TACT-ful: Multi-Channel Terrain Affordance and Compliance Training for Payload-Robust Perceptive Humanoid Locomotion

Anonymous Author(s)
Affiliation
Address
email

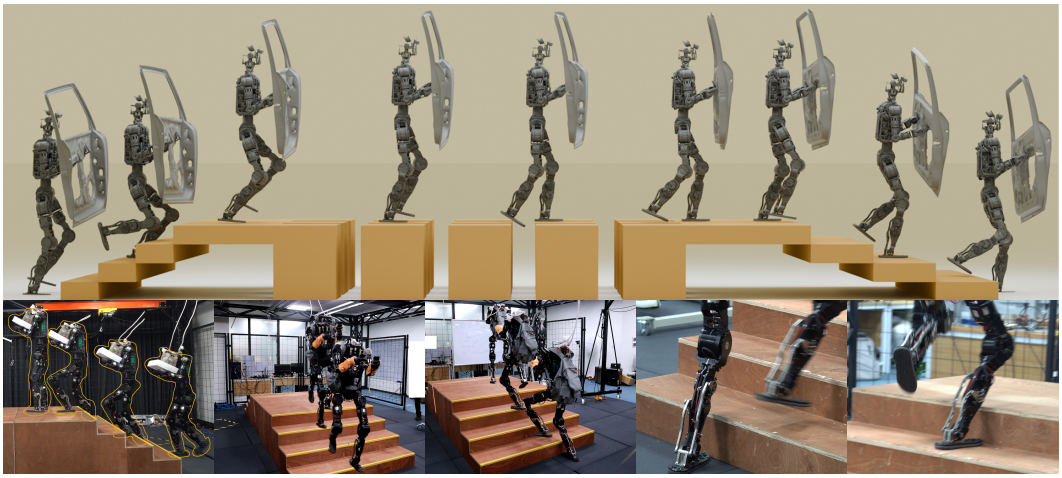

Figure 1: A service humanoid, trained on TACT-ful, traverses structured terrain while carrying heavy payload.

Abstract: Foothold selection on structured terrain requires explicit reasoning about contact planarity, surface steepness, and kinematic reachability, properties not captured by a single height-based terrain signal. We propose a multi-channel terrain cost combining flatness, steepness, and velocity-aware height feasibility, plus a forward climb reward, that simultaneously drives a GPU-parallel divergent component of motion (DCM) foothold planner and shapes a dense per-step affordance reward for an asymmetric actor-critic policy trained with proximal policy optimization (PPO) from depth images. A Bézier swing trajectory with adaptive apex bias extends foothold tracking to joint position-and-orientation, using the arc tangent to guide sole orientation through riser crossings and tread landings. To support payload tasks, we introduce a lower-body compliance training procedure in which a virtual wrench is injected at a sampled load attachment point, generating physically consistent force and moment; wrench-aware compliance targets replace rigid pose penalties, and the policy learns to yield to load-induced perturbations without force sensing. The full system trains end-to-end with standard PPO, no distillation, and no teacher-student staging, and is deployed on a humanoid directly from simulation with configuration changes only. In simulation, the policy reaches $1.0\ \mathrm{m/s}$ on stairs with risers up to $0.20\ \mathrm{m}$ and improves payload robustness up to $\sim 15\ \mathrm{kg}$ centered load and for moment-dominated wrist loads without fine-tuning. We also provide a qualitative hardware demonstration on structured terrain. Project website: https://fai-rl-tech.github.io/tact-locomotion.github.io/

Submitted to the 10th Conference on Robot Learning (CoRL 2026). Do not distribute.

1 Introduction

Service humanoids operating in human environments must traverse structured non-flat terrain, including staircases, stepped platforms, and ramps, while carrying tools or cargo. Foothold placement on such terrain is consequential: a sole that straddles a step edge bears load on a line contact, narrowing the friction cone and generating lateral impulses that scale with robot mass. The capture-point [1] divergence rate is independent of mass, but the ground reaction force needed to intercept a divergent trajectory scales with total weight, so that early, terrain-quality-aware foothold selection [2] becomes more consequential as platform mass grows. Payload carrying adds a second challenge: a carried load shifts the effective neutral pelvis pose and imposes a time-varying tug on the torso, requiring the locomotion controller to absorb these perturbations without loss of balance.

Existing approaches address subsets of this problem. Model-based foothold planners select discrete landing targets subject to stability and reachability constraints [3, 4, 5], but their outputs are not used to shape the reinforcement-learning (RL) policy reward. Perceptive RL methods for humanoids consume elevation maps or depth images to adapt gait on rough terrain [6, 7, 8], but typically encode terrain through a single height channel and do not reason explicitly about contact planarity or kinematic overreach. Compliance under payload is largely treated separately, either through dedicated force controllers, online payload estimators, or manual gait re-tuning, rather than as a behavior the locomotion policy acquires within a single terrain-aware training run.

The contributions of this paper are:

1. **Multi-channel terrain cost with terrain-adaptive swing and tangent-guided foot orientation.** Flatness, steepness, velocity-aware height feasibility, and a climb bonus jointly drive a GPU-parallel DCM foothold planner and serve as a dense terrain-affordance learning signal for the RL policy. A Bézier swing trajectory with adaptive apex bias and clearance extends foothold tracking to joint position-and-orientation references, using the arc tangent to guide sole orientation through riser crossings and tread landings. Unlike the closest predecessor, which couples a single-channel DCM planner to a position-only foothold reward under a multi-stage curriculum [7], our cost adds explicit flatness, velocity-aware feasibility, and speed-gated climb channels, and the tangent schedule extends the swing reference to sole *orientation*.
2. **Lower-body compliance training for payload-robust locomotion.** A virtual wrench injected at a sampled load attachment point generates physically consistent force and moment; together with wrench-aware compliance targets for pelvis height and trunk orientation, it teaches the policy to yield to payload-induced perturbations without force sensing, improving robustness up to $\sim 15\,\mathrm{kg}$ centered load and for moment-dominated wrist loads without payload-specific fine-tuning.
3. **An integrated end-to-end pipeline with a qualitative hardware demonstration.** The system trains with standard PPO in a single run (no distillation, no teacher-student staging) and combines terrain-aware foothold selection with payload compliance, and is deployed zero-shot from simulation on a service humanoid traversing structured terrain while carrying payload.

2 Related Work

Foothold planning and terrain-conditioned reinforcement learning. Model-based foothold planners select discrete landing targets under LIPM capturability constraints: Acosta and Posa [3] formulate MIQP over convex foothold regions, while Xiang et al. [4] jointly optimize step timing and placement under DCM bounds. Kim et al. [5] replace the MIQP with an RL-trained footstep policy mapping elevation maps to 3D targets, decoupled from end-to-end training. Lee et al. [9] use LIP-derived footstep targets as a partial RL reward, and Liu et al. [7] reformulate DCM foothold search as GPU-parallel discrete optimization used as an explicit RL reward with a multi-stage curriculum, the most direct predecessor to our approach. Wang et al. [10] show that a purely learned dense foothold reward with a double-critic curriculum can traverse sparse footholds without an explicit planner.

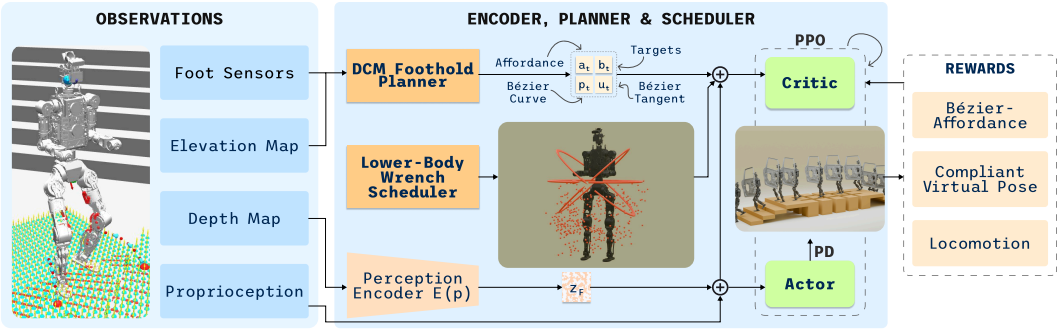

Figure 2: **Overview of the proposed framework.** Two parallel modules feed into the training reward buffer. *(i) DCM foothold planner:* at every control step, a pelvis-mounted elevation map is consumed by the GPU-parallel DCM foothold planner, which selects terrain-optimal landing targets and produces a Bézier swing trajectory reference; these targets define the foothold-tracking and terrain-specific reward terms and are provided as privileged observations to the critic. *(ii) Wrench compliance scheduler:* a virtual wrench is injected at a sampled load attachment point, generating force that shifts the pelvis and moment that tilts the trunk; wrench-aware compliance targets replace rigid pelvis-height and orientation penalties, training the policy to absorb payload-induced perturbations without explicit force sensing.

Perceptive humanoid locomotion and terrain representation. Agarwal et al. [11], Miki et al. [12] pioneered the privileged-training-to-depth-distillation paradigm, decoupling policy and perception learning for zero-shot real-world deployment, which has become standard for perceptive legged locomotion. Building on this, Song et al. [6] reconstruct a dense egocentric height map from depth and jointly output joint targets and adaptive gait frequency; Radosavovic et al. [13] train a Transformer policy on diverse outdoor terrain without perceptive reward shaping; Long et al. [14] achieve stair climbing via LiDAR elevation maps; Ben et al. [8] extend perception to full 3D structure via a voxel-grid representation; Zhang et al. [15] distil terrain-specific experts into a unified transformer policy with multi-view depth fusion and velocity-scaled features; and Sun et al. [16] train end-to-end from raw stereo depth images using terrain-specific multi-critic shaping. Vision-based whole-body policies demonstrate perceptive parkour, stair climbing, gap crossing, and platform jumping, on humanoids and quadrupeds [17, 18, 19]. Standard terrain representations include probabilistic elevation maps [20], traversability scoring [21], and multi-layer grid maps [22].

Payload-carrying and compliant loco-manipulation. Large-scale domain randomization yields implicit payload tolerance [23]; Kumar et al. [24] achieve payload and terrain robustness on quadrupeds through rapid online system identification; Zhang et al. [25] demonstrate balance recovery under large external perturbations via a multi-phase curriculum; Fu et al. [26] make payload awareness explicit on the Tiangong humanoid via a history-based estimator in a decoupled lower-body RL architecture, evaluated on flat terrain. Pasricha et al. [27] extend payload limits to $3\times$ nominal via compliant trajectory diffusion. For force-adaptive compliance without force sensors, Xu et al. [28] train legged robots to absorb collision impulses and sustain payload pulls up to $\frac{2}{3}$ body weight by tracking a virtual impedance reference; Zhi et al. [29] unify position and force control for contact-rich loco-manipulation. Terrain property estimation from vision [30] and friction identification [31] connect perception to contact mechanics. None of these works couple terrain-quality-aware foothold selection with compliance training in a single end-to-end RL formulation, leaving the interaction between payload mass, step-height feasibility, and terrain planarity unaddressed.

3 Terrain Affordance and Compliance Training

Figure 2 Overview the framework: a DCM foothold planner and a wrench compliance scheduler feed a shared reward for an asymmetric actor-critic policy. All hyperparameters and implementation details are located in the Appendices.

3.1 Capture-Point Stability

We use the divergent component of motion (DCM) / capture-point framework [1, 32] with a constant natural frequency $\omega_0 = \sqrt{g/z_0}$, where g is gravitational acceleration and z_0 is the nominal center of mass (CoM) height [33]. The DCM is defined as $\xi = x_{\mathrm{CoM}} + \dot{x}_{\mathrm{CoM}}/\omega_0$ and propagates as $\xi(T) = \xi_0\, e^{\omega_0 T}$ over a swing of duration T (step time); applied component-wise in the horizontal plane this gives the 2D DCM $\boldsymbol{\xi}_T \in \mathbb{R}^2$. The nominal step offset $\mathbf{b}_{\mathrm{nom}} = (b_x, b_y)$ keeps the divergent mode in steady state [34]; we adopt the linear-CoM-height reduction of Liu et al. [7] and set $\omega = \omega_0$ throughout.

3.2 Multi-Channel Terrain Cost

The planner selects the cell i that minimizes the total cost

$$\mathcal{J}_i = \alpha_{\mathrm{pos}}\, d_{\mathrm{pos},i} + \alpha_{\mathrm{dcm}}\, d_{\mathrm{dcm},i} + \alpha_E\, E_i + \alpha_Q\, Q_i + \alpha_M\, M_i - \alpha_{\mathrm{climb}}\, b_i, \tag{1}$$

where $d_{\mathrm{pos},i} = \left(x_i - v_x T\right)^2 + \beta\left(y_i - (v_y T + \mathrm{sign}(f)\, l_p)\right)^2$ is an asymmetric position residual from the nominal stride target ($\beta > 1$ penalizes lateral deviation more than sagittal to limit foot spread; $\mathbf{p}_i = (x_i, y_i)^\top$ is the 2D candidate foothold, (v_x, v_y) the commanded planar velocity, l_p the nominal lateral inter-foot distance, and $\mathrm{sign}(f) \in \{+1, -1\}$ selects the swinging leg), and $d_{\mathrm{dcm},i} = \|\boldsymbol{\xi}_T - \mathbf{p}_i - \mathbf{b}_{\mathrm{nom}}\|^2$ is the capture-point stability residual. These are computed from the elevation map $\mathcal{H}$ in one GPU-batched forward pass.

Flatness cost Q. A sole straddling a height discontinuity tilts by $\theta_i \approx \arctan(Q_i/L_{\mathrm{foot}})$, reducing admissible friction to $\mu_{\mathrm{eff},i} \le \mu_0 \cos\theta_i - \sin\theta_i$, which collapses to zero at $Q_i/L_{\mathrm{foot}} = \mu_0 \approx 0.6$, precisely the failure regime on standard stairs. We use the elevation range over a footprint-sized kernel as a GPU-efficient continuous proxy:

$$Q_i = \max_{j \in \mathcal{N}_i}(z_j) - \min_{j \in \mathcal{N}_i}(z_j). \tag{2}$$

Sampling-based foothold rewards [10] evaluate *contact support coverage*, suited to surface voids but blind to planarity loss. On structured terrain, a tread-edge landing scores zero under that criterion yet incurs $Q_i > 0$; the two costs target distinct failure modes and are complementary.

Steepness cost E. An edge contact on a vertical riser produces a lateral impulse. Sobel-filtered gradients $g_{x,j}$, $g_{y,j}$ are max-pooled, deliberately propagating the worst gradient within the footprint:

$$E_i = \max_{j \in \mathcal{N}_i} \sqrt{g_{x,j}^2 + g_{y,j}^2 + \varepsilon}. \tag{3}$$

Height-feasibility cost M. A step too tall to reach kinematically or dynamically is worse than a sub-optimal flat placement. The quadratic penalty

$$M_i = \max(|\Delta z_i| - h^*_{\mathrm{eff}},\ 0)^2, \tag{4}$$

where Δz_i is the stance foot height difference, uses a *velocity-aware effective maximum step height*:

$$h^*_{\mathrm{eff}} = h^*_{\mathrm{min}} + (h^*_{\mathrm{max}} - h^*_{\mathrm{min}})\ \mathrm{clip}\Big(\frac{v_x}{v^*},\, 0,\, 1\Big), \tag{5}$$

where h^*_{min} and h^*_{max} bound the reachable step height at standstill and rated forward speed v^*, respectively. At low speed the robot lacks the leg momentum needed to assist a step-up, so the effective kinematic reach is reduced; at high speed the dynamic leg drive can clear the full height. Below a minimum speed threshold a hard override replaces the planned target with the current foot position, and $h^*_{\mathrm{eff}} \to h^*_{\mathrm{min}}$ ensures the cost still disfavors high steps in the transition region.

Forward climb bonus b. To encourage the robot to step *up* onto reachable treads, a speed-gated capped bonus $b_i = \min\big(\max(\Delta z_i, 0),\ h^*_{\mathrm{eff}}\big) \cdot \mathbf{1}[v_x > v_{\mathrm{min}}]$ rewards upward steps proportionally to their height, capped at h^*_{eff}.

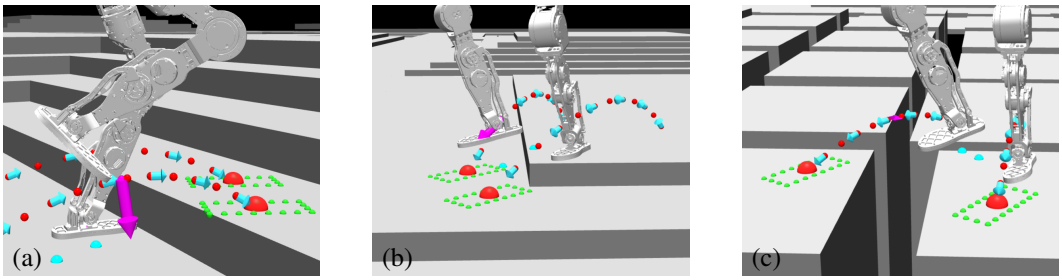

Figure 3: **Bézier swing arcs with adaptive apex bias.** (a) Step-up: apex biased toward the landing target, keeping the peak over the riser face. (b) Step-down: apex biased toward lift-off, extending horizontal travel before descent. (c) Gap: behave analogously, with clearance scaled by $|\Delta z|$.

3.3 Bézier Swing Trajectory and Tangent-Guided Foot Orientation

After selecting per-foot landing target $\mathbf{p}_f^*$, the swing foot tracks a quadratic Bézier arc whose tangent simultaneously determines foot orientation:

$$\begin{cases} \mathbf{p}(u) = (1-u)^2\,\mathbf{p}_\ell + 2(1-u)u\,\mathbf{p}_{\text{apex}} + u^2\,\mathbf{p}_f^*, \\ \dot{\mathbf{p}}(u) = 2(1-u)(\mathbf{p}_{\text{apex}} - \mathbf{p}_\ell) + 2u(\mathbf{p}_f^* - \mathbf{p}_{\text{apex}}), \end{cases} \quad u \in [0,1], \tag{6}$$

where $\mathbf{p}_\ell$ is the lift-off position and $\dot{\mathbf{p}}(u)$ points in the direction of instantaneous foot travel. The apex xy is biased toward whichever endpoint is higher:

$$\text{bias} = \text{clip}\left(0.5 + \kappa\,\frac{\Delta z}{h^*_{\max}},\ b_{\min},\ b_{\max}\right), \quad \mathbf{p}_{\text{apex},xy} = (1-\text{bias})\,\mathbf{p}_{\ell,xy} + \text{bias}\,\mathbf{p}^*_{f,xy}, \tag{7}$$

For step-up ($\Delta z > 0$), bias > 0.5 places the trajectory peak over the riser face, preventing premature descent onto the vertical surface; for step-down ($\Delta z < 0$), bias < 0.5 extends horizontal travel before descent, reducing the landing impact angle (fig. 3a-b). A phase-conditional schedule derived from $\dot{\mathbf{p}}(u)$ guides sole orientation through riser crossings. Both position proximity and foot heading are jointly optimized during swing via an exponential proximity kernel:

$$r_{\text{foothold}} = \sum_f I_{\text{swing},f}\,\exp\!\big(-\sigma_p\,\|\mathbf{p}_f - \mathbf{p}_{\text{Béz}}(t)\|^2 - \sigma_d\,\|\hat{\mathbf{d}}_f - \hat{\mathbf{t}}_f(u)\|^2\big), \tag{8}$$

where $\hat{\mathbf{d}}_f = R_f\hat{\mathbf{e}}_x$ is the foot forward axis and $\sigma_d = 0$ recovers the position-only form when orientation is not used.

The apex height adapts to terrain relief:

$$c = \min\!\big(c_{\min} + s\,|\Delta z|,\ c_{\max}\big), \quad z_{\text{apex}} = 2\big(\max(z_\ell,\ z_f^*) + c\big) - \frac{1}{2}(z_\ell + z_f^*), \tag{9}$$

where s is a clearance scale factor and $c_{\max}$ caps the swing peak. The tangent is vertical at

$$u_{\text{peak}} = \frac{z_\ell - z_{\text{apex}}}{z_\ell - 2z_{\text{apex}} + z_f^*}, \tag{10}$$

equal to 0.5 for level steps and shifting toward the higher endpoint on transitions. Two windows around u_{peak} define the reference unit vector $\hat{\mathbf{t}}_f(u)$: in the *pre-apex* window $[u_{\text{peak}}{-}\delta_\ell^-,\ u_{\text{peak}}{-}\delta_\ell^+)$, $\dot{\mathbf{p}}$ is rotated $90°$ in the sagittal plane ($(t_x,t_y,t_z)\mapsto(t_z,t_y,-t_x)$) and normalized, yielding a forward-downward direction that guides the sole through the riser and reduces toe-strike risk; in the *post-apex* window $(u_{\text{peak}}{+}\delta_r^-,\ u_{\text{peak}}{+}\delta_r^+]$, $\hat{\mathbf{t}}_f = \dot{\mathbf{p}}/\|\dot{\mathbf{p}}\|$ aligns the foot with the forward travel direction for tread landing; elsewhere $\hat{\mathbf{t}}_f = \mathbf{0}$ (no orientation constraint).

3.4 Lower Body Compliance Training

A payload exerts a wrench $\mathbf{W} = (\mathbf{F}, \boldsymbol{\tau})$ on the robot, where the moment $\boldsymbol{\tau} = \mathbf{r}_{\text{load}} \times \mathbf{F}$ scales with the CoM-to-load arm $\mathbf{r}_{\text{load}}$ independently of mass. $\boldsymbol{\tau}_{\text{ext}}$ tilts the trunk by φ, displacing the CoM by $\Delta x_{\text{CoM}} \approx d_{\text{CoM}}\varphi$; under LIPM this offset grows by $e^{\omega_0 T_{\text{step}}} > 2$ over a typical step, so

moment-induced tilt perturbs the capture point more than an equivalent translational offset, and both components of $\mathbf{W}$ require explicit coverage in the disturbance distribution. We emulate the wrench by applying a spring-damper at the *virtual load attachment point* $\mathbf{p}_{\text{load}} = \mathbf{p}_{\text{CoM}} + \mathbf{r}_{\text{load}}$:

$$\mathbf{F}_{\text{ext}}(t) = k\big(\mathbf{p}_a - \mathbf{p}_{\text{load}}(t)\big) - c\,\dot{\mathbf{p}}_{\text{load}}(t), \tag{11}$$

where $\mathbf{p}_a$ is a fixed world-frame anchor drawn once per episode, with stiffness k and damping c set so the spring force at maximum displacement R_a reaches a fixed fraction of nominal weight. Applying the force at $\mathbf{p}_{\text{load}}$ rather than the CoM generates $\boldsymbol{\tau}_{\text{ext}} = \mathbf{r}_{\text{load}} \times \mathbf{F}_{\text{ext}}$ automatically. At each reset, $\mathbf{r}_{\text{load}}$ is drawn from a two-component mixture ($\mathbf{r}_{\text{load}} = \mathbf{r}_{\text{body}}$ with prob. ρ_{body}, else $\mathbf{r}_{\text{arm}}$) that jointly covers both carry scenarios: *Body-attached* with $\mathbf{r}_{\text{body}} \sim \text{Ball}(R_{\text{close}})$, direction biased toward $-\hat{\mathbf{z}}$; and *Arm-extended* with $\mathbf{r}_{\text{arm}}$ from a shoulder-frame half-ellipsoid, biased forward. Both components use cosine-power lobes to concentrate sampling:

$$\begin{cases} p(\hat{\mathbf{d}}) \propto \cos^n \theta, \quad \theta \in [0, \pi/2] & \text{(body-attached direction)}, \\ p(\mathbf{r}_{\text{arm}}) \propto \cos^{n_r}\big(\angle(\mathbf{r}_{\text{arm}}, \hat{\mathbf{x}}_{\text{shoulder}})\big) & \text{(arm-extended bias)}. \end{cases} \tag{12}$$

where $\hat{\mathbf{d}}$ is the unit direction of $\mathbf{r}_{\text{body}}$ and θ its polar angle from $-\hat{\mathbf{z}}$. A weight-ϵ isotropic component blends in lateral and overhead samples. We replace the height and orientation penalties with wrench-aware compliance targets. For the height,

$$z^*_{\text{virt}}(t) = h^{\text{terrain}}_{\text{pelvis}}(t) + h^* + \alpha_z\big(z_a(t) - h^{\text{terrain}}_{\text{pelvis}}(t) - h^*\big)\,, \tag{13}$$

where h^* is the commanded base height, $z_a(t) = [\mathbf{p}_{\text{load}}(t)]_z$ is the world-frame z-coordinate of the virtual attachment point, and $\alpha_z \in [0, 1]$ is the height compliance gain. For the trunk orientation, the virtual target is derived from the induced moment:

$$\varphi^*_{\text{virt}} = \alpha_\varphi \frac{\tau_y}{k_{\text{rot}}}, \qquad \psi^*_{\text{virt}} = \alpha_\psi \frac{\tau_x}{k_{\text{rot}}}, \tag{14}$$

where τ_x, τ_y are the roll and pitch components of $\boldsymbol{\tau}_{\text{ext}}$, k_{rot} $[\text{N}{\cdot}\text{m}/\text{rad}]$ is a rotational stiffness converting induced moment to a tilt angle, and $\alpha_\varphi, \alpha_\psi \in [0, 1]$ are orientation compliance gains. The combined compliance reward is

$$r_{\text{comply}} = -\big(z_{\text{pelvis}} - z^*_{\text{virt}}\big)^2 - \big(\varphi_B - \varphi^*_{\text{virt}}\big)^2 - \big(\psi_B - \psi^*_{\text{virt}}\big)^2, \tag{15}$$

where φ_B, ψ_B are pelvis pitch and roll from the projected gravity vector. When $\alpha_z = \alpha_\varphi = \alpha_\psi = 0$ the reward collapses to the original rigid height and orientation penalties; at nonzero gains the policy learns to yield to the full wrench, acquiring wrench-aware compliance in both translation and rotation as a learned behavior without explicit force or torque sensing.

3.5 Policy and Training

An asymmetric Actor-Critic Encoder is used. The actor concatenates a convolutional embedding of a temporally stacked depth image with proprioceptive state and passes the result through a feedforward multilayer perceptron (MLP). The critic shares the same encoder but also receives privileged training-only signals. Actions are joint position targets relative to the default pose, converted to torques via proportional-derivative (PD) control and clipped to mechanical limits. Following the gait-adaptive extension [6], the policy also outputs a scalar gait frequency f_t that advances an internal phase clock via an exponential-moving-average (EMA) smoothed update:

$$\hat{f}_t = (1 - \alpha_{\text{EMA}})\hat{f}_{t-1} + \alpha_{\text{EMA}}\ \text{clip}(f_t,\, f_{\min},\, f_{\max}), \quad \phi_{t+1} = (\phi_t + \Delta t\, \hat{f}_t) \bmod 1, \tag{16}$$

with $(\hat{f}_t, \sin 2\pi\phi_t, \cos 2\pi\phi_t)$ appended to the actor observation. Training uses standard PPO [35] with adaptive Kullback–Leibler (KL) scheduling, generalized advantage estimation (GAE), and massively-parallel joint-velocity and terrain-difficulty curricula [36].

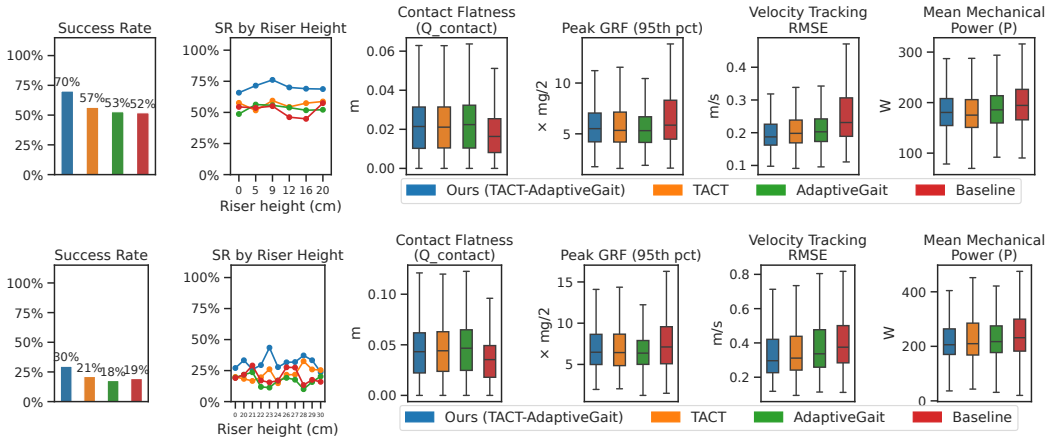

Figure 4: Terrain traversal ablation on **standard terrain** (top) and **hard terrain** (out-of-distribution).

4 Results

Experimental setup. Four variants are compared: *TACT + Adaptive Gait (Ours)*, *TACT-only*, *Adaptive Gait only*, and *Baseline* (a standard depth-map perceptive policy with no terrain-cost channels and no privileged elevation-map input to the critic). Each variant is evaluated at iteration 20 k across 4096 environments in MuJoCo [37] on stairs, slopes, and rough terrain; a second, unseen hard-terrain sweep covers risers 0.20–0.30 m, over which we report the success rate $\mathrm{SR}_{\mathrm{hard}}$ computed on the strict interior $[0.22, 0.28]$ m.

4.1 Terrain Traversal Ablation

Ours (TACT + Adaptive Gait) leads on both standard and hard terrain (70 % standard SR and 30 % $\mathrm{SR}_{\mathrm{hard}}$), outperforming TACT-only by 13 and 9 percentage points (pp) and Adaptive Gait only by 17 and 12 pp (Fig. 4); Adaptive Gait only falls below the Baseline on hard terrain (18 % vs. 19 %), confirming that frequency re-timing without terrain-quality guidance is actively harmful when risers approach kinematic limits. The Baseline's 52 % and 19 % SR coincide with the highest 95th-percentile ground reaction force (GRF) ($\approx$6.7× and $8.0 \times mg/2$), consistent with edge-biased contacts from blind foothold placement. On efficiency, Ours achieves lower velocity-tracking root-mean-square error (RMSE) (0.22 vs. 0.23 m/s) and mechanical power (229 vs. 241 W) than TACT-only on hard terrain; Q_{c} is identical between the two (0.023), isolating terrain cost channels as the mechanism for tread-centered landings. Fig. 5a shows that the SR gap relative to the Baseline is speed-invariant (71 / 68 / 69 % vs. $\approx$20 pp lower at 0.3 / 0.6 / 1.0 m/s), confirming foothold quality, not speed, drives the difference; Fig. 5b shows that removing all TACT channel weights raises foot-target distance 2.8× (0.088 → 0.251 m).

4.2 Payload Generalization

Fig. 5c evaluates payload generalization across three conditions (pelvis +15 kg, pelvis +20 kg, wrist +10 kg) on terrain and compares Ours against the Baseline. At moderate centered load (pelvis +15 kg), Ours achieves 50 % SR versus the Baseline's 38 % at lower power (247 W vs. 277 W), consistent with compliance training absorbing the downward wrench and suppressing impulsive GRF recovery. At high centered load (pelvis +20 kg), both policies degrade substantially to $\approx$37 % and $\approx$35 % SR respectively, nearly indistinguishable, indicating the load magnitude approaches the boundary of the training distribution; power rises for both, with Ours remaining 27 W below the Baseline (293 vs. 320 W), confirming that partial compliance is retained even as balance recovery deteriorates. Wrist-mounted mass (+10 kg), which generates a large moment rather than a direct CoM shift, yields the largest margin across all three conditions: Ours achieves 65 % SR versus the

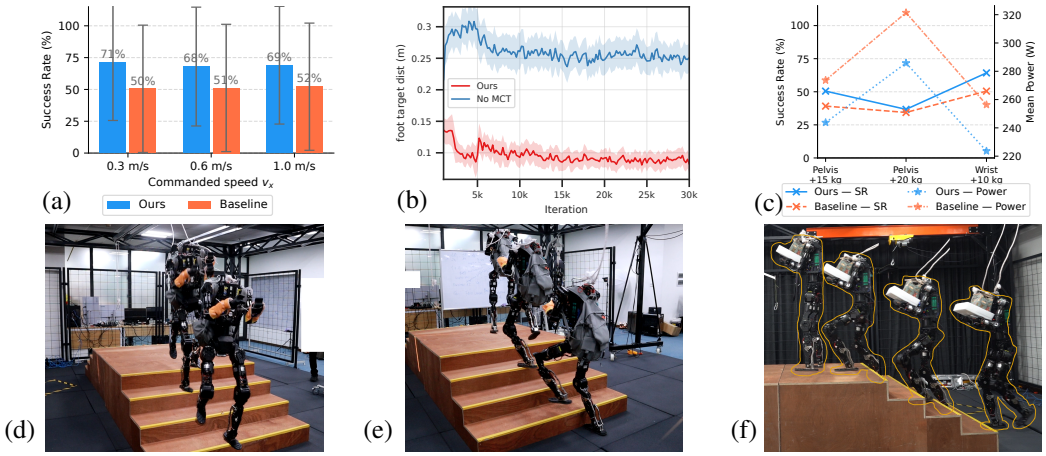

Figure 5: (a) Speed-conditioned SR; (b) foot-target distance; (c) SR (%) and mean power (W). (d–f) Qualitative hardware payload demonstrations, separate from the simulated ablations (d–e: 20 kg backpack; f: 7 kg tray).

Baseline's 50 % at 223 W versus 259 W, consistent with the arm-extended wrench samples in the disturbance force-field explicitly targeting moment-dominated loads.

5 Limitations

Simulation-only evaluation. All quantitative ablations are conducted in MuJoCo; hardware results are qualitative demonstrations, with domain randomization as the sole sim-to-real bridge, so absolute success rates on real structured terrain remain uncharacterized and unmodeled contact phenomena will shift them. Even in simulation, the best variant reaches only 30 % $\mathrm{SR}_{\mathrm{hard}}$ on hard terrain (interior risers 0.22–0.28 m), a ceiling insufficient for reliable deployment on tall staircases.

Compliance and perception scope. The compliance benefit vanishes near 20 kg of centered load (Ours ≈37 % vs. Baseline ≈35 % SR at pelvis +20 kg): the sampled wrench distribution and leg-only actuation set an effective payload ceiling, and lateral or distributed loads fall outside the sampled space. The DCM planner depends on accurate elevation-map registration with no fallback under localization drift, and the forward-facing depth camera gives no lateral coverage. Finally, the terrain cost channels are collectively necessary but individually non-critical, removing any one keeps foot-target distance within 5 %, and their weights are hand-tuned rather than derived. Extended analysis of each limitation (adaptive-gait fragility, channel redundancy, upper-body coupling, and deformable/low-friction terrain) is given in Appendix.

6 Conclusion

We presented a physics-informed terrain affordance learning system for payload-robust perceptive humanoid locomotion. A multi-channel terrain cost drives a GPU-parallel DCM foothold planner and provides a dense terrain-affordance learning signal to the RL policy; a terrain-adaptive Bézier swing with tangent-guided foot orientation extends foothold tracking to joint position-and-orientation references. Because the cost is defined from sole geometry rather than learned per embodiment [12], it is platform-agnostic by construction. Lower-body compliance training via a cosine-power-weighted virtual spring-damper yields a learned wrench-dependent impedance (conceptually related to virtual impedance tracking [28] but without a separate impedance reference, and avoiding the online payload estimation of Kumar et al. [24]), improving payload robustness up to ∼15 kg centered load and for moment-dominated wrist loads without retraining. The full system trains with standard PPO in a single run (no teacher-student distillation [11, 15]) and is deployed zero-shot from simulation on a humanoid; quantitative real-world evaluation and a terrain × payload coupling study are the main remaining steps toward reliable deployment.

References

[1] Jerry Pratt, John Carff, Sergey Drakunov, and Ambarish Goswami. Capture point: A step toward humanoid push recovery. In *2006 6th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 200–207, 2006. doi:10.1109/ICHR.2006.321385.

[2] Eric Whitman and Gina Christine Fay. Terrain aware step planning system. U.S. Patent Application Publication US20200117198A1, assigned to Boston Dynamics, Inc., April 2020. URL https://patents.google.com/patent/US20200117198A1/en. Published Apr. 16, 2020; granted as US11287826B2.

[3] Brian Acosta and Michael Posa. Perceptive mixed-integer footstep control for underactuated bipedal walking on rough terrain. *IEEE Transactions on Robotics*, 41:4518–4537, 2025. doi:10.1109/TRO.2025.3587998.

[4] Zhaoyang Xiang, Upama Pant, and Ayonga Hereid. Perceptive variable-timing footstep planning for humanoid locomotion on disconnected footholds, 2026. URL https://arxiv.org/abs/2603.07400.

[5] Minku Kim, Brian Acosta, Pratik Chaudhari, and Michael Posa. Learning a vision-based footstep planner for hierarchical walking control. *2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids)*, pages 1–8, 2025. URL https://arxiv.org/abs/2508.06779.

[6] Haolin Song, Hongbo Zhu, Tao Yu, Yan Liu, Mingqi Yuan, Wengang Zhou, Hua Chen, and Houqiang Li. Gait-adaptive perceptive humanoid locomotion with real-time under-base terrain reconstruction. *IEEE Robotics and Automation Letters*, 11(4):4969–4976, 2026. doi:10.1109/LRA.2026.3664167.

[7] Yan Liu, Tao Yu, Haolin Song, Hongbo Zhu, Nianzong Hu, Yuzhi Hao, Xiuyong Yao, Xizhe Zang, Hua Chen, and Jie Zhao. FastStair: Learning to run up stairs with humanoid robots, 2026. URL https://arxiv.org/abs/2601.10365.

[8] Qingwei Ben, Botian Xu, Kailin Li, Feiyu Jia, Wentao Zhang, Jingping Wang, Jingbo Wang, Dahua Lin, and Jiangmiao Pang. Gallant: Voxel grid-based humanoid locomotion and local-navigation across 3D constrained terrains, 2025. URL https://arxiv.org/abs/2511.14625.

[9] Ho Jae Lee, Seungwoo Hong, and Sangbae Kim. Integrating model-based footstep planning with model-free reinforcement learning for dynamic legged locomotion. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11248–11255, 2024. doi:10.1109/IROS58592.2024.10801468.

[10] Huayi Wang, Zirui Wang, Junli Ren, Qingwei Ben, Tao Huang, Weinan Zhang, and Jiangmiao Pang. BeamDojo: Learning agile humanoid locomotion on sparse footholds. In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2025. doi:10.15607/RSS.2025.XXI.068. URL https://www.roboticsproceedings.org/rss21/p068.html.

[11] Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Legged locomotion in challenging terrains using egocentric vision, 2022. URL https://arxiv.org/abs/2211.07638.

[12] Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 7(62):eabk2822, 2022. doi:10.1126/scirobotics.abk2822. URL https://doi.org/10.1126/scirobotics.abk2822.

[13] Ilija Radosavovic, Sarthak Kamat, Trevor Darrell, and Jitendra Malik. Learning humanoid locomotion over challenging terrain, 2024. URL https://arxiv.org/abs/2410.03654.

[14] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model, 2024. URL https://arxiv.org/abs/2411.14386.

[15] Yuanhang Zhang, Younggyo Seo, Juyue Chen, Yifu Yuan, Koushil Sreenath, Pieter Abbeel, Carmelo Sferrazza, Karen Liu, Rocky Duan, and Guanya Shi. RPL: Learning robust humanoid perceptive locomotion on challenging terrains, 2026. URL https://arxiv.org/abs/2602.03002.

[16] Wandong Sun, Yongbo Su, Leoric Huang, Alex Zhang, Dwyane Wei, Mu San, Daniel Tian, Ellie Cao, Baoshi Cao, Yang Liu, Finn Yan, Ethan Xie, and Zongwu Xie. Now You See That: Learning end-to-end humanoid locomotion from raw pixels, 2026. URL https://arxiv.org/abs/2602.06382.

[17] Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. Humanoid parkour learning, 2024. URL https://arxiv.org/abs/2406.10759.

[18] Zhen Wu, Xiaoyu Huang, Lujie Yang, Yuanhang Zhang, Xi Chen, Pieter Abbeel, Rocky Duan, Angjoo Kanazawa, Carmelo Sferrazza, Guanya Shi, and C. Karen Liu. Perceptive Humanoid Parkour: Chaining dynamic human skills via motion matching, 2026. URL https://arxiv.org/abs/2602.15827.

[19] David Hoeller, Nikita Rudin, Dhionis Sako, and Marco Hutter. Anymal parkour: Learning agile navigation for quadrupedal robots, 2023. URL https://arxiv.org/abs/2306.14874.

[20] Péter Fankhauser, Michael Bloesch, and Marco Hutter. Probabilistic terrain mapping for mobile robots with uncertain localization. *IEEE Robotics and Automation Letters*, 3(4):3019–3026, 2018. doi:10.1109/LRA.2018.2849506. URL https://doi.org/10.1109/LRA.2018.2849506.

[21] David D. Fan, Kyohei Otsu, Yuki Kubo, Anushri Dixit, Joel Burdick, and Ali-akbar Agha-mohammadi. STEP: Stochastic traversability evaluation and planning for risk-aware off-road navigation. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi:10.15607/RSS.2021.XVII.021. URL https://www.roboticsproceedings.org/rss17/p021.html.

[22] Péter Fankhauser and Marco Hutter. A universal grid map library: Implementation and use case for rough terrain navigation. In Anis Koubaa, editor, *Robot Operating System (ROS): The Complete Reference (Volume 1)*, volume 625 of *Studies in Computational Intelligence*, chapter 5, pages 99–120. Springer, Cham, 2016. doi:10.1007/978-3-319-26054-9_5. URL https://doi.org/10.1007/978-3-319-26054-9_5.

[23] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning, 2023. URL https://arxiv.org/abs/2303.03381.

[24] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi:10.15607/RSS.2021.XVII.011. URL https://www.roboticsproceedings.org/rss17/p011.html.

[25] Tong Zhang, Boyuan Zheng, Ruiqian Nai, Yingdong Hu, Yen-Jen Wang, Geng Chen, Fanqi Lin, Jiongye Li, Chuye Hong, Koushil Sreenath, and Yang Gao. HuB: Learning extreme humanoid balance, 2025. URL https://arxiv.org/abs/2505.07294.

[26] Lequn Fu, Yijun Zhong, Xiao Li, Yibin Liu, Zhiyuan Xu, Jian Tang, and Shiqi Li. Load-aware locomotion control for humanoid robots in industrial transportation tasks, 2026. URL https://arxiv.org/abs/2603.14308.

[27] Anuj Pasricha, Joewie Koh, Jay Vakil, and Alessandro Roncone. Dynamics-compliant trajectory diffusion for super-nominal payload manipulation, 2025. URL https://arxiv.org/abs/2508.21375.

[28] Botian Xu, Haoyang Weng, Qingzhou Lu, Yang Gao, and Huazhe Xu. Facet: Force-adaptive control via impedance reference tracking for legged robots, 2025. URL https://arxiv.org/abs/2505.06883.

[29] Peiyuan Zhi, Peiyang Li, Jianqin Yin, Baoxiong Jia, and Siyuan Huang. Learning a unified policy for position and force control in legged loco-manipulation, 2025. URL https://arxiv.org/abs/2505.20829.

[30] Jiaqi Chen, Jonas Frey, Ruyi Zhou, Takahiro Miki, Georg Martius, and Marco Hutter. Identifying terrain physical parameters from vision - towards physical-parameter-aware locomotion and navigation. *IEEE Robotics and Automation Letters*, 9(11):9279–9286, 2024. doi:10.1109/LRA.2024.3455788. URL https://doi.org/10.1109/LRA.2024.3455788.

[31] Hajun Kim, Dongyun Kang, Min Gyu Kim, Gijeong Kim, and Hae Won Park. Online friction coefficient identification for legged robots on slippery terrain using smoothed contact gradients. *IEEE Robotics and Automation Letters*, 10(4):3150–3157, 2025. doi:10.1109/LRA.2025.3541428. URL https://doi.org/10.1109/LRA.2025.3541428.

[32] Johannes Englsberger, Christian Ott, and Alin Albu-Schäffer. Three-dimensional bipedal walking control based on divergent component of motion. *IEEE Transactions on Robotics*, 31(2):355–368, 2015. doi:10.1109/TRO.2015.2405592.

[33] Majid Khadiv, Alexander Herzog, S. Ali. A. Moosavian, and Ludovic Righetti. Walking control based on step timing adaptation. *IEEE Transactions on Robotics*, 36(3):629–643, 2020. doi:10.1109/TRO.2020.2982584.

[34] Twan Koolen, Tomas De Boer, John Rebula, Ambarish Goswami, and Jerry Pratt. Capturability-based analysis and control of legged locomotion, part 1: Theory and application to three simple gait models. *The International Journal of Robotics Research*, 31:1094–1113, 07 2012. doi:10.1177/0278364912452673.

[35] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017. URL https://arxiv.org/abs/1707.06347.

[36] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In Aleksandra Faust, David Hsu, and Gerhard Neumann, editors, *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pages 91–100. PMLR, 08–11 Nov 2022. URL https://proceedings.mlr.press/v164/rudin22a.html.

[37] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5026–5033, 2012. doi:10.1109/IROS.2012.6386109.

Appendix

A Implementation Details

DCM derivation (§3.1). Liu et al. [7] show that for a linear CoM height profile $z(t) = k_z t + z_0$ during swing, the Variable-Height Inverted Pendulum dynamics reduce to $\dot{\xi} \approx \omega(t)\,\xi$ when $a = 1 + k_z/(2\sqrt{g z_0}) \approx 1$; we adopt this result and set $\omega = \omega_0$. The steady-state nominal step offset that keeps the divergent mode bounded [34] is $b_x = v_x T/(e^{\omega_0 T} - 1)$ and $b_y = \mathrm{sign}(f)\, l_p/(1 + e^{\omega_0 T})$, where l_p is the nominal lateral inter-foot distance and $\mathrm{sign}(f) \in \{+1, -1\}$ selects the swinging leg.

Table 1: DCM foothold planner configuration.

Parameter	Value	Notes
z_0	$0.90\,\mathrm{m}$	Nominal CoM height
T_{swing} (= T in §3.1)	$0.45\,\mathrm{s}$	Half gait period
l_p	$0.20\,\mathrm{m}$	Lateral inter-foot distance
κ	0.4	Bézier apex bias gain
$b_{\min},\ b_{\max}$	$0.25,\ 0.75$	Bézier bias clamp bounds
$c_{\min}$	$0.05\,\mathrm{m}$	Base swing clearance
s	0.5	Clearance scale (m per m step height)
$c_{\max}$	$0.20\,\mathrm{m}$	Maximum swing clearance
$\delta_\ell^-, \delta_\ell^+$	$0.30,\ 0.05$	Lift phase: swing fractions before/after apex where tangent orientation is blended
δ_r^-, δ_r^+	$0.05,\ 0.25$	Reach phase: swing fractions before/after apex where landing orientation is blended
Elevation map	$37 \times 25, 0.05\,\mathrm{m}$	$1.80 \times 1.20\,\mathrm{m}$ at pelvis
Foot footprint kernel	$0.25 \times 0.10\,\mathrm{m}$	For Q_i, E_i pooling
Search window	$\pm 0.30\,\mathrm{m}$ (x and y)	Around nominal stride; fallback to nominal at mean valid height if empty
$h^*_{\min}$	$0.05\,\mathrm{m}$	Effective step height at standstill
$h^*_{\max}$	$0.28\,\mathrm{m}$	Effective step height at rated speed
v^*	$0.5\,\mathrm{m/s}$	Rated forward speed for h^*_{eff} scaling (Eq. (5))
$v_{\min}$	$0.05\,\mathrm{m/s}$	Standing override: planned target replaced by current foot position
Lateral penalty	$\beta = 2.5$	Lateral vs. sagittal asymmetry
$\alpha_{\mathrm{pos}},\ \alpha_{\mathrm{dcm}}$	$1.0,\ 0.5$	Position and capture-point stability
$\alpha_E,\ \alpha_Q,\ \alpha_M,\ \alpha_{\mathrm{climb}}$	$0.6,\ 4.0,\ 6.0,\ 1.5$	Terrain channels (Eq. (1))
Gait-adaptive extension		
$f_{\min},\ f_{\max}$	$1.0,\ 1.5\,\mathrm{Hz}$	Gait frequency bounds
α_{EMA}	0.2	Frequency smoothing weight

B Policy Learning Details

Observation space (actor). Base angular velocity, projected gravity vector, velocity command, joint positions and velocities relative to default, last action, binary foot contacts, gait phase ($\sin\phi$, $\cos\phi$), and a stacked depth image (four temporally delayed frames at $30\,\mathrm{Hz}$, blurred, range-clipped, and normalized). Proprioceptive signals are maintained as a 5-frame history at the $50\,\mathrm{Hz}$ control frequency; the depth image stack is updated at the camera pipeline rate ($30\,\mathrm{Hz}$).

Observation space (critic, training only). Actor observations plus: true CoM velocity, noiseless elevation map (ray misses filled with a sentinel value), ground-truth foot contact forces and heights, foot air times, and planner landing targets $\mathbf{p}_f^*$ for each foot.

Training hyperparameters. PPO with clip parameter $\epsilon = 0.2$, learning rate 10^{-3} with adaptive KL-based scheduling, GAE $\lambda = 0.95$, discount $\gamma = 0.99$, 5 learning epochs per rollout, 4 mini-batches, 24 steps per environment per rollout.

Table 2: Lower-body compliance training parameters (§3.4).

Parameter	Value	Notes
Spring-damper		
R_a	$0.40\,\mathrm{m}$	Anchor ball radius
k	$200\,\mathrm{N/m}$	Virtual spring stiffness
c	$20\,\mathrm{N{\cdot}s/m}$	Virtual damping
Wrench sampling		
ρ_{body}	0.6	Body-attached component probability
R_{close}	$0.10\,\mathrm{m}$	Body-attached ball radius
n	2	Cosine-power exponent (bias toward $-\hat{\mathbf{z}}$)
n_r	3	Cosine-power exponent (arm-extended forward bias)
$(a_{\mathrm{fwd}},\, a_{\mathrm{lat}},\, a_{\mathrm{vert}})$	$(0.50,\, 0.40,\, 0.30)\,\mathrm{m}$	Arm-extended half-ellipsoid semi-axes
ϵ	0.1	Isotropic mixture weight
Compliance targets (Eq. (13), (14))		
k_{leg}	$300\,\mathrm{N/m}$	Effective leg stiffness; gives $\alpha_z = k/(k + k_{\mathrm{leg}}) = 0.40$
α_z	$k/(k + k_{\mathrm{leg}}) = 0.40$	Height compliance gain
k_{rot}	$50\,\mathrm{N{\cdot}m/rad}$	Rotational stiffness (moment-to-tilt conversion)
$\alpha_\varphi,\, \alpha_\psi$	0.5	Orientation compliance gains (pitch, roll)

Table 3: Reward terms. Terrain-specific terms are highlighted; all other terms are standard locomotion rewards. I_f denotes foot contact, F_f contact force, L_c centroidal angular momentum.

Term	w	Definition
Velocity tracking		
$r_{\mathrm{vel},xy}$	$+3.5$	$\exp(-\lVert e_{xy}\rVert^2/0.25)$
$r_{\mathrm{vel},z}$	$+3.0$	$\exp(-e_z^2/0.5)$
Gait quality		
r_{gait}	$+2.0$	Phase match: stance/swing vs. gait clock
r_{airtime}	$+2.0$	Single-stance duration reward
r_{orient}	-7.0	$\lVert \mathbf{g}_{xy}\rVert^2$ (base tilt)
r_{pelvis}	-3.0	$\lVert \mathbf{g}_{xy}\rVert^2$ (pelvis)
r_{action}	-0.8	$\lVert a_t - a_{t-1}\rVert^2$
$r_{\mathrm{collision}}$	-2.0	Self-collision force $> 10\,\mathrm{N}$
r_{limits}	-1.0	Joint-limit violation
r_{slip}	-0.4	Tangential contact velocity
r_L	-0.001	$\lVert L_c\rVert^2$ angular momentum
Terrain-specific (contribution)		
r_{foothold}	$\mathbf{+2.1}$	Eq. (8); $\sigma_p = 10,\ \sigma_d \in \{0,\, 5.0\}$
$r_{\mathrm{clearance}}$	$\mathbf{-0.5}$	$\sum_f \lvert h_f^{\mathrm{foot}} - h_f^{\mathrm{terrain}} - d^*\rvert \cdot \lVert v_{f,xy}\rVert,\ d^* = 0.06\,\mathrm{m}$
r_{stumble}	$\mathbf{-0.5}$	$\mathbf{1}[\lVert F_f^{xy}\rVert > 4\lvert F_f^z\rvert \ \wedge\ \lVert v_f^{xy}\rVert > 0.15\,\mathrm{m/s}]$
r_{comply}	$\mathbf{-1.5}$	Eq. (15); $h^* = 0.90\,\mathrm{m},\ \alpha_z = k/(k + k_{\mathrm{leg}})$

Table 4: Domain randomization ranges applied at episode startup.

Parameter	Range	Mode
Waist-link mass offset	$[-2, +2]\,\mathrm{kg}$	add
Other body masses	$\times[0.95, 1.05]$	scale
Foot friction	$[0.3, 1.6]$	absolute
PD gains (K_p, K_d)	$\times[0.7, 1.1]$	scale
Joint stiffness / damping	$\times[0.7, 1.3]$	scale
Joint armature	$\times[0.2, 5.0]$	scale
CoM offset (pelvis)	$\pm 0.05\,\mathrm{m}$	add
Encoder bias	$\pm 0.015\,\mathrm{rad}$ per joint	add
Depth camera pose	pos x, y $\pm\,0.01\,\mathrm{m}$, z $(-0.03, +0.01)\,\mathrm{m}$; rot roll/yaw $\pm 2^\circ$, pitch $\pm 10^\circ$	add

Curriculum. A two-stage velocity curriculum starts at $[0, 0.5]\,\mathrm{m/s}$ and expands to $[0, 1.0]\,\mathrm{m/s}$ after 120,000 steps. The terrain curriculum uses stairs with 0.05–0.20 m risers and 0.30–0.60 m treads, open-width stairs, and flat sections.

Table 5: Actor network specification.

Component	Specification
Depth encoder	Conv2d: channels $[32, 32]$, kernel 3×3, max-pool FC: $[128, 64]$, output embedding $\in \mathbb{R}^{64}$
Actor MLP	Input: embed (64) $\oplus$ proprioception; hidden $(1024, 512, 256, 128)$, ELU activation
Action	Joint position increments, clipped to limits
History	5-frame obs history at 50 Hz

Gait-adaptive frequency action. The policy outputs a scalar gait frequency $f_t \in [f_{\min}, f_{\max}]$ as an additional action dimension appended after joint targets. Each physics step the gait phase advances as $\phi_t = (\phi_{t-1} + \Delta t_{\text{phys}} \cdot f_t) \bmod 1.0$. The raw policy output is clipped to $[f_{\min}, f_{\max}]$ and smoothed by an EMA before use: $f_t = (1-\alpha_{\text{EMA}})\, f_{t-1} + \alpha_{\text{EMA}}\ \text{clip}(f_t^{\text{raw}}, f_{\min}, f_{\max})$. The phase and its sinusoidal encoding $(\sin\phi_t, \cos\phi_t)$ are included in the actor observation; $f_{\min}$, $f_{\max}$, and α_{EMA} are listed in the DCM planner table (§A).

Auxiliary reward terms. *Clearance*: $\sum_f |h_f^{\text{foot}} - h_f^{\text{terrain}} - d^*| \cdot \|v_{f,xy}\|$ penalizes swing proximity to the terrain, velocity-weighted so edge collisions at high swing speed incur proportionally larger cost. *Stumble*: fires when $\|F_f^{xy}\| > 4|F_f^z|$ and $\|v_f^{xy}\| > 0.15\,\text{m/s}$; the criterion $\mu_{\text{eff}} > 4$ reliably distinguishes riser edge-strikes from high-force landings. *Compliant base height*: $(z_{\text{pelvis}} - z^*_{\text{virt}})^2$ tracks a virtual compliant target (eq. (13)); at zero compliance gain it reduces to the standard terrain-relative penalty, preserving the LIPM eigenfrequency $\omega_0 = \sqrt{g/z_0}$ through elevation changes.

C Experimental Details

Training setup. All policies are trained on a single NVIDIA H100 SXM5 80 GB GPU with 8192 parallel MuJoCo environments. Each iteration collects 24 simulation steps per environment. The PPO rollout and network update run on the same GPU; physics simulation is CPU-parallelized across the host node's available cores. All training uses the `mjlab` framework with MuJoCo 3.8.0 as the physics backend. Total wall-clock training time per variant is approximately 10–12 h.

Training terrain configurations. Figure 6 shows the eight terrain types used during curriculum training. Stair variants (a–d) are the primary training signal for the multi-channel terrain cost and foothold reward; slope and rough terrain (e–h) provide gradient and contact-stability diversity. The policy is never trained with riser heights above $0.20\,\text{m}$; the hard-terrain evaluation extends risers to $0.30\,\text{m}$ and is fully out-of-distribution.

Evaluation setup. All ablation variants are evaluated at iteration 20 k with 4096 environments using different random seeds in MuJoCo on stairs, slopes, and rough terrain. Standard evaluation terrain spans the same riser range as training: stair flights with risers $0.05\,\text{m}$ to $0.20\,\text{m}$ and treads $0.25\,\text{m}$ to $0.55\,\text{m}$, including 3-step and 5-step flights, open-width stairs (no side walls), slopes up to 23°, and random rough height fields 0–0.15 m. Hard-terrain evaluation uses riser heights 0.20–0.30 m with configurations unseen during training. An episode succeeds if the robot traverses $D_{\text{target}} = 10\,\text{m}$ without triggering fall termination (base tilt $> 70°$, self-collision, or joint-limit violation). SR_{hard} is computed over the strict interior riser range $[0.22,\ 0.28]\,\text{m}$, excluding the boundary cases at the edges of the hard sweep.

Ablation variants. All variants share the base training setup; only cost weights, reward terms, and observation inputs differ. Each is evaluated at iteration 20 k with 4096 environments over different random seeds.

TACT + Adaptive Gait (Ours): full system with all four terrain-cost channels ($\alpha_E, \alpha_Q, \alpha_M, \alpha_{\text{climb}} > 0$; $\beta > 1$ controls lateral asymmetry in d_{pos}), foothold tracking reward, and gait-adaptive frequency head [6].

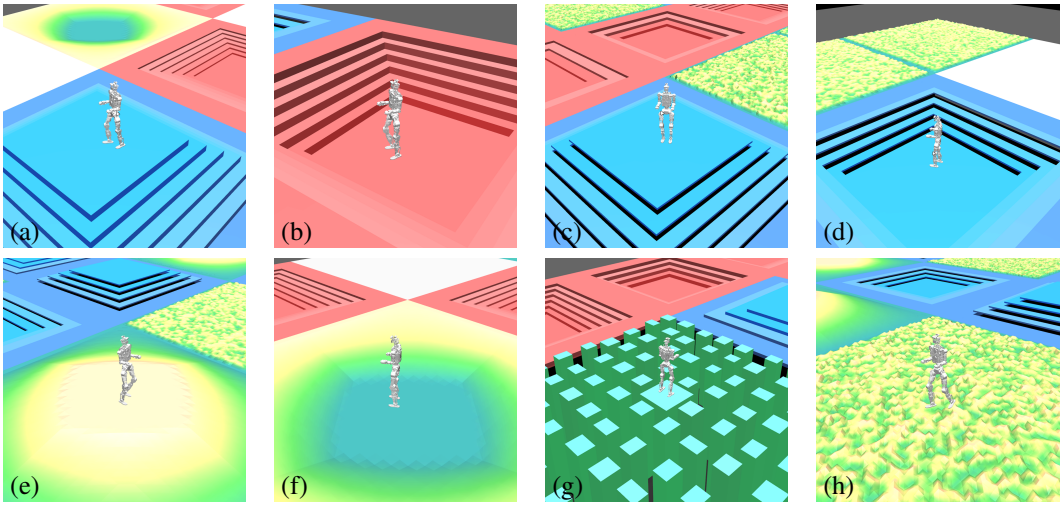

Figure 6: **Training terrain configurations.** (a) Pyramid stairs, ascending. (b) Pyramid stairs, descending. (c) Open-width stairs, ascending (no side walls). (d) Open-width stairs, descending. (e) Pyramid slope, ascending. (f) Pyramid slope, descending. (g) Stepping stones. (h) Gravel (random rough height field). Stair risers span $0.05\,\mathrm{m}$ to $0.20\,\mathrm{m}$ with treads $0.25\,\mathrm{m}$ to $0.55\,\mathrm{m}$; slopes up to 23°; rough field height 0–$0.15\,\mathrm{m}$. Hard-terrain evaluation extends risers to $0.30\,\mathrm{m}$ (out-of-distribution).

TACT-only: all terrain-cost channels and foothold tracking reward active; gait-adaptive frequency head removed.

Adaptive Gait only: gait-adaptive frequency head added; all terrain-cost channel weights and foothold tracking reward set to zero ($\alpha_Q{=}\alpha_E{=}\alpha_M{=}\alpha_{\text{climb}}{=}0$, $w_{\text{foothold}}{=}0$); no privileged height-map input to the critic.

Baseline: standard depth-map perceptive policy; neither TACT channels nor gait adaptation; no privileged critic input.

Payload generalization (terrain). Three attachment conditions are evaluated: pelvis-mounted $+15\,\mathrm{kg}$ and $+20\,\mathrm{kg}$ (CoM offset; tests balance under downward wrench) and wrist-mounted $+10\,\mathrm{kg}$ (large moment arm; tests trunk-tilt compliance). Mass is added at test time with no policy modification. A no-mass-DR baseline is trained identically to the full policy but without waist-link mass offset randomization. Each condition uses 100 episodes at $v_x = 0.5\,\mathrm{m\,s^{-1}}$ on a 5-step staircase with $0.20\,\mathrm{m}$ risers.

Planned evaluations (future work). The following targeted protocols are specified for completeness; their quantitative results are deferred to future work and not claimed here: non-nominal pelvis-height tracking ($h^* \in \{h_0{-}6,\ h_0{-}3,\ h_0{+}3\}\,\mathrm{cm}$, recording h^*-RMSE), and a maximum sustained downward pull sweep ($F_{\max}$ at $\mathrm{SR} > 90\,\%$).

Metric definitions.

- SR: success rate (%), fraction of episodes that traverse D_{target}.
- $\mathrm{SR}_{\text{hard}}$: SR restricted to riser heights $\Delta z \in [0.22, 0.28]\,\mathrm{m}$; isolates performance on the most demanding subset of the hard sweep (strict interior, excluding boundary cases at 0.20 and 0.30 m).
- Q_{c}: mean flatness cost Q_i (Eq. (2)) at the elevation-map cell under each foot at touchdown, averaged over all touchdown events.
- F^{95}: $\mathrm{pct}_{95}(\|F_f\|/(mg/2))$, 95th-percentile normalized ground reaction force at touchdown events.
- E_v: velocity-tracking RMSE, $\sqrt{N^{-1}\sum_{t=1}^{N}\|v_{xy,t} - v_{\mathrm{cmd},t}\|^2}$ (m/s).

- P: mean absolute mechanical power, $N^{-1}\sum_{t=1}^{N}\|\boldsymbol{\tau}_t \odot \dot{\mathbf{q}}_t\|_1$ (W); lower values indicate more efficient gait [16].
- **Foot-target distance**: mean planar Euclidean distance between the executed touchdown centroid and the nearest DCM-planned foothold at swing initiation, averaged over all touchdown events in the episode (the "foot-target distance" reported in §4.1).

D Hardware and Deployment

Sensor architecture. The actor policy receives a single perceptual stream at runtime: a noisy depth image from a forward-facing RGB-D (color-and-depth) camera. A raycaster-based elevation map mounted at the pelvis is available during training only and is used exclusively by the privileged critic; it is not available at deployment. Additional per-foot proximity rays and a pelvis height sensor supply scalar terrain signals used exclusively in reward computation and privileged critic observations. The simulation models each sensor with geometry-matched raycasters; the depth camera adopts the intrinsics of a RealSense D435 and injects pose domain randomization to bridge the sim-to-real gap.

Depth camera processing pipeline. Raw depth frames are captured at 30 Hz, cropped to remove border artifacts, resized to 30×30 px, Gaussian-blurred to suppress raycaster aliasing, and normalized to $[0, 1]$. Four temporally spaced frames are stacked into a history buffer and fed to the depth encoder at each 50 Hz control step. Camera intrinsics are matched to a RealSense D435; extrinsic pose domain randomization (table 4) covers mounting uncertainty. The identical pipeline runs at deployment with no modification.

Deployment. The trained actor (depth encoder + policy MLP) is exported to ONNX and executed at 50 Hz on the robot's onboard computer via a joint-position PD interface. No real-world fine-tuning is performed; the sim-to-real gap is closed entirely by domain randomization during training.

E Additional Results

Flat-Terrain Payload Generalization

Fig. 7 evaluates payload generalization across three conditions (pelvis +15 kg, pelvis +25 kg, wrist +15 kg) on flat terrain and compares Ours against the Baseline, isolating compliance behavior from terrain traversal difficulty. At moderate load (pelvis +15 kg, wrist +15 kg), Ours maintains 76–79 % SR against the Baseline's 67–76 % while consuming 9–20 % less power, consistent with compliance training suppressing impulsive GRF recovery. At high centered load (pelvis +25 kg), Ours drops to 47.5 % (below the Baseline's 60.0 %), indicating that a large CoM shift exceeds the compliance training distribution and causes over-compensation; power remains lower (257 vs. 283 W), confirming the policy still reduces impulsive contacts but can no longer maintain balance. Wrist-mounted mass, which generates a large moment rather than a direct CoM offset, is handled comparably in SR but with a clear power advantage for Ours (186 vs. 231 W), consistent with the arm-extended wrench samples in the disturbance force-field.

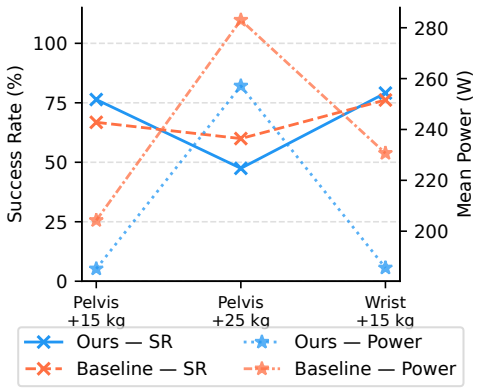

Figure 7: SR (%) and mean power (W) vs. payload mass on flat terrain.

Cross-Embodiment Generalization

Fig. 8 shows the ablation evaluated on two platforms (Platform-A (H1-2 class) and Unitree G1) using identical reward weights and hyperparameters, with only the robot asset and joint configuration changed. The relative ordering of variants is preserved across embodiments, supporting the platform-agnostic claim in §6.

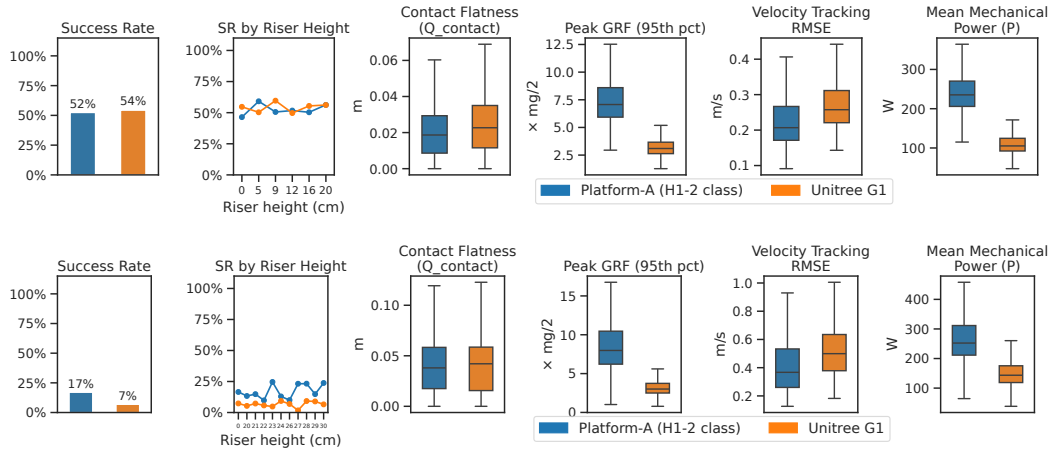

Figure 8: Cross-embodiment ablation on Platform-A (H1-2 class) and Unitree G1: **standard terrain** (top) and **hard terrain, out-of-distribution** (bottom). Identical reward weights and hyperparameters; only robot asset and joint configuration differ.

F Extended Limitations

Adaptive-gait fragility under kinematic stress. The *Adaptive Gait only* variant falls *below* the Baseline on hard terrain (18 % vs. 19 % SR): EMA-smoothed frequency re-timing without terrain-quality guidance modifies stance-to-swing departure timing with no corresponding adjustment to the landing target, producing foot-target mismatches that the foothold reward would otherwise suppress. Adaptive gait frequency is therefore an auxiliary capability contingent on accurate foothold guidance, not an independent locomotion primitive, and its benefit is confined to regimes where kinematic margins are non-trivial. A natural remedy is to co-condition the frequency head on the terrain-cost signal so that re-timing decisions are aware of contact quality, or to gate frequency updates on a kinematic feasibility check before committing to the new landing target.

Terrain-cost channel redundancy. Removing any single terrain cost channel leaves foot-target distance within 5 % of the full system (0.085–0.093 m vs. 0.088 m), while removing all three raises it 2.8× to 0.251 m. The channels are thus collectively necessary but individually non-critical, and the manually tuned weighting (α_Q=4.0, α_E=0.6, α_M=6.0) does not follow from principled minimality. A compressed two-channel formulation and a weight-sensitivity sweep across platforms are left to future work. More broadly, replacing hand-tuned weights with a differentiable meta-learning or Bayesian optimization pass over the cost coefficients could yield principled, platform-adaptive weightings without requiring per-robot re-tuning.

Compliance distribution and upper-body coupling. Near 20 kg of centered CoM shift, two factors compound. First, the spring-damper sampling concentrates on body-attached downward-biased and arm-extended forward-biased scenarios; lateral carry, two-point distributed loads, and dynamically swinging payloads produce wrenches outside this sampled space. Second, the policy controls leg joints only: at high payload masses, upper-body inertia shifts the effective CoM in ways the lower-body policy cannot observe, as the pelvis inertial measurement unit (IMU) and joint encoders provide no direct signal of trunk angular momentum. Together these set the effective payload ceiling. Both failure modes have tractable remedies: broadening the wrench distribution to include lateral

and multi-point loads during training, and extending the controlled joints to include the torso or arm would directly address the sampling gap and the observability gap respectively. An online payload estimator feeding a residual compliance target could further extend the effective range without requiring a larger wrench training distribution.

Tangent-guided orientation failure on slopes. The Bézier swing reference orients the sole using the arc tangent of the trajectory at the landing point, which on discrete stair steps closely approximates the riser-face normal and drives the foot to land with the sole parallel to the tread. On smooth slopes this mechanism fails: the swing trajectory approaches the inclined surface from above along a path whose touchdown tangent is approximately horizontal, independent of terrain slope angle. The resulting orientation reference does not rotate the sole to match the surface normal, so the foot arrives level on a tilted surface. On ascending grades the surface rises toward the direction of travel, placing the toe geometrically closer to the ground than the heel at the moment of touchdown; the foot therefore contacts toe-first, concentrating GRF on the toe edge. On descending grades the inverse holds, producing heel-first contact with the toe elevated. Both deviate from the designed flat-sole landing, narrow the friction cone, and raise F^{95}; on grades above roughly 15° the resulting ankle-roll moment is frequently sufficient to trigger fall termination.

A complementary failure arises in the foothold planner. The flatness cost Q_i measures height variance within a local neighbourhood around each candidate cell: a uniformly inclined surface produces near-zero inter-cell variance, so $Q_i \approx 0$ regardless of slope angle, and the planner assigns high contact quality to steep cells. The steepness cost E_i is intended to compensate, but with the current weighting ($\alpha_E = 0.6$ vs. $\alpha_M = 6.0$) it is the weakest channel and is insufficient to steer footholds away from high-gradient cells when flatness and reachability are simultaneously satisfied. The combination of incorrect sole orientation and inadequate slope-avoidance pressure jointly accounts for the observed degradation on slope terrain relative to stairs of comparable height gain per step. A corrected planner would supplement Q_i with a surface-normal tilt term and extend the swing orientation reference from the trajectory tangent to the estimated surface normal at the target cell, analogous to normal-aligned foot placement used in model-based planners.

Additional Bézier-specific failures emerge on slopes. The apex height is set relative to the take-off elevation; on ascending grades the landing point is higher than take-off by $\Delta z = l_{\text{stride}} \sin\theta$, reducing effective clearance over the landing surface by the same amount. At steep grades and long strides the foot can contact the slope during mid-swing before the apex is reached, a form of early collision absent on flat terrain and stairs. On descending grades the inverse produces excess apex clearance, which extends swing duration and can desynchronize foot arrival with the DCM-timed stance transition. Addressing both the orientation and clearance failures requires replacing the trajectory-tangent reference with a surface-normal estimate at the target cell (available from the elevation map) and setting the apex height relative to the landing elevation rather than the take-off elevation, a straightforward modification to the existing Bézier parameterization.

Elevation-map dependency and terrain scope. The DCM planner consumes a pelvis-mounted elevation map requiring accurate extrinsic odometry; localization drift degrades foothold selection monotonically with no fallback planner, and the forward-facing depth camera provides no lateral or rearward coverage. Highly deformable surfaces (sand, gravel, foam) violate the rigid-terrain assumption underlying both the Bézier clearance computation and the flatness cost Q_i, causing systematic under-clearance, and wet or icy contacts fall outside the domain-randomization friction range ($\mu \in [0.3, 1.6]$). Terrain with moving obstacles is not modeled. Robustness to localization drift could be improved by incorporating a fallback stance-hold controller that activates when odometry confidence is low, and by fusing the depth camera into a local elevation map that does not depend on global pose. Extending the friction domain randomization range and adding deformable-terrain simulation (e.g., via compliant contact models in MuJoCo) are the most direct paths to covering low-friction and soft-surface deployments. Handling dynamic obstacles would require integrating a short-horizon obstacle prediction model into the foothold search window.