Hierarchical Decomposition of Terrain Adaptation for Quadrupedal Locomotion

Ji-Hyeon Yoo, Kartikeya Singh, Karthik Dantu

***Abstract*—Quadrupedal locomotion over heterogeneous terrain requires adaptation not only to terrain geometry, but also to changing physical interaction properties such as friction, compliance, contact conditions, payload, and robot dynamics. Existing learning-based methods address parts of this problem through perception, history-based adaptation, preference conditioning, and structured locomotion planning. However, these roles are often combined within a policy or latent representation, making it difficult to determine what should be adapted and where that adaptation should occur. We introduce TRACER, a hierarchical architecture that separates objective reasoning, robot–terrain interaction representation, gait-level strategy, and model-based physical execution.**

As an initial study of this hierarchy, we evaluate the objective-to-execution pathway using an objective-conditioned high-level policy and a fixed model-based low-level controller. Objective conditioning changes the high-level strategy, but reward-side improvements are not always preserved in episode-level physical outcomes. Because the architecture separates objective reasoning, strategy generation, and execution, these mismatches can be examined at specific interfaces rather than treated as a single policy failure. The study shows how the proposed decomposition helps separate different failure sources and motivates future work on interaction representation, richer gait-level strategy, and layer-specific online adaptation.

I. Introduction

Quadrupedal robots are expected to operate across environments with diverse geometric and physical properties, ranging from stairs and rough terrain to sand, mud, slippery surfaces, and deformable ground. Across these environments, the same velocity command or locomotion policy does not necessarily produce the same physical outcome, because locomotion depends not only on terrain geometry but also on physical interaction factors such as friction, ground compliance, contact conditions, payload, and changes in robot dynamics [1], [2]. Heterogeneous-terrain locomotion is therefore not merely a problem of recognizing terrain or learning terrain-specific behaviors. The same point-goal task may call for different trade-offs among fast progress, stable locomotion, and low energy use. Likewise, an unexpected failure may come from an unsuitable objective, an inaccurate estimate of the robot–terrain interaction, insufficient gait-level authority, or a low-level execution limit. When these roles are coupled inside one policy, these causes can be difficult to distinguish.

Recent learning-based locomotion methods address these challenges through perceptive control and history-based dynamics adaptation [1]–[3], preference-conditioned learning and terrain-adaptive objectives [4]–[6], and hierarchical or structured planning [7]–[9]. These mechanisms adapt different parts of the locomotion system. For example, RMA infers hidden dynamics from recent proprioceptive history [1], preference-conditioned methods vary objective trade-offs [5], and RLOC separates learned footstep planning from model-based execution [7]. In several learning-based systems, however, multiple adaptation roles remain coupled within a policy or latent representation. This can make it difficult to identify whether a failure comes from objective reasoning, interaction representation, insufficient gait-level authority, or low-level feasibility. This motivates a design question: how should different forms of adaptation be organized and connected within a locomotion hierarchy?

To investigate this question, we propose TRACER, a hierarchical framework that distinguishes four functional roles: Objective, which determines what should be prioritized; Interaction, which characterizes the current robot–terrain interaction; Strategy, which selects a locomotion strategy; and Execution, which realizes that strategy under robot dynamics and contact constraints. These four levels are not four required learned subpolicies. Each role may be implemented by a learned model, a fixed mapping, or a model-based controller depending on the function it serves. TRACER further distinguishes an externally meaningful physical preference η from an internal policy-conditioning variable β. The former represents the desired trade-off among Motion, Stability, and Energy, whereas the latter modulates the high-level locomotion policy. We do not assume that η and β have the same physical meaning; instead, their relationship is treated as a calibration problem. High-level decisions are expressed through high-level meta-actions and passed as physically meaningful references to a downstream model-based controller.

Our design hypothesis is that separating Objective, Interaction, Strategy, and Execution through defined interfaces makes it easier to identify where an adaptation failure occurs and which level should be corrected.

This paper tests the Objective–Strategy–Execution part of the hypothesis with a fixed low-level controller and a current 3-D Strategy action $[v_x, \omega_z, h_{\mathrm{body}}]$; richer gait variables are held fixed. The results show that objective conditioning changes the Strategy output, but these changes are not always preserved in physical outcomes. Within this 3-D Strategy space, most objective-induced action variation lies along a common aggressive–conservative direction. These observations illustrate how the proposed decomposition can help localize different sources of adaptation failure.

The contributions of this work are:

- A hierarchical architecture for terrain adaptation that separates the Objective, Interaction, Strategy, and Execution

roles.

- Defined interfaces linking Objective conditioning, Interaction representation, Strategy action, and Execution while keeping physical preference and physical outcome distinct from internal variables, allowing different adaptation failures to be localized within the hierarchy.
- An initial objective-to-execution study showing that objective information can alter high-level strategy without guaranteeing the same ordering in physical outcomes, which helps identify where mismatches appear in the hierarchy.

Ultimately, this work asks not only how to improve a locomotion policy, but where different forms of terrain adaptation should reside and what information should be preserved across module boundaries.

II. Related Work

A. Terrain-Aware Representation and Online Adaptation

Recent quadrupedal locomotion studies have increasingly used perception and interaction history to infer unobserved environmental properties and adapt locomotion behavior accordingly. Rapid Motor Adaptation separates a base policy from an adaptation module and infers latent information about hidden factors such as terrain properties, payload, and actuator conditions from recent proprioceptive history [1]. GRAM further extends this line of work toward robustness under both in-distribution and out-of-distribution dynamics [2]. Rather than explicitly estimating terrain parameters, these approaches transform interaction history into latent information that is useful for policy adaptation.

A related line of work focuses on improving the representation of the environment itself. World-model-based locomotion methods, such as WMP, learn predictive representations from vision and proprioception and use them to anticipate future robot behavior [3]. This suggests that predictive representations can provide interaction-relevant information beyond handcrafted terrain descriptors.

TRACER is closely related to these approaches at the interaction level, but asks a related architectural question. Rather than only learning a latent representation for direct policy adaptation, TRACER investigates where such interaction information should enter the locomotion hierarchy, distinguishing its role from objective reasoning, gait-level strategy, and physical execution.

B. Reward Adaptation and Preference-Conditioned Locomotion

Another direction adapts the locomotion objective itself according to terrain or task requirements. Terrain-adaptive reward methods adjust reward terms or penalty weights based on environmental characteristics so that a single locomotion policy can express different behaviors across terrain conditions [4]. Preference-conditioned and multi-objective reinforcement learning similarly allow a policy to represent different trade-offs among competing objectives [5], while recent preference-learning approaches learn locomotion rewards from semantic comparisons over trajectories [6].

These studies provide important evidence that reward weighting and preference conditioning can shape locomotion behavior. TRACER also considers trade-offs among Motion, Stability, and Energy, but explicitly distinguishes a desired physical preference from the internal variable used to condition the locomotion policy. Specifically, the external preference η represents the desired trade-off among physical objectives, whereas the internal reward-weight coordinate β modulates policy behavior. TRACER therefore does not assume $\eta = \beta$

because the semantics of an internal reward weight need not coincide with its realized physical effect. Instead, their relationship is treated as a physical calibration problem. This distinction is motivated by our initial observations that changes in β can induce behavioral specialization without consistently producing the corresponding physical outcome.

C. Hierarchical and Structured Locomotion Planning

Hierarchical locomotion methods provide another important foundation for TRACER by separating high-level decision making from low-level physical execution. RLOC, for example, uses a reinforcement-learning policy to generate footstep plans from sensory information and desired motion commands, while a model-based motion controller is responsible for realizing those plans [7]. Such separation demonstrates that learned locomotion planning and model-based execution can be treated as distinct functional layers.

Structured planning approaches further show that learning need not operate directly in the joint-level action space. EEMP exposes physically meaningful planner parameters, such as foot-placement regions, swing timing, step height, and body height, as the variables to be optimized [8]. This motivates the use of semantic locomotion parameters as an interface between learning-based high-level reasoning and downstream model-based control.

Continuous gait representations provide an additional perspective on structured action spaces. Gaitor learns a continuous latent gait space in which multiple gait patterns and transitions can be represented within a common policy [9]. Such approaches suggest that gait adaptation can potentially be expressed through a structured and continuous strategy representation rather than through direct low-level motor commands. This direction is particularly relevant to the future expansion of gait-level action authority in TRACER.

D. Positioning of TRACER

Prior work has demonstrated methods for rapid dynamics adaptation, predictive environment representation, terrain-dependent reward adaptation, preference-conditioned locomotion, structured gait planning, and model-based execution. These methods address important but often distinct subproblems of adaptive locomotion. In contrast, TRACER focuses on how physical task preference, inferred robot–terrain interaction, gait-level strategy, and physical execution should be organized and connected within a common locomotion hierarchy.

The goal of TRACER is therefore not to claim a new mechanism for each individual form of adaptation. Instead,

it explores a modular decomposition in which objective, interaction, strategy, and execution are assigned to distinct levels, allowing the interfaces between them—and allowing the interfaces between them—and mismatches across those interfaces—to be analyzed directly.

III. Problem Formulation and Hierarchical Architecture

A. Problem Formulation

For a given objective-conditioning variable β, we model high-level locomotion as a member of an objective-conditioned family of partially observable decision processes,

$$\mathcal{M}_\beta = (\mathcal{S}, \mathcal{A}, \mathcal{O}, p, \Omega, R_\beta, \rho_0, \gamma),$$

where $\mathcal{S}$ is the underlying robot–environment state space, including robot state and environment or contact properties that govern the dynamics; $\mathcal{A}$ is the high-level Strategy action space with $\Theta_t \in \mathcal{A}$; and $\mathcal{O}$ is the observation space available to the high-level planner. The transition kernel p describes the high-level state evolution induced by the robot dynamics and the Execution controller, Ω is the observation model, R_β is the reward under objective conditioning β, ρ_0 is the initial-state distribution, and γ is the discount factor.

At each high-level decision step, the planner receives $o_t \in \mathcal{O}$, including robot-state estimates and available terrain or context observations, and outputs $\Theta_t \in \mathcal{A}$. The task g_t and physical preference η_t are external conditioning variables. The Objective level may use context features c_t derived from available task or environment information to select β_t. Physical interaction properties such as friction and compliance need not be identifiable from a single o_t, which motivates the history-based Interaction representation introduced below.

Within this decision process, TRACER separates objective selection, interaction inference, strategy generation, and physical execution as distinct functional roles. To describe the interfaces among them, we distinguish the high-level task specification, physical preference, internal objective-conditioning variable, high-level locomotion strategy, and physical outcome.

At time t, we denote the high-level task specification by g_t, which may include mission-level requirements such as a goal position, commanded velocity, and heading:

$$g_t = \left\{ p_{\mathrm{goal}}, v^{\mathrm{cmd}}, \omega^{\mathrm{cmd}}, \dots \right\}.$$

In the initial experiments considered in this paper, we use a short point-goal locomotion task as a minimal task specification, while focusing the evaluation on objective-conditioned behavior rather than general mission-level planning.

While the task specifies *what the robot should accomplish*, the physical preference η_t specifies *which physical properties should be prioritized* while accomplishing the task. We represent this preference as a trade-off among Motion, Stability, and Energy:

$$\eta_t = [\eta_M, \eta_S, \eta_E], \qquad \eta_i \geq 0, \qquad \sum_i \eta_i = 1.$$

For example, the same locomotion task may correspond to different preferences depending on whether motion performance, stability, or energy efficiency is emphasized.

In contrast, β_t denotes an internal objective-conditioning variable used to modulate the high-level locomotion policy:

$$\beta_t = [\beta_M, \beta_S, \beta_E], \qquad \beta_i \geq 0, \qquad \sum_i \beta_i = 1.$$

Importantly, we do not assume that η_t and β_t have the same physical meaning. The former represents a desired physical trade-off, whereas the latter controls the relative importance of reward components inside the policy. Therefore, $\eta_t \neq \beta_t$, and increasing a particular reward weight does not directly guarantee an improvement in the physical quantity associated with the same objective label.

The physical outcome after locomotion is represented as

$$\mathbf{J}_t = [J_M, J_S, J_E],$$

where J_M, J_S, and J_E denote metrics associated with realized motion performance, physical stability, and energy consumption, respectively. The objective-to-outcome flow examined in the initial study can therefore be summarized as

$$(g_t, \eta_t) \rightarrow \beta_t \rightarrow \Theta_t \rightarrow \mathbf{J}_t.$$

Thus, the desired physical preference, internal objective conditioning, locomotion strategy, and physical outcome are treated as separate variables whose relationships can be analyzed separately.

B. Hierarchical Decomposition of Terrain Adaptation

TRACER decomposes heterogeneous-terrain adaptation into four levels: *objective*, *interaction*, *strategy*, and *execution*. Figure 1 summarizes the architecture and the corresponding module mappings.

Objective and Interaction are distinct upstream roles rather than sequential stages. The Objective level determines what should be prioritized through β_t, whereas the Interaction level represents the experienced robot–terrain dynamics through z_t^{int}. The Strategy level combines these inputs into high-level locomotion references, and the model-based Execution level realizes those references on the robot.

The purpose of this decomposition is to keep these adaptation roles separate. This makes it possible to examine where information changes or is lost across the hierarchy, and it forms the basis of the interface study in Section IV.

Objective-level adaptation determines which objective should be emphasized given the current task, physical preference, and contextual information. Conceptually,

$$(g_t, \eta_t, c_t) \rightarrow \beta_t,$$

where c_t denotes contextual information related to the terrain and robot state.

Interaction-level adaptation characterizes the dynamics of the current robot–terrain interaction. Even under the same terrain geometry or semantic label, the realized robot response may

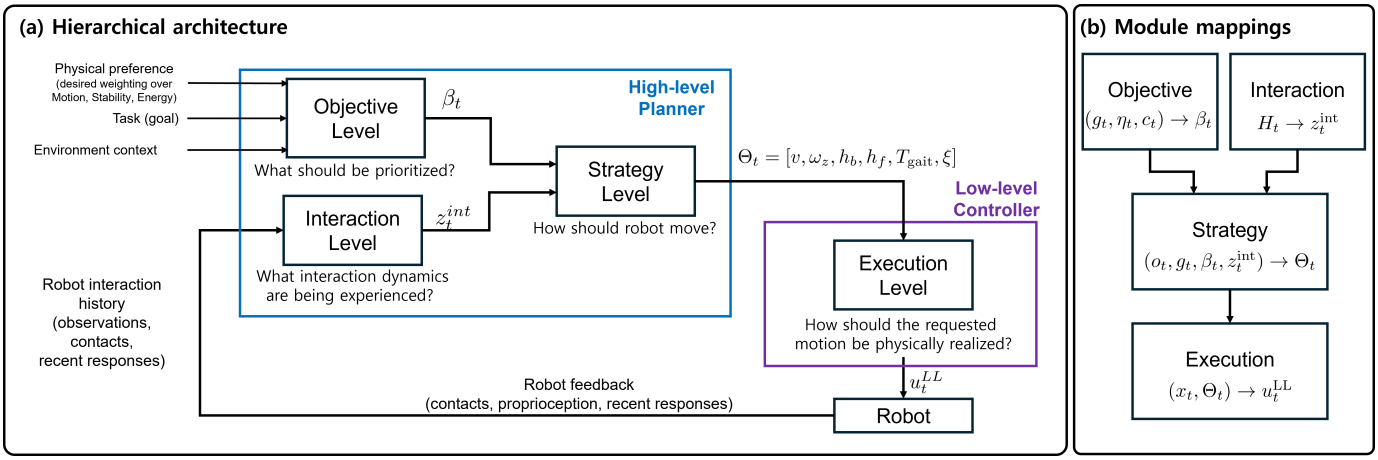

Fig. 1. TRACER hierarchy and module mappings. (a) Objective, Interaction, Strategy, and Execution are separated as functional roles. (b) Explicit input–output mappings at the module boundaries. The present study evaluates the Objective–Strategy–Execution path with a fixed Execution controller; the Interaction path is not instantiated in the present study.

vary due to friction, ground compliance, payload, or contact conditions. We therefore consider a recent interaction history

$$H_t = \{o_{t-K:t}, \Theta_{t-K:t-1}\},$$

from which an interaction representation z_t^{int} may be inferred:

$$z_t^{\text{int}} = E(H_t).$$

In this paper, the Interaction level is part of the intended architecture, while a concrete predictive representation and online adaptation mechanism are left for future extensions.

Strategy-level adaptation determines a high-level locomotion strategy from the current observation, task, objective conditioning, and interaction information. We use π_{HL} to denote the learned high-level policy:

$$\Theta_t = \pi_{\text{HL}}\left(o_t, g_t, \beta_t, z_t^{\text{int}}\right).$$

The intended Θ_t is not a low-level motor command such as joint torque or joint position. Instead, it is a high-level meta-action composed of interpretable locomotion parameters, including commanded translational and yaw motion, body height, swing-foot clearance, gait period, and duty factor:

$$\Theta_t = [v_x, \omega_z, h_{\text{body}}, h_{\text{swing}}, T_{\text{gait}}, \xi, \ldots], \tag{1}$$

where the current study uses only the forward component v_x of the translational velocity command; planar translation can be added in later implementations.

Finally, the execution level realizes the requested locomotion strategy under the robot dynamics and contact constraints:

$$u_t^{\text{LL}} = \mathcal{C}_{\text{LL}}\left(x_t, \Theta_t\right).$$

Here, x_t denotes the low-level robot state estimate, $\mathcal{C}_{\text{LL}}$ denotes the model-based low-level controller, and u_t^{LL} denotes its control output. The Execution level is responsible for tracking, contact-force realization, and stabilization.

The present study evaluates only part of this hierarchy: the low-level Execution module is fixed, and the current Strategy action contains only $[v_x, \omega_z, h_{\text{body}}]$. Swing clearance, gait period, and duty factor remain fixed. This isolates whether objective conditioning changes high-level behavior before richer gait variables are introduced.

C. Preference, Conditioning, and Physical Calibration

The internal variable β conditions the policy; it does not directly specify a physical outcome. For a high-level policy $\Theta_t = \pi_{\text{HL}}(o_t, g_t, \beta_t, z_t^{\text{int}})$, different values of β generate different trajectories τ_β and therefore different physical outcomes,

$$\mathbf{J}(c, \beta) = [J_M(\tau_\beta), J_S(\tau_\beta), J_E(\tau_\beta)].$$

A larger weight β_i therefore need not minimize the corresponding physical metric J_i.

Given a desired physical preference η, the Objective level can instead be viewed as a calibration problem:

$$\beta^*(c, \eta) = \arg\min_{\beta \in \mathcal{B}(c)} \Omega_\eta(\mathbf{J}(c, \beta)). \tag{2}$$

Here, $\mathcal{B}(c)$ is the set of conditioning vectors that yield feasible locomotion under context c, and Ω_η evaluates the resulting trade-off. TRACER therefore treats η and β as distinct variables and asks whether the intended preference is preserved across the downstream interfaces.

D. Initial Objective-to-Execution Study

The current study isolates the objective-to-execution pathway while keeping the Execution level fixed:

$$\beta_t \to \pi_{\text{HL}} \to \Theta_t^{\text{3D}} \to \mathcal{C}_{\text{LL}}.$$

The objective-related reward is

$$r_t = -(\beta_M C_{M,t} + \beta_S C_{S,t} + \beta_E C_{E,t}) - C_{\text{aux},t},$$

where $C_{\text{aux},t}$ is independent of β. The current study uses the 3-D Strategy output

$$\Theta_t^{\text{3D}} = [\, v_x,\ \omega_z,\ h_{\text{body}}\,],$$

while the remaining gait variables in Eq. (1) are fixed. This setup tests two questions: whether objective information reaches the Strategy level, and whether the resulting behavior preserves the intended physical trade-off after execution.

IV. Initial Study of Hierarchical Interfaces

We examine two interfaces in the objective-to-execution pathway: $\mathcal{I}_1 : \beta \to \Theta^{3D}$, which tests whether objective conditioning changes the Strategy level, and $\mathcal{I}_2 : \Theta^{3D} \to \mathbf{J}$, which tests whether the resulting change is preserved in physical outcomes. The purpose of the experiment is to use the hierarchy as a diagnostic structure, not to validate the complete architecture.

A. Experimental Protocol

We evaluate the study described in Section III-D using a fixed Quadruped-PyMPC execution layer [10]. A single trained high-level policy is evaluated on nine held-out rough Perlin terrains. The richer gait variables remain fixed at nominal values, so all comparisons share the same low-level controller and gait schedule.

For controlled evaluation, we use four conditioning anchors:

$$\begin{aligned}\beta^{\mathrm{bal}} &= [\,1/3,\ 1/3,\ 1/3\,],\\ \beta^{M} &= [\,0.70,\ 0.15,\ 0.15\,],\\ \beta^{S} &= [\,0.15,\ 0.70,\ 0.15\,],\\ \beta^{E} &= [\,0.15,\ 0.15,\ 0.70\,].\end{aligned}$$

These are internal conditioning settings, not physical preference vectors η. We compare Fixed, a constant command equal to the mean Balanced command over the training set; Balanced, the learned policy with β^{bal}; and Objective-conditioned, the same policy with β^{M}, β^{S}, or β^{E}.

B. Objective Metrics and Comparisons

For rollout-level comparison, each objective cost is averaged over the N high-level decisions,

$$\bar{C}_i(\tau) = \frac{1}{N}\sum_{t=1}^{N} C_{i,t}, \qquad i \in \{M, S, E\}.$$

Task-feasibility and hard safety terms remain independent of β.

a) Motion cost: Motion uses realized progress toward the point goal:

$$\begin{aligned}v_{\mathrm{prog},t} &= \frac{d_{t-1} - d_t}{\Delta t},\\ C_{M,t} &= \frac{1}{2}\left[1 - \mathrm{clip}\left(\frac{v_{\mathrm{prog},t}}{v_{\max}}, -1, 1\right)\right],\end{aligned}$$

with $v_{\max} = 0.40$ m/s.

b) Stability cost: Stability combines posture and traction burdens,

$$C_{\mathrm{posture},t} = \max\left\{\left(\frac{|\phi_t|}{\phi_{\mathrm{unsafe}}}\right)^2, \left(\frac{|\theta_t|}{\theta_{\mathrm{unsafe}}}\right)^2\right\},$$

$$c_{\mathrm{tr}}(v_{\mathrm{slip}}) = \left[\frac{\max(0, v_{\mathrm{slip}} - v_{\mathrm{db}})}{v_{\mathrm{unsafe}} - v_{\mathrm{db}}}\right]^2,$$

where $v_{\mathrm{db}} = 0.002$ m/s and $v_{\mathrm{unsafe}} = 0.150$ m/s. The decision-level traction cost is the contact-time average over established stance contacts, and

$$C_{S,t} = \max(C_{\mathrm{posture},t}, C_{\mathrm{traction},t}).$$

c) Energy cost: Energy is based on measured absolute actuator mechanical work,

$$C_{E,t} = \frac{\Delta E_{\mathrm{abs},t}}{P_{\mathrm{ref}}\,\Delta t_{\mathrm{HL}}^{\mathrm{nom}}\,s_E},$$

with fixed normalization constants across all experiments. Because the reward-side comparison uses mean interval cost $\bar{C}_E$ while the external metric is total episode energy, the two need not have the same ordering.

For each held-out context c, we define a directional improvement for objective i when

$$\bar{C}_i(c, \beta^{(i)}) < \bar{C}_i(c, \beta^{\mathrm{bal}}).$$

We also record whether the intended conditioning produces the minimum target cost among all four conditioning settings. The first comparison tests direction relative to Balanced; the second tests whether the intended conditioning is the best of the four for its target cost.

C. Interface I: Objective Conditioning to Strategy

Objective conditioning changes the Strategy level. Relative to Balanced, Motion-oriented conditioning reduces $\bar{C}_M$ in $9/9$ held-out contexts, Energy-oriented conditioning reduces $\bar{C}_E$ in $8/9$, and Stability-oriented conditioning reduces $\bar{C}_S$ in $5/9$. The Motion and Energy directional results give exact paired sign-test values of $p = 0.0039$ and $p = 0.0391$, respectively.

Across all four conditioning settings, the intended conditioning produces the minimum target cost in $7/9$ Motion contexts, $8/9$ Energy contexts, and $1/9$ Stability contexts. Balanced also does not consistently outperform the Fixed baseline. Together, these results show that changing β reaches the Strategy level, especially for Motion and Energy, but the three objectives do not produce equally distinct strategies.

D. Interface II: Strategy to Physical Outcome

Table I compares reward-side improvement with episode-level physical improvement under the same held-out contexts.

Motion and Energy show consistent reward-side improvement, but their episode-level metrics do not preserve the same ordering. For Motion, a change in progress cost need not move completion time by a full 0.20 s decision step. For Energy, $\bar{C}_E$ is a mean interval cost while the external metric is total episode energy. Stability is weaker at both interfaces, and C_S

TABLE I
INTERFACE-LEVEL COMPARISON ACROSS NINE HELD-OUT CONTEXTS. ENTRIES COUNT CONTEXTS WITH IMPROVEMENT RELATIVE TO BALANCED CONDITIONING.

Objective	Reward-side improvement	Physical-outcome improvement
Motion	9/9	2/9 completion time (7 ties)
Stability	5/9	1/9 pitch RMS 4/9 roll RMS
Energy	8/9	3/9 episode energy

also includes traction while roll and pitch RMS measure attitude only. The hierarchy therefore separates two questions that would otherwise be mixed together: whether objective information changes the Strategy level, and whether that change has the intended physical effect.

E. Strategy Authority in the Current 3-D Space

We next examine whether the current 3-D Strategy space limits separation among objectives. PCA of the mean normalized actions $[v_x, \omega_z, h_{\mathrm{body}}]$ shows that PC1 explains 96.09% of the objective-induced variation; PC2 and PC3 explain 3.68% and 0.23%. The PC1 direction is approximately $[0.688, -0.336, -0.643]$. Relative to Balanced, the objective-induced directions have $\cos(M, S) = 0.790$, $\cos(M, E) = -0.950$, and $\cos(S, E) = -0.940$.

Most variation therefore lies along one aggressive–conservative direction: Motion and Stability are positively aligned, while Energy points in the opposite direction. Because the current 3-D Strategy space does not expose swing clearance, gait period, or duty factor, Stability may lack an independent contact-structural control channel. This analysis does not prove that the 3-D action space is the only cause, but it motivates expanding the Strategy level with additional gait variables.

Overall, the initial study shows that objective transmission, physical alignment, and strategy authority are different interface-level problems within the hierarchy.

V. ARCHITECTURAL IMPLICATIONS AND FUTURE DIRECTIONS

The results motivate the main architectural premise of TRACER: different adaptation problems should be assigned to different levels, and each module boundary should have a defined role.

A. Physical Preference Calibration

The gap between reward-side and physical outcomes supports keeping the desired preference η separate from the internal conditioning coordinate β. Rather than directly predicting a reward-weight vector from terrain context, the Objective level can evaluate candidate conditionings according to their expected physical trade-offs under the current context.

One possible implementation is a context-conditioned regret model over a feasible set of candidate β values. Given a desired preference η, the selector can choose

$$\beta^* = \arg\min_{\beta \in \mathcal{B}(c)} \hat{R}(c, \beta; \eta),$$

where $\hat{R}$ estimates the regret of each internal conditioning relative to the desired physical trade-off. This is a learned approximation of the calibration objective in Eq. (2). It keeps the external preference separate from the internal policy coordinate and allows their mapping to change with robot–terrain context.

B. Interaction Representation

The full hierarchy assigns robot–terrain interaction inference to a separate Interaction level. Recent observations and executed strategies can be used to summarize the current interaction, and prediction error can indicate when the observed response differs from expectation.

A direct mapping from one learned latent to another does not define what information should be preserved at the module boundary. We instead propose to train the interaction representation to predict action-dependent robot responses. This gives the representation a testable role even if individual latent dimensions do not have physical labels.

C. Strategy Authority

The current 3-D Strategy space expresses most objective variation along one direction. Expanding toward the gait variables in Eq. (1) can provide direct control over contact timing and gait shape, which may help separate Stability-oriented behavior from Motion and Energy. A future gait-selection module can use the interaction representation together with the selected objective to choose first among discrete gait regimes and later among continuous contact-schedule variables such as gait period, duty factor, and phase offsets.

These extensions share a design rule: separating the system into modules is not sufficient by itself. A direct context-to-β mapping and a direct interaction-latent-to-gait-latent mapping are useful baselines, but neither defines what information should be preserved at the module boundary. TRACER therefore treats interface design as a separate problem: the Objective interface uses predicted physical trade-offs, while the Interaction–Strategy interface uses action-dependent response prediction.

D. Layer-Specific Online Correction

The same decomposition provides a simple view of online correction: interaction-model errors belong at the Interaction level, preference-to-outcome errors at the Objective level, insufficient locomotion authority at the Strategy level, and persistent tracking or contact-model errors at the Execution level. This makes it possible to ask where a mismatch occurs instead of treating every mismatch as a generic policy failure.

VI. Conclusion

TRACER proposes a modular hierarchy for terrain-adaptive quadrupedal locomotion with four roles: Objective, Interaction, Strategy, and Execution. The main question is not only where each form of adaptation should occur, but also what information should cross each module boundary.

The initial objective-to-execution study shows why this separation is useful. Objective conditioning changes the high-level Strategy, but the resulting physical outcome does not always preserve the same objective ordering, and the current 3-D Strategy space limits separation among objectives. These observations do not validate the complete hierarchy. They show how a modular decomposition with defined interfaces can make different failure sources easier to identify and provide a clear direction for extending each level.

References

[1] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid motor adaptation for legged robots," in *Proc. Robotics: Science and Systems (RSS)*, 2021.

[2] J. Queeney, X. Cai, A. Schperberg, R. Corcodel, M. Benosman, and J. P. How, "GRAM: Generalization in deep RL with a robust adaptation module," *IEEE Robot. Autom. Lett.*, vol. 11, no. 2, pp. 2218–2225, 2026.

[3] H. Lai, J. Cao, J. Xu, H. Wu, Y. Lin, T. Kong, Y. Yu, and W. Zhang, "World model-based perception for visual legged locomotion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 11531–11537, 2025.

[4] Y. Dong, J. Ma, Y. Lu, J. Zhang, W. Li, Y. Chen, T. Zhang, X. Chen, Z. Yu, and P. Lu, "MGDP: Mastering a generalized depth perception model for quadruped locomotion," *Advanced Science*, vol. 13, no. 34, Art. no. e24345, 2026.

[5] T. Basaklar, S. Gumussoy, and U. Y. Ogras, "PD-MORL: Preference-driven multi-objective reinforcement learning algorithm," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2023.

[6] X. Chen, M. Lin, S. Shi, and J. Campbell, "MAPL: Multi-objective preference learning for robot locomotion," *arXiv preprint arXiv:2606.25398*, 2026.

[7] S. Gangapurwala, M. Geisert, R. Orsolino, M. Fallon, and I. Havoutis, "RLOC: Terrain-aware legged locomotion using reinforcement learning and optimal control," *IEEE Trans. Robot.*, vol. 38, no. 5, pp. 2908–2927, 2022.

[8] A. Schperberg, M. Menner, and S. Di Cairano, "Energy-efficient motion planner for legged robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 18888–18895, 2025.

[9] A. L. Mitchell, W. Merkt, A. Papatheodorou, I. Havoutis, and I. Posner, "Gaitor: Learning a unified representation across gaits for real-world quadruped locomotion," in *Proc. Conf. Robot Learn. (CoRL)*, vol. 270, pp. 780–793, 2025.

[10] M. Elobaid, G. Turrisi, L. Rapetti, G. Romualdi, S. Dafarra, T. Kawakami, T. Chaki, T. Yoshiike, C. Semini, and D. Pucci, "Adaptive non-linear centroidal MPC with stability guarantees for robust locomotion of legged robots," *IEEE Robot. Autom. Lett.*, vol. 10, no. 3, pp. 2806–2813, 2025.