IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

MILD: Tractable Terrain Modeling for Learning Improved Bipedal Locomotion on Deformable Surfaces

Zeren Luo[1], Jiahui Zhang[1], Zhe Xu[2], Wanyue Li[1], Xinqi Li[1], Xuechao Chen[2], Zhangguo Yu[2], Annan Tang[3,†], Peng Lu[1,†]

***Abstract*— Enabling robots to walk on yielding terrain is vital for applications ranging from disaster response to planetary exploration. While bipedal robots hold immense potential, their locomotion on deformable surfaces remains limited as current simulators fail to capture the spatiotemporal heterogeneity of such yielding substrates. We present MILD, featuring a physics-grounded discrete-element contact solver that accurately simulates spatially varying foot-terrain interactions. Complementing this model, we train a terrain-aware locomotion controller via deep reinforcement learning with latent modulation and proprioceptive estimation. Quantitative comparisons against state-of-the-art methods show our approach generates more diverse and realistic contact scenarios during training, resulting in controllers that exhibit natural adaptation on real deformable surfaces. Through hardware experiments, we demonstrate the system's capability for online terrain identification and adaptation across a wide range of surface stiffness.**

***Index Terms*— Contact modelling, Yielding terrain, Bipedal Robots, Reinforcement Learning**

I. INTRODUCTION

In recent years, a notable upsurge has been witnessed in the bipedal robotics sector, as it shows great potential for application in deformable and unstructured terrain, which constitutes a significant portion of the Earth's surface. Current bipedal robot locomotion controllers are predominantly optimized for rigid ground, as mainstream simulators [1] [2] [3] rely on rigid-body dynamics that cannot capture the spatiotemporal heterogeneity of deformable terrains. This fundamental limitation prevents accurate modeling of soft-surface interactions, ultimately restricting the generation of realistic data for controller design.

A. Soft Surface Simulation and Modeling

In the robotics physics community, exploring methods that can accurately and efficiently simulate the deformation and stress of soft substrates upon contact with robots has always been a highly valuable research direction. The study begins with granular media (GM), complex systems composed of discrete particles. Particle-based simulations can model grain interactions but are computationally expensive, often requiring days to simulate just one second of real-time behavior [4]. To elucidate the mechanical principles of GM interactions with robotic entities, researchers propose simplified contact mechanics such as linear viscoelastic models [5], [6] or rate-dependent plastic models [7], [8]. While computationally efficient, these models oversimplify the hydrodynamic-like nature of GM, leading to significant deviations from real-world behavior. For a more accurate description of GM mechanics, the resistive force theory (RFT) is introduced [9], [10], which is an empirical model that describes quasistatic reaction forces. It is applicable to biped quasistatic locomotion when combined with a ZMP-based controller [11]. The theory maintains its validity when extended to three-dimensional robots intrusion scenarios with GM [12], [13].

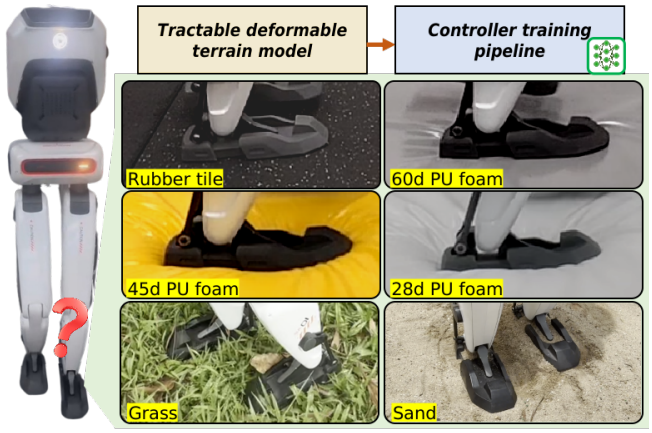

Fig. 1: The bipedal robot as a multi-articulated floating-base system interacting with compliant surfaces through integrated physics-grounded modeling and learning-based control. The locomotion is demonstrated across terrains with parametrically varied stiffness and deformability (d – density of the polyurethane foam. Unit: kg/m^3).

Despite this, RFT is primarily applicable to slow interactions failing to capture reaction forces during high-dynamic penetration. For high-speed penetration, researchers propose impulsive force models [14], [15], simplifying forces as instantaneous actions. Another study reveals limitations in these models, as high-speed penetration compresses and accumulates grains beneath the object, inducing an added mass effect that generates additional forces on the contact surface [16]. This developing cone principally reduces the formulation, and its tractable solution enhances computational efficiency in simulations. In this work, we extend it to a more generalized model, taking into account scenarios involving multiple target areas' interactions, which are overlooked in the original model.

B. Legged Locomotion on Deformable Terrain

Early research in this field is largely restricted to reduced-degree-of-freedom systems, such as unidirectional hoppers

This work was supported by General Research Fund under Grant No. 17204222. ([†]Corresponding author lupeng@hku.hk, tang@jsk.imi.i.u-tokyo.ac.jp).

[1]The authors are with the Adaptive Robotic Controls Lab (ArcLab), Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. [2]The authors are with School of Mechanical and Electrical Engineering, Beijing Institute of Technology, Beijing, China. [3]The author is with Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan.

[17], [5], [16] or simplified bipedal models [11], where the limited kinematic complexity enables tractable analysis and control design. Conventional model-based controllers fundamentally struggle with the coupled dynamics between the robot and deforming surface - particularly the violation of static contact assumptions that underpin their stability guarantees [18]. Passivity-based control frameworks [19], [20] resolve this limitation through an impedance wrench that regulates the center of mass and end-effector trajectory, enabling stable whole-body locomotion on compliant terrain.

Recent advances in reinforcement learning enable legged robots to traverse various challenging terrains, including deformable surfaces like grass and sand [21], [22]. These methods often combine implicit system identification with domain randomization over rigid terrain geometries to achieve generalization. [23] specifically simulates terrain compliance by randomizing parameters of MuJoCo's built-in mass-spring-damper contact model. While certain parameter combinations may partially capture characteristics of soft substrate interactions, this approximation remains fundamentally limited compared to real-world terrain penetration. The learned controllers's performance deteriorates considerably due to the lack of exposure to such data.

Recent studies attempt to establish more realistic models of yielding terrain in simulation for quadrupeds, aiming to provide higher-fidelity training data for RL-based controllers and bridge the reality gap [24], [25]. A key limitation in these works is their assumption that the reaction force is uniformly distributed across the contact area, which disregards the influence of insertion posture on contact force computation [7]. However, for bipedal robots, this assumption becomes invalid due to their large footplates, where contact forces vary significantly across different regions. In this work, the aim is to solve the aforementioned issues regarding the modeling accuracy and efficiency in an RL formulation. We propose **MILD**: a tractable **M**odel achieving **I**mproved bipedal **L**ocomotion on **D**eformables.

In summary, our walking controller introduces innovations in threefold: (i) We propose a tractable foot-terrain interaction model for substantial contact areas, overcoming the prevailing assumption of uniform force distribution in prior work (e.g., quadrupeds and 1-D hoppers). This model explicitly accounts for eccentric insertion and spatiotemporal heterogeneous penetration depths, enabling accurate force prediction for biped-scale footplates. (ii) Building on this, we develop a novel RL-based training framework with an encoder-modulation architecture that implicitly adapts to terrain compliance by dynamically encoding the properties into latent representations. (iii) Through extensive simulations and hardware validations, we demonstrate that the proposed model outperforms state-of-the-art methods in robustness and energy efficiency, and the controller's capability of online identifying and handling abrupt terrain stiffness transitions.

II. METHOD

Generalized contact solver for robot-deformable terrain interactions extends physics-grounded models through discrete-element segmentation with continuity constraints. Integrated into the simulators, it enables RL-based bipedal locomotion training. Section II-A - II-B illustrates the contact solver's derivation, and Section II-C details the RL problem formulation.

A. Discrete Element-based Granular Media Model

Prior work models terrain stiffness via depth-dependent forces on a developing cone, initially for circular contacts [16] and later extends to rectangular approximations [25]. However, both assume axisymmetric penetration and uniform force distribution—limitations that stem from analyzing only small foot areas.

To accurately capture contact force distribution in bipedal locomotion, we discretize the foot envelope surface into elements (Fig. 2(A)) whose size aligns with granular interaction assumptions. Based on preliminary experiments, a resolution of 154 elements (0.0195 m per unit) optimally balances contact richness and computational efficiency [1] . Each element undergoes independent penetration, forming distinct jamming cones that naturally yield non-uniform force distributions. For each penetrating element $e^{\{i\}}$, a growing compaction cone forms in the granular medium. The contact area consists of two components:

$$S_{flat}^{\{i\}}(h^{\{i\}}) = \pi(R - \nu\frac{h^{\{i\}} - h_0^{\{i\}}}{\tan\alpha})^2 \tag{1}$$

$$S_{cone}^{\{i\}}(h^{\{i\}}) = \frac{\pi R^2 - S_{flat}(h^{\{i\}})}{\cos\alpha}, \tag{2}$$

where $h^{\{i\}}$ is the penetration depth, α is the shear band angle, and ν is the recruitment rate. R is the span of the discrete element at the intersected point $h_0^{\{i\}}$ (subscript $\{i\}$ is omitted in the following equations for brevity). Added-mass m_A describes the accumulated effect of the compacted grains inside the jammed cone, which is given as follows:

$$m_A = -c_g\phi\rho\nu\int_{h_0}^{h} S_{flat}(z)dz, \tag{3}$$

where ϕ is the volume fraction, which indicates the degree of looseness in the interparticle connections. ρ and c_g are the grain density and surrounding mass scaling factor, respectively. Granular reaction force F_{GM} in the normal direction is composed of quasistatic force F_g and the force induced by the momentum change caused by collision with the virtual added mass $d(m_A\dot{h})$.

$$F_g = \frac{k_{pen}}{s^{\{i\}}}\int_{h_0}^{h} S_{flat}(z)dz + \sigma_{cone}\int_{h_0}^{h} S_{cone}(z)dz \tag{4}$$

$$F_{GM} = F_g - c_d\frac{dm_A}{dt}\dot{\boldsymbol{v}}_z - m_A\ddot{\boldsymbol{v}}_z \tag{5}$$

where k_{pen} is the penetration stiffness, c_d is the inertial drag scaling factor and σ_{cone} is the depth-dependent conical stress. Differential added mass $\frac{dm_A}{dt}\dot{\boldsymbol{v}}_z$ can be represented

[1]Beyond this resolution, the marginal gain in contact richness is outweighed by a prohibitive increase in computation time, which severely compromises training scalability.

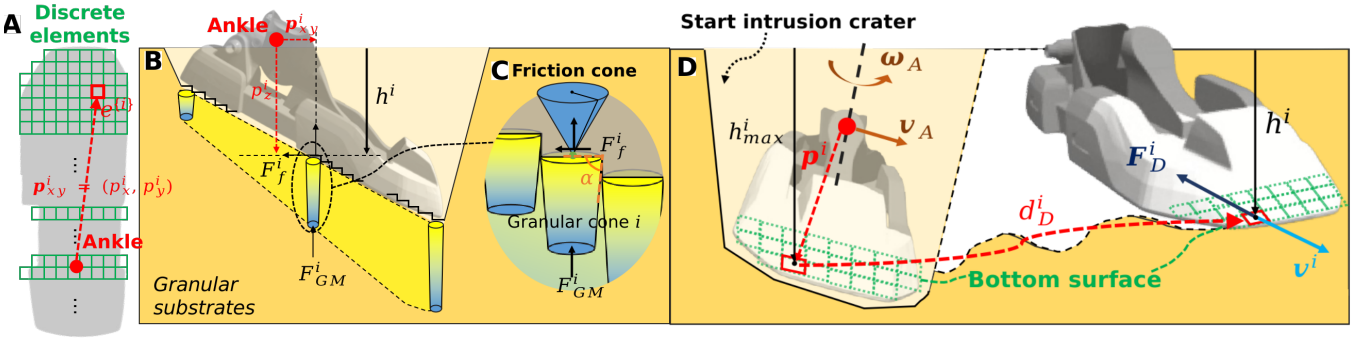

Fig. 2: Eccentric penetration mechanics of granular media (GM) under arbitrary-shaped feet. The footplate is discretized into elements (DEs), each modeling localized GM interactions via independent developing cones (Eq. (1) - (5)). Deeper intrusion amplifies the cone's added mass effect. The model requires each element to satisfy individual friction cone constraints while maintaining continuity constraints between element pairs (Eq. (11) - (13)). Drift-resistive forces are opposite to the element-wise motion and are scaled by the accumulated post-intrusion moving distance.

as $\frac{dm_A}{dh}\dot{\boldsymbol{v}}_z^2 = -c_g\phi\rho\nu S_{flat}\dot{\boldsymbol{v}}_z^2$ according to the chain rule. Besides, the velocity of $e^{\{i\}}$ w.r.t the world frame is accessible given the ankle's velocity $\boldsymbol{v}_A$, $\boldsymbol{\omega}_A$, and the vector from the ankle point to $e^{\{i\}}$, i.e. $\boldsymbol{v}^{\{i\}} = \boldsymbol{v}_A + \boldsymbol{\omega}_A \times \boldsymbol{p}^{\{i\}}$. It is worthwhile to note that the two integrations of Eqs. (4 - 5) possess closed-form expressions as functions of h and h_0. Our model generates the terrain configurations for different parameter combinations shown in Fig. 3.

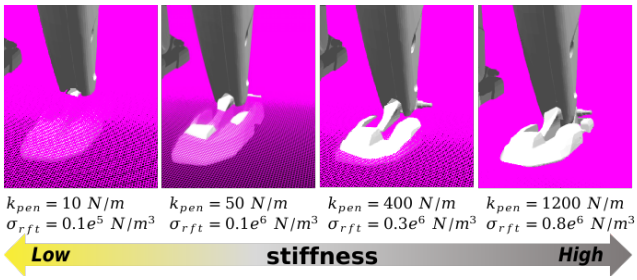

Fig. 3: Illustration of different penetration stiffness and depth-dependent resistive stress that lead to different foot penetration scenarios. The color grids indicate the terrain surface h_0.

B. Whole-body Dynamics and Kinematic-aware Contact Solver

1) System modeling: The vertical force computed via a discrete GM model determines the total contact count N_c, which initializes the tangential force solver. For all N_c contacts, the contact solver aims to drive the contact velocity to zero ($\boldsymbol{v}_{t+1}^{\{i\}} \to 0$) by iteratively adjusting tangential forces based on the predicted velocity deviation. To establish this, we bridge the system dynamics with contact interactions through the floating-based formulation.

The bipedal robot, as depicted in Fig. 1, is essentially a poly-articulated floating-base system that can be modeled as a series of interconnected rigid bodies. The resulting dynamics are therefore governed by the following Newton-Euler equation,

$$\boldsymbol{M}(\boldsymbol{q})\dot{\boldsymbol{\nu}} + \boldsymbol{b}(\boldsymbol{q},\boldsymbol{\nu}) = \boldsymbol{\tau} + \boldsymbol{J}_c^T(\boldsymbol{q})\boldsymbol{F}_c, \tag{6}$$

where $\boldsymbol{M}$ is the mass matrix, $\boldsymbol{\tau}$ is a generalized force, and $\boldsymbol{q}, \boldsymbol{\nu} \in \mathbb{R}^{6+n_j}$ is the generalized coordinate and velocity, which involves the base linear velocity $\boldsymbol{v}_b$, angular velocity $\boldsymbol{\omega}_b$ and the joint velocity. $\boldsymbol{b}$ represents the nonlinear effects (i.e. Coriolis, centrifugal and gravitational terms). By discretizing and mapping it to the world frame via jacobian $\boldsymbol{J}_c$, we obtain the velocity change at each contact point:

$$\boldsymbol{J}_c\Delta\boldsymbol{\nu} = \Delta\boldsymbol{v} = \boldsymbol{J}_c\boldsymbol{M}^{-1}\boldsymbol{J}_c^T\boldsymbol{I}_t + \boldsymbol{J}_c\boldsymbol{M}^{-1}\Delta t(\boldsymbol{\tau} - \boldsymbol{b}), \tag{7}$$

where $\boldsymbol{I}_t$ is the impulse generated by contact during interval Δt. Next, to distinguish the rotational effects on different end-effectors e, we introduce the notation $\{e, i\}$ to represent elements. We then generalize Eq. (7) to account for multiple contact interactions by superposing the velocity changes resulting from all active contacts. The contribution of $\boldsymbol{I}_t$ to both linear and angular velocity can be expressed as:

$$\begin{aligned}\boldsymbol{v}_{t+1}^{\{e,i\}} = \boldsymbol{v}_t^{\{e,i\}} &+ \sum_{j=1}^{N_c} \boldsymbol{J}_{c,\mathcal{L}}^{\{e,i\}}\boldsymbol{M}^{-1}\boldsymbol{J}_{c,\mathcal{L}}^{T\{e,j\}}\boldsymbol{I}_t^{\{e,j\}} \\ &+ \boldsymbol{J}_{c,\mathcal{L}}^{\{e,i\}}\boldsymbol{M}^{-1}\Delta t(\boldsymbol{\tau} - \boldsymbol{b})\end{aligned} \tag{8}$$

$$\begin{aligned}\boldsymbol{\omega}_{t+1}^{\{e\}} = \boldsymbol{\omega}_t^{\{e\}} &+ \sum_{j=1}^{N_c} \boldsymbol{J}_{c,\mathcal{R}}^{\{e\}}\boldsymbol{M}^{-1}\boldsymbol{J}_{c,\mathcal{L}}^{T\{e,j\}}\boldsymbol{I}_t^{\{e,j\}} \\ &+ \boldsymbol{J}_{c,\mathcal{R}}^{\{e\}}\boldsymbol{M}^{-1}\Delta t(\boldsymbol{\tau} - \boldsymbol{b})\end{aligned} \tag{9}$$

where $\boldsymbol{J}_{c,\mathcal{L}}$ and $\boldsymbol{J}_{c,\mathcal{R}}$ are the translational and rotational Jacobians that respectively map generalized velocities to translational and angular motion at contact points. $\boldsymbol{\omega}_{t+1}^{\{e\}}$ is the shared angular velocity of the end-effector's body (COM-referenced). It is noteworthy that in the Isaac Gym, only the Jacobian at the start of a rigid body, i.e., ankle point $\boldsymbol{J}_A$ is accessible. The Jacobian at each discrete element's center is obtained via the deduced rigid body transformation ($[\cdot]_\times$ denotes the skew matrix of the given vector).

$$\boldsymbol{J}^{\{i\}} = \begin{bmatrix}\boldsymbol{I}_{3\times3} & -[\boldsymbol{p}^{\{i\}}]_\times \\ 0_{3\times3} & \boldsymbol{I}_{3\times3}\end{bmatrix}\boldsymbol{J}_A \tag{10}$$

2) Physical constraint: The predicted velocities of individual elements at time $t+1$ are calculated in isolation. This ignores physical interactions between elements, causing discontinuous motion. To ensure consistency, a kinematic constraint is applied during impulse updates. Considering any two points i and j on the feet with a relative distance

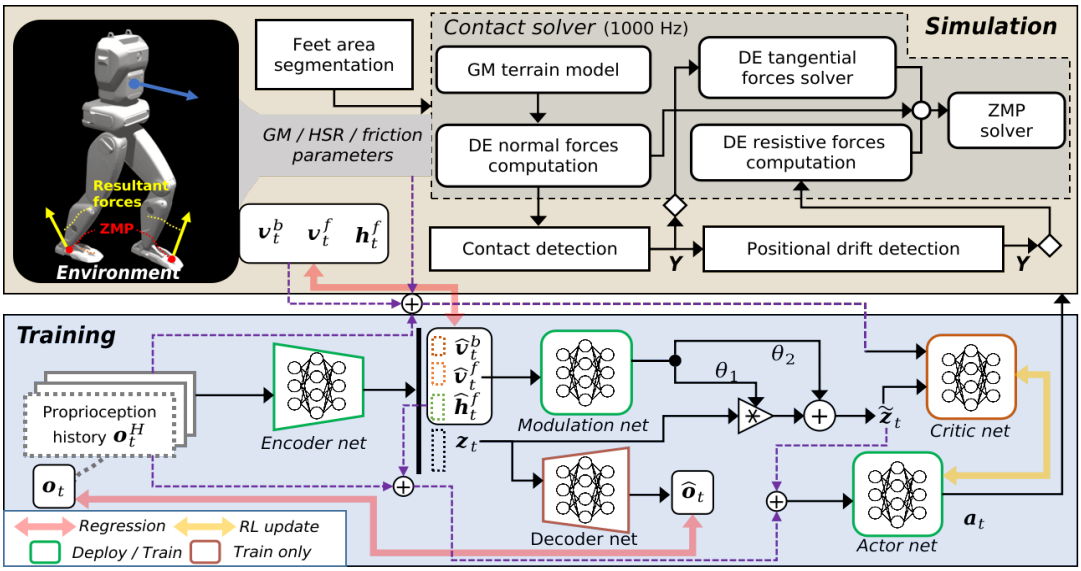

Fig. 4: Schematic of the simulation and training pipeline. The contact solver integrates depth-dependent granular dynamics and drifting resistance to model heterogeneous foot-terrain interactions. In the asymmetric actor-critic (AAC) training pipeline, a VAE-based estimator reconstructs privileged states from proprioception, while the modulation net dynamically adjusts latent terrain representations using estimated feet velocity to enable implicit adaptation.

$\boldsymbol{r}^{\{ij\}}$, the relative velocity error $\boldsymbol{\delta}_{rb}^{\{ij\}}$ between them can be expressed by the following formula:

$$\boldsymbol{\delta}_{rb}^{\{ij\}} = \|\boldsymbol{v}_{t+1}^{\{i\}} - \boldsymbol{v}_{t+1}^{\{j\}} - \boldsymbol{\omega}_{t+1} \times \boldsymbol{r}^{\{ij\}}\|, \{\forall i, j \in N_c\} \quad (11)$$

The average constraint violation $\boldsymbol{\delta}_{rb}^{\{i\}}$ is computed for element i by considering its relative motion with respect to all preceding elements, quantifying the kinematic inconsistency at the current optimization step.

$$\boldsymbol{\delta}_{rb}^{\{i\}} = \frac{1}{i-1} \sum_{j<i} \boldsymbol{\delta}_{rb}^{\{ij\}} \quad (12)$$

Moreover, the friction cone condition checks whether the tangential impulse $\boldsymbol{I}_{xy,t}^{\{i\}}$ of the current step — plus two corrective terms — lies within the friction cone formed by the normal impulse. The corrective terms are: (1) $\beta \boldsymbol{v}_{t+1}^{\{i\}}$, to drive the desired velocity to zero; and (2) $\gamma \boldsymbol{\delta}_{rb}^{\{i\}}$, to maintain kinematic coherency. This constraint is written as $\mathcal{C} : \mu \boldsymbol{I}_{z,t}^{\{i\}} > \|\boldsymbol{I}_{xy,t}^{\{i\}} - \gamma \boldsymbol{\delta}_{rb}^{\{i\}} - \beta \boldsymbol{v}_{t+1}^{\{i\}}\|$, which can be finally formalized as

$$\begin{cases} \boldsymbol{I}_{xy,t,next}^{\{i\}} = \boldsymbol{I}_{xy,t}^{\{i\}} - \gamma \boldsymbol{\delta}_{rb}^{\{i\}} - \beta \boldsymbol{v}_{t+1}^{\{i\}}, & \text{if } \mathcal{C} \\ \boldsymbol{I}_{xy,t,next}^{\{i\}} = \mu \boldsymbol{I}_{xy,t}^{\{i\}} \frac{\boldsymbol{I}_{xy,t}^{\{i\}} - \gamma \boldsymbol{\delta}_{rb}^{\{i\}} - \beta \boldsymbol{v}_{t+1}^{\{i\}}}{\|\boldsymbol{I}_{xy,t}^{\{i\}} - \gamma \boldsymbol{\delta}_{rb}^{\{i\}} - \beta \boldsymbol{v}_{t+1}^{\{i\}}\|}, & \text{Otherwise.} \end{cases} \quad (13)$$

Together, these terms ensure that the solver balances local contact physics with consistency across the footplate, avoiding unphysical artifacts like partial slipping or motion tearing. The complete implementation of the contact solver is outlined in Algorithm 1.

3) Drifting Resistance: Additionally, a drift-resistive force counteracts horizontal motion, scaling with both penetration depth and horizontal displacement. We find that defining the travel distance as the accumulated path length from the starting intrusion position to the current position, rather than the direct Euclidean distance between these two points as in [25], yields better performance, i.e. $d_D = \sum_t \|\boldsymbol{s}_t^{\{i\}} - \boldsymbol{s}_{t-1}^{\{i\}}\|$.

Algorithm 1 GM Terrain Model Multi-contact Solver at Timestep t

Input: $\boldsymbol{J}^{\{i\}}, \boldsymbol{M}, h_t^{\{i\}}, \dot{h}_t^{\{i\}}, \boldsymbol{s}_{t-1}^{\{i\}}, \boldsymbol{s}_t^{\{i\}}, \boldsymbol{v}_t^{\{i\}}, d_{D,t-1}^{\{i\}}, i = 1, ..., N$
Initialization:
$N_c = 0$
for $i = 1 : N$ **do**
  $\boldsymbol{I}_{xy,t}^{\{i\}} = \{0, 0\}$, $\boldsymbol{I}_z^{\{i\}} = \Delta t \boldsymbol{F}_{GM}(h_t^{\{i\}}, \dot{h}_t^{\{i\}})$
  contact-detection (N_c, $\boldsymbol{I}_z^{\{i\}}$)
end for
Main:
while $e \geq e_c$ **do** // Iterative solution loop
  $e = 0$
  for $i = 1 : N_c$ **do**
    $\boldsymbol{v}_{t+1}^{\{i\}}$ $\boldsymbol{\omega}_{t+1}$ = dynamics-update(N_c, $\boldsymbol{v}_t^{\{i\}}$, $\boldsymbol{J}^{\{i\}}$, $\boldsymbol{M}$, $\boldsymbol{I}^{\{1:N_c\}}$) // Eqs. (8-9)
    $\boldsymbol{\delta}_{rb}^{\{i\}}$ = continuity-error($\boldsymbol{v}_{t+1}^{\{i\}}$, $\boldsymbol{\omega}_{t+1}$) // Eqs. (11-12)
    $\boldsymbol{I}_{xy,t,next}^{\{i\}}$ = friction-constraint($\boldsymbol{\delta}_{rb}^{\{i\}}$, $\boldsymbol{v}_{t+1}^{\{i\}}$, $\boldsymbol{I}^{\{i\}}$) // Eq. (13)
    $e = e + \|\boldsymbol{I}_{xy,t,next}^{\{i\}} - \boldsymbol{I}_{xy,t}^{\{i\}}\|$
    $\boldsymbol{I}^{\{i\}} = \{\boldsymbol{I}_{xy,t,next}^{\{i\}}, \boldsymbol{I}_{z,t}^{\{i\}}\}$
  end for
end while
for $i = 1 : N_c$ **do** // Resistive force
  Calculate $d_{D,t}^{\{i\}} = d_{D,t-1}^{\{i\}} + \|\boldsymbol{s}_t^{\{i\}} - \boldsymbol{s}_{t-1}^{\{i\}}\|$
  Calculate $\boldsymbol{F}_D^{\{i\}}(d_D^{\{i\}}, h_t^{\{i\}}, \dot{\boldsymbol{v}}_t^{\{i\}})$ // Eq. (14)
end for
Return:
$\boldsymbol{F}_t^{\{i\}} = \{\boldsymbol{F}_{D,t}^{\{i\}} + \frac{\boldsymbol{I}_{t,next}^{\{i\}}}{\Delta t}, \frac{\boldsymbol{I}_z^{\{i\}}}{\Delta t}\}$

The drift-resistive force can be obtained in the following form

$$\boldsymbol{F}_D = -(k_D d_D h_t + b_D \|\boldsymbol{v}_t\|)\hat{\boldsymbol{v}}_t \quad (14)$$

where $\hat{\boldsymbol{v}}_t$ is the unit vector of $\boldsymbol{v}_t$, while k_D and b_D denote the stiffness and damping factor. Finally, we solve the zero moment points (ZMP) given the resultant force $\boldsymbol{F}_t = \sum_{i=1}^{N} \boldsymbol{F}_t^{\{i\}}$ and the calculated resultant torque $\boldsymbol{\tau}_t = \sum_{i=1}^{N} \boldsymbol{F}_t^{\{i\}} \times \boldsymbol{p}^{\{i\}}$.

$$\boldsymbol{r}_{zmp} = [\boldsymbol{F}_t]_{\times}^{+} \cdot \boldsymbol{\tau}_t, \tag{15}$$

where $\boldsymbol{r}_{zmp}$ is the distance from the ZMP to the ankle points, and $[\boldsymbol{F}_t]_{\times}^{+}$ is the Moore-Penrose pseudoinverse of the skew matrix of the $\boldsymbol{F}_t$.

C. Terrain-aware Reinforcement Learning Pipeline

Our RL pipeline is depicted in Fig. 4. This control problem can be formulated as a Markov Decision Process (MDP) defined by $\{\mathcal{S}, \mathcal{A}, r, p, \gamma\}$, where $\mathcal{S}$ is the state space and $\mathcal{A}$ is the action space. At state $\boldsymbol{s}_t$ of time step t, a policy π performs an action $\boldsymbol{a}_t$ to forward the environment to the next state $\boldsymbol{s}_{t+1}$ with transition probability $p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t)$ meanwhile receiving reward $r_t = r(\boldsymbol{s}_t, \boldsymbol{a}_t)$. RL aims to find the optimal parameter θ that maximizes the discounted return: $J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{\infty}(\gamma^t r_t)\right]$.

1) Observation-action space: An asymmetric actor-critic framework [26] is employed where the actor and the critic receive different inputs (Fig. 4). The actor takes H-step historical proprioceptive states $\boldsymbol{o}_t^H$ as its input, which includes base angular velocity $\boldsymbol{\omega}$, commanded velocity $\boldsymbol{v}^{cmd}$, projected gravity $\boldsymbol{g}$, joint angles $\boldsymbol{q}$, joint velocities $\dot{\boldsymbol{q}}$, and the action of the last step $\boldsymbol{a}_{t-1}$. In addition to this, the actor also receives a series of estimates of the robot's key states and a latent representation (Section II-C.3). The critic additionally takes in the ground-truth states that are only accessible in simulation, as well as the real parameters of the terrain. The action $\boldsymbol{a}_t \in \mathbb{R}^{12}$ is chosen as the desired deviation of the joint angle from a time-invariant nominal pose $\mathring{\boldsymbol{q}}$, i.e. $\boldsymbol{a}_t = \boldsymbol{q}_t^{des} - \mathring{\boldsymbol{q}}$.

TABLE I: Major reward terms for training policy. $(\cdot)^{cmd}$ and $(\cdot)^{des}$ is the desired and commanded physical values. $\mathcal{C}$ represents the feet contact states. $\mathbb{I}$ is the indicator function that returns 1 when the condition is true and 0 otherwise.

Term	Equation	Weight
Lin. velocity tracking	$\exp\{-5(\boldsymbol{v}_{xy}^{cmd} - \boldsymbol{v}_{xy})^2\}$	1.4
Ang. velocity tracking	$\exp\{-5(\boldsymbol{\omega}_{z}^{cmd} - \boldsymbol{\omega}_{z})^2\}$	1.1
Feet clearance	$\|\boldsymbol{h}_t^{f,des} - \boldsymbol{h}_t^{f}\| \cdot (1 - \mathcal{G}_p)$	2.4
Gait alignment	$1.3\sum_{i=1}^{2}\mathbb{I}(\mathcal{C}_i = \mathcal{G}_{p,i}) - 0.3$	1.4
Feet slippery speed	$\boldsymbol{v}_t^f \cdot \mathcal{C}$	-0.1
Base acceleration	$\exp\{-3\|\boldsymbol{v}_t^b - \boldsymbol{v}_{t-1}^f\|\}$	0.2
Action smoothness	$(\boldsymbol{a}_t - \boldsymbol{a}_{t-1})^2$ $+(\boldsymbol{a}_t - 2\boldsymbol{a}_{t-1} + \boldsymbol{a}_{t-2})^2$	-0.003

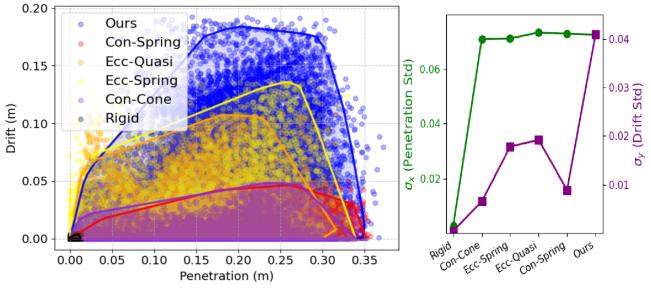

Fig. 5: The distribution of positional drift versus penetration for different methods and their standard deviations along the x and y directions.

2) Reward engineering: Our controller learns a velocity-tracking periodic gait $\mathcal{G}_p$ defined by a sinusoidal function that schedules time-varying contact events. We optimize foot contacts to closely follow this gait schedule. Some regularization terms are introduced to make the motion style more aesthetically pleasing and ensures stability during deployment. The detailed reward terms and their weights can be found in Table I.

3) Terrain Estimator and Modulation Net: Previous work demonstrates the benefits of using latent variables for reasoning about the terrain [21] and an online estimation of robot states to improve robustness [27]. To this end, we introduce a multi-head variational autoencoder (VAE) [28] as the first adaptive module. It performs estimation of base velocity, feet velocities and feet heights $\{\boldsymbol{v}_t^b, \boldsymbol{v}_t^f, \boldsymbol{h}_t^f\}$, and outputs a latent variable $\boldsymbol{z}_t$ encoding implicit robot-environment information. The network is jointly optimized by regression and PPO policy loss:

$$\begin{aligned} Loss &= \epsilon \cdot loss_{reg} + (1-\epsilon) \cdot loss_{policy} \\ loss_{reg} &= \mathcal{L}_{VAE} + \sum_{x \in \{\boldsymbol{v}_t^b, \boldsymbol{v}_t^f, \boldsymbol{h}_t^f\}} \mathbf{MSE}(x, \hat{x}) \end{aligned} \tag{16}$$

where hyperparameter $\epsilon \in [0, 1]$ ($\epsilon = 0.3$ in our case). The loss $\mathcal{L}_{VAE}$ aims to reconstruct the observation to $\hat{\boldsymbol{o}}_t$ and minimize the KullbackLeibler (KL) divergence. $p(\boldsymbol{z}_t|\boldsymbol{o}_t^H)$ is the posterior distribution of $\boldsymbol{z}_t$ given $\boldsymbol{o}_t^H$. $p(\boldsymbol{z}_t)$ is the prior distribution of $\boldsymbol{z}_t$ which is assume to be $\mathcal{N}(0, \mathbf{I})$.

$$\mathcal{L}_{VAE} = \mathbf{MSE}(\boldsymbol{o}_t, \hat{\boldsymbol{o}}_t) + \varphi \mathbf{D}_{KL}(p(\boldsymbol{z}_t|\boldsymbol{o}_t^H) \| p(\boldsymbol{z}_t)) \tag{17}$$

The reduced stiffness of compliant terrain alters contact restitution, causing foot velocity patterns distinct from high-stiffness surfaces. These dynamics are encoded into the latent variable $\boldsymbol{z}_t$ to enhance terrain awareness. We propose to use a modulation network [29] to form another adaptive module. This net dynamically adjusts $\boldsymbol{z}_t$ through a parametric transformation governed by $\boldsymbol{v}_t^f$. This transformation layer processes $\boldsymbol{v}_t^f$ to generate two adaptive scaling factors (θ_1, θ_2), which then perform an affine operation on $\boldsymbol{z}_t$ expressed as $\tilde{\boldsymbol{z}}_t = \theta_1 \cdot \boldsymbol{z}_t + \theta_2$.

III. Experimental Results

A. Implementation details

We train the controller on 4096 agents in parallel on the Isaac Gym simulator for 20000 episodes. The linear velocity

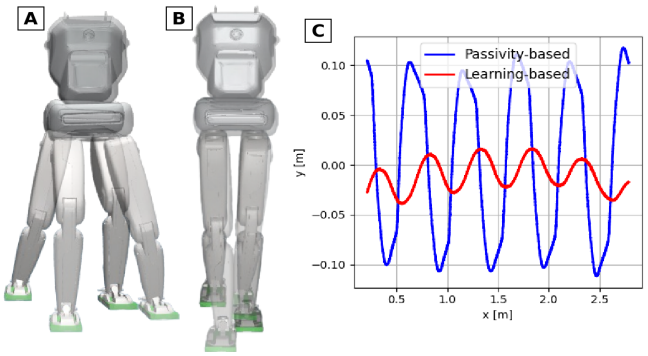

Fig. 6: Performance of (**A**) the passivity-based controller and (**B**) the learning-based controller. (**C**) Center of Mass (CoM) trajectory comparison.

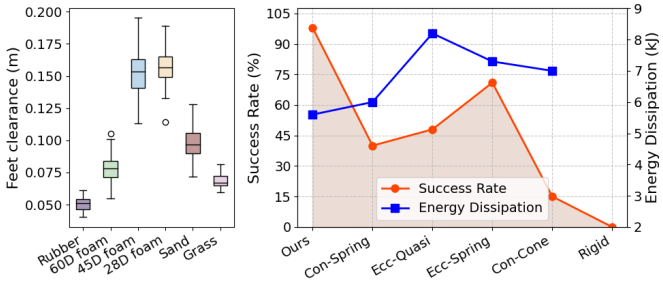

Fig. 7: MILD controller's feet clearance across all terrain types (**Left**) and policy performance of the compared contact models on 25D foam (**Right**).

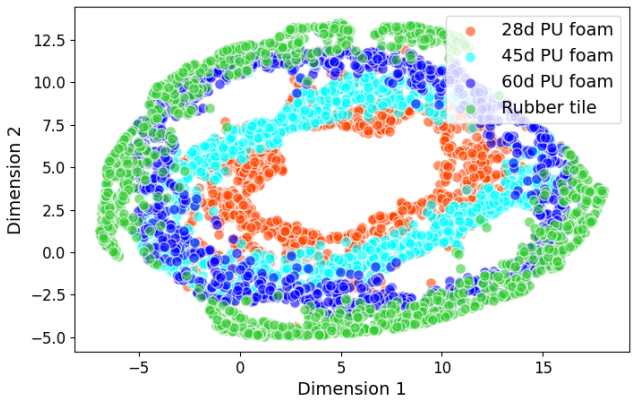

Fig. 8: The t-SNE classification of the encoded variable $\tilde{\boldsymbol{z}}_t$ across terrains with varying stiffness levels, where the dataset comprises 15-second walking trials on each terrain type.

command in the forward direction is set as $[0.0, 1.2]\ m/s$, and the angular velocity command as $[-1.5, 1.5]\ rad/s$. Different contact scenarios on deformable terrain are modeled through randomized parameters, which are listed alongside training parameters in Table II. Our pipeline consists of five MLP components: an encoder, a decoder, a modulation net, an actor, and a critic, which are updated concurrently by the gradients of corresponding losses. The entire training is performed on a desktop PC with an NVIDIA RTX 4080 GPU, which costs approximately 42.4 hours.

TABLE II: Randomization range of critical parameters

Category	Parameters	Range	Unit
Modeling	Shear band angle α	[0.698, 1.310]	Rad
	Compact rate ν	[0.8, 2.4]	-
	Surr. mass scale c_g	[-1.4, 4.6]	-
	Inertial drag scale c_d	[-6.0, 28.0]	-
	Volume fraction ϕ	[0.50, 0.64]	-
	Friction coefficient μ	[0.2, 1.3]	-
	Pen. stiffness k_{pen}	[8, 1400]	N/m
	Conical stress σ_{cone}	[0.01, 0.24]	$10^6\ N/m^3$
Training	Added base mass	[-4, 4]	Kg
	Center of mass offset	[-6, 6]	mm
	Motor latency	[1, 10]	ms
	Motor offset	[-0.035, 0.035]	Rad
	IMU latency	[1, 10]	ms
	Command latency	[1, 10]	ms

As shown in Fig. 1, the hardware validation test is carried out on the EngineAI SA01 bipedal robot [30], which has a total of 12 degrees of freedom (DOF), with 6 DOF for each leg. We employ a PD controller as the joint-level tracking of the desired joint angles. It operates at a frequency of 100 Hz, with $K_p = [50, 50, 70, 70, 20, 20], K_d = [5.0, 5.0, 7.0, 7.0, 0.2, 0.2]$ for the joints on each leg.

B. Simulation Experiments

To assess the advantages of our approach, we compare it with existing state-of-the-art methods for deformable terrain modeling. The data is collected from the policies trained with the same episodes on their respective models:

- **Con-Spring** [5][6]: This model characterizes foot-substrate interaction as a concentric spring system that generates unidirectional penetration forces.
- **Con-Cone** [25]: A developing cone model captures the added mass effect of the underlying substrate through concentric interaction at the geometric center.
- **Rigid**: The baseline policy that is trained via the simulation's built-in rigid contact terrain.
- **Ecc-Quasi** (RFT) [10] [11]: Uniformly distributed contact stress independent of penetration depth, neglecting inertial drag from substrate mass variation.
- **Ecc-Spring** [5][6]: The surface-type foot is discretized into multiple unidirectional spring elements.

It is worthwhile to mention that **Ecc-Spring** is implemented to ensure comparison fairness, where **Con-Spring** is extended to the eccentric intrusion scenarios. As depicted in Fig. 5, the two concentric single-point-contact cases exhibit the most limited performance (after rigid surface), as they fail to capture complex edge interactions and severely underestimate contact area. In contrast, our approach achieves deeper penetration (+38 % vs. Ecc-Spring) and longer slip distances, demonstrating broader contact interaction characteristics. It outperforms the other two surface-expanded eccentric compression models because it fully accounts for substrate mass variation during penetration at different compression points, leading to a broad range of inertial drag. The improvements are more beneficial for the RL agent by helping the agent witness a wider range of randomization during training, enhancing the policy's adaptability in real-world situations.

We further compare our method against a passivity-based whole-body controller [20] designed for compliant terrain, which requires computationally expensive online optimization to prioritize stability. In contrast, MILD encodes robust dynamics through pre-optimized RL policies, resulting in significantly smaller Center of Mass (CoM) oscillations and greater stability during dynamic walking (Fig. 6).

C. Real-World Experiments

1) Implicit terrain adaptability: Different from prior work testing on unquantified grass/sand, we evaluate system performance on standardized materials with defined stiffness grades. Specifically, we consider six distinct terrains, including four manufactured surfaces—rubber tiles, 60d, 45d, and 28d polyurethane (PU) foam—and two natural surfaces: grass and sand, as shown in Fig. 1. The rubber tile provides slight deformability through its porous structure composed of granular elastomeric particles, while the PU foams exhibit a stiffness gradient based on their density grades, where

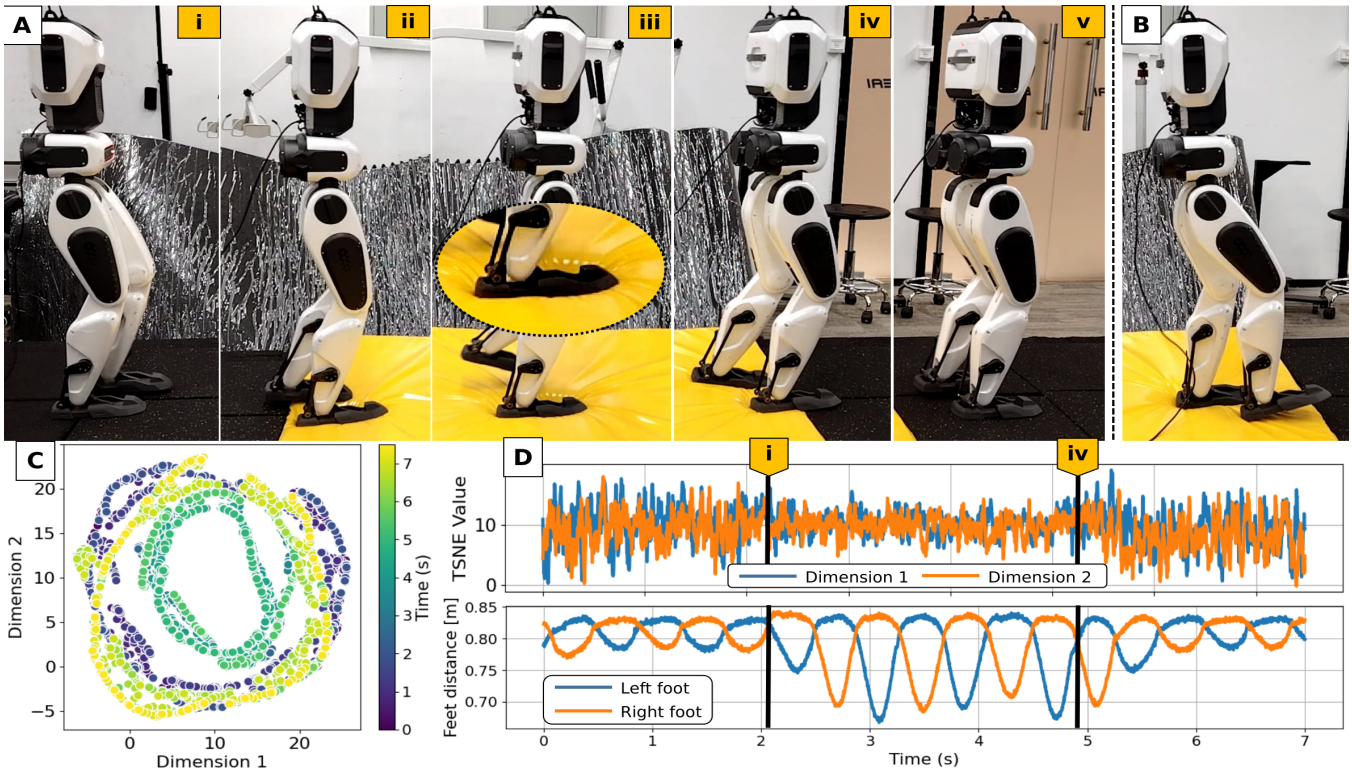

Fig. 9: Robot motion analysis during sudden transition to foam penetration while moving forward (**A**) and backward (**B**), with corresponding t-SNE visualization of modulated latent space variations (**C**-**D**). The latent representations show distinct clustering patterns during terrain transitions, demonstrating the controller's real-time adaptation to varying surface compliance.

smaller values indicate softer, more malleable surfaces. Our controller demonstrates clear adaptability across these terrains (Fig. 7), automatically increasing foot clearance on softer surfaces to compensate for ground deformation. Compared to other approaches, our fine-grained contact modeling achieves the highest success rate and energy efficiency.

To analyze the system's terrain implicit perception, we visualize the affine-transformed latent state $\tilde{z}_t$ using t-SNE, as shown in Fig. 8. The visualization reveals distinct clusters corresponding to the terrain's stiffness-deformability spectrum, with softer terrains occupying cluster interiors and stiffer terrains distributed peripherally. This clustering demonstrates our model's ability to encode terrain information that enables adaptive behavior generation across deformation levels.

2) Comparative analysis: This capability is verified through extensive trials across seven terrain types (including rigid planes). The robot successfully completes 10 forward/backward walking cycles at 1.2 m/s on each terrain without a single failure. For comparison, we also deploy controllers trained with baseline contact models (Section III-B) on 45d foam. Due to the restricted training domains for contact conditions, these controllers adopt overly conservative gaits characterized by insufficient foot clearance during swing phases. This results in either joint limit violations or stability loss - failures that highlight the importance of proper leg lifting.

We further compare against two state-of-the-art learning-based methods: *HT-2* [22], a transformer-based history-conditioned policy, and *Clock* [23], a clock-signal guided controller. Both methods achieve generalization through terrain geometry randomization on rigid surface without explicit modeling of terrain mechanics. As shown in Table III, MILD shows an overwhelming performance compared to these methods in energy efficiency (COT), velocity tracking, and joint torque across various commanded velocities. This fact suggests that the integration of a high-fidelity compliant contact model significantly enhances the controller's performance.

TABLE III: Performance comparison. Cost of Transport is calculated as COT $=$ Power$/$(Weight $\cdot$ Velocity) with Power $= \sum_{actuators} \|\tau\dot{q}\|$

Commanded Velocity $\boldsymbol{v}_x$(m/s)		0.3	0.6	0.9	1.2
Cost of Transport (Hardware)	*MILD*	**0.82**	**0.75**	**0.78**	**0.83**
	HT-2	1.05	0.96	1.20	1.08
	Clock	1.38	1.30	1.35	1.42
Maximum torque (Nm) (Hardware)	*MILD*	**58.2**	**63.2**	**73.3**	**88.4**
	HT-2	62.9	75.8	96.0	91.6
	Clock	67.6	82.5	102.8	115.2
Measured Velocity (m/s) (Simulation)	*MILD*	**0.28**	0.57	**0.84**	**1.15**
	HT-2	**0.28**	**0.61**	0.98	1.28
	Clock	0.23	0.54	0.81	1.09

3) Online identification and transition: The system demonstrates emergent terrain adaptation when the robot transitions between surfaces of varying stiffness without prior knowledge of this change (Fig. 9). As observed in the linked video, when the robot moves from a high-stiffness to a low-stiffness surface using an initially conservative gait, the controller automatically increases stride length to maintain

dynamic stability. This adaptation results from variations in the latent state, which in turn modulate the generated motion. The t-SNE visualization clearly captures this process, exhibiting distinct latent space trajectories during the complete rubber $\Rightarrow$ foam $\Rightarrow$ rubber transition cycle. Remarkably, our pipeline maintains effective adaptation even when encountering abrupt terrain changes that are never experienced during training, indicating that the latent representation learns fundamental physical properties of ground stiffness rather than merely memorizing specific training conditions.

IV. CONCLUSIONS

We presented a high-fidelity compliant contact model for bipedal locomotion on deformable terrain, focusing on footplates with substantial contact areas. The model captures key interaction dynamics—eccentric penetration, high-speed impacts, and spatially varying forces. We paired this with an RL-based controller featuring online estimation and identification for compliance adaptation. Through systematic comparisons against state-of-the-art models, our approach demonstrated superior performance in both simulated and real-world experiments. Hardware experiments validated the system's capability to rapidly adapt to stiffness changes, as demonstrated by seamless transitions between distinct deformable levels, while accurately classifying their mechanical properties.

The proposed framework could be extended by incorporating visual terrain deformation perception for improved adaptability. Future work of this model may also explore multi-gait generalization, including jumping and running maneuvers, and their implementation on model-based control architectures.

REFERENCES

[1] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.

[2] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," http://pybullet.org, 2016–2021.

[3] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv:2108.10470*, 2021.

[4] T. Pöschel and T. Schwager, *Computational granular dynamics: models and algorithms*. Springer Science & Business Media, 2005.

[5] D. J. Lynch, K. M. Lynch, and P. B. Umbanhowar, "The soft-landing problem: Minimizing energy loss by a legged robot impacting yielding terrain," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3658–3665, 2020.

[6] D. J. Lynch, J. L. Pusey, S. W. Gart, P. B. Umbanhowar, and K. M. Lynch, "Efficient, responsive, and robust hopping on deformable terrain," *IEEE Transactions on Robotics*, 2024.

[7] L. Ding, H. Gao, Z. Deng, J. Song, Y. Liu, G. Liu, and K. Iagnemma, "Foot–terrain interaction mechanics for legged robots: Modeling and experimental validation," *The International Journal of Robotics Research*, vol. 32, no. 13, pp. 1585–1606, 2013.

[8] V. Vasilopoulos, I. S. Paraskevas, and E. G. Papadopoulos, "Compliant terrain legged locomotion using a viscoplastic approach," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 4849–4854.

[9] C. Li, P. B. Umbanhowar, H. Komsuoglu, D. E. Koditschek, and D. I. Goldman, "Sensitive dependence of the motion of a legged robot on granular media," *Proceedings of the National Academy of Sciences*, vol. 106, no. 9, pp. 3029–3034, 2009.

[10] C. Li, T. Zhang, and D. I. Goldman, "A terradynamics of legged locomotion on granular media," *science*, vol. 339, no. 6126, pp. 1408–1412, 2013.

[11] X. Xiong, A. D. Ames, and D. I. Goldman, "A stability region criterion for flat-footed bipedal walking on deformable granular terrain," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 4552–4559.

[12] L. K. Treers, C. Cao, and H. S. Stuart, "Granular resistive force theory implementation for three-dimensional trajectories," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1887–1894, 2021.

[13] X. Chen, A. Anikode, J. Yi, and T. Liu, "Foot shape-dependent resistive force model for bipedal walkers on granular terrains," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 13 093–13 099.

[14] H. Katsuragi and D. J. Durian, "Unified force law for granular impact cratering," *Nature physics*, vol. 3, no. 6, pp. 420–423, 2007.

[15] L. Tsimring and D. Volfson, "Modeling of impact cratering in granular media," *Powders and grains*, vol. 2, pp. 1215–1223, 2005.

[16] J. Aguilar and D. I. Goldman, "Robophysical study of jumping dynamics on granular media," *Nature Physics*, vol. 12, no. 3, pp. 278–283, 2016.

[17] C. M. Hubicki, J. J. Aguilar, D. I. Goldman, and A. D. Ames, "Tractable terrain-aware motion planning on granular media: An impulsive jumping study," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 3887–3892.

[18] 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 *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2018, pp. 1–9.

[19] B. Henze, M. A. Roa, and C. Ott, "Passivity-based whole-body balancing for torque-controlled humanoid robots in multi-contact scenarios," *The International Journal of Robotics Research*, vol. 35, no. 12, pp. 1522–1543, 2016.

[20] G. Mesesan, J. Englsberger, G. Garofalo, C. Ott, and A. Albu-Schäffer, "Dynamic walking on compliant and uneven terrain using dcm and passivity-based whole-body control," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 25–32.

[21] 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.

[22] I. Radosavovic, S. Kamat, T. Darrell, and J. Malik, "Learning humanoid locomotion over challenging terrain," *arXiv preprint arXiv:2410.03654*, 2024.

[23] R. P. Singh, M. Morisawa, M. Benallegue, Z. Xie, and F. Kanehiro, "Robust humanoid walking on compliant and uneven terrain with deep reinforcement learning," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*. IEEE, 2024, pp. 497–504.

[24] J. Chen, J. Frey, R. Zhou, T. Miki, G. Martius, and M. Hutter, "Identifying terrain physical parameters from vision-towards physical-parameter-aware locomotion and navigation," *IEEE Robotics and Automation Letters*, 2024.

[25] 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.

[26] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," *arXiv preprint arXiv:1710.06542*, 2017.

[27] 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.

[28] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *International conference on learning representations*, 2017.

[29] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[30] Engineai. [Online]. Available: https://www.engineai.com.cn/product_three