2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Puppeteer and Marionette: Learning Anticipatory Quadrupedal Locomotion Based on Interactions of a Central Pattern Generator and Supraspinal Drive

Milad Shafiee[1], Guillaume Bellegarda[1], and Auke Ijspeert[1]

***Abstract*— Quadruped animal locomotion emerges from the interactions between the spinal central pattern generator (CPG), sensory feedback, and supraspinal drive signals from the brain. Computational models of CPGs have been widely used for investigating the spinal cord contribution to animal locomotion control in computational neuroscience and in bio-inspired robotics. However, the contribution of supraspinal drive to anticipatory behavior, i.e. motor behavior that involves planning ahead of time (e.g. of footstep placements), is not yet properly understood. In particular, it is not clear whether the brain modulates CPG activity and/or directly modulates muscle activity (hence bypassing the CPG) for accurate foot placements. In this paper, we investigate the interaction of supraspinal drive and a CPG in an anticipatory locomotion scenario that involves stepping over gaps. By employing deep reinforcement learning (DRL), we train a neural network policy that replicates the supraspinal drive behavior. This policy can either modulate the CPG dynamics, or directly change actuation signals to bypass the CPG dynamics. Our results indicate that the direct supraspinal contribution to the actuation signal is a key component for a high gap crossing success rate. However, the CPG dynamics in the spinal cord are beneficial for gait smoothness and energy efficiency. Moreover, our investigation shows that sensing the front feet distances to the gap is the most important and sufficient sensory information for learning gap crossing. Our results support the biological hypothesis that cats and horses mainly control the front legs for obstacle avoidance, and that hind limbs follow an internal memory based on the front limbs' information. Our method enables the quadruped robot to cross gaps of up to 20 cm (50% of body-length) without any explicit dynamics modeling or Model Predictive Control (MPC).**

I. Introduction and Related Work

Quadruped animals can perform highly agile motions including running, jumping over hurdles, and leaping over gaps. Performing such anticipatory motor behaviors at high speeds requires complex interactions between the supraspinal drive, spinal cord dynamics, and sensory feedback [1]. Through recent advances in machine learning and optimal control, animals' legged robot counterparts are becoming increasingly capable of traversing complex terrains [2]. Robots can also be used as scientific tools to investigate biological hypotheses of animal adaptive behavior [3], and conversely we can take inspiration from the underlying mechanisms of animal locomotion to develop robotic systems that approach the agility of animals [4], [5]. In this paper, by leveraging recent robotics tools, we investigate the interaction between the supraspinal drive, the central pattern generator (CPG), and sensory information to generate anticipatory locomotion control for a gap crossing task. We propose a hierarchical biologically-inspired framework, where higher control centers in the brain (represented by an artificial neural network) send supraspinal drive signals to either modulate the CPG dynamics, or directly output actuation signals which bypass the CPG dynamics.

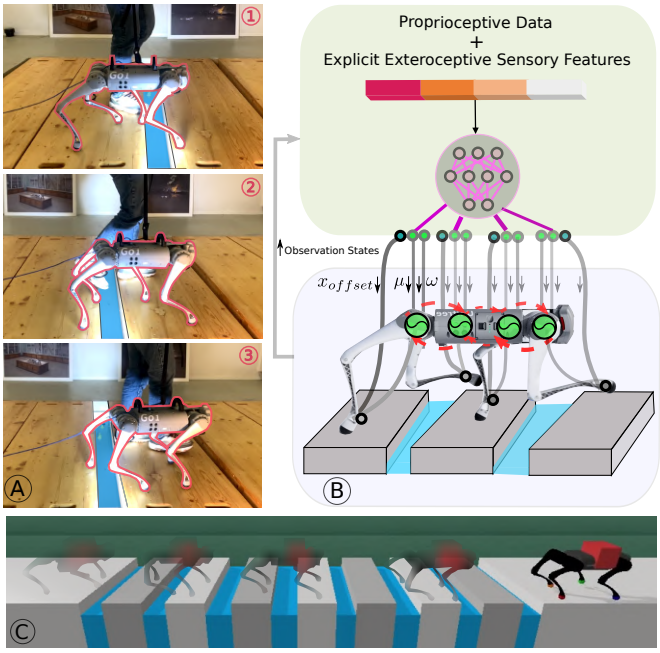

Fig. 1: **A**: Crossing a 15 cm gap with Unitree Go1. **B**: We represent the motor control system as a Puppeteer and Marionette, where the supraspinal higher control centers work as a Puppeteer to manipulate the movement of the body (Marionette) with limited strings. The supraspinal drive controls movement by either modulating the frequency and amplitude of the CPG oscillators, or directly sending actuation signals to bypass the CPG dynamics. **C**: Testing policy robustness by crossing variable gaps between 14 and 20 cm with an unknown 5 kg load. Videos: `https://miladshafiee.github.io/puppeteer-and-marionette/`

A. Central Pattern Generators

It is widely accepted that the mammalian spinal cord contains a central pattern generator (CPG) that can produce basic locomotor rhythm in the absence of input from supraspinal drive and peripheral sensory feedback [6]. In robotics, abstract models of CPGs are commonly used for locomotion pattern generation [4], [7]–[9], as well as to investigate biological hypotheses [10], [11]. Besides the intrinsic oscillatory behavior of CPGs, several other properties such as robustness and implementation simplicity make CPGs desirable for locomotion control [12]. For legged robots, the

[1] This research is supported by the Swiss National Science Foundation (SNSF) as part of project No.197237. The authors are with the BioRobotics Laboratory, Ecole Polytechnique Federale de Lausanne (EPFL). (e-mail: firstname.lastname@epfl.ch)

CPG is usually designed for feedforward rhythm generation, and dynamic balancing is achieved with optimization [4] or hand-tuned feedback [7], [13]. CPGs also provide an intuitive formulation for specifying different gaits [14], and spontaneous gait transitions can arise by increasing descending drive signals and incorporating contact force feedback [15], [16] or vestibular feedback [17]. In addition to proprioceptive feedback, incorporating exteroceptive feedback information allows for CPG-based locomotion over uneven terrains [18], [19] and navigation in complex environments [20], [21].

Most of these studies consider the CPG to be isolated from higher control centers located in the brain. However, the interaction between supraspinal drive and the Central Pattern Generator during learning and planning leads to fascinating anticipatory locomotion in animals (i.e. motor behavior that involves planning ahead of time). In this article, we investigate the roles of supraspinal drive and the CPG in learning such anticipatory behaviors.

B. Learning Legged Locomotion

Deep Reinforcement Learning (DRL) has emerged as a powerful approach for training robust legged locomotion control policies in simulation, and deploying these sim-to-real on hardware [2], [22]–[31]. To facilitate this sim-to-real transfer, a variety of techniques can be employed such as online parameter adaptation [26], [27], learned state estimation modules [28], teacher-student training [2], [24], [27], and careful Markov Decision Process choices [30]–[33]. Most of these works view the trained artificial neural network (ANN) as a "brain" which has full access to complete whole-body proprioceptive sensing, which it queries at a high rate to update motor commands. Therefore, different gaits emerge through the combination of reward function tuning (i.e. minimizing energy consumption [34]), incorporating phase biases [35], [36], or imitating animal reference data to replicate bio-inspired movements [26], [37]. However, here and in CPG-RL [31], we represent the ANN as a higher level control center which sends descending drive signals to modulate the central pattern generator in the spinal cord, and map this rhythm generation network to a pattern formation layer. Moreover, here we study the interplay between the ANN modulating the CPG directly, or bypassing the CPG to directly control lower level circuits.

Beyond "blind" terrain locomotion, recent works incorporate exteroceptive sensing in the learning loop, for example for obstacle avoidance [38] or walking over rough terrain by employing height maps [2], [39]. Gap crossing has been demonstrated by employing MPC and dynamic models for motion planning during learning [40]–[43]. For difficult simulation tasks, curriculum learning is helpful for surmounting increasingly challenging terrain [44], and jumping over large hurdles has been demonstrated by employing a mentor during the learning process [45].

C. Contribution

Despite advances in understanding motor control of mammalian locomotion [1], little is known about the emergence of anticipatory locomotion skills through the interaction of supraspinal drive and CPGs. In this work, we investigate two broad neuroscience research questions:

- What are the plausible contributions of supraspinal drive and CPG circuits in the spinal cord for producing anticipatory locomotion skills?
- What are the necessary sensory feedback features for learning anticipatory locomotion skills?

Although these research questions are broad, we view this work as a starting point for leveraging robotics techniques and biological inspiration to investigate the interaction of supraspinal drive (from the brain) with the CPG. We employ CPG models and a neural network (NN) policy trained with deep reinforcement learning to investigate this interaction for a gap-crossing task. For the first question, our results indicate that the direct supraspinal contribution to the actuation signal is a key component for a high success rate. However, the CPG dynamics in the spinal cord are beneficial for gait smoothness and energy efficiency.

Regarding the second question, our investigation shows that the front foot distance to the gap is the most important visually-extracted sensory information to successfully cross variable gaps. Our results show that front limb information is sufficient for learning gap-crossing, and that DRL can learn to create and encode an internal kinematic model by combining proprioceptive information with the internal CPG states to modulate the hind leg motion for gap crossing. This supports the biological hypothesis that cats and horses control their front legs for obstacle avoidance, and that hind legs follow an internal memory based on front foot information [46], [47]. Furthermore, in contrast to previous robotics works, to the best of our knowledge, this is the first learning-based framework with gap-crossing abilities which does not have a dynamical model, MPC, curriculum, or mentor in the loop. This illustrates the versatility of the proposed framework, which requires minimum expert knowledge (i.e. no model of the system dynamics or more traditional optimal control).

The rest of this paper is organized as follows. Section II describes the CPG topology. Section III details the DRL framework and design of the Markov Decision Process. Section IV presents results and analysis regarding the two mentioned research questions, and a brief conclusion is given in Section V.

II. Central Pattern Generators

The locomotor system of vertebrates is organized such that the spinal CPGs are responsible for producing basic rhythmic patterns, while higher-level centers (i.e. the motor cortex, cerebellum, and basal ganglia) are responsible for modulating the resulting patterns according to environmental conditions [1]. Rybak et al. [48] propose that biological CPGs have a two-level functional organization, with a half-center rhythm generator (RG) that determines movement frequency, and pattern formation (PF) circuits that determine the exact shapes of muscle activation signals. Similar organizations have also been used in robotics, for example in our previous work [31], and in [49].

A. Rhythm Generator (RG) Layer

We employ amplitude-controlled phase oscillators to model the RG layer of the CPG circuits in the spinal cord. Such oscillators have been successfully used for locomotion control of legged robots [4], [10], [31] with the following dynamics:

$$\ddot{r}_i = \alpha\left(\frac{\alpha}{4}(\mu_i - r_i) - \dot{r}_i\right) \tag{1}$$

$$\dot{\theta}_i = \omega_i + \sum_j r_j w_{ij} \sin(\theta_j - \theta_i - \phi_{ij}) \tag{2}$$

where r_i is the amplitude of the oscillator, θ_i is the phase of the oscillator, μ_i and ω_i are the intrinsic amplitude and frequency, α is a positive constant representing the convergence factor. Couplings between oscillators are defined by the weights w_{ij} and phase biases ϕ_{ij}. In this paper, we use the oscillators without neural coupling ($w_{ij} = 0$), and gaits (i.e. phase relationships between limbs) are thus determined by the supraspinal control policy. As in [31], we will investigate the modulation of the intrinsic amplitude and frequency (μ_i and ω_i) for each limb as control signals for the CPG.

B. Pattern Formation (PF) Layer

To map from the RG layer to joint commands, we first compute corresponding desired foot positions, and then calculate the desired joint positions with inverse kinematics. The desired foot position coordinates are formed as follows:

$$x_{i,\text{foot}} = \quad x_{off,i} - L_{step}(r_i)\cos(\theta_i) \tag{3}$$

$$z_{i,\text{foot}} = \begin{cases} z_{off,i} - h + L_{clrnc}\sin(\theta_i) & \text{if } \sin(\theta_i) > 0 \\ z_{off,i} - h + L_{pntr}\sin(\theta_i) & \text{otherwise} \end{cases} \tag{4}$$

where L_{step} is the step length, h is the nominal leg length, L_{clrnc} is the max ground clearance during swing, L_{pntr} is the max ground penetration during stance, and x_{off} and z_{off} are set-points that change the equilibrium point of oscillation in the x and z directions. Modulating the foot horizontal offset x_{off} and vertical offset z_{off} represents direct supraspinal control of the general position of the limb, bypassing the rhythm generation layer. A description and visualization of the foot trajectory is illustrated in Figure 2.

III. Hierarchical Bio-Inspired Learning of Anticipatory Gap Crossing Tasks

In this section we describe our hierarchical bio-inspired learning framework for learning anticipatory gap crossing abilities for quadruped robots. We represent the supraspinal controller as an artificial neural network which is trained with DRL to modulate both the feet positions (CPG offsets) and/or the intrinsic frequencies and amplitudes of oscillation for each limb to produce anticipatory behavior. The problem is represented as a Markov Decision Process (MDP), and we describe each of its components below.

A. Action Space

We consider one RG layer for each limb based on Equations (1) and (2), where the RG output will be used in a PF layer to generate the spatio-temporal foot trajectories in Cartesian space (Equations (3) and (4)). We do not consider any explicit neural coupling (i.e. $w_{ij} = 0$), with the intuition that inter-limb coordination will be managed by the supraspinal drive.

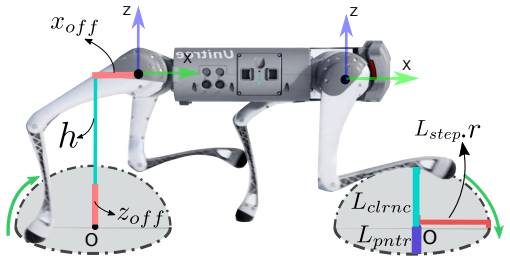

Fig. 2: Visualization of the task space foot trajectories generated by the PF layer. The oscillatory trajectory is built around a central point O. The offsets x_{off} and z_{off} are used to change the central point of oscillation. x_{off} is a horizontal offset between the oscillation set-point and the center of the hip coordinate, and controlled directly by the supraspinal drive, bypassing the CPG dynamics. $z_{off} + h$ is the vertical distance between the oscillation set-point, O, and the center of the hip coordinate. $L_{step}r$ is the step length multiplied by the oscillator amplitude, h is the nominal leg length, L_{clrnc} is the max ground clearance during leg swing phase, and L_{pntr} is the max ground penetration during stance.

As in [31], our action space modulates the intrinsic amplitudes and frequencies of the CPG, by continuously updating μ_i and ω_i for each leg. However, unlike [31], we also consider modulating the oscillation set-points by directly learning foot Cartesian offsets x_{off_i}, z_{off_i} for each leg. Thus, our action space can be summarized as $\mathbf{a} = [\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{x_{off}}, \boldsymbol{z_{off}}] \in \mathbb{R}^{16}$. We divide the descending drive modulation into two categories: oscillatory components of the CPG dynamics $\mathbf{a}_{osc} = [\boldsymbol{\mu}, \boldsymbol{\omega}] \in \mathbb{R}^8$, and offset components $\mathbf{a}_{off} = [\boldsymbol{x_{off}}, \boldsymbol{z_{off}}] \in \mathbb{R}^8$, shown in Equations (1)-(4). This separation allows us to investigate how gap crossing can best be accomplished, i.e. by modulating CPG activity (by changing μ_i and ω_i) and/or by directly updating the limb posture (by changing x_{off_i} or z_{off_i}). Based on this investigation, we use $\mathbf{a} = [\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{x_{off}}] \in \mathbb{R}^{12}$ for analyzing the roles of sensory feedback features in Section IV-B. The agent selects these parameters at 100 Hz, which will therefore vary during each step according to sensory inputs. We use the following limits for each input during training: $\mu \in [0.5, 4]$, $\omega \in [0, 5]$ Hz, $x_{off_i} \in [-7, 7] cm$, $z_{off_i} \in [-7, 7] cm$.

B. Observation Space

We consider two different observation space types based on (1) only proprioceptive sensing (enough for locomotion on flat terrain) and (2) also including exteroceptive anticipatory features. Various exteroceptive anticipatory features will be investigated to understand the roles and importance of different sensory quantities.

Flat terrain observation: We consider body orientation, body linear and angular velocity, joint positions and velocities, foot contact booleans, the previous action chosen by the policy network, and CPG states $\{\boldsymbol{r}, \dot{\boldsymbol{r}}, \boldsymbol{\theta}, \dot{\boldsymbol{\theta}}\}$ as the flat terrain (proprioceptive sensing) observation space.

Exteroceptive Feedback Features: We assume that the visual system and brain can extract important geometrical information such as foot distance to a gap, and we call such information *exteroceptive feedback features*. We are

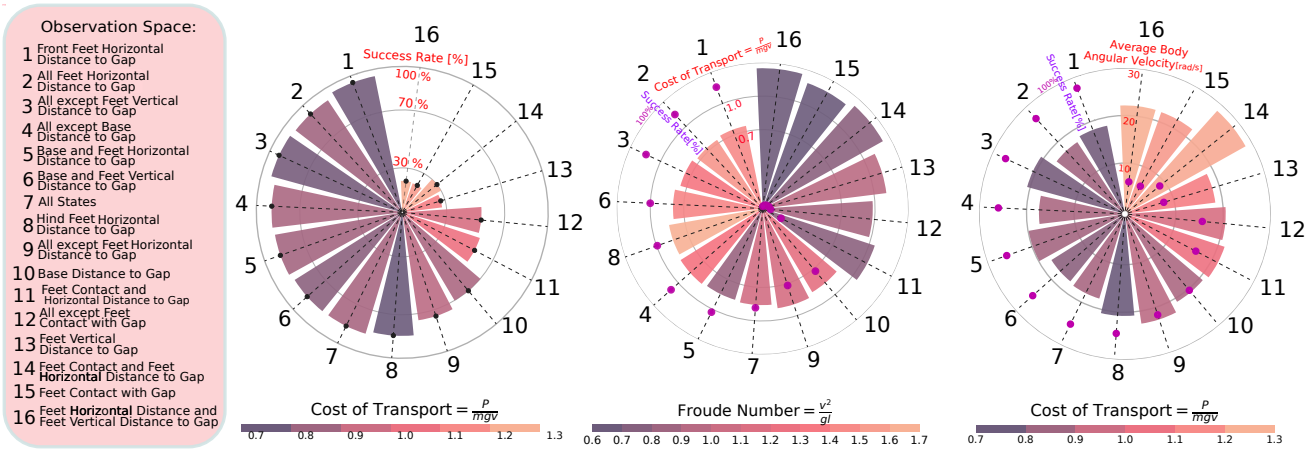

Fig. 3: Quantitative results from testing 16 policies trained with different combinations of exteroceptive feedback features consisting of feet distance to the gap, base distance to the gap, feet vertical distance with the ground, and contact/penetration with the gap. All policies also include the flat terrain (proprioceptive sensing) observation, and the action space consists of both oscillatory and offset terms. We report the average results of testing the polices for 4000 samples each (100 attempts of crossing 7 gaps with randomized lengths between $[14,20]$ cm). We characterize the viability performance of the system by the success rate, which is the proportion of the gaps in front of the robot that it could successfully cross. Energy efficiency is characterized by the Cost of Transport (CoT), mean velocity is characterized by the Froude number, and gait smoothness is evaluated by the mean angular velocity.

interested in investigating which exteroceptive feedback features are most useful for the emergence of anticipatory locomotion skills. To reverse engineer this process, we divide the exteroceptive feedback features into two categories: *predictive* and *instantaneous* feedback features. *Predictive* features consist of foot distance and/or base distance to the beginning and end of a gap. *Instantaneous* feedback features consist of boolean indicators of stepping into a gap (foot contact/penetration into the gap), as well as vertical distance of the foot with the ground (larger values over a gap). These features are instantaneous feedback, so they cannot be used to predict information about upcoming gaps.

C. Reward Function

Our reward function promotes forward progress, gap crossing ability, maintaining base stability, and energy efficiency with the following terms:

$$r = \alpha_1 \cdot \min(f_x, d_{max}) + (S_{gap} + n_{gap}) + \alpha_3 \cdot |y_{base}| + \alpha_4 \cdot ||\boldsymbol{o}_{base} - \boldsymbol{o}_{zero}|| + \alpha_5 \cdot |\boldsymbol{\tau} \cdot (\dot{\boldsymbol{q}}_t - \dot{\boldsymbol{q}}_{t-1})|$$

- *Forward progress*: In the first term, f_x corresponds to forward progress in the world (along the x-direction). We limit this term to avoid exploiting simulator dynamics and achieving unrealistic speeds, where d_{max} is the maximum distance the robot will be rewarded for moving forward during each control cycle ($\alpha_1 = 2$).
- *Gap reward*: The agent receives a sparse reward ($S_{gap} = 3$) if it crosses a gap, and a negative reward of $n_{gap} = -0.03$ for each control cycle during which a foot penetrates below ground level into a gap.
- *Base y direction penalty*: The third term penalizes lateral deviation of the body ($\alpha_3 = -0.05$).
- *Base orientation penalty*: The fourth term penalizes non-zero body orientation ($\alpha_4 = -0.02$).
- *Power*: The fifth term penalizes power in order to find energy efficient gaits, where $\boldsymbol{\tau}$ and $\dot{\boldsymbol{q}}$ are joint torques and velocities ($\alpha_5 = -0.00008$).

TABLE I: PPO Hyperparameters and neural network architecture.

Parameter	Value	Parameter	Value
Batch size	4096	SGD Iterations	10
SGD Mini-bach size	128	Discount factor	0.99
Desired KL-divergence kl^*	0.01	Learning rate α	0.0001
GAE discount factor	0.95	Hidden Layers	2
Clipping Threshold	0.2	Nodes	[256,256]
Entropy coefficient	0.01	Activation	tanh

D. Training details

We use PyBullet [50] as our physics engine for training and simulation purposes, and the Unitree A1 and Go1 quadruped robots [51]. To train the policies, we use Proximal Policy Optimization (PPO) [52], and Table I lists the PPO hyperparameters and neural network architecture. The control frequency of the policy is 100 Hz, and the torques computed from the desired joint positions are updated at 1 kHz. The equations for each of the oscillators (Equations 1 and 2) are thus also integrated at 1 kHz. The joint PD controller gains are $K_p = 100$ and $K_d = 2$. All policies are trained for 3.5×10^7 samples.

IV. Results

In this section, we present results of our proposed framework for quadruped gap crossing scenarios. In Section IV-A, we investigate the role of the supraspinal drive and CPG based on three criteria: success rate, energy efficiency, and gait smoothness. Furthermore, we investigate the effects of including varying exteroceptive sensory features on these criteria in Section IV-B. Section IV-C presents results for a more challenging task of crossing successive narrowly-spaced gaps, and Section IV-D discusses sim-to-real hardware results. The reader is encouraged to watch the supplementary video for clear visualizations of all discussed experiments.

A. Contribution of CPG and Supraspinal Drive to Actuation

In this section we train locomotion policies for the following scenarios and action spaces:

1. Flat terrain. CPG only in xz directions.
2. Gap terrain. CPG only in xz directions.

3. Gap terrain. CPG in z direction and offset in x direction.
4. Gap terrain. CPG in xz and offset in x direction.
5. Gap terrain. Offsets only in xz directions.
6. Gap terrain. CPG and offsets both in xz directions.

The CPG in these cases means the agent (supraspinal drive) modulates the frequency and amplitude of Equations (1) and (2). The offset terms are considered as a part of the actuation signal applied directly by the supraspinal drive, bypassing the spinal cord dynamics. Cases 5 and 6 are the only cases in which we modulate the offset in the z direction.

We train policies for each case for 3.5×10^7 samples for episodes of $10\ s$ on terrains with 7 consecutive gaps (except for Case 1, which is trained on flat terrain only) with all exteroceptive feedback features in the observation space. Each gap length is randomized in $[14{,}20]\ cm$ during both training and test time, with $1.4\ m$ distances between gaps. An episode terminates early because of a fall, i.e. if the body height drops below $15\ cm$. We define the success rate as the number of gaps successfully crossed out of the total number of gaps. In order to test the six policies, we perform 30 policy rollouts on a test environment of locomoting over 7 randomized gaps. Table II summarizes the results from investigating how supraspinal drive can modulate locomotion in these six cases.

1) Gap Crossing Success Rate: Case 5 has the highest success rate of 99%, indicating the benefit of direct supraspinal actuation in anticipatory scenarios. Case 4, with both oscillatory and offset terms in the x direction, has the second highest success rate of 97%. The third highest success rate is for Case 6 with both oscillatory and offset terms in both x and z directions. The fourth highest success rate is Case 3 (with only the offset component in the x direction), and Case 2 (with only the CPG components) has the fifth best success rate of 17%. These results show that direct supraspinal actuation of the foot offset/position is critical for successful gap-crossing, though the CPG can contribute to a high success rate in the absence of z offset modulation.

2) Gait Smoothness: To compare the gait smoothness between policies, we analyze the robot body oscillations during locomotion, and in particular the average angular velocity of the robot body $\bar{\omega}_{Body} = (\sum_{t=1}^{N} |\omega_{x,t}| + |\omega_{y,t}| + |\omega_{z,t}|)/(3N)$.

Body orientation deviations are penalized in the reward function, as high (absolute) angular velocities tend to correspond to shaky gait patterns. As shown in Table II, the first case has the smoothest gait. This is expected since it corresponds to steady-state locomotion behavior on flat terrain. A comparison of the third and fourth cases indicates a 45% reduction in body oscillation when the agent can also modulate CPG amplitudes. The gait smoothness of case 5 is drastically reduced by removing the CPG dynamics. This result shows the importance of spinal cord dynamics (limit-cycle oscillatory dynamics) for obtaining smooth locomotion.

3) Cost of Transport (CoT) and Froude number: We investigate gait efficiency by comparing the CoT, and mean velocity with the Froude number [4]. We observe that the fourth case (with both oscillatory and offset components), has the best combined CoT, Froude number, and gait smoothness (low $\bar{\omega}_{Body}$). This demonstrates the benefit of having both supraspinal drive and CPG dynamics for coordinating locomotion. Case 3, which has the CPG oscillatory component in z and offset in x, has the lowest CoT, and we observe significant added energy expenditure by removing the CPG dynamics (Case 5). Case 6 shows that having both oscillatory and offset terms leads to the highest Froude number, but also a high CoT and $\bar{\omega}_{Body}$, suggesting that overparameterizing the action space can make it difficult for the agent to converge to an optimal policy.

TABLE II: Testing policies trained with different combinations of oscillatory and offset terms in the action space. $\bar{\omega}_{Body}$ is the average body angular velocity. Case 1 is for walking on flat terrain without gaps. We only show mean values since the standard deviations are small (i.e. less than 10% of the means).

Case	x_{osc}	x_{off}	z_{osc}	z_{off}	Success[%]	CoT	Froude	$\bar{\omega}_{Body}$
1	✓	×	✓	×	×	0.92	0.34	**0.36**
2	✓	×	✓	×	17	1.45	0.29	0.73
3	×	✓	✓	×	60	**0.84**	0.29	0.77
4	✓	✓	✓	×	97	0.94	0.55	0.42
5	×	✓	×	✓	**99**	1.24	0.56	0.96
6	✓	✓	✓	✓	93	1.35	**0.88**	0.85

B. Roles of Feedback Features for Anticipatory Locomotion

In this section we investigate which exteroceptive sensory feedback information is necessary and sufficient for learning and planning anticipatory tasks. As shown in Figure 3, we consider 16 different combinations of predictive and instantaneous feedback features as described in Section III-B. We train 16 different policies with the Case 4 action space from Section IV-A (i.e. with CPG and x offset modulation).

For evaluation, we rollout each policy 100 times and present mean results across all tests. Fig. 3-A shows the gap crossing success rate (by bar height) and CoT (by color). Our results show that policy 1, which has the front feet distances to the gap in the observation space, has both one of the best success rates as well as one of the lowest CoTs. These results show that front leg information is sufficient for learning the gap-crossing task. As the agent is explicitly blind about hind leg positions, this forces it to learn an internal kinematic model by combining exteroceptive front feet positions to the gap, internal CPG states, and proprioceptive sensing to modulate the hind leg motions for gap crossing. This result supports the biological hypothesis that cats and horses control the front legs for obstacle avoidance, and that hind legs follow based on an internal kinematic memory [46], [47].

Notably, as could be expected, the 8 policies with the highest success rates contain the feet and/or base distances to the gap in the observation space, which indicates the importance of predictive feedback features in the observation space. Interestingly, policy 7, which observes all discussed exteroceptive sensing in the observation space, has a slightly lower success rate with respect to the first 6 policies with subsets of the full sensing. This suggests that including supererogatory information in the observation space may not necessarily improve the quality of the learned policy, and may even prevent the convergence of the RL algorithm in finding the optimal policy.

Figure 3-B shows the CoT as the bar height, with the color indicating the Froude number. The lowest CoT and

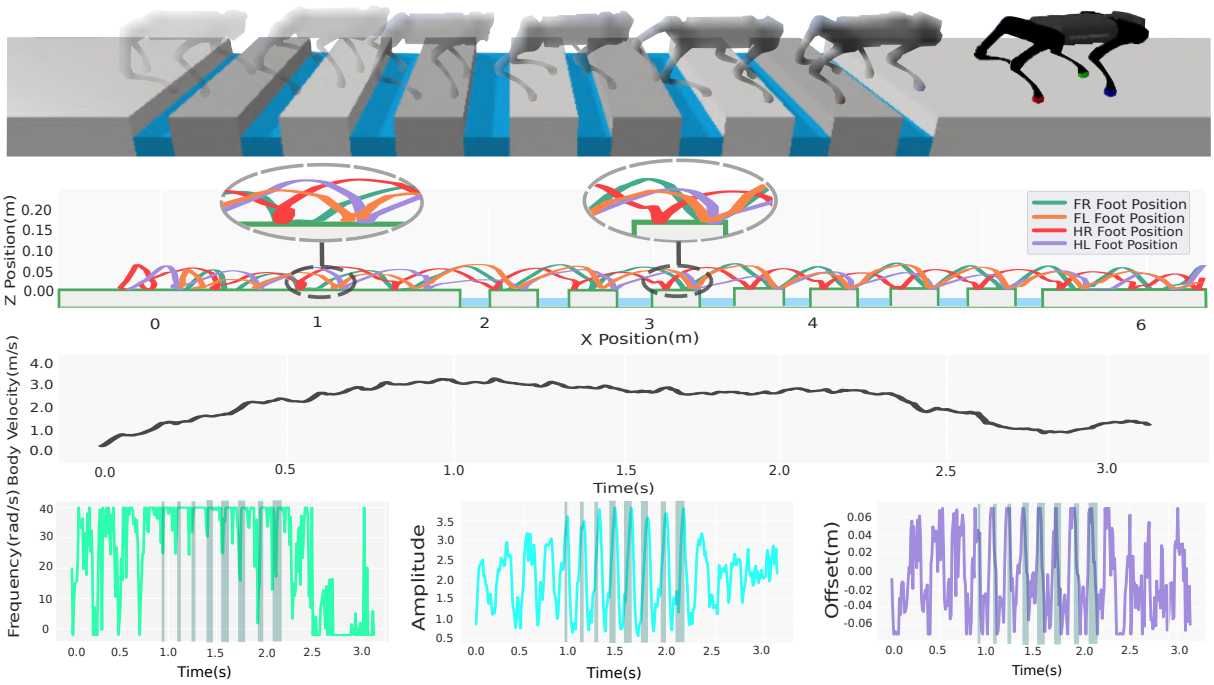

Fig. 4: Crossing 8 gaps with randomized lengths between [14,20] cm, with only 30 cm contact surfaces. **Top:** simulation snapshots. **Middle:** body velocity and foot positions in the XZ plane. **Bottom:** CPG frequency, amplitude, and offset for the front left limb. The shadow bars indicate when the foot is over a gap.

highest Froude number are for the first three policies, which include the front feet positions in the observation space. The first 10 policies show that having predictive information in the observation space helps to learn an energy-efficient gait for the gap-crossing task.

Figure 3-C shows the average body angular velocity to investigate gait smoothness. We observe that having feet distances to the gap in the observation space leads to lower average body angular velocities, and as a result smoother gaits.

C. Training for a More Challenging Gap-Crossing Scenario

In this section, we train the robot to cross 8 gaps with the front feet distances to the gap in the observation space (the same as policy 1 from Section IV-B), with 30 cm platforms between each gap. The gaps have randomized lengths between $[14,20]$ cm, and the first gap position is randomized between $[1.25,2.25]$ m. As shown in Figure 4, the supraspinal drive increases the velocity of the robot by increasing the frequency of the CPG. The desired velocity for the robot is 1 m/s, however, the agent has learned to increase the velocity of the robot to up to 3 m/s to overcome the gaps. We observe that the policy increases the limb frequency to near its maximum limits for all legs as soon as it reaches the first gap. This indicates that the supraspinal drive modulates the locomotion speed by increasing the CPG frequencies. The oscillation amplitude also changes the step position and reaches maximum values for each foot to cross the gaps.

As seen in bottom right of Figure 4, the offset term is an important component for inter-limb coordination and can explicitly modulate the step position. On average it has the highest and lowest values when the foot starts and stops crossing a gap. Figure 4 (middle) interestingly shows the robot places its HL limb approximately where the FL limb was located in the previous stride.

D. Hardware Experiment

We perform a sim-to-real transfer of policy 1 from Section IV-B to the Go1 hardware for a two gap scenario with widths of 15 cm and 7 cm. We simplify the sim-to-real transfer by using a trained neural network to capture the actuator dynamics [23], [53], and we assume knowledge of the relative gap distance to the robot from an equivalent scenario completed in simulation. Figure 1-A shows snapshots of trotting over the gaps with a mean velocity of 0.7 m/s.

V. Conclusion

In this work, we have proposed a framework to investigate the interactions between supraspinal drive and the CPG to generate anticipatory quadruped locomotion in gap crossing scenarios. Our results show that supraspinal drive is critical for high success rates for gap crossing, but CPG dynamics are beneficial for energy efficiency and gait smoothness. Moreover, our results show that the front foot distance to the gap is the most important and sufficient visually-extracted sensory information for learning gap crossing scenarios. This supports the biological hypothesis that cats and horses control their front legs for obstacle avoidance, and that hind legs follow an internal memory based on the front feet information [46], [47]. This shows that DRL is able to create and encode an internal kinematic model with proprioceptive sensing to modulate the hind leg motion for gap crossing. Furthermore, in contrast to previous work, to the best of our knowledge, this is the first RL framework with gap-crossing capability without having a dynamical model, MPC, curriculum, or mentor in the loop.

Acknowledgements

We would like to thank Alessandro Crespi for assisting with hardware setup.

REFERENCES

[1] S. Grillner and A. El Manira, "Current principles of motor control, with special reference to vertebrate locomotion," *Physiological reviews*, vol. 100, no. 1, pp. 271–320, 2020.

[2] T. Miki, J. Lee, J. Hwanbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, 2022.

[3] A. J. Ijspeert, "Biorobotics: Using robots to emulate and investigate agile locomotion," *science*, vol. 346, no. 6206, pp. 196–203, 2014.

[4] A. Spröwitz, A. Tuleu, M. Vespignani, M. Ajallooeian, E. Badri, and A. J. Ijspeert, "Towards dynamic trot gait locomotion: Design, control, and experiments with cheetah-cub, a compliant quadruped robot," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 932–950, 2013.

[5] D. J. Hyun, S. Seok, J. Lee, and S. Kim, "High speed trot-running: Implementation of a hierarchical controller using proprioceptive impedance control on the mit cheetah," *The International Journal of Robotics Research*, vol. 33, no. 11, pp. 1417–1445, 2014.

[6] S. Grillner and S. Rossignol, "On the initiation of the swing phase of locomotion in chronic spinal cats," *Brain research*, vol. 146, no. 2, pp. 269–277, 1978.

[7] M. Ajallooeian, S. Pouya, A. Sproewitz, and A. J. Ijspeert, "Central pattern generators augmented with virtual model control for quadruped rough terrain locomotion," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 3321–3328.

[8] S. Aoi, P. Manoonpong, Y. Ambe, F. Matsuno, and F. Wörgötter, "Adaptive control strategies for interlimb coordination in legged robots: a review," *Frontiers in neurorobotics*, vol. 11, p. 39, 2017.

[9] H. Kimura, Y. Fukuoka, and A. H. Cohen, "Adaptive dynamic walking of a quadruped robot on natural ground based on biological concepts," *The International Journal of Robotics Research*, vol. 26, no. 5, pp. 475–490, 2007.

[10] 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.

[11] 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, 2021.

[12] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots: A review," *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008, robotics and Neuroscience.

[13] L. Righetti and A. J. Ijspeert, "Pattern generators with sensory feedback for the control of quadruped locomotion," in *2008 IEEE International Conference on Robotics and Automation*. IEEE, 2008, pp. 819–824.

[14] 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, pp. 1–10, 2019.

[15] 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.

[16] Y. Fukuoka, Y. Habu, and T. Fukui, "A simple rule for quadrupedal gait generation determined by leg loading feedback: a modeling study," *Scientific reports*, vol. 5, no. 1, pp. 1–11, 2015.

[17] T. Fukui, H. Fujisawa, K. Otaka, and Y. Fukuoka, "Autonomous gait transition and galloping over unperceived obstacles of a quadruped robot with cpg modulated by vestibular feedback," *Robotics and Autonomous Systems*, vol. 111, pp. 1–19, 2019.

[18] A. A. Saputra, J. Botzheim, A. J. Ijspeert, and N. Kubota, "Combining reflexes and external sensory information in a neuromusculoskeletal model to control a quadruped robot," *IEEE Transactions on Cybernetics*, 2021.

[19] S. Gay, J. Santos-Victor, and A. Ijspeert, "Learning robot gait stability using neural networks as sensory feedback function for central pattern generators," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 194–201.

[20] M. Thor and P. Manoonpong, "Versatile modular neural locomotion control with fast learning," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 169–179, 2022.

[21] M. Thor, T. Kulvicius, and P. Manoonpong, "Generic neural locomotion control framework for legged robots," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 9, pp. 4013–4025, 2020.

[22] A. Iscen, K. Caluwaerts, J. Tan, T. Zhang, E. Coumans, V. Sindhwani, and V. Vanhoucke, "Policies modulating trajectory generators," in *Conference on Robot Learning*. PMLR, 2018, pp. 916–926.

[23] 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, 2019.

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

[25] J. Siekmann, K. Green, J. Warila, A. Fern, and J. Hurst, "Blind bipedal stair traversal via sim-to-real reinforcement learning," *arXiv preprint arXiv:2105.08328*, 2021.

[26] X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," 2020.

[27] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.

[28] G. Ji, J. Mun, H. Kim, and J. Hwangbo, "Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4630–4637, 2022.

[29] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," *arXiv preprint arXiv:2205.02824*, 2022.

[30] G. Bellegarda, Y. Chen, Z. Liu, and Q. Nguyen, "Robust high-speed running for quadruped robots via deep reinforcement learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 10 364–10 370.

[31] G. Bellegarda and A. Ijspeert, "CPG-RL: Learning central pattern generators for quadruped locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 547–12 554, 2022.

[32] G. Bellegarda and K. Byl, "Training in task space to speed up and guide reinforcement learning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 2693–2699.

[33] G. Bellegarda and Q. Nguyen, "Robust quadruped jumping via deep reinforcement learning," *arXiv preprint arXiv:2011.07089*, 2020.

[34] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing energy consumption leads to the emergence of gaits in legged robots," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 928–937.

[35] 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.

[36] Y. Yang, T. Zhang, E. Coumans, J. Tan, and B. Boots, "Fast and efficient locomotion via learned gait transitions," in *Conference on Robot Learning*. PMLR, 2022, pp. 773–783.

[37] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[38] R. Yang, M. Zhang, N. Hansen, H. Xu, and X. Wang, "Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers," in *International Conference on Learning Representations*, 2022.

[39] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 91–100.

[40] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, "Visual-locomotion: Learning to walk on complex terrains with vision," in *Conference on Robot Learning*. PMLR, 2022, pp. 1291–1302.

[41] Z. Xie, X. Da, B. Babich, A. Garg, and M. van de Panne, "Glide: Generalizable quadrupedal locomotion in diverse environments with a centroidal model," *arXiv preprint arXiv:2104.09771*, 2021.

[42] K.-H. Lee, O. Nachum, T. Zhang, S. Guadarrama, J. Tan, and W. Yu, "Pi-ars: Accelerating evolution-learned visual-locomotion with predictive information representations," *arXiv preprint arXiv:2207.13224*, 2022.

[43] G. B. Margolis, T. Chen, K. Paigwar, X. Fu, D. Kim, S. b. Kim, and P. Agrawal, "Learning to jump from pixels," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 164. PMLR, 08–11 Nov 2022, pp. 1025–1034.

[44] Z. Xie, H. Y. Ling, N. H. Kim, and M. van de Panne, "Allsteps: Curriculum-driven learning of stepping stone skills," in *Computer Graphics Forum*, vol. 39, no. 8. Wiley Online Library, 2020, pp. 213–224.

[45] A. Iscen, G. Yu, A. Escontrela, D. Jain, J. Tan, and K. Caluwaerts, "Learning agile locomotion skills with a mentor," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 2019–2025.

[46] D. McVea and K. Pearson, "Contextual learning and obstacle memory in the walking cat," *Integrative and Comparative Biology*, vol. 47, no. 4, pp. 457–464, 2007.

[47] I. Q. Whishaw, L.-A. R. Sacrey, and B. Gorny, "Hind limb stepping over obstacles in the horse guided by place-object memory," *Behavioural brain research*, vol. 198, no. 2, pp. 372–379, 2009.

[48] I. A. Rybak, N. A. Shevtsova, M. Lafreniere-Roula, and D. A. McCrea, "Modelling spinal circuitry involved in locomotor pattern generation: insights from deletions during fictive locomotion," *The Journal of physiology*, vol. 577, no. 2, pp. 617–639, 2006.

[49] 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.

[50] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," http://pybullet.org, 2016–2019.

[51] Unitree Robotics. (2021, February) A1. [Online]. Available: https://www.unitree.com/products/a1/

[52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017.

[53] G. B. Margolis and P. Agrawal, "Walk these ways: Tuning robot control for generalization with multiplicity of behavior," *arXiv preprint arXiv:2212.03238*, 2022.