{cmd}}$, and the realised end-effector tacking of the robot in all 6 DoF. The next three subplots depict the error magnitudes in base velocity command tracking, contact schedule tracking, and base angular velocity. The last plots detail the realised feet contact states and commanded gait.

Table 6.1: Mean errors for gait switching and end-effector trajectory experiments.

$\boldsymbol{U}^{\text{cmd}}$ Type	Terrain	p^{err} [m]	θ^{err} [rad]	v_B^{err} [m/s]	$c_{\text{avg}}^{\text{err}}$ [%]	ω_B^{err} [rad/s]
Constant	Flat (Fig. 6.3)	0.035	0.075	0.13	6.4	0.50
v_x^{cmd}	Rough (Fig. 6.4)	0.039	0.067	0.17	14	0.55
Sinusoidal	Flat (Fig. 6.5)	0.024	0.064	0.19	6.8	0.54
v_x^{cmd}, ω_z^{cmd}	Rough (Fig. 6.6)	0.028	0.062	0.20	15	0.64

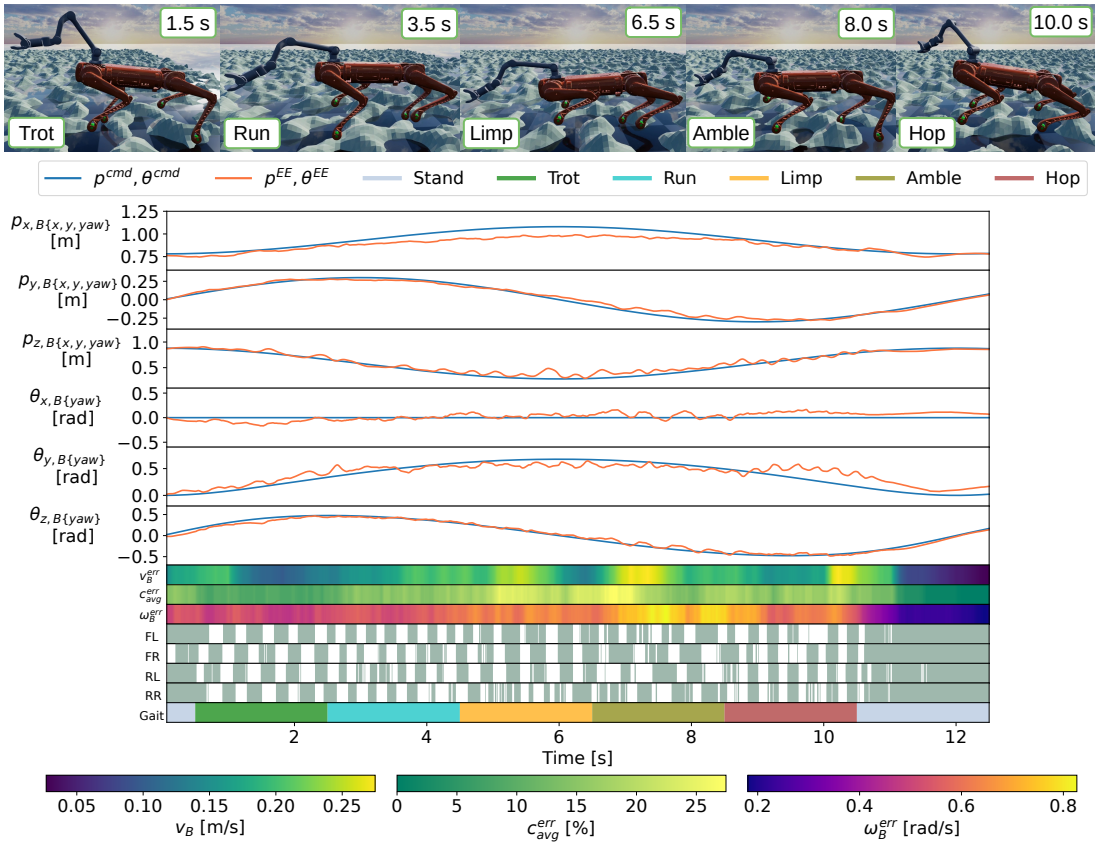


Figure 6.4: Plots depicting the same experiment to that in Fig. 6.3 (using the same graph format) but on rough terrain, featuring a maximum terrain height of 54% of the nominal robot base height. Numerical comparison between rough and flat terrain performance is presented in Table 6.1.

This end-effector trajectory tracking proficiency can be attributed to the whole-body motions the π_{LM}-IK-WBC pair is able to realise, as depicted by the snapshots of the experiments at the top of Fig. 6.3 and Fig. 6.4; the IK-WBC is able to utilise all robot DoF to enable the realisation of the end-effector trajectory, in accordance with the motions and findings presented in Chapter 3, but through combining it with π_{LM} the robot is able to exhibit these whole-body motions while also locomotion, such as pitching down the robot base to help the end-effector reach trajectories tending towards the ground. Furthermore, despite the introduction of the flat terrain in Fig. 6.4 and π_{LM} only observing flat terrain during training, the increase in $\boldsymbol{p}^{\text{cmd}}$ and $\boldsymbol{\theta}^{\text{cmd}}$ tracking errors is minimal, only increasing the average error by 0.004 m, in turn demonstrating

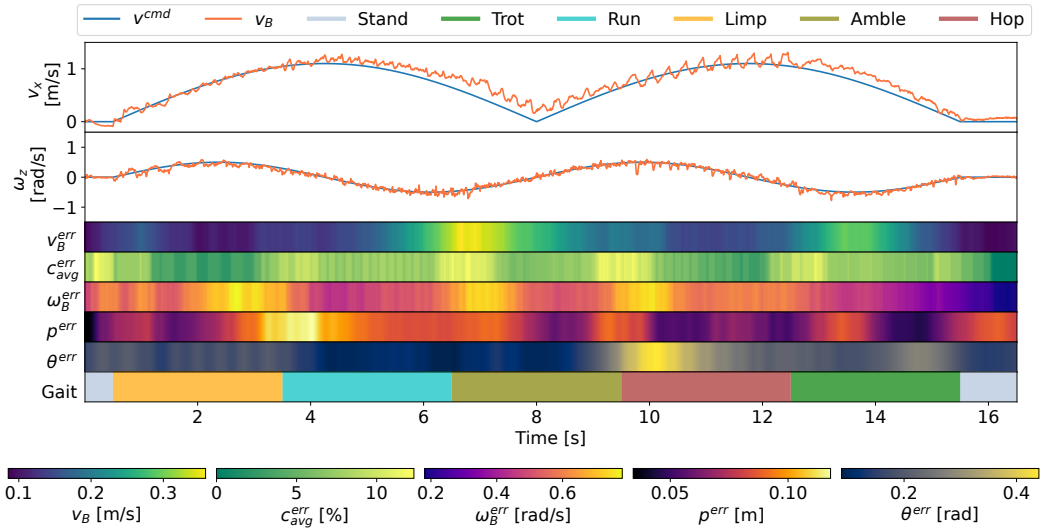


Figure 6.5: The top two plots illustrate the tracking of the robot's base to the commanded base velocity in x and yaw, following a sinusoidal trajectory. The following six plots, in order, depict the base velocity tracking error, contact schedule tracking error, base angular velocity tracking error, end-effector position tracking error, end-effector orientation tracking error, and the commanded gait.

the exemplary adaptability this framework is able to provide.

In the second set of experiments, the same end-effector trajectory, terrains and gait selection methods that were used but with $\boldsymbol{U}^{\text{cmd}}$ now providing sinusoidal base velocity commands in both x and yaw, with amplitudes of 1.1 m/s and 0.5 rad/s respectively, in order to assess if this previously observed proficiency is upheld with more complex base velocities. From inspecting the data presented in Fig. 6.5 for flat terrain and Fig. 6.6 for rough terrain, despite spikes in $\boldsymbol{p}^{\text{cmd}}$ and $\boldsymbol{\theta}^{\text{cmd}}$ tracking errors, p^{err} and θ^{err} respectively, occurring at the $\boldsymbol{U}^{\text{cmd}}$ maximum amplitude the overall performance proficiency difference is negligible to that in the previous experiment. This is highlighted in Table 6.1, where both p^{err} and θ^{err} are roughly 0.01 m lower than in the previous experiment. In turn, this demonstrates that not only can the framework realise a diverse set of gaits, but also complete challenging end-effector trajectory tasks during demanding base velocity commands, all the while completing gait transitions and remaining stable on rough terrain. Furthermore, the magnitudes of the average errors detailed in Table 6.1 are very similar to that collected during the assessment of π_L^{bio} in Fig. 5.6, further supporting the notion that π_{LM} has attained the same level of

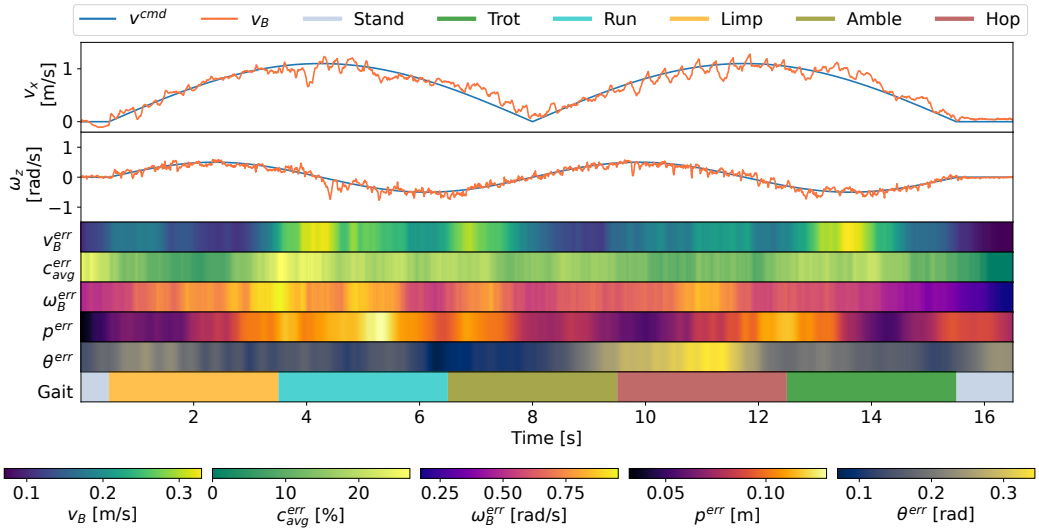


Figure 6.6: Plots depicting the same experiment to that in Fig. 6.5 (using the same graph format) but on rough terrain, featuring a maximum terrain height of 54% of the nominal robot base height. Numerical comparison between rough and flat terrain performance is presented in Table 6.1.

proficiency as to that of π_L^{bio}.

With the IK-WBC directly realising the joint position references generated by π_{LM}, determining the proficiency of these references can be directly observed. However, the effectiveness of the $\Delta \boldsymbol{p}^{\text{ref}}$ part of the π_{LM} action is not immediately clear; why not just use $\boldsymbol{p}^{\text{cmd}}$ directly as the reference for the end-effector Cartesian IK-WBC task? To address this, the experiments in Fig. 6.3 and Fig. 6.4 are repeated, but with the IK-WBC using $\boldsymbol{p}^{\text{cmd}}$ directly as the reference for the end-effector Cartesian task. This data is then plotted in Fig. 6.7 along with the end-effector data from Fig. 6.3 and Fig. 6.4 to reveal and inspect the proficiency gain through utilising $\Delta \boldsymbol{p}^{\text{ref}}$ generated by π_{LM}. From the data presented in Fig. 6.7, it is clear that through utilising $\Delta \boldsymbol{p}^{\text{ref}}$ there is a dramatic reduction in tracking error in the resultant performance, demonstrating that during training π_{LM} learns to effectively utilise $\Delta \boldsymbol{p}^{\text{ref}}$ to improve command tracking; taking an average across the data in Fig. 6.7, tracking is improved by 53% and 40% for flat and rough terrain respectively. An additional notable observation from Fig. 6.7 is that the use of $\boldsymbol{p}^{\text{cmd}}$ has better performance on rough terrain compared to its performance on flat terrain. Specifically, the highest spike in tracking error occurs during a hopping gait, which itself results in large variations in base height. With

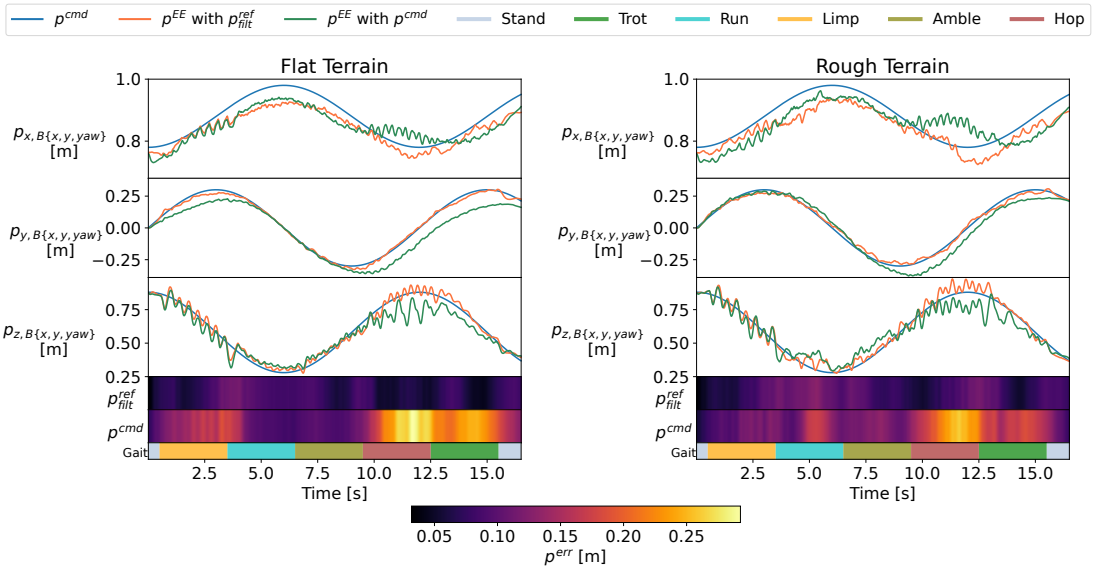


Figure 6.7: A comparison between using the π_{LM} action $\Delta \boldsymbol{p}^{\text{ref}}$ and just $\boldsymbol{p}^{\text{cmd}}$ with no adjustment term as the input reference trajectory for the end-effector Cartesian task within the IK-WBC on flat and rough terrain. The top plots show the end-effector command, when using $\Delta \boldsymbol{p}^{\text{ref}}$, and when using $\boldsymbol{p}^{\text{cmd}}$ Cartesian trajectories, while the bottom heatmaps show the normalised (to the maximum and minimum value across all heatmaps) end-effector position error magnitude.

rough terrain occasionally reducing the distance between the base and ground, this reduces the magnitude of base height oscillations, and in turn inflicts less adverse motions on the robot arm. Consequently, with the tracking error during hopping being significantly reduced when using $\Delta \boldsymbol{p}^{\text{ref}}$, this demonstrates that π_{LM} can not only learn to facilitate challenging end-effector trajectories, but also offset the tracking errors that perturbations would otherwise inflict.

Overall, from these experiments and the data collected, it can be determined that the combination of π_{LM} and the IK-WBC provide the proficiency in locomotion and manipulation required to form the foundation of the LTA-M framework. However, it is yet to be determined if the same animal adaptability achieved in Chapter 5 has been instilled within this framework and if loco-manipulation tasks are indeed achievable, which in turn forms the focus of the rest of this results section.

6.3.2 Validation of Animal Gait Strategies Successfully Transferred

With the retraining of π_G^{uni} for the B1Z1 legged manipulator, it is vital to validate the successful instillation of the same animal gait strategies that emerged in Chapter 5 to ensure the locomotion proficiencies gained through the development of the LTA framework are also present within the LTA-M framework. To maintain clarity, the gait selection policy that was trained in Chapter 5 and deployed on the A1 will be referred to as $\pi_{G,\text{A1}}^{\text{uni}}$, and the retrained gait selection policy to be used in the locomanipulation framework and deployed on the B1Z1 will be referred to as $\pi_{G,\text{B1Z1}}^{\text{uni}}$. To achieve this, the same experimental setup that was used in Section 5.3.3 in order to plot a comparable graph to compare the robot data presented in Fig. 5.7 of $\pi_{G,\text{A1}}^{\text{uni}}$ with data collected from using $\pi_{G,\text{B1Z1}}^{\text{uni}}$, as presented in Fig. 6.8. With $\pi_{G,\text{A1}}^{\text{uni}}$ being validated to successfully realise animal gait strategies within Chapter 5, it stands to reason that if the performance of $\pi_{G,\text{B1Z1}}^{\text{uni}}$ matches $\pi_{G,\text{A1}}^{\text{uni}}$ then it strongly indicates the successful instillation of animal gait strategies.

To collect this data, taking the same approach as in Fig. 5.7, the framework is run with increasing base velocity commands up to 4 m/s in increments of 0.13 m/s. At each velocity command interval, 1000 samples are collected for each metric, which are then taken as an average to form each data point within Fig. 6.8. With Fig. 6.8 featuring data from two systems of considerably different size and weight, the Froude number is used on the x-axis rather than base velocity for improved data visualisation for comparison. Similarly to Fig. 5.7, a core observation from the performance of $\pi_{G,\text{B1Z1}}^{\text{uni}}$ is that its gait transition region (where no single gait occupies more than 75% of the gaits used at a specific Froude number) occurs over a range of Froude numbers rather than a discrete point. This behaviour is reflected in both animals [121, 113, 161] and in the $\pi_{G,\text{A1}}^{\text{uni}}$ data. Consequently, due to this observation and that the transition region for $\pi_{G,\text{B1Z1}}^{\text{uni}}$ and $\pi_{G,\text{A1}}^{\text{uni}}$ are almost identical in respect to the Froude number, this supports the notion that this aspect of animal locomotion has been successfully transferred.

Energy Expenditure – Cost of Transport

As discussed in Chapter 5, animals are known to minimise CoT through transitioning to more energy-efficient gaits [109], which is also reflected by $\pi_{G,\text{A1}}^{\text{uni}}$. This behaviour is also successfully exhibited by $\pi_{G,\text{B1Z1}}^{\text{uni}}$, as shown within Fig. 6.8, where at lower

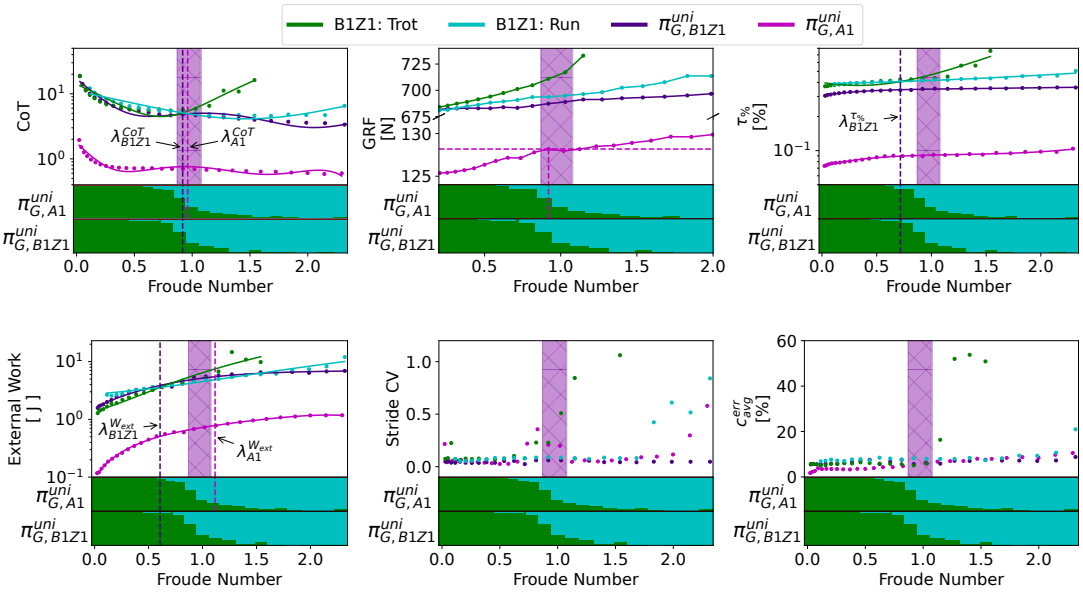


Figure 6.8: Comparing the performance and locomotion characteristics of the retrained π^{uni}_{G} on B1Z1, $\pi^{\text{uni}}_{G,\text{B1Z1}}$, with π^{uni}_{G} on A1, $\pi^{\text{uni}}_{G,\text{A1}}$, from Chapter 5 to validate successful transfer of animal gait strategies. The A1 data was taken directly from Fig. 5.7. The magenta and purple highlighted regions of the graph represent the gait transition regions (where no single gait occupies more than 75% of the gaits used at a specific Froude number) for $\pi^{\text{uni}}_{G,\text{A1}}$ and $\pi^{\text{uni}}_{G,\text{B1Z1}}$ receptively. It should be noted that distinguishing between the transition regions of each robot is rather unclear as they are almost exactly the same.

Froude numbers a trotting gait is used and at higher Froude numbers $\pi^{\text{uni}}_{G,\text{B1Z1}}$ switches to a running gait to preserve efficiency. Furthermore, similar to the strategy used by $\pi^{\text{uni}}_{G,\text{A1}}$, $\pi^{\text{uni}}_{G,\text{B1Z1}}$ has its transition phase centered around its OTP, $\lambda^{\text{CoT}}_{\text{B1Z1}}$, a gait characteristic also know to be present in animals [123]. It should also be noted that $\lambda^{\text{CoT}}_{\text{B1Z1}}$ and the OTP for $\pi^{\text{uni}}_{G,\text{A1}}$, $\lambda^{\text{CoT}}_{\text{A1}}$, occur at very similar Froude number values. This aligns with the dynamic similarity hypothesis [112] that underpins Chapter 4 and this chapter's findings, in turn further strengthening this project's claim that leveraging the Froude number to govern gait transitions provides a general and efficient gait transition method.

Actuator-structural Forces – Foot Contact Forces

Animals [118] and $\pi^{\text{uni}}_{G,\text{A1}}$ both minimise actuator-structural forces and improve stability, as detailed in Chapter 5, and can be characterised by a reduction of the ground reaction force (GRF) of the feet. However, as shown in Fig. 6.8, the trotting gait of B1Z1 never exhibits a lower GRF than a running gait, most likely due to the additional disturbance forces of the mounted Z1 arm. In turn, this results in $\pi^{\text{uni}}_{G,\text{B1Z1}}$ starting to select a running gait at lower Froude numbers compared to $\pi^{\text{uni}}_{G,\text{A1}}$ in order to improve stability with a gait of higher stride frequency [121], albeit the use of the running gait is still limited at lower Froude number values. However, when inspecting the torque saturation[1], $\tau_{\%}$, plot for $\pi^{\text{uni}}_{G,\text{B1Z1}}$, there is indeed an OTP in terms of $\tau_{\%}$, $\lambda^{\tau_{\%}}_{\text{B1Z1}}$, which coincides with the increased use of the running gait just before the transition phase starts. In turn, this supports the notion that $\pi^{\text{uni}}_{G,\text{B1Z1}}$ is indeed minimising actuator-structural and improving stability through switching gaits, just as animals do, demonstrating that this animal gait strategy has been successfully embedded.

Mechanical Work – External Work

Animals have been shown to transition between gaits to preserve mechanical work [123], a characteristic that is also present in $\pi^{\text{uni}}_{G,\text{A1}}$ but with reduced impact compared to the other metrics. Although $\pi^{\text{uni}}_{G,\text{B1Z1}}$ does indeed minimise mechanical work through gait transitions, as shown in Fig. 6.8, contrary to the OTP for $\pi^{\text{uni}}_{G,\text{A1}}$, $\lambda^{W_{\text{ext}}}_{\text{A1}}$, the $\pi^{\text{uni}}_{G,\text{B1Z1}}$ OTP for mechanical work, $\lambda^{W_{\text{ext}}}_{\text{B1Z1}}$, occurs at a much earlier Froude number than the transition phase. This is a direct result of the instability the addition of the Z1 arm to the B1 quadruped induces; considering that mechanical work was discovered to characterise stability and that gaits of increase stride frequency have improved stability [121], this early OTP reflects the earlier rise in instability and the requirement for $\pi^{\text{uni}}_{G,\text{B1Z1}}$ to start using a running gait. This behaviour is also reflected in the plot of $\tau_{\%}$ (which also characterises stability), where the metric's OTP coincides with the initial use of the running gait. However, with this metric seeing relaxed penalisation, as with $\pi^{\text{uni}}_{G,\text{A1}}$, the impact of this metric on $\pi^{\text{uni}}_{G,\text{B1Z1}}$ appears to be minimal.

[1] As discussed in Chapter 5, torque saturation is a better metric for characterising actuator-structural forces in robots.

Stability – Stride Duration Coefficient of Variation

It has been discovered that animals preserve their stability through reducing their stride duration CV [121], a behaviour successfully instilled within $\pi^{\text{uni}}_{G,\text{A1}}$. This behaviour is replicated by $\pi^{\text{uni}}_{G,\text{B1Z1}}$, where through gait selection the stride CV is minimised, resulting in stability being successfully preserved as shown in Fig. 6.8. However, within the stride CV animal data in presented in Fig. 5.7 and the $\pi^{\text{uni}}_{G,\text{A1}}$ data in Fig. 6.8 a clearly identifiable rise in stride CV is observed just before the gait transition phase, yet in the stride CV data for $\pi^{\text{uni}}_{G,\text{B1Z1}}$ this initial rise never occurs. This is a result of the increased risk of critical instability being reached due to the Z1 arm, which is mitigated by $\pi^{\text{uni}}_{G,\text{B1Z1}}$ aggressively minimsing stride CV before any spike is observed (with preemptive indications of decrease stability most likely being indicated by $\tau_{\%}$ and W_{ext}); the increased stride CV observed by the trot gait of the B1Z1 at higher Froude numbers is never exhibited by $\pi^{\text{uni}}_{G,\text{B1Z1}}$ due to it initiating a gait transition. However, when inspecting the plot for $c^{\text{err}}_{\text{avg}}$, which provides an indication of stability at every timestep rather than every gait cycle, both $\pi^{\text{uni}}_{G,\text{A1}}$ and $\pi^{\text{uni}}_{G,\text{B1Z1}}$ exhibit very similar characteristics and no spike in the metric is observed. In turn, this suggests that although the data of $\pi^{\text{uni}}_{G,\text{A1}}$ and $\pi^{\text{uni}}_{G,\text{B1Z1}}$ vary in exact stride CV minimisation characteristics, they both indeed realise the high-level animal gait strategy of maintaining stability through transitioning gaits to preserve stability.

Overall, although there are some slight variations between $\pi^{\text{uni}}_{G,\text{A1}}$ and $\pi^{\text{uni}}_{G,\text{B1Z1}}$ in their exact behaviour in terms of the locomotion metrics in question, mostly due to the addition of the Z1 arm, they both exhibit the same overarching animal gait strategies of minimising the metrics to preserve stability and efficiency through gait transitions and phases. In turn, this validates the successful instillation of the animal gait strategies achieved in Chapter 5 within $\pi^{\text{uni}}_{G,\text{B1Z1}}$.

6.3.3 Dynamic Loco-manipulation Tasks on Different Terrain

With the LTA-M framework validated to successfully exhibit whole body motions, deploy a diverse set of gaits, and utilise animal gait strategies for optimal gait selection, only one question remains: how can this framework utilise these acquired proficiencies to complete tasks that mirror real-world scenarios, in turn demonstrating its versatility? To achieve this, two loco-manipulation tasks are designed. The first focuses on manipulation accuracy during locomotion, while the second focuses on stability during

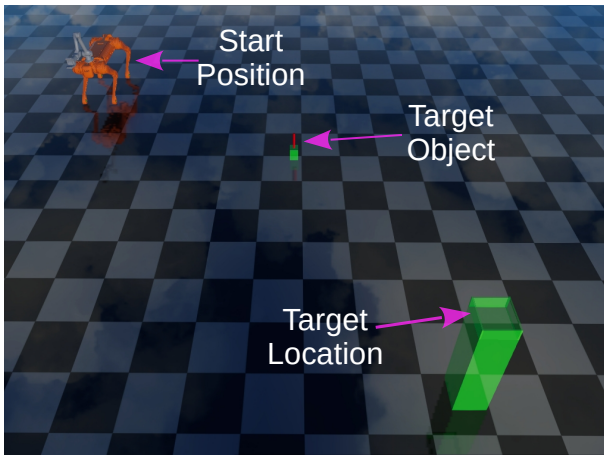


Figure 6.9: The layout of the pick-and-place experiment where from the starting position the robot is to pick up the target object and place it at the target location.

loco-manipulation tasks, with both being completed on two different terrains. To clarify, across all experiments no changes are made to the framework detailed in Fig. 6.2 and no retraining is completed to add the terrains in question to the trained observation space of π_{LM} nor $\pi^{\text{uni}}_{G,\text{B1Z1}}$; zero-shot deployment is achieved across all experiments.

Dynamic Pick-and-Place

To exhibit and assess the LTA-M framework's proficiency in terms of accuracy and its ability to fully utilise DoF redundancy to realise complex and dynamic motions, a pick-and-place task is set up as shown in Fig. 6.9. In this experiment, from the starting position, the robot's objective is to locomote to the target object, pick up the target object, locomote close to the target location, drop the target object at the target location, and then move clear of the area, all while exhibiting continuous locomotion to present an efficient dynamic pick-and-place task. The velocity command used is of a sinusoidal profile in both x and yaw, with amplitudes of 1.1 ms/ and 0.5 rad/s respectively, while the target object and target locations are at a height of 26% and 160% of the nominal robot base height, hence requiring the LTA-M framework to operate at its workspace limits; only through highly proficient redundancy utilisation can this task be completed successfully. It should be noted that between $t = 8.1$ s and $t = 9.9$ s only the yaw velocity component of the base velocity command is used to move the robot out of the way of the target location platform; hence at $t = 9.9$ s

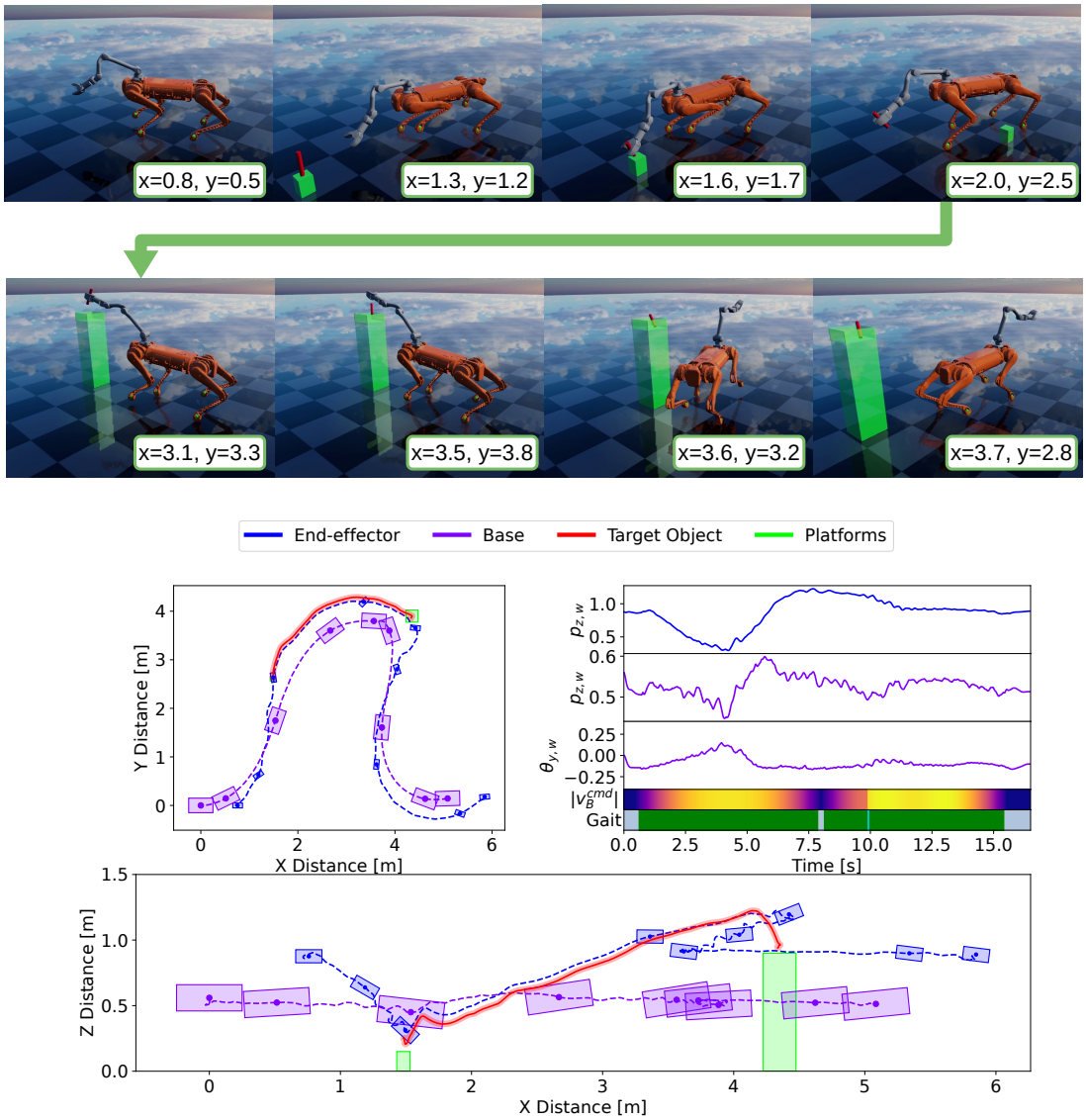


Figure 6.10: The collected data for the pick-and-place task on flat terrain. The top snapshots depict how the robot is able to leverage its redundancy to complete the task. The plots on the left and bottom detail the Cartesian position data of the robot base, end-effector, target object and platforms, while the plot on the right details the timeseries data throughout the experiment to provide an insight into the relationship between the motions of the end-effector and base and the relationship between the base velocity command the selected gait by $\pi^{\text{uni}}_{G,\text{B1Z1}}$. Yaw and pitch orientations are shown for the robot base and end-effector on the x-y and x-z Cartesian plots respectively with 1.8 s intervals as indicated by the boxes drawn at each point.

there is a point of high acceleration, the x component of the velocity command being reactivated. This experiment is completed on both flat and rough terrain, with data for these experiments being presented in Fig. 6.10 and Fig. 6.11 respectively and the rough terrain being generated through fractal noise of a maximum amplitude of 35% of the nominal base height.

As depicted by the data and experiment snapshots presented in Fig. 6.10, the framework is able to generate whole-body motions to reach down to the target object and up to the target location, as presented by the pitch and height of the base throughout the experiment, while still completing a smooth base velocity profile despite the complexity of the generated motions. This method of expanding the arm's workspace through adjusting the pose of the robot's base during locomotion is evident in the time series plots, whereas the end-effector decreases in height to reach the target object the base pitches down and reduces its height to enable this. The inverse is also realised by the framework when the end-effector needs to reach up to the height of the target location, a height that would otherwise be unreachable without adjustments to the base's pose; this would otherwise result in an increased risk of singularities and missing the target location through poor reference tracking. In turn, this results in the successful grasping of the target object and placing it at the target location, all while continuously moving for time-efficient task completion. For the majority of the experiment, only a trotting gait is selected due to the simplicity of flat terrain and moderate base velocity command. The only exception is at $t = 9.9$ s, where, due to the high acceleration, a brief transition to a running gait is observed to preserve stability and efficiency.

Similar success is seen when completing the same experiment on rough terrain, as shown in the Cartesian data and snapshots of Fig. 6.11. Although the robot is subject to increased disturbances, the robustness, adaptability and stability of the framework ensure the success of the task even when encountering a terrain never seen during training. This is supported by the time series data in Fig. 6.11, where it can be observed that through continuous adaption of the base height and pitch, the end-effector is able to maintain a relatively smooth trajectory. This proficiency is a direct result of the IK-WBC being able to realise the locomotion commands of π_{LM} to the required accuracy while also ensuring that the end-effector is achieving its target trajectory, even when encountering significant disturbances. In terms of the autonomous gait selection of π_G^{uni}, there is still limited gait transition events due to the moderate velocity command, with

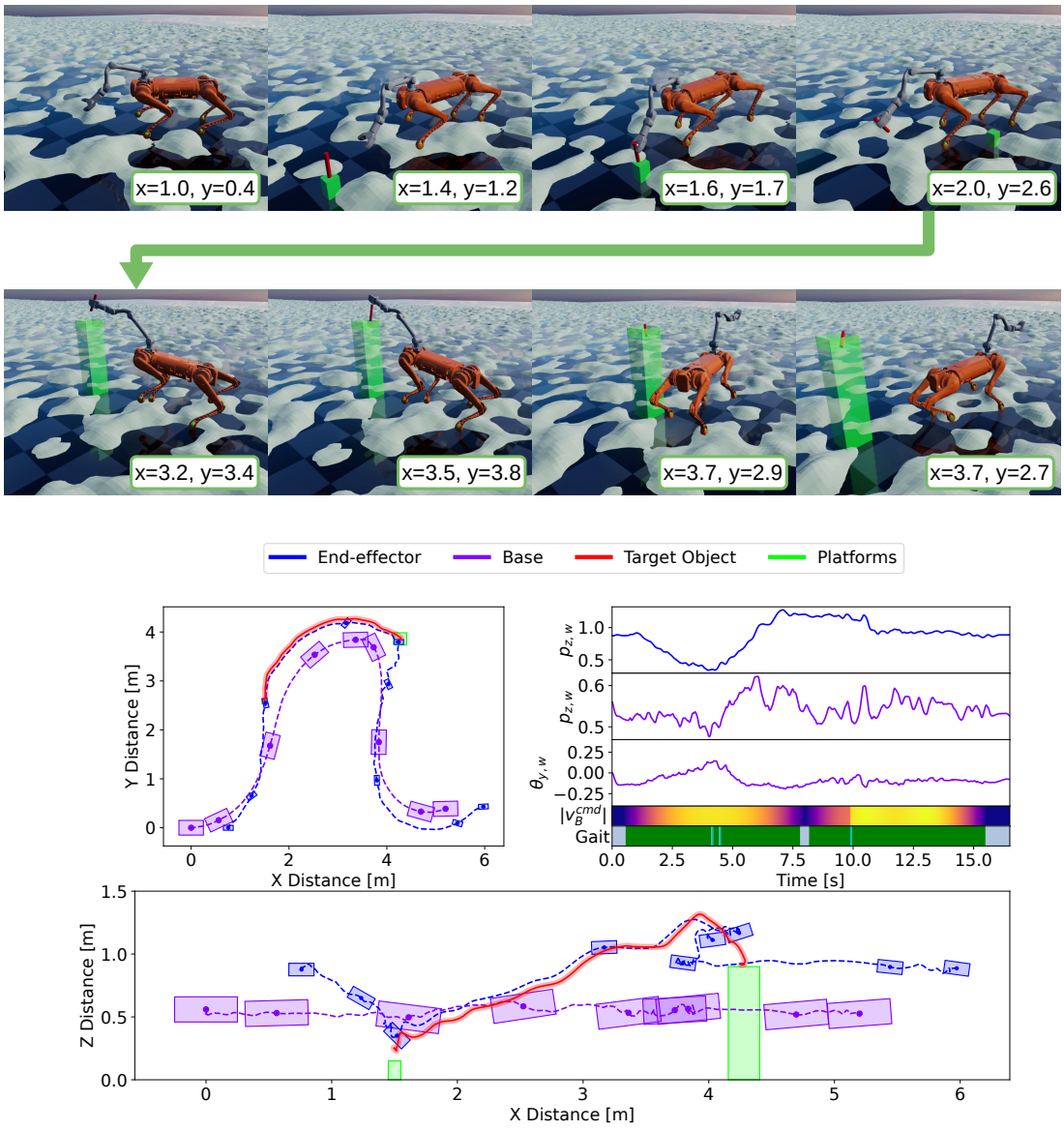


Figure 6.11: This figure details the same experiments and insights to that of Fig. 6.10 but on rough terrain with a maximum terrain height of 35% of the nominal base height.

the same transition seen that was present on flat terrain during the high acceleration region. However, there are several additional transitions between trot and run gaits at the first peak of the velocity command, where the end-effector is also at its lowest position in z. Through these transitions, the robot would have gained the stability required to locomote at this relatively high speed while completing a challenging end-effector trajectory and simultaneously overcoming rough terrain.

Overall, from the two experiments completed in Fig. 6.10 and Fig. 6.11, whether on flat or rough terrain through the combination of π_{LM} and the IK-WBC the framework is able to complete complex and dynamic loco-manipulation tasks on simple and challenging terrains without losing proficiency; on flat terrain an average end-effector tracking error of 0.043 m and 0.106 rad was observed, while on rough terrain an average of 0.04 m and 0.1 rad was observed.

Clearing the Path of Obstacles

To thoroughly test and evaluate the stability performance of the framework during loco-manipulation tasks and reject large external disturbances, an experiment is designed in which the robot is commanded to locomote with a constant x velocity of 1.5 m/s while moving increasingly heavy blocks out of its path. The setup of this experiment is presented in Fig. 6.12, where the weight of the blocks increases from 1 kg, to 5 kg, and finally 10 kg.

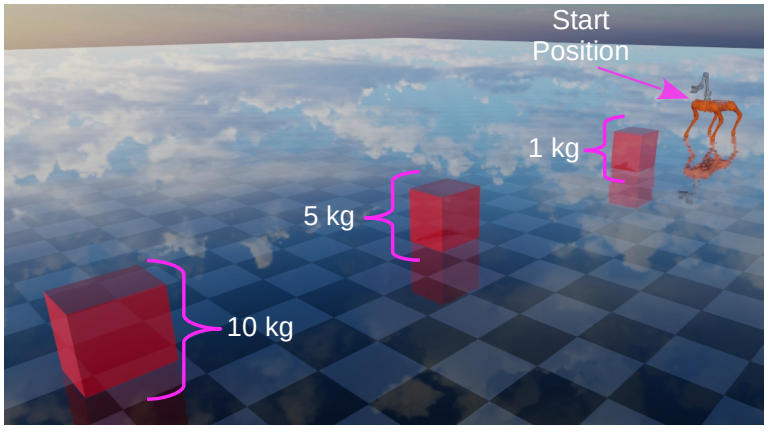


Figure 6.12: The layout of the path-clearing experiment where from the starting position the robot is commanded to locomote forward while clearing the path of the box obstacles of increasing weight from 1 kg to 10 kg.

To assess the framework's ability to complete this task across different challenging terrains, this experiment is completed on low-friction flat terrain with a friction coefficient of 0.2 (simulates the contact between rubber and ice at $-2°$C [167]) and rough terrain with a maximum height 54% of the nominal base height, for which data is presented in Fig. 6.13 and Fig. 6.14 respectively.

Additionally, to highlight the effectiveness of the π_G^{uni} policy's ability to maintain stability during these stability-critical tasks, the experiment is repeated but using a fixed trot gait. It should be noted that, as a base *pitch* velocity is actually desirable within this task, the base velocity error, ω_B^{err}, is updated to be calculated from base *roll*, *y*, and *yaw* velocities.

Across both experiments, the framework is able to complete the task successfully, clearing the obstacles from the path while maintaining stability. During both experiments, π_G^{uni} initiates frequent gait transitions between trotting and running to preserve stability and efficiency, with slightly more transition events occurring on rough terrain due to the increased instability encountered on this terrain (as reflected in higher $c_{\text{avg}}^{\text{err}}$ being observed on rough terrain as detailed in Table 6.2). Additionally, when the robot reaches the heaviest block, in both experiments π_G^{uni} utilises the auxiliary gaits to recover from states of critical stability; on the low-friction terrain, limping, and hopping are used, while on rough terrain just limping is used but ultimately both scenarios result in stability being preserved.

Table 6.2: The average gait metric values for the experiments detailed in Fig. 6.13 and Fig. 6.14 to compare the performance of using π_G^{uni} for autonomous gait selection compared to a fixed trotting gait.

Gait Selection Method	Terrain	CoT	W_{ext} [J]	$\tau_{\%}$ [%]	$c_{\text{avg}}^{\text{err}}$ [%]	ω_B^{err} [rad/s]
π_G^{uni}	Low-friction	6.53	395	28.3	10.1	0.68
Fixed Trot	(Fig. 6.13)	7.78	400	29.9	12.2	0.71
π_G^{uni}	Rough Terrain	6.95	404	30.1	16.7	0.73
Fixed Trot	(Fig. 6.14)	8.25	408	31.7	18.3	0.76

To demonstrate that stability would have otherwise been lost without the utilisation of the auxiliary gaits, these experiments are run again, but using only a fixed trotting

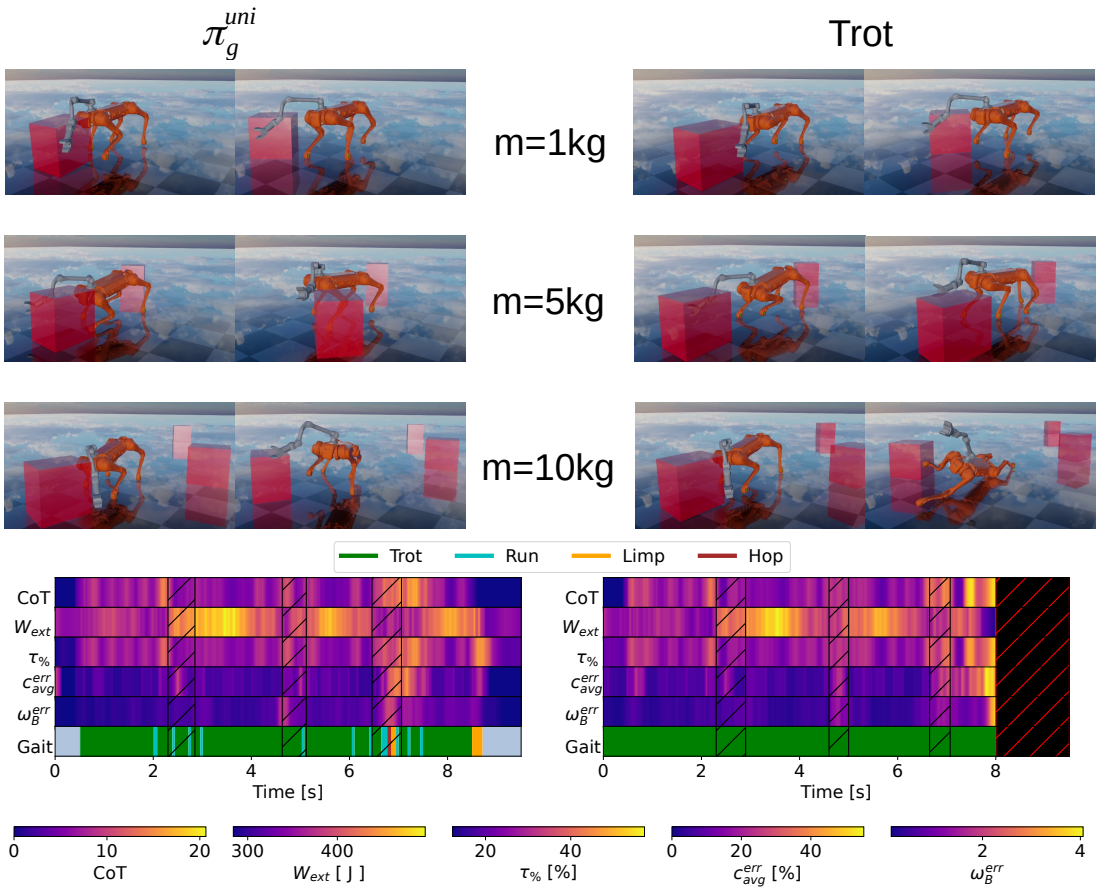


Figure 6.13: Plots and snapshots of the path-clearing experiment on low-friction terrain, with the autonomous gait selection data on the left and fixed trot gait on the right. The heat map plots focus on the same gait metrics seen in Fig. 5.11, with the transparent hatched areas indicating the duration which the end-effector is in contact with a block, while the blacked-out hatched region of the trot heat maps indicate where the robot failed.

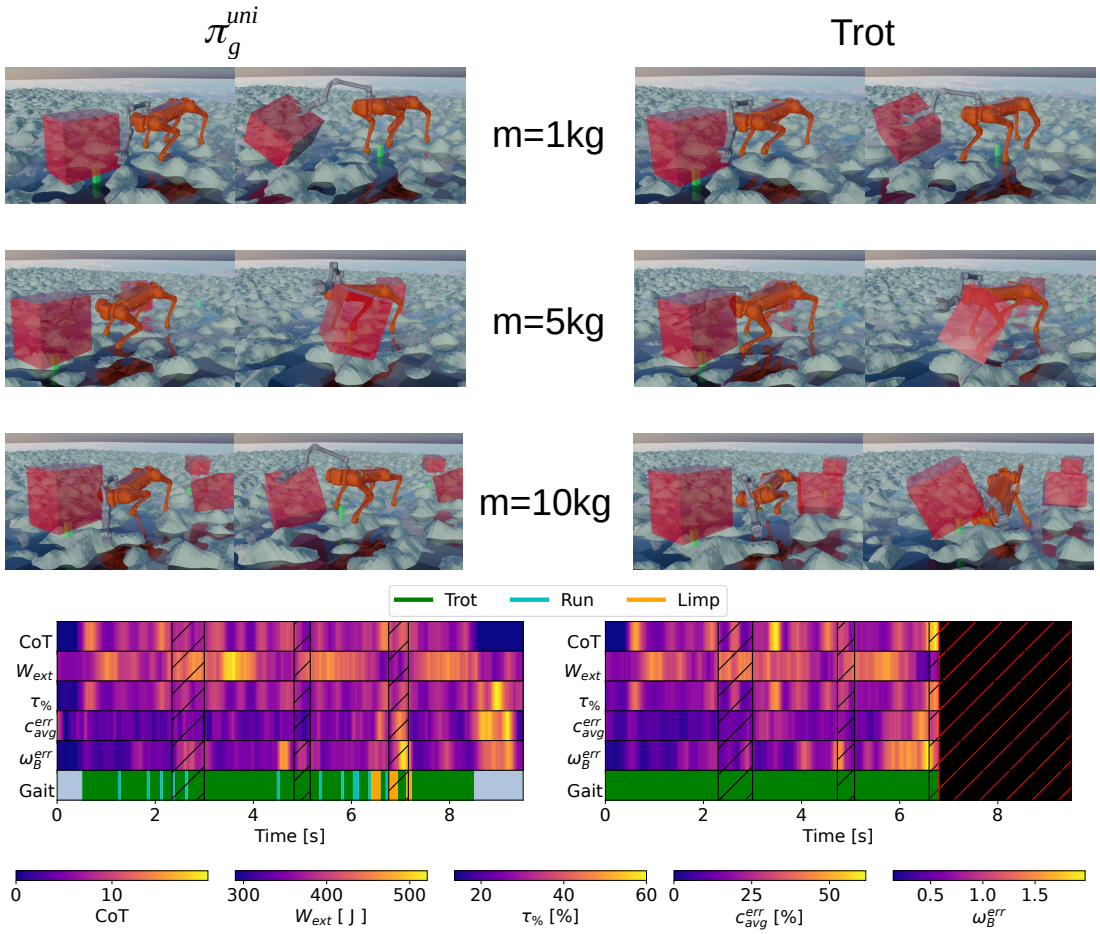


Figure 6.14: This figure details the same experiments and insights to that of Fig. 6.13 but on rough terrain with a maximum terrain height of 54% of the nominal base height. A numerical comparison between rough and low-friction terrain performance is detailed in Table 6.2.

gait. As presented in both Fig. 6.13 and Fig. 6.14, shortly after or during contact between the end-effector and the heaviest block, the robot is seen to fail when only using the fixed trotting gait. Furthermore, even before this point of failure, through the gait transitions seen throughout the experiments, π_G^{uni} enables the framework to unanimously improve metric performance in comparison to using a fixed trotting gait as detailed in Table 6.2. In turn, this demonstrates that through the optimal gait selection of π_G^{uni}, stability and efficiency can be effectively preserved in challenging loco-manipulation tasks. Furthermore, similar to the findings in Chapter 5, it appears that the same metrics spikes in $\tau_{\%}$ and $c_{\text{avg}}^{\text{err}}$ preempt and trigger a transition either between trotting and running to maintain nominal stability and efficiency, while also switching to the auxiliary gaits during critical stability events. In turn, this further validates the successful transfer of the bio-inspired gait strategies gained in Chapter 5 to the final framework for improved proficiency in loco-manipulation tasks.

6.4 Discussion

From the extensive suite of experiments completed in Section 6.3, it can be determined that the pairing of π_{LM} and the IK-WBC successfully enables the full utilisation of the robot's redundancy through whole-body motions to complete complex and challenging loco-manipulation tasks while utilising a diverse set of gaits that enables π_G^{uni} to leverage these gaits to preserve stability and efficiency. This exhibited proficiency is a direct result of how π_{LM} and the IK-WBC interact; π_{LM} is able to generate the joint commands that generate all learnt gaits and the adjustment term $\Delta \boldsymbol{p}^{\text{ref}}$ to overcome SE error build up, while the IK-WBC realises these references while respecting the mechanical limits of the robot and modulating the motions of the robot to optimise between achieving reference tracking accuracy and feasible motion, in turn generating safe, stable, and refined control. In turn, based on the success of these experiments and the demonstrated framework proficiency, it is clearly evident that the control modules developed across chapters 3 through 5 have been successfully unified while preserving and complementing their respective proficiencies.

Additionally, as outlined in Chapter 1, for legged manipulators to find extensive use and application within the real-world, they must be able to overcome and adapt to complex terrain. Through the rough and low-friction experiments completed in Section 6.3, not only does the framework demonstrate its ability to adapt to these terrains

never observed during training, but it also does so while completing the task and maintaining the same proficiencies that each control module provides as originally seen on flat terrain. In turn, this demonstrates that this developed framework lays a strong foundation towards enabling legged manipulators to be used in real-world applications and achieving project Objective 4.

However, despite the exhibited proficiency of the final framework, some aspects remain unresolved. One limitation is that although many aspects of the final control framework have been tested on hardware, the final form of the framework has not been due to time and resource limitations of the project, leaving sim-to-real success ambiguous. That being said, the success of all previous chapters in their hardware validation, and differences between the simulation and hardware in this final framework being the same as its predecessors, provide strong confidence that the framework can achieve successful sim-to-real transfer with minimal amendments. This assertion is attributed to three core factors:

- All observations and states are derived from simulated IMU, joint, and foot contact sensors, which include realistic simulated sensor noise. This method of enabling sim-to-real transfer has been verified in Chapter 5.
- The framework's ability to adapt to the immediate environment, irrespective of if the environment is been previously observed, is due to the online foot reference adaption of the BGS and automatic gait selection policy π_G^{uni}. The proficiency of this adaption to the real-world has been verified in Chapter 5.
- The ability of the IK-WBC to fully leverage robot redundancy for whole-body motions has been validated to successfully transfer to the real robot in Chapter 3, even without the additional aid from π_{LM}, which provides whole-body dynamics knowledge to an otherwise dynamics-naive controller.

Accounting for these factors, this provides strong confidence in the framework's ability to be successfully deployed on hardware. Another current limitation is that for successful loco-manipulation tasks to be completed, the framework relies on manual input command trajectories; the framework relies on the user to know where the targets are in space and direct the robot accordingly. However, integrating a system that combines computer vision and trajectory planning to enable autonomous loco-manipulation tasks, such as the one presented in [168], should be a simple process due to the hierarchical

structure of the framework; with the framework having an operational frequency 5 times that of [168], hence computation efficiency not presenting an operation bottleneck, all that is required is to generate $\boldsymbol{U}^{\text{cmd}}$ and $\boldsymbol{M}^{\text{cmd}}$ through this computer vision-planner system. It should be noted that in high-risk applications, this limitation could also be resolved through applying the teleoperation framework developed in Chapter 3 where a human-in-the-loop is critical for task accountability.

A final limitation that has carried over from Chapter 5, is that currently Froude-free motions, such as climbing stairs, have not been targeted by this framework; without a vision system or terrain height information, traversing over such terrain would be highly unstable. However, as outlined in Section 5.4.3, adding Froude-free locomotion capabilities is just a case of integrating this framework into a Froude-free focused framework, such as [100], to replace their Froude-characterised locomotion method. A theoretical combined framework has a high probability of being feasible, as not only can the framework developed in this chapter run at 10 times the frequency of the locomotion controller in [100], but it is highly unlikely that loco-manipulation tasks will be required during Froude-free locomotion, and consequently this limitation will not affect the Froude-free components of this theoretical framework. Overall, despite these discussed limitations of the final framework, the further development required to mitigate them is not only possible but is highly feasible due to the proficiency of the methods used in its development, its efficiency, and its modular architecture.

6.5 Chapter Conclusion

Within this chapter, all control modules developed in prior chapters are successfully unified to form a highly efficient and adaptable LTA-M framework for legged manipulators. The proficiency of this LTA-M framework, and the preservation of the proficiencies gained in previous chapters, was exhaustively tested and assessed in a series of loco-manipulation experiments. The results collected from these experiments clearly exhibited how not only is the framework capable of completing complex and dynamic loco-manipulation tasks, but also that this performance is preserved across a range of challenging terrains, in turn achieving project Objective 4. Indeed, there remain several limitations to the framework, namely the lack of hardware validation due to project constraints. Yet, the path to overcoming these limitations is perfectly feasible due to the extensive hardware validation seen in previous chapters, the same sim-to-real meth-

ods being carried forward, and the modular hierarchical architecture of the framework (enabling its integration with other technologies). With the final framework developed and its performance analysed, all that remains is to determine the extent to which the project objectives have been met, resolve any remaining ambiguities, and propose clear directions for future work, all of which will be covered in the next and final chapter.

Chapter 7

General Discussion and Conclusions

This chapter serves to bring the findings of the project together, analyse how well the project objectives have been met, and assess how the developed control framework serves to further the field's overarching objective of deploying robotic solutions in real-world applications. This chapter and the project will conclude with proposing several routes for future work and summarises with some concluding remarks.

7.1 Assessment of Research Objectives

As outlined in Section 1.3, the success of the project is measured through the extent to which the project's objectives have been met. The degree to which this has been achieved is detailed as follows, with each main objective being denoted by its number and its sub-objective by its alphanumeric character:

1. *Complete an extensive literature review on current state-of-the-art legged robot control frameworks, techniques and relevant technologies, along with the relevant findings within biomechanics that uncover the intrinsics of terrestrial animal locomotion strategies and behaviour.*

In Chapter 2, a detailed and critical literature review was completed, covering optimal control methods (*1.a*), DRL methods and control modules (*1.b*), gait schedulers and state estimators (*1.c*), and the relevant findings within biomechanics that have potential to augment legged robot control (*1.d*). The findings of this literature review were critically analysed to identify how this project could advance state-of-the-art legged robot control and develop the first draft of the loco-manipulation framework,

hence structuring its development moving forward.

2. *Develop a WBC for a legged manipulator with an attached arm to facilitate safe and dexterous manipulation.*

In Chapter 3, a generic IK-WBC we developed, featuring a range of dynamically interchangeable and modular tasks and constraints, that enables the control of a legged manipulator while fully leveraging its redundancy (*2.a*). To test and evaluate the IK-WBC, a teleoperation control framework was developed to generate the input commands and trajectories (*2.b*), which in turn was used to control the IK-WBC in a set of complex manipulation and basic locomotion tasks in simulation and on hardware (*2.c*). The success of these experiments validated the IK-WBC's proficiency and suitablility to be later integrated within the loco-manipulation framework.

3. *Develop a bio-inspired locomotion controller for optimal and adaptable locomotion.*

Initial work on this objective is completed in Chapter 4, in which the initial version of the BGS is developed from findings within biomechanics to generate stable, efficient and automatic gait transitions (*3.a, 3.b*). This initial version of the BGS is then tested on hardware and validated to uphold the targeted animal gait transition strategy (*3.d, 3.e*). With the first version of the BGS complete, Chapter 5 not only augments the BGS to account for a wider range of gaits, but also develops the LTA framework (*3.c*). The LTA framework takes the relevant findings from biomechanics to instill high-level animal locomotion proficiency attributes within it (*3.a*). This is achieved through coupling the BGS with DRL locomotion controller π_L^{bio} for adaptive motion adjustments, the BGS acts as pseudo gait procedural memory, and through training π_G^{uni} by utlising metrics that characterise animal gait strategies this same stategies are instilled within the framework for optimal gait seleciton (*3.b*). The LTA framework is then deployed in a range of simulation and hardware experiments to evaluate its performance in challenging environments (*3.d*) and to investigate and verify that the targeted animal proficiency attributes have been successfully instilled within the framework (*3e.*).

4. *Develop a loco-manipulation framework that unifies both WBC and bio-inspired locomotion controllers to simultaneously exhibit the proficiencies of all controllers.*

In Chapter 6, all control modules developed in previous chapters are brought together and unified to formulate the final loco-manipulation control framework (*4.a*).

For successful unification and broadening of several modules' operational scope to include manipulation, the relevant augmentation of the control modules was completed (*4.b*). Specifically, the IK-WBC had its tasks and constraints modified to accommodate the realisation of locomotion from the input reference trajectories of π_{LM}, while π_{LM} was adapted directly from π_L^{bio} to accommodate generating $\Delta \boldsymbol{p}^{\text{ref}}$ alongside locomotion suitable for gait generation while completing manipulation tasks (*4.b*). To evaluate the performance of the loco-manipulation framework, and if indeed the past proficiencies of each control module is preserved after unification, a set of complex loco-manipulation experiments were completed on a variety of challenging terrains (*4.c, 4.d*).

7.2 General Discussion

With quadrupedal legged manipulators (QLMs) being capable of dynamic locomotion and dexterous manipulation, they possess high potential in replacing humans in hazardous and laborious jobs. However, for effective deployment of QLMs in such applications, harnessing their natural versatility and innate utility is paramount. Recalling the criteria for effective deployment from Chapter 1:

1. Exhibit optimal locomotion in terms of efficiency and stability through seamlessly adapting to the immediate environment.

2. Able to interact with the environment to complete complex manipulation tasks, facilitated through effective deployment of the system's utility.

3. Successfully undertake hybrid loco-manipulation tasks while exhibiting the same proficiencies detailed above.

As unveiled in Chapter 2, existing control frameworks for legged robots are unable to achieve all three criteria; existing control methods, modules, and architectures tend to excel at specific criteria yet fail to achieve all three. End-to-end DRL frameworks are able to successfully complete dynamic loco-manipulation tasks yet struggle to generalise to complex or unobserved environments [13, 34], while hierarchical optimal control frameworks exhibit accurate manipulation during locomotion but are limited to tasks of simple dynamics due to model inaccuracies [41, 42]. Furthermore, both framework types are unable to exhibit optimal locomotion in terms of stability and efficiency due

to their limited deployable skill set; often only a trotting gait is deployable [100] or no methods of optimal skill selection is present [49].

In proposing a method of overcoming the limitations of past works and realising the defined criteria (and the project objectives derived from them), Chapter 2 presents a multifaceted approach. Not only should the framework take on a hierarchical architecture to refine the operational scope of each control module, but it should combine bio-inspired, DRL policies, and optimal control methods. This results in a framework that can leverage the strengths of DRL and optimal control methods to offset each other's limitations, while utilising the findings from biomechanics to instill animal locomotion proficiencies for adaptable and optimal performance. This lead to the creation of the draft architecture presented in Fig. 2.12, in turn setting out the control modules to be developed in this project.

The IK-WBC was developed in Chapter 3, which through its highly generic formulation enables dynamic interchangeable tasks and constraints, in turn facilitating the full operation over the robot's utility. The generality that the IK-WBC provides enables it to adjust and interface with the upstream DRL loco-manipulation policy despite being developed in advance. Furthermore, it offers a significantly higher level of controllability and modularity than existing WBCs, where their formulation is static and provides limited utility [42, 96]. Additionally, a teleoperation framework is also developed in Chapter 3. In the frame of this project, its role is solely to facilitate the testing of the IK-WBC, however it not only enables intuitive control of highly complex systems, it also provides a level of granular control and task orchestration that is unmatched by existing teloperation frameworks [75, 81].

Initial work in the development of the bio-inspired gait scheduler (BGS) was conducted in Chapter 4, in which a gait scheduler is augmented to generate bio-inspired gait transitions. Through leveraging the Froude number and the same energy preserving strategy animals exhibit for gait selection, the BGS is able to generate stable, efficient, and generic autonomous gait transitions. This method significantly outperforms existing gait generation modules in terms of generality, resource efficiency, and independence [67, 66] due to the generality the Froude number provides when leveraged to govern gait transitions. In turn, this provided initial proof of the effectiveness of applying findings within biomechanics for improving legged robot control. However, with animal locomotion proficiency centered around preserving not only efficiency [116]

but also stability [121], further development of the BGS in Chapter 5 was required to expand its functional scope. This included enabling gait-to-gait agnostic transitions, expanding the set of deployable gaits, and online generation of swing foot trajectories based on the robot's state. These augmentations only further improved the proficiency of this gait scheduler compared to existing methods, forming the foundation of the exemplary gait adaptability that real-world deployment requires.

The locomotion and gait selection DRL policies were developed in Chapter 5, in which animal locomotion proficiencies are instilled within the control modules of the learning to adapt (LTA) framework (including the further development of the BGS). In reference to Fig. 5.1, proficiencies were broken down and allocated so that:

1. Animal gait transition strategies were instilled with the gait selection policy for optimal efficiency and stability, even in cases of critical stability loss.

2. Gait procedural memory was encapsulated by the BGS to enable a diverse set of gaits to be deployable and selected by the gait selection module to adapt to the immediate scenario.

3. Adaptive motion adjustment was achieved through the BGS passing online-generated gait references to the locomotion policy that adapts with the state of the robot, enabling natural adaption to previously unobserved terrains.

This bio-inspired approach resulted in significantly enhanced adaptability, efficiency, and stability. This gain in proficiency was so substantial that it enabled the LTA framework to zero-shot deploy on complex real-world terrains never previously observed, outperforming all similar existing works, as detailed in Table 5.4. Explicitly, no other framework is capable of simultaneously achieving all three animal locomotion proficiencies which are essential for proficient real-world deployment. In turn this also provided further evidence of the effectiveness of leveraging findings within biomechanics for enhanced legged robot control.

To assemble the final loco-manipulation framework, referred to as the learning to adapt with manipulation (LTA-M) framework, in Chapter 6 all developed control modules are augmented to enable their unification within the final framework presented in Fig. 6.2. What could have been a challenging task became relatively straightforward, as both the BGS and IK-WBC were designed to be highly generic; no additional changes were made to the BGS and the main change to the IK-WBC was simply

swapping out foot Cartesian tasks for joint tasks. The only other change was the conversion of locomotion policy to a loco-manipulation policy through adding an additional reward for end-effector task tracking and expanding the action space to include an end-effector adjustment term to mitigate IK-WBC model inaccuracies. Through extensive simulation-based testing, focusing on manipulation accuracy and stability, the LTA-M framework demonstrated the following proficiencies:

- A diverse set of gaits are deployable while completing whole-body motions to facilitate successful end-effector task tracking (Fig. 6.3 and Fig. 6.4).
- The loco-manipulation policy is able to offset the model inaccuracies of the IK-WBC with the adjustment term effectively across all gaits (Fig. 6.7).
- The animal gait strategies instilled within the gait selection policy, for optimal efficiency and stability, are preserved despite migrating from a quadruped to a QLM platform (Fig. 6.8).
- The framework is able to successfully complete accuracy- and stability-critical loco-manipulation tasks, where utilising optimal gait selection, adaptable motions, and whole-body motions are all simultaneously required for success (Figures 6.10 to 6.14).
- All aforementioned proficiencies are upheld when deployed on challenging terrain, including rough and low-friction terrain.

From achieving these proficiencies, not only does this validate the successful unification of the control modules, but also demonstrates the successful realisation of the criteria for successful QLM deployment for loco-manipulation tasks. This presents a significant landmark towards QLM deployment in real-world tasks through achieving a level of versatility not previously observed in existing frameworks. Explicitly, while end-to-end DRL [13, 34], hierarchical optimal control [41, 42], and hybrid frameworks [50, 52] remain limited in adaptability, dynamic performance, and operational scope respectively, the LTA-M framework developed in this project overcomes these limitations through unifying bio-inspiration, DRL, and optimal control.

However, the core question remains: is this LTA-M framework deployment-ready to use within real-world tasks? Despite presenting significant progress towards this goal, without hardware testing and validation this question remains uncertain. That being

said, as discussed within Section 6.4, the case for successful sim-to-real transfer of the framework is strong, as all control modules have been previously successfully deployed on hardware and demonstrated exceptional robustness. Providing further confidence, the training environment features the same sim-to-real transfer techniques that were used in the development of the LTA framework; this was implemented at the sensor level, hence this method is independent from robot morphology.

7.3 Future Work

This project has presented significant progress towards deploying QLMs in real-world applications, however there are still some validation steps that need to be taken before this can be achieved. Furthermore, there are several promising directions that could expand the functional scope of the framework, improve performance, or even involve applying the contributions of this project to different systems and applications.

7.3.1 Hardware Validation and Further Investigations

It has been established that to definitively conclude whether the developed framework could enable QLMs to be used in real-world applications hardware validation is required. In practical terms, with access to the B1Z1 robot this would be a relatively straight forward process of deployment with the same setup that was used in Chapter 5. For the experiments themselves, similar experiments to the those featured in Chapter 6 would be setup with the simulated challenging terrains substituted for complex and unpredictable natural terrain. Additional experiments could include those similar to that in Chapter 5 to induce states of critical instability, to further investigate if critical stability recovery is achievable during hardware deployment, and if the animal gait strategies are upheld. Another valuable experiment would be to decouple control between locomotion and manipulation to investigate the performance gain through utilising whole-body motions during loco-maniplation tasks. This would involve using a purely locomotion DRL policy for controlling the quadruped joints and the IK-WBC for controlling the arm, with the only shared state information being the base state.

Although there is strong evidence to suggest successful hardware deployment, one aspect that has limited supporting evidence of its successful transfer to hardware is the interaction between the DRL loco-manipulation policy and the IK-WBC. An IK-WBC

was selected due to its computational efficiency and ease of integration with the DRL policy, hypothesising that the DRL loco-manipulation policy will be able to mitigate the limitations of the IK-WBC. This holds true in simulation, however due to the chaotic nature of the real-world, without the WBC being able to account for robot dynamics itself this may impact real-world deployment performance. This would be particularly prevalent in manipulation accuracy as the DRL loco-manipulation policy has a reduced role in controlling the end-effector trajectory. If this does indeed prove to be an issue during hardware deployment, a potential solution could be to use an ID-WBC in place of the IK-WBC. However, this would also require a comparison study between these two controllers in loco-manipulation tasks to assess the performance gain or loss.

7.3.2 Expanding Locomotion Capabilities

Looking beyond the typical applications for QLMs, despite the exemplary performance of the framework in Froude-characterised locomotion, the current framework and system is not suitable for applications that require Froude-Free locomotion. Although Froude-free tasks are not the focus of this project and lie outside its scope, this does present a limitation in the functional scope of the framework. To resolve this limitation, this framework could be integrated with other works that focus on this type of locomotion, with the strongest candidate being [100] due to its modularity, to achieve both optimal Froude-characterised locomotion and proficient Froude-free locomotion.

Currently, the framework is able to seamlessly adapt to natural environments through utilising only interoceptive sensors, demonstrating its exemplary adaptability and robustness. To expand its capabilities, the QLM could be outfitted with exteroceptive sensors, such as LiDAR and stereo cameras, to inform the framework of the environment's features. In turn, this would allow the framework to not only adjust its behaviour reactively based on its internal state, but also preemptively to avoid hazards. For example, if there was a hole in the ground, the current framework would only react once already inside the hole, increasing the risk of failure. With exteroceptive data, the framework would be able to switch to either a pronking or bounding gait preemptively to ensure the the robot never comes into contact with the hole, hence completely mitigating the risk.

7.3.3 Autonomous Tasks

To control the developed LTA-M framework, either manual command trajectory design or online teleoperation is required. This is sufficient for many applications, particularly where the main goal is to remove humans from hazardous environments or where the tasks is highly repeatable. However, if the framework could be operated fully autonomously, this would expand its potential applications to include those with labour shortages, such as working on a factory floor. Through equipping the system with exteroceptive sensors and having the LTA-M framework handle motion generation, locomotion path planning [169] coupled with manipulation planning [170] modules can enable the LTA-M framework to operate autonomously. These planners are often formulated around training diffusion models via imitation learning which require high-quality training data for proficient real-world deployment. With the highest quality data being real-world data, this presents the the ideal opportunity to leverage the teleoperation framework developed in Chapter 3 to gather this data in complex loco-manipulation tasks; the utility and intuitive control it provides would be more than adequate in this application.

7.3.4 Expanding to Humanoids

While QLMs offer inherent stability, manipulation capabilities, and a rich set of locomotion skills, when it comes to bimanual tasks they often fall short. Even if carrying two arms, their morphology prevents optimal utilisation. Although less stable, humanoid robots offer a great solution for such tasks. However, the current adaptability of humanoid control frameworks remain limited in complex environments and optimal Froude-characterised locomotion. Specifically, current frameworks tend to only have one deployable gait [171], can only locomote on simple terrain [172], and do not account for optimal skill deployment [173]. In turn, as the developed framework provides the subject robotic system with enhanced adaptability and optimal gait deployment, this presents humanoid robot control as an ideal application to take full advantage of the proficiencies the LTA-M framework. Transferring from a quadruped to a biped locomotion may require a slight re-design of the animal gait transition strategies. However, not only do all Froude number-based equations and findings still hold true for bipeds

[109], the study of human biomechanics is a very well established field[1]. Some provoking studies have worked towards understanding the mechanics of nominal optimal gait selection [174, 175, 176], off-nominal gait utilisation (such as skipping) [177], and how humans leverage gait characteristics and arm swing to preserve stability [178, 179, 180]. With such a rich selection of studies to drawn from, the LTA-M has great potential in leveraging human gait strategies for the application of humanoid loco-manipulation control. In practical terms, the BGS would just need to be augmented for two foot gait generation, maintaining the same phased-based control method[2], and an additional Cartesian task would be added to the IK-WBC. Due to the highly generic nature of the LTA-M framework, this remains a very feasible set of augmentations to achieve.

7.4 Concluding Remarks

QLMs appear to be a prime candidate for replacing humans undertaking tasks in hazardous and laborious jobs, yet due to their complexity, current control frameworks are limited in their performance. Specifically, they fall short in simultaneously realising highly adaptive and optimal locomotion while completing tasks that require whole-body manipulation. Looking to the natural world, animals exhibit no such limitations, where they are seen to seamlessly adapt to new terrains and interact with their environment due to a set of locomotion proficiencies. This presents the ideal opportunity to instil animal gait strategies within a learning-based loco-manipulation framework for enhanced adaptability and optimal gait selection.

This project aimed to do exactly this through the development of a gait selection DRL policy for bio-inspired optimal gait selection, a BGS for gait memory and stable gait transitions, a loco-manipulation DRL policy for adaptive motion adjustments, and an IK-WBC that realises whole-body manipulation and refines the output of the upstream DRL policies. In terms of locomotion, the LTA framework has demonstrated this adaptability and efficiency through zero-shot deployment on challenging real-world terrains, while the IK-WBC has also been deployed on hardware to complete complex manipulation tasks. Through the unification of all developed control modules within the loco-manipulation framework, LTA-M, it was able to undertake a suite of dynamic

[1] As instructions can be provided to humans and they are easier to outfit with sensors, more complex studies of the mechanics and strategies of human locomotion can be completed.

[2] Arm swing could also be controlled via this method too.

loco-manipulation tasks on a variety of challenging terrains, all while preserving the respective proficiencies of each control module. Although the LTA-M framework has yet to be deployed on hardware, there is strong supporting evidence for this to be successful. Reflecting on the proficiencies of the LTA-M framework presents a diverse set of future research directions and applications that could bring this framework closer to real-world application deployment.

References

[1] J.-S. Fiset, M. Effati, and K. Skonieczny, "Power and energy consumption of skid-steer rovers turning on loose soil," *Journal of Field Robotics*, vol. 39, no. 9, 2022.

[2] Behind the Robot: HITTs Construction Site Monitoring Husky UGV. RoboticsTomorrow. [Online]. Available: https://www.roboticstomorrow.com/article/2021/10/behind-the-robot-hitts-construction-site-monitoring-husky-ugv/17628

[3] F. Jenelten, T. Miki, A. E. Vijayan, M. Bjelonic, and M. Hutter, "Perceptive Locomotion in Rough Terrain - Online Foothold Optimization," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5370–5376, 2020.

[4] Q. Li, Q. Meng, Y. Qin, J. Chen, X. Ding, and K. Xu, "Dynamic Interaction Control in Legged Mobile Manipulators: A Decoupled Approach," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 1406–1412.

[5] N. A. Curtin, H. L. Bartlam-Brooks, T. Y. Hubel, J. C. Lowe, A. R. Gardner-Medwin, E. Bennitt, S. J. Amos, M. Lorenc, T. G. West, and A. M. Wilson, "Remarkable muscles, remarkable locomotion in desert-dwelling wildebeest," *Nature*, vol. 563, no. 7731, pp. 393–396, 2018.

[6] T. Y. Hubel, K. A. Golabek, K. Rafiq, J. W. McNutt, and A. M. Wilson, "Movement patterns and athletic performance of leopards in the okavango delta," *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1877, p. 20172622, 2018.

[7] S. Wilshin, M. A. Reeve, G. C. Haynes, S. Revzen, D. E. Koditschek, and A. J. Spence, "Longitudinal quasi-static stability predicts changes in dog gait on rough terrain," *Journal of Experimental Biology*, vol. 220, no. 10, pp. 1864–1874, 05 2017.

[8] S. Zimmermann, R. Poranne, and S. Coros, "Go Fetch! - Dynamic Grasps using Boston Dynamics Spot with External Robotic Arm," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4488–4494.

[9] J.-R. Chiu, J.-P. Sleiman, M. Mittal, F. Farshidian, and M. Hutter, "A Collision-Free MPC for Whole-Body Dynamic Locomotion and Manipulation," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 4686–4693.

[10] M. Aractingi, P.-A. Léziart, T. Flayols, J. Perez, T. Silander, and P. Souères, "Controlling the solo12 quadruped robot with deep reinforcement learning," *scientific Reports*, vol. 13, no. 1, p. 11945, 2023.

[11] X. Zhang, Z. Xiao, Q. Zhang, and W. Pan, "SYNLOCO: Synthesizing central pattern generator and reinforcement learning for quadruped locomotion," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 2640–2645.

[12] G. Bledt, M. J. Powell, B. Katz, J. Di Carlo, P. M. Wensing, and S. Kim, "MIT Cheetah 3: Design and Control of a Robust, Dynamic Quadruped Robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2245–2252.

[13] Z. Fu, X. Cheng, and D. Pathak, "Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion," in *6th Annual Conference on Robot Learning*, 2022. [Online]. Available: https://openreview.net/forum?id=zldI4UpuG7v

[14] N. Kashiri, L. Baccelliere, L. Muratore, A. Laurenzi, Z. Ren, E. M. Hoffman, M. Kamedula, G. F. Rigano, J. Malzahn, S. Cordasco, P. Guria, A. Margan, and N. G. Tsagarakis, "CENTAURO: A Hybrid Locomotion and High Power Resilient Manipulation Platform," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1595–1602, 2019.

[15] Y. Gong, R. Hartley, X. Da, A. Hereid, O. Harib, J.-K. Huang, and J. Grizzle, "Feedback Control of a Cassie Bipedal Robot: Walking, Standing, and Riding a Segway," in *2019 American Control Conference (ACC)*, 2019, pp. 4559–4566.

[16] M. Schwartz, J. Sim, J. Ahn, S. Hwang, Y. Lee, and J. Park, "Design of the Humanoid Robot TOCABI," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, 2022, pp. 322–329.

[17] J. Lee, M. Bjelonic, A. Reske, L. Wellhausen, T. Miki, and M. Hutter, "Learning robust autonomous navigation and locomotion for wheeled-legged robots," *Science Robotics*, vol. 9, no. 89, p. eadi9641, 2024.

[18] L. Amatucci, G. Turrisi, A. Bratta, V. Barasuol, and C. Semini, "Accelerating model predictive control for legged robots through distributed optimization," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 12 734–12 741.

[19] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine, "Neural Network Dynamics for Model-Based Deep Reinforcement Learning with Model-Free Fine-Tuning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 7559–7566.

[20] P. Wu, A. Escontrela, D. Hafner, K. Goldberg, and P. Abbeel, "DayDreamer: World Models for Physical Robot Learning," 2022.

[21] Y. Yang, K. Caluwaerts, A. Iscen, T. Zhang, J. Tan, and V. Sindhwani, "Data Efficient Reinforcement Learning for Legged Robots," in *Proceedings of the Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, L. P. Kaelbling, D. Kragic, and K. Sugiura, Eds., vol. 100. PMLR, 11 2020, pp. 1–10.

[22] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. Pieter Abbeel, and W. Zaremba, "Hindsight experience replay," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[23] C. Szepesvári, "Algorithms for reinforcement learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 4, no. 1, pp. 1–103, 2010.

[24] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," 2017.

[25] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015.

[26] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, "Trust Region Policy Optimization," 2015. [Online]. Available: https://arxiv.org/abs/1502.05477

[27] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 7 2018, pp. 1861–1870.

[28] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 7 2018, pp. 1587–1596.

[29] Y. Jin, X. Liu, Y. Shao, H. Wang, and W. Yang, "High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning," *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1198–1208, 2022.

[30] 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.

[31] 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.

[32] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Manyquadrupeds: Learning a single locomotion policy for diverse quadruped robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 3471–3477.

[33] V. Atanassov, J. Ding, J. Kober, I. Havoutis, and C. Della Santina, "Curriculum-based reinforcement learning for quadrupedal jumping: A reference-free design," *IEEE Robotics & Automation Magazine*, 2024.

[34] G. Pan, Q. Ben, Z. Yuan, G. Jiang, Y. Ji, S. Li, J. Pang, H. Liu, and H. Xu, "RoboDuet: Learning a Cooperative Policy for Whole-Body Legged Loco-Manipulation," *IEEE Robotics and Automation Letters*, vol. 10, no. 5, pp. 4564–4571, 2025.

[35] D. Kang, S. Zimmermann, and S. Coros, "Animal Gaits on Quadrupedal Robots Using Motion Matching and Model-Based Control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8500–8507.

[36] C. D. Bellicoso, F. Jenelten, C. Gehring, and M. Hutter, "Dynamic Locomotion Through Online Nonlinear Motion Optimization for Quadrupedal Robots," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2261–2268, 2018.

[37] M. Bjelonic, C. D. Bellicoso, Y. de Viragh, D. Sako, F. D. Tresoldi, F. Jenelten, and M. Hutter, "Keep Rollin' - Whole-Body Motion Control and Planning for Wheeled Quadrupedal Robots," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2116–2123, 2019.

[38] M. Bjelonic, R. Grandia, O. Harley, C. Galliard, S. Zimmermann, and M. Hutter, "Whole-Body MPC and Online Gait Sequence Generation for Wheeled-Legged Robots," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8388–8395.

[39] J.-P. Sleiman, F. Farshidian, M. V. Minniti, and M. Hutter, "A Unified MPC Framework for Whole-Body Dynamic Locomotion and Manipulation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4688–4695, 2021.

[40] Q. Nguyen, M. J. Powell, B. Katz, J. D. Carlo, and S. Kim, "Optimized Jumping on the MIT Cheetah 3 Robot," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 7448–7454.

[41] A. Xie, T. Chen, G. Zhang, Y. Li, and X. Rong, "Manipulability Enhancement of Legged Manipulators by Adaptive Motion Distribution," *IEEE Transactions on Industrial Electronics*, vol. 72, no. 1, pp. 724–733, 2025.

[42] J.-P. Sleiman, F. Farshidian, and M. Hutter, "Versatile multicontact planning and control for legged loco-manipulation," *Science Robotics*, vol. 8, no. 81, p. eadg5014, 2023.

[43] D. H. Kim and J. H. Park, "Reduced model predictive control toward highly dynamic quadruped locomotion," *IEEE Access*, vol. 12, pp. 20 003–20 018, 2024.

[44] B. Liu, F. Meng, S. Gu, X. Chen, Z. Yu, and Q. Huang, "Variational-Based Geometric Nonlinear Model Predictive Control for Robust Locomotion of Quadruped Robots," *IEEE Transactions on Automation Science and Engineering*, 2025.

[45] G. Romualdi, S. Dafarra, G. L'Erario, I. Sorrentino, S. Traversaro, and D. Pucci, "Online Non-linear Centroidal MPC for Humanoid Robot Locomotion with Step Adjustment," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 10 412–10 419.

[46] S. Gangapurwala, M. Geisert, R. Orsolino, M. Fallon, and I. Havoutis, "RLOC: Terrain-Aware Legged Locomotion Using Reinforcement Learning and Optimal Control," *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 2908–2927, 2022.

[47] J. Cheng, D. Kang, G. Fadini, G. Shi, and S. Coros, "RAMBO: RL-augmented Model-based Optimal Control for Whole-body Loco-manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2504.06662

[48] D. Kang, J. Cheng, M. Zamora, F. Zargarbashi, and S. Coros, "RL + Model-Based Control: Using On-Demand Optimal Control to Learn Versatile Legged Locomotion," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6619–6626, 2023.

[49] F. Jenelten, J. He, F. Farshidian, and M. Hutter, "DTC: Deep Tracking Control," *Science Robotics*, vol. 9, no. 86, p. eadh5401, 2024.

[50] H. Duan, J. Dao, K. Green, T. Apgar, A. Fern, and J. Hurst, "Learning Task Space Actions for Bipedal Locomotion," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1276–1282.

[51] Y. Ma, F. Farshidian, T. Miki, J. Lee, and M. Hutter, "Combining Learning-Based Locomotion Policy With Model-Based Manipulation for Legged Mobile Manipulators," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2377–2384, 2022.

[52] S. Lyu, H. Zhao, and D. Wang, "A Composite Control Strategy for Quadruped Robot by Integrating Reinforcement Learning and Model-Based Control," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 751–758.

[53] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, "Contact-aided invariant extended Kalman filtering for robot state estimation," *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.

[54] G. Fink and C. Semini, "Proprioceptive Sensor Fusion for Quadruped Robot State Estimation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 914–10 920.

[55] R. Grandia, F. Farshidian, E. Knoop, C. Schumacher, M. Hutter, and M. Bächer, "DOC: Differentiable Optimal Control for Retargeting Motions onto Legged Robots," *ACM Trans. Graph.*, vol. 42, no. 4, Jul. 2023.

[56] H. Li, W. Yu, T. Zhang, and P. M. Wensing, "Zero-Shot Retargeting of Learned Quadruped Locomotion Policies Using Hybrid Kinodynamic Model Predictive Control," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 971–11 977.

[57] T. Yoon, D. Kang, S. Kim, J. Cheng, M. Ahn, S. Coros, and S. Choi, "Spatio-Temporal Motion Retargeting for Quadruped Robots," *IEEE Transactions on Robotics*, pp. 1–20, 2025.

[58] S.-W. Lee, X. Kang, and Y.-L. Kuo, "Diff-Dagger: Uncertainty Estimation With Diffusion Policy for Robotic Manipulation," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 4845–4852.

[59] W. Liu, J. Wang, Y. Wang, W. Wang, and C. Lu, "ForceMimic: Force-Centric Imitation Learning with Force-Motion Capture System for Contact-Rich Manip-

ulation," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 1105–1112.

[60] X. Sun, J. Li, A. V. Kovalenko, W. Feng, and Y. Ou, "Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 3, pp. 1735–1744, 2023.

[61] D. Owaki and A. Ishiguro, "A Quadruped Robot Exhibiting Spontaneous Gait Transitions from Walking to Trotting to Galloping," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[62] T. Fukui, S. Matsukawa, Y. Habu, and Y. Fukuoka, "Gait Transition from Pacing by a Quadrupedal Simulated Model and Robot with Phase Modulation by Vestibular Feedback," *Robotics*, vol. 11, no. 1, 2022.

[63] A. Fukuhara, D. Owaki, T. Kano, R. Kobayashi, and A. Ishiguro, "Spontaneous gait transition to high-speed galloping by reconciliation between body support and propulsion," *Advanced Robotics*, vol. 32, no. 15, pp. 794–808, 2018.

[64] C. Boussema, M. J. Powell, G. Bledt, A. J. Ijspeert, P. M. Wensing, and S. Kim, "Online Gait Transitions and Disturbance Recovery for Legged Robots via the Feasible Impulse Set," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1611–1618, 2019.

[65] Y. Yang, T. Zhang, E. Coumans, J. Tan, and B. Boots, "Fast and Efficient Locomotion via Learned Gait Transitions," in *Conference on Robot Learning*, 2022, pp. 773–783.

[66] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing Energy Consumption Leads to the Emergence of Gaits in Legged Robots," in *Conference on Robot Learning (CoRL)*, 2021.

[67] C. Gehring, S. Coros, M. Hutter, M. Bloesch, M. A. Hoepflinger, and R. Siegwart, "Control of Dynamic Gaits for a Quadrupedal Robot," in *IEEE ICRA*, 2013, pp. 3287–3292.

[68] J. Wang, A. Laurenzi, and N. Tsagarakis, "Autonomous Behavior Planning For Humanoid Loco-manipulation Through Grounded Language Model," in *2024*

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024, pp. 10 856–10 863.

[69] H. Kolvenbach, D. Wisth, R. Buchanan, G. Valsecchi, R. Grandia, M. Fallon, and M. Hutter, "Towards autonomous inspection of concrete deterioration in sewers with legged robots," *Journal of Field Robotics*, vol. 37, no. 8, pp. 1314–1327, 2020.

[70] A. Coelho, Y. Sarkisov, X. Wu, H. Mishra, H. Singh, A. Dietrich, A. Franchi, K. Kondak, and C. Ott, "Whole-Body Teleoperation and Shared Control of Redundant Robots with Applications to Aerial Manipulation," *Journal of Intelligent & Robotic Systems*, vol. 102, 05 2021.

[71] G. Xin, J. Smith, D. Rytz, W. Wolfslag, H.-C. Lin, and M. Mistry, "Bounded haptic teleoperation of a quadruped robotâ€™s foot posture for sensing and manipulation," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 1431–1437.

[72] S. Wang and J. Ramos, "Dynamic Locomotion Teleoperation of a Reduced Model of a Wheeled Humanoid Robot Using a Whole-Body Human-Machine Interface," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1872–1879, 2022.

[73] H. Xing, L. Ding, H. Gao, W. Li, and M. Tavakoli, "Dual-User Haptic Teleoperation of Complementary Motions of a Redundant Wheeled Mobile Manipulator Considering Task Priority," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–13, 2022.

[74] K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, A. Chavez, and D. Pucci, "Whole-Body Geometric Retargeting for Humanoid Robots," in *IEEE-RAS International Conference on Humanoid Robots*, 2019, pp. 679–686.

[75] M. Risiglione, J.-P. Sleiman, M. V. Minniti, B. Ã‡izmeci, D. Dresscher, and M. Hutter, "Passivity-based control for haptic teleoperation of a legged manipulator in presence of time-delays," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021, pp. 5276–5281.

[76] T.-C. Lin, A. U. Krishnan, and Z. Li, "Physical Fatigue Analysis of Assistive Robot Teleoperation via Whole-body Motion Mapping," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019, pp. 2240–2245.

[77] C. Yang, J. Chen, and F. Chen, "Neural learning enhanced teleoperation control of baxter robot using IMU based Motion Capture," in *International Conference on Automation and Computing*, 2016, pp. 389–394.

[78] Y. Wu, P. Balatti, M. Lorenzini, F. Zhao, W. Kim, and A. Ajoudani, "A Teleoperation Interface for Loco-Manipulation Control of Mobile Collaborative Robotic Assistant," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3593–3600, 2019.

[79] K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. J. A. Chavez, and D. Pucci, "Whole-Body Geometric Retargeting for Humanoid Robots," in *IEEE-RAS International Conference on Humanoid Robots*, 2019, pp. 679–686.

[80] E. Dalin, I. Bergonzani, T. Anne, S. Ivaldi, and J.-B. Mouret, "Whole-body Teleoperation of the Talos Humanoid Robot: Preliminary Results," in *ICRA Workshop on Teleoperation of Dynamic Legged Robots in Real Scenarios*, 2021.

[81] L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, and S. Ivaldi, "A Multimode Teleoperation Framework for Humanoid Loco-Manipulation: An Application for the iCub Robot," *IEEE Robotics Automation Magazine*, vol. 26, no. 4, pp. 73–82, 2019.

[82] J. Li, H. Gao, Y. Wan, J. Humphreys, C. Peers, H. Yu, and C. Zhou, "Whole-Body Control for a Torque-Controlled Legged Mobile Manipulator," *Actuators*, vol. 11, no. 11, 2022.

[83] S. Xin and S. Vijayakumar, "Online Dynamic Motion Planning and Control for Wheeled Biped Robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 3892–3899.

[84] Y. You, S. Xin, C. Zhou, and N. Tsagarakis, "Straight Leg Walking Strategy for Torque-controlled Humanoid Robots," in *IEEE International Conference on Robotics and Biomimetics*, Qingdao, China, December 3–7 2016, pp. 2014–2019.

[85] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim, "The MIT Humanoid Robot: Design, Motion Planning, and Control for Acrobatic Behaviors," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots*, 2021, pp. 1–8.

[86] C. Dario Bellicoso, C. Gehring, J. Hwangbo, P. Fankhauser, and M. Hutter, "Perception-less terrain adaptation through whole body control and hierarchical optimization," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 558–564.

[87] C. Dario Bellicoso, F. Jenelten, P. Fankhauser, C. Gehring, J. Hwangbo, and M. Hutter, "Dynamic locomotion and whole-body control for quadrupedal robots," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3359–3365.

[88] C. D. Bellicoso, K. KrÃ¤mer, M. StÃ¤uble, D. Sako, F. Jenelten, M. Bjelonic, and M. Hutter, "ALMA - Articulated Locomotion and Manipulation for a Torque-Controllable Robot," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8477–8483.

[89] N. Ramuzat, G. Buondonno, S. Boria, and O. Stasse, "Comparison of Position and Torque Whole-Body Control Schemes on the Humanoid Robot TALOS," in *2021 20th International Conference on Advanced Robotics (ICAR)*, 2021, pp. 785–792.

[90] J. A. Castano, E. M. Hoffman, A. Laurenzi, L. Muratore, M. Karnedula, and N. G. Tsagarakis, "A Whole Body Attitude Stabilizer for Hybrid Wheeled-Legged Quadruped Robots," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 706–712.

[91] A. Heins, M. Jakob, and A. P. Schoellig, "Mobile Manipulation in Unknown Environments with Differential Inverse Kinematics Control," in *2021 18th Conference on Robots and Vision (CRV)*, 2021, pp. 64–71.

[92] P. D. Lillo, F. Pierri, F. Caccavale, and G. Antonelli, "Experiments on whole-body control of a dual-arm mobile robot with the set-based task-priority inverse kinematics algorithm," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 9096–9101.

[93] S. Caron, A. Kheddar, and O. Tempier, "Stair Climbing Stabilization of the HRP-4 Humanoid Robot using Whole-body Admittance Control," in *2019 International Conference on Robotics and Automation*, 2019, pp. 277–283.

[94] C. Zhou, C. Fang, X. Wang, Z. Li, and N. Tsagarakis, "A generic optimization-based framework for reactive collision avoidance in bipedal locomotion," in *IEEE International Conference on Automation Science and Engineering*, 2016, pp. 1026–1033.

[95] S. Yang, H. Chen, L. Zhang, Z. Cao, P. M. Wensing, Y. Liu, J. Pang, and W. Zhang, "Reachability-based Push Recovery for Humanoid Robots with Variable-Height Inverted Pendulum," in *2021 IEEE International Conference on Robotics and Automation*, 2021, pp. 3054–3060.

[96] M. Murooka, K. Chappellet, A. Tanguy, M. Benallegue, I. Kumagai, M. Morisawa, F. Kanehiro, and A. Kheddar, "Humanoid Loco-Manipulations Pattern Generation and Stabilization Control," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5597–5604, 2021.

[97] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, 1 2019.

[98] 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.

[99] 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.

[100] 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.

[101] G. Christmann, Y.-S. Luo, and W.-C. Chen, "Expert Composer Policy: Scalable Skill Repertoire for Quadruped Robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 9727–9734.

[102] G. Feng, H. Zhang, Z. Li, X. B. Peng, B. Basireddy, L. Yue, Z. SONG, L. Yang, Y. Liu, K. Sreenath, and S. Levine, "GenLoco: Generalized Locomotion Controllers for Quadrupedal Robots," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 1893–1903.

[103] Y. Shao, Y. Jin, X. Liu, W. He, H. Wang, and W. Yang, "Learning Free Gait Transition for Quadruped Robots Via Phase-Guided Controller," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1230–1237, 2022.

[104] G. B. Margolis and P. Agrawal, "Walk These Ways: Tuning Robot Control for Generalization with Multiplicity of Behavior," in *6th Annual Conference on Robot Learning*, 2022.

[105] J. Zhang, S. Heim, S. H. Jeon, and S. Kim, "Learning Emergent Gaits with Decentralized Phase Oscillators: on the role of Observations, Rewards, and Feedback," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 3426–3433.

[106] C. Vanden Hole, J. Goyens, S. Prims, E. Fransen, M. Ayuso Hernando, S. Van Cruchten, P. Aerts, and C. Van Ginneken, "How innate is locomotion in precocial animals? A study on the early development of spatio-temporal gait variables and gait symmetry in piglets," *Journal of Experimental Biology*, vol. 220, no. 15, pp. 2706–2716, 08 2017.

[107] E. Avital and E. Jablonka, *Animal traditions: Behavioural inheritance in evolution*. Cambridge University Press, 2000.

[108] A. N. Wimberly, G. J. Slater, and M. C. Granatosky, "Evolutionary history of quadrupedal walking gaits shows mammalian release from locomotor constraint," *Proceedings of the Royal Society B: Biological Sciences*, vol. 288, no. 1957, p. 20210937, 2021.

[109] R. M. Alexander, "The gaits of bipedal and quadrupedal animals," *Int J Rob Res*, vol. 3, no. 2, pp. 49–59, 1984.

[110] J. G. Fleagle and R. A. Mittermeier, "Locomotor behavior, body size, and comparative ecology of seven surinam monkeys," *American Journal of Physical Anthropology*, vol. 52, no. 3, pp. 301–314, 1980.

[111] H. Pontzer and R. W. Wrangham, "Climbing and the daily energy cost of locomotion in wild chimpanzees: implications for hominoid locomotor evolution," *Journal of Human Evolution*, vol. 46, no. 3, pp. 315–333, 2004.

[112] R. M. Alexander and A. S. Jayes, "A dynamic similarity hypothesis for the gaits of quadrupedal mammals," *J. Zool.*, vol. 201, no. 1, pp. 135–152, 1983.

[113] T. Griffin, R. Kram, S. J. Wickler, and D. F. Hoyt, "Biomechanical and energetic determinants of walk-trot transition in horses," *J Exp Biol*, vol. 207, no. 24, pp. 4215–4223, 2004.

[114] E. Muybridge, L. Brown, L. Brown, L. Brown, and A. History, *Animals in motion.* Dover Publications, 1957.

[115] F. Righini, M. Carpineti, F. Giavazzi, and A. Vailati, "Pronking and bounding allow a fast escape across a grassland populated by scattered obstacles," *Royal Society Open Science*, vol. 10, no. 9, p. 230587, 2023.

[116] F. J. Diedrich and W. H. W. Jr., "The Dynamics of Gait Transitions: Effects of Grade and Load," *Journal of Motor Behavior*, vol. 30, no. 1, pp. 60–78, 1998.

[117] S. J. Wickler, D. F. Hoyt, E. A. Cogger, and G. Myers, "The energetics of the trot-gallop transition," *Journal of Experimental Biology*, vol. 206, no. 9, pp. 1557–1564, 05 2003.

[118] C. T. Farley and C. R. Taylor, "A Mechanical Trigger for the Trot-Gallop Transition in Horses," *Science*, vol. 253, no. 5017, pp. 306–308, 1991.

[119] A. A. Biewener and C. R. Taylor, "Bone Strain: A Determinant of Gait and Speed?" *Journal of Experimental Biology*, vol. 123, no. 1, pp. 383–400, 07 1986.

[120] A. A. Biewener, "Muscle-tendon stresses and elastic energy storage during locomotion in the horse," *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, vol. 120, no. 1, pp. 73–87, 1998.

[121] M. C. Granatosky, C. M. Bryce, J. Hanna, A. Fitzsimons, M. F. Laird, K. Stilson, C. E. Wall, and C. F. Ross, "Inter-stride variability triggers gait transitions in mammals and birds," *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1893, p. 20181766, 2018.

[122] G. A. Cavagna, N. C. Heglund, R. K. Taylor, C. Richard, and T. Mechanical, "Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure." *The American journal of physiology*, vol. 233 5, pp. R243–61, 1977.

[123] A. E. Minetti, L. P. Ardigo, E. Reinach, and F. Saibene, "The relationship between mechanical work and energy expenditure of locomotion in horses," *Journal of Experimental Biology*, vol. 202, no. 17, pp. 2329–2338, 09 1999.

[124] F. Saibene and A. E. Minetti, "Biomechanical and physiological aspects of legged locomotion in humans," *European journal of applied physiology*, vol. 88, pp. 297–316, 2003.

[125] M. Lemieux, N. Josset, M. Roussel, S. Couraud, and F. Bretzner, "Speed-Dependent Modulation of the Locomotor Behavior in Adult Mice Reveals Attractor and Transitional Gaits," *Frontiers in Neuroscience*, vol. 10, 2016.

[126] J. R. Usherwood, "An extension to the collisional model of the energetic cost of support qualitatively explains trotting and the trotâ€"canter transition," *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, vol. 333, no. 1, pp. 9–19, 2020.

[127] M. A. Daley, A. J. Channon, G. S. Nolan, and J. Hall, "Preferred gait and walkâ€"run transition speeds in ostriches measured using gps-imu sensors," *Journal of Experimental Biology*, vol. 219, no. 20, pp. 3301–3308, 10 2016.

[128] A. J. Ijspeert, A. Crespi, D. Ryczko, and J.-M. Cabelguen, "From swimming to walking with a salamander robot driven by a spinal cord model," *science*, vol. 315, no. 5817, pp. 1416–1420, 2007.

[129] R. Thandiackal, K. Melo, L. Paez, J. Herault, T. Kano, K. Akiyama, F. Boyer, D. Ryczko, A. Ishiguro, and A. J. Ijspeert, "Emergence of robust self-organized

undulatory swimming based on local hydrodynamic force sensing," *Science robotics*, vol. 6, no. 57, p. eabf6354, 2021.

[130] S. Dutta, A. Parihar, A. Khanna, J. Gomez, W. Chakraborty, M. Jerry, B. Grisafe, A. Raychowdhury, and S. Datta, "Programmable coupled oscillators for synchronized locomotion," *Nature communications*, vol. 10, no. 1, p. 3299, 2019.

[131] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Viability leads to the emergence of gait transitions in learning agile quadrupedal locomotion on challenging terrains," *Nature Communications*, vol. 15, no. 1, p. 3073, 2024.

[132] D. Kang, F. De Vincenti, N. C. Adami, and S. Coros, "Animal Motions on Legged Robots Using Nonlinear Model Predictive Control," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 955–11 962.

[133] L. Han, Q. Zhu, J. Sheng, C. Zhang, T. Li, Y. Zhang, H. Zhang, Y. Liu, C. Zhou, R. Zhao *et al.*, "Lifelike agility and play in quadrupedal robots using reinforcement learning and generative pre-trained models," *Nature Machine Intelligence*, vol. 6, no. 7, pp. 787–798, 2024.

[134] J. Li, H. Gao, Y. Wan, H. Yu, and C. Zhou, "A real-time planning and control framework for robust and dynamic quadrupedal locomotion," *Journal of Bionic Engineering*, vol. 20, no. 4, pp. 1449–1466, 2023.

[135] M. Arduengo, A. Arduengo, A. Colome, J. Lobo-Prat, and C. Torras, "Human to Robot Whole-Body Motion Transfer," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots*, 2021, pp. 299–305.

[136] J. Nakanishi, R. Cory, M. Mistry, J. Peters, and S. Schaal, "Operational Space Control: A Theoretical and Empirical Comparison," *International Journal of Robotics Research*, vol. 27, no. 6, pp. 737–757, Jun. 2008.

[137] P.-A. Leziart, T. Flayols, F. Grimminger, N. Mansard, and P. Soueres, "Implementation of a Reactive Walking Controller for the New Open-Hardware Quadruped Solo-12," in *2021 IEEE International Conference on Robotics and Automation*, 2021, pp. 5007–5013.

[138] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," 2016.

[139] J. Humphreys. High Utility Teleoperation Framework for Legged Manipulators Through Leveraging Whole-Body Control. Youtube. [Online]. Available: https://youtu.be/iL_K_CVA0pU

[140] ——. Bio-inspired Gait Transitions for Quadruped Locomotion. Youtube. [Online]. Available: https://www.youtube.com/watch?v=vhA0tiHbSSk

[141] J. Choi, "Multi-Phase Joint-Angle Trajectory Generation Inspired by Dog Motion for Control of Quadruped Robot," *Sensors*, vol. 21, no. 19, 2021.

[142] J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim, "Dynamic locomotion in the MIT Cheetah 3 through convex model-predictive control," in *IEEE/RSJ IROS*, 2018, pp. 1–9.

[143] J. A. Vilensky, J. Njock Libii, and A. M. Moore, "Trot-gallop gait transitions in quadrupeds," *Physiology & Behavior*, vol. 50, no. 4, pp. 835–842, 1991.

[144] D. Kim, J. Di Carlo, B. Katz, G. Bledt, and S. Kim, "Highly Dynamic quadruped Locomotion via Whole-Body Impulse Control and Model Predictive Control," *arXiv preprint arXiv:1909.06586*, 2019.

[145] V. A. Tucker, "The energetic cost of moving about," *American Scientist*, vol. 63, no. 4, pp. 413–419, 1975.

[146] T.-C. Huang, Y.-J. Huang, and W.-C. Lin, "Real-time horse gait synthesis," *Computer Animation and Virtual Worlds*, vol. 24, no. 2, pp. 87–95, 2013.

[147] F. J. Diedrich and W. H. Warren, "Why change gaits? dynamics of the walk-run transition," *Journal of experimental psychology. Human perception and performance*, vol. 21 1, pp. 183–202, 1995.

[148] J. Humphreys. Learning to Adapt through Bio-Inspired Gait Strategies for Versatile Quadruped Locomotion. Youtube. [Online]. Available: https://www.youtube.com/watch?v=ecplBINQ3Tg

[149] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[150] I. M. Aswin Nahrendra, B. Yu, and H. Myung, "DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5078–5084.

[151] O. Kiehn, "Decoding the organization of spinal circuits that control locomotion," *Nature Reviews Neuroscience*, vol. 17, no. 4, pp. 224–238, 2016.

[152] K. Takakusaki, "Functional neuroanatomy for posture and gait control," *Journal of movement disorders*, vol. 10, no. 1, p. 1, 2017.

[153] B. R. Noga and P. J. Whelan, "The Mesencephalic Locomotor Region: Beyond Locomotor Control," *Frontiers in Neural Circuits*, vol. 16, 2022.

[154] S. M. Morton and A. J. Bastian, "Cerebellar control of balance and locomotion," *The neuroscientist*, vol. 10, no. 3, pp. 247–259, 2004.

[155] M. H. Raibert, *Legged robots that balance*. MIT press, 1986.

[156] Z. Afelt, J. Błaszczyk, and C. Dobrzecka, "Speed control in animal locomotion: transitions between symmetrical and nonsymmetrical gaits in the dog," *Acta neurobiologiae experimentalis*, vol. 43, no. 4-5, pp. 235–250, 1983.

[157] J. Hwangbo, J. Lee, and M. Hutter, "Per-contact iteration method for solving contact dynamics," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 895–902, 2018. [Online]. Available: www.raisim.com

[158] J. Humphreys. Deployment of Bio-inpsired Locomotion Framework on Real-world Terrain. Youtube. [Online]. Available: https://www.youtube.com/watch?v=NwHoB7pErYQ

[159] ——. Comparison Study between Our Bio-inspired Locomotion Policy and Standard Locomotion Policies. Youtube. [Online]. Available: https://www.youtube.com/watch?v=-DfkDFA3KkI

[160] ——. Comparison Study between Static Gait Deployment and Our Bio-inspired Gait Selection Policy. Youtube. [Online]. Available: https://www.youtube.com/watch?v=y4KnzMEdf78

[161] D. F. Hoyt and C. R. Taylor, "Gait and the energetics of locomotion in horses," *Nature*, vol. 292, no. 5820, pp. 239–240, 1981.

[162] J. Humphreys. Deployment of Bio-inspired Locomotion Policy on Grassy Terrain. Youtube. [Online]. Available: https://www.youtube.com/watch?v=I02DQ1RGdyw

[163] ——. Stability Recovery Study of Bio-inspired Locomotion Framework. Youtube. [Online]. Available: https://www.youtube.com/watch?v=f6CqJ7gb3ZM

[164] T. Olsen. INSANE Rock crawling- Extreme Mule Riding. Youtube. [Online]. Available: https://www.youtube.com/watch?v=7rhJU6qpaQE

[165] H. Boukamcha, "3D Kalman Filter - C++ Implementation," https://github.com/hamdiboukamcha/3DKalmanFilter, 2024.

[166] T. Lee, "Geometric tracking control of the attitude dynamics of a rigid body on SO(3)," in *Proceedings of the 2011 American Control Conference*, 2011, pp. 1200–1205.

[167] O. Lahayne, B. Pichler, R. Reihsner, J. Eberhardsteiner, J. Suh, D. Kim, S. Nam, H. Paek, B. Lorenz, and B. N. Persson, "Rubber friction on ice: experiments and modeling," *Tribology letters*, vol. 62, no. 2, p. 17, 2016.

[168] Y. Ma, A. Cramariuc, F. Farshidian, and M. Hutter, "Learning coordinated badminton skills for legged manipulators," *Science Robotics*, vol. 10, no. 102, p. eadu3922, 2025.

[169] Z. Yang, J. Mao, Y. Du, J. Wu, J. B. Tenenbaum, T. Lozano-Pérez, and L. P. Kaelbling, "Compositional Diffusion-Based Continuous Constraint Solvers," in *Conference on Robot Learning*, 2023.

[170] M. Stamatopoulou, J. Liu, and D. Kanoulas, "DiPPeST: Diffusion-based Path Planner for Synthesizing Trajectories Applied on Quadruped Robots," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 7787–7793.

[171] G. Mesesan, R. Schuller, J. Englsberger, M. A. Roa, J. Lee, C. Ott, and A. Albu-Schaffer, "Motion Planning for Humanoid Locomotion: Applications to Homelike

Environments," *IEEE Robotics and Automation Magazine*, vol. 32, no. 1, pp. 35–48, 2025.

[172] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, vol. 9, no. 89, p. eadi9579, 2024.

[173] W. Cui, S. Li, H. Huang, B. Qin, T. Zhang, h. hanjinchao, L. Zheng, Z. Tang, C. Hu, N. Yan, J. Chen, and Z. Jiang, "Adapting Humanoid Locomotion over Challenging Terrain via Two-Phase Training," in *Proceedings of The 8th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 270. PMLR, 06–09 Nov 2025, pp. 3439–3453.

[174] S. M. Kung, P. W. Fink, S. J. Legg, A. Ali, and S. P. Shultz, "What factors determine the preferred gait transition speed in humans? a review of the triggering mechanisms," *Human movement science*, vol. 57, pp. 1–12, 2018.

[175] M. Ackermann and A. J. Van den Bogert, "Optimality principles for model-based prediction of human gait," *Journal of biomechanics*, vol. 43, no. 6, pp. 1055–1060, 2010.

[176] N. S. Baker, L. Long, and M. Srinivasan, "Realistic overground gait transitions are not sharp but involve gradually changing walk-run mixtures as per energy optimality," *arXiv preprint arXiv:2501.00720*, 2025.

[177] A. E. Minetti, "The biomechanics of skipping gaits: a third locomotion paradigm?" *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1402, pp. 1227–1233, 1998.

[178] J. S. Matthis, J. L. Yates, and M. M. Hayhoe, "Gaze and the control of foot placement when walking in natural terrain," *Current Biology*, vol. 28, no. 8, pp. 1224–1233, 2018.

[179] T. J. Dick, L. K. Punith, and G. S. Sawicki, "Humans falling in holes: adaptations in lower-limb joint mechanics in response to a rapid change in substrate height during human hopping," *Journal of the Royal Society interface*, vol. 16, no. 159, p. 20190292, 2019.

[180] S. M. Bruijn, O. G. Meijer, P. J. Beek, and J. H. Van Dieen, "The effects of arm swing on human gait stability," *Journal of experimental biology*, vol. 213, no. 23, pp. 3945–3952, 2010.