

Information-Directed Active Calibration of Unknown Velocity Dynamics

Daniel Okafor • Priya Raghunathan • Hana Sato

Advanced Robotics Group, Toyota Research Institute, Palo Alto, CA, USA

PUBLISHED

August 2, 2026

REGISTERED

August 8, 2026

RECORD TYPE

Preprint

ABSTRACT

Navigation stacks for legged robots usually command a vendor locomotion controller through an opaque velocity interface whose gains, saturation limits, and latency differ from their nominal values and must be calibrated before a planner can trust its motion model. Calibration requires choosing probe commands, and uniform excitation spends much of its motion on directions whose response is already well determined. This paper formulates probe selection as information-directed sampling. A particle posterior is maintained over a six-parameter state-delay model of the interface, comprising three per-axis gains, a planar-speed and a yaw-rate saturation limit, and a transport delay, and each probe is chosen to maximize the expected information gain about these parameters per unit of commanded path length. The information gain is evaluated through the saturated model, and a hierarchical shrinkage prior couples the per-axis gains so that weakly excited directions borrow strength from well-excited ones. The method is compared with uniform sweep excitation, random probing under a motion budget, and the planner-conditioned probes of GAUGE over 200 simulated calibration sessions of a Unitree B2 in Isaac Lab, with validation on five real-robot sessions. Information-directed probing attains a worst-direction calibration error of 8.9 percent against 9.5 to 10.3 percent for the baselines, reaches navigation-grade accuracy in 38 s against 40 to 42 s, and uses 11.4 m of commanded motion against 11.8 to 13.0 m. Ablations indicate that cost weighting, saturation awareness, and shrinkage are each needed to outperform uniform excitation.

INDEX TERMS active calibration, information-directed sampling, Bayesian experimental design, legged robots, velocity interface, system identification

KAGG NOTE Calibration motion is best spent where the posterior over the interface is most uncertain per meter of commanded travel, not where a fixed sweep or the planner's habits happen to place it.

01 Introduction

Navigation stacks on commercial quadrupeds rarely command joint torques. They issue body-frame twist commands to a vendor locomotion controller through a velocity interface, and the local planner, whether a dynamic-window search [1], a sampling-based model-predictive controller [2], or a learned local policy [3], predicts how the robot will move in response. The controller behind the interface may be a convex model-predictive controller [4] or a learned policy [5, 6]; in either case its internal gains, rate limits, saturation, and latency are rarely documented in a form a planner can use, and they shift with firmware, gait mode, and terrain. The planner's motion model is therefore only as good as the calibration of an interface that it cannot inspect.

Calibrating such an interface means commanding probe motions and fitting a command-to-motion model to the measured response. Probing is not free. The robot must move inside a bounded workspace, every meter of probing costs time and wear, and probes near the saturation boundary engage the vendor's safety logic. Uniform excitation, which sweeps a fixed grid of directions and amplitudes, treats every direction as equally uncertain and therefore spends motion on directions whose response is already well determined. GAUGE [7] takes a different view: it conditions probes on the commands the navigation planner is expected to issue, which concentrates

effort on frequently used directions but can leave rarely used ones, such as lateral, combined, and yaw-heavy commands, comparatively poorly identified. These are precisely the commands a planner falls back on during recovery maneuvers in clutter.

Neither strategy selects a probe according to what the current estimate is uncertain about, and neither weighs that uncertainty against the motion each probe costs. Bayesian experimental design [8, 9] supplies the first principle, and information-directed sampling [10] supplies a template for trading information against cost. What remains is to make the principle concrete for an interface that is nonlinear, saturating, and delayed, and whose probes must respect a motion budget and a safety layer.

This paper develops such a procedure. The interface is modeled by a six-parameter state-delay model with per-axis gains, a planar-speed and a yaw-rate saturation limit, and a transport delay. A particle posterior over these parameters is updated after each probe, and the next probe is chosen from a finite candidate set to maximize a saturation-aware estimate of expected information gain divided by the probe's commanded path length. The contributions are:

1. A formulation of opaque-interface calibration as cost-normalized sequential Bayesian experimental design over an explicit state-delay model, together with directional error metrics (worst-direction and median-direction error) that match how a navigation planner uses the calibrated model.

2. An information-directed probe-selection method that combines a hierarchical shrinkage prior over the per-axis gains, a moment-matched information-gain estimate evaluated through the saturated model, selection by information per unit of commanded motion, and a reachability filter for workspace safety.
3. A comparison against uniform sweep excitation, random probing under a motion budget, and GAUGE's planner-conditioned probes over 200 simulated Unitree B2 calibration sessions and five real-robot sessions, in which the proposed method attains the lowest worst-direction error (8.9% against 9.5–10.3%), the shortest time to navigation-grade accuracy (38 s against 40–42 s), and the smallest motion budget (11.4 m against 11.8–13.0 m).
4. An ablation and failure analysis showing that removing any one of cost weighting, saturation awareness, or shrinkage raises worst-direction error above that of uniform excitation, and documenting the failure modes observed under a degenerate prior and at the saturation boundary on hardware.

02 Related Work

2.1 Calibration of opaque velocity interfaces

GAUGE [7] is the work closest to ours. It treats the quadruped's velocity controller as an opaque interface rather than attempting to model the underlying locomotion policy: it issues probe commands conditioned on the planner, so that probing reflects the command distribution the downstream planner will actually produce, and it fits a calibrated model of the command-to-motion response online, which the planner then consumes. Two of its design decisions shape this paper. The first, treating the interface as a black box between planner and gait controller, is adopted here unchanged. The second, choosing probes by planner relevance, is replaced here by posterior uncertainty and motion cost. Planner conditioning is efficient when the deployment command distribution is known and stationary; it under-samples commands outside that distribution, and the worst-direction error in Section 5 reflects this. The two criteria are complementary rather than exclusive, and Section 6 discusses how planner relevance could weight the information objective.

Calibration of command-to-motion models has a long history for wheeled robots. Borenstein and Feng [11] introduced a prescribed square-path test that separates systematic odometry errors of differential-drive robots; the excitation is fixed and designed by insight into the error sources. Censi et al. [12] calibrate odometry and exteroceptive sensor parameters jointly and analyze which trajectories render the parameters observable. Both works establish that the choice of excitation determines what can be identified. The present setting differs in that the interface is a legged gait controller with saturation and delay, and probes are chosen sequentially from the current posterior rather than prescribed in advance.

2.2 Bayesian experimental design and information-directed sampling

Chaloner and Verdinelli [8] review Bayesian experimental design and the utilities on which it rests, among them expected information gain, which for linear-Gaussian models reduces to D-optimality. Rainforth et al. [9] survey modern sequential and adaptive design, with emphasis on estimators of expected

information gain for implicit and nonlinear models. Foster et al. [13] propose variational bounds that make information-gain estimation tractable when nested Monte Carlo is too expensive; the interface model used here is cheap enough to simulate that a moment-matched estimate suffices. Deep adaptive design [14] amortizes the sequential design policy into a network trained offline, which yields fast online decisions; this paper keeps online greedy optimization because the candidate set is small and the safety constraints change with the robot's position, although amortization is a natural extension. Houlisby et al. [15] express mutual information between parameters and predictions as disagreement between posterior samples; the estimator in Section 4 follows the same idea, measuring the spread of particle predictions relative to measurement noise.

Information-directed sampling (IDS) [10] selects actions that minimize the ratio of squared expected regret to expected information gain, balancing the cost of an action against what it reveals. Kirschner and Krause [16] extend IDS to heteroscedastic observation noise, which is relevant here because gait-induced velocity fluctuations grow with commanded speed. Kirschner et al. [17] apply IDS to linear partial monitoring, where the observations an action produces need not reveal its loss directly; calibration has an analogous structure, since the quantity paid (motion) is distinct from the quantity learned (interface parameters). Srinivas et al. [18] connect regret to maximum information gain in Gaussian-process bandits and thereby link bandit exploration to experimental design. The criterion used here is a pure-exploration specialization of IDS in which regret is replaced by commanded path length.

2.3 Active system identification and exploration

Mehra [19] surveys optimal input design for parameter estimation in dynamic systems, typically by maximizing a scalar function of the Fisher information for linear models. Wagenmaker and Jamieson [20] show that adapting inputs to the current estimate yields instance-optimal identification rates for linear dynamical systems, and Mania et al. [21] give guarantees for active identification of nonlinear systems that are linear in known features by planning trajectories that excite those features. Saturation breaks the assumptions of these analyses: the information a probe carries about a gain collapses once the command crosses the saturation boundary, and the Fisher information becomes strongly input-dependent. Buisson-Fenet et al. [22] actively learn Gaussian-process dynamics models by planning entropy-maximizing trajectories, a nonparametric alternative to the low-dimensional parametric model adopted here.

In model-based reinforcement learning, Shyam et al. [23] drive exploration by disagreement within an ensemble of learned models, and Pathak et al. [24] use ensemble disagreement as an intrinsic reward; Sukhija et al. [25] propose optimistic active exploration of dynamical systems with guarantees stated in terms of information gain. These methods target rich models and long horizons, whereas calibration involves a small parameter vector and tight motion budgets. Memmel et al. [26] are close in spirit: they learn an exploration policy that maximizes Fisher information about physical parameters of a manipulation scene and use the identified parameters for sim-to-real transfer. The present method differs in operating online on the robot, with an explicit motion cost. Lew et al. [27] combine active dynamics learning with safety constraints in a sequential exploration-exploitation framework; here safety

is handled by a simpler reachability filter in combination with the vendor's safety layer.

2.4 Legged locomotion controllers and sim-to-real adaptation

Learned locomotion controllers expose exactly the kind of velocity interface calibrated in this paper. Hwangbo et al. [5], Lee et al. [28], and Miki et al. [6] train velocity-tracking policies in simulation and deploy them on quadrupeds over challenging terrain; Rudin et al. [29] show that massively parallel simulation makes such training fast. GPU simulators such as Isaac Gym [30] and the Orbit framework [31], from which Isaac Lab descends, provide the simulation environment used in Section 5. Model-based controllers such as convex MPC [4] also track commanded twists, with tracking behavior shaped by horizon, constraints, and gait schedule.

Sim-to-real research mostly addresses the controller rather than the planner's model of it. Domain randomization [32] and dynamics randomization [33] train policies that are robust across parameter ranges instead of identifying the parameters. Yu et al. [34] pair a universal policy with an online system-identification module, and RMA [35] estimates an extrinsics vector from proprioceptive history so that the policy adapts implicitly. BayesSim [36] infers a posterior over simulator parameters from real trajectories, and Chebotar et al. [37] iteratively adapt the randomization distribution with real-world experience. These methods adapt the controller or the simulator; in contrast, the present work treats the controller as fixed and calibrates the planner's model of it, with explicit design of which data to collect.

2.5 Navigation and safety over velocity interfaces

The dynamic window approach [1] searches velocity space within a window derived from acceleration limits and therefore depends directly on an accurate velocity and acceleration model. Model predictive path integral control [2] evaluates sampled command sequences by rolling out a dynamics model, so systematic model error biases every rollout. Learned navigation systems, including end-to-end locomotion and local navigation [3], self-supervised navigation from experience [38], and online traversability learning [39], still act through a commanded twist and inherit any mismatch between command and response. Control barrier functions [40] filter commands to maintain safety, but their margins assume a model of how the commanded input is realized; an unmodeled gain or delay erodes those margins. These dependencies motivate the use of worst-direction rather than average calibration error as the primary metric in this paper.

03 Problem Formulation

3.1 Velocity interface model

At each control step t with period $\Delta t = 10$ ms, the planner sends a body-frame command $\mathbf{u}_t = [v_{x,t}^c, v_{y,t}^c, \omega_t^c]^\top \in \mathcal{U}$, where $\mathcal{U} = \{\mathbf{u} : |v_x^c| \leq \bar{u}_x, |v_y^c| \leq \bar{u}_y, |\omega^c| \leq \bar{\omega}\}$ is the box of admissible commands. The state estimator returns a measured body twist $\mathbf{y}_t \in \mathbb{R}^3$. The interface is modeled as a delayed, saturated first-order lag:

$$\mathbf{r}_t = G \mathbf{u}(t\Delta t - d), \quad G = \text{diag}(g_x, g_y, g_\omega), \quad (1)$$

$$\sigma_\theta(\mathbf{r}) = \begin{bmatrix} \mathbf{r}_{xy} \min(1, \bar{v}/\|\mathbf{r}_{xy}\|) \\ \text{clip}(r_\omega, -\bar{\omega}, \bar{\omega}) \end{bmatrix}, \quad (2)$$

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \frac{\Delta t}{\tau} (\sigma_\theta(\mathbf{r}_t) - \mathbf{x}_t), \quad \mathbf{y}_t = \mathbf{x}_t + \epsilon_t. \quad (3)$$

Here $\mathbf{r}_t$ is the delayed, scaled reference, $\mathbf{u}(t\Delta t - d)$ is the command history evaluated at a delay d by linear interpolation between control steps, $\mathbf{r}_{xy}$ and r_ω are the planar and yaw components of $\mathbf{r}_t$, $\bar{v}$ is the planar-speed saturation limit, $\bar{\omega}$ is the yaw-rate saturation limit, $\mathbf{x}_t$ is the realized body twist, τ is a fixed lag time constant, and ϵ_t is zero-mean measurement noise. The unknown parameter vector is

$$\theta = (\log g_x, \log g_y, \log g_\omega, \log \bar{v}, \log \bar{\omega}, d) \in \mathbb{R}^6. \quad (4)$$

Log coordinates keep gains and limits positive and make multiplicative errors symmetric. The planar limit acts on the norm of $\mathbf{r}_{xy}$ rather than per axis because legged gait controllers typically cap total stepping speed. The steady-state response to a constant command is $f_\theta(\mathbf{u}) = \sigma_\theta(G\mathbf{u})$, which is the quantity a planner's motion model needs most.

3.2 Directional calibration error

Navigation-relevant error is directional. Let $\mathcal{D} = \{-1, 0, 1\}^3 \setminus \{\mathbf{0}\}$ be the 26 sign patterns of the command axes, let $\bar{U} = \text{diag}(\bar{u}_x, \bar{u}_y, \bar{u}_\omega)$, and let $\mathbf{u}_j(\alpha) = \alpha \bar{U} \mathbf{e}_j$ for $\mathbf{e}_j \in \mathcal{D}$ and amplitude $\alpha \in \mathcal{A}$. Let $f^*(\mathbf{u})$ be the measured steady-state response of the true interface to a held command $\mathbf{u}$. For an estimate $\hat{\theta}$, the error in direction j is

$$\varepsilon_j(\hat{\theta}) = \frac{1}{|\mathcal{A}|} \sum_{\alpha \in \mathcal{A}} \frac{\|W(f_{\hat{\theta}}(\mathbf{u}_j(\alpha)) - f^*(\mathbf{u}_j(\alpha)))\|}{\|Wf^*(\mathbf{u}_j(\alpha))\|}, \quad (5)$$

where $W = \text{diag}(1, 1, \ell)$ converts yaw rate into the equivalent tangential speed at a characteristic radius ℓ . The worst-direction error is $E_{\max} = \max_j \varepsilon_j$ and the median-direction error is $E_{\text{med}} = \text{median}_j \varepsilon_j$, both reported in percent. The worst-direction error is the primary metric because a planner that relies on a poorly calibrated direction during an evasive maneuver is not protected by accuracy in the other 25 directions.

3.3 Probes, cost, and the selection problem

A probe $a = (\mathbf{e}, \alpha, T_h)$ is a command profile $\mathbf{u}^a(s)$, $s \in [0, T_a]$, that ramps along direction $\bar{U}\mathbf{e}$ to amplitude α with a raised-cosine ramp of duration T_r , holds for T_h , and ramps back to zero, so $T_a = T_h + 2T_r$. Its cost is the commanded path length

$$c(a) = \int_0^{T_a} \|W\mathbf{u}^a(s)\| ds, \quad (6)$$

which equals the planar path length for translational probes and charges yaw probes at the characteristic radius ℓ . After k probes the data are $\mathcal{H}_k = \{(a_i, \mathbf{y}^{(i)})\}_{i=1}^k$. Let $\hat{\theta}_t$ be the estimate at time t and define the navigation-grade time T_{nav} as the first time after which the direction-averaged error $\frac{1}{26} \sum_j \varepsilon_j(\hat{\theta}_t)$ stays below a fixed tolerance ε_{nav} for the remainder of the session. The motion budget L is the sum of $c(a_i)$ over all probes and recentering segments executed before T_{nav} . A probe-selection policy maps $\mathcal{H}_k$ to a_{k+1} ; ideally it minimizes the expected motion budget needed to reach navigation-grade accuracy while keeping $E_{\max}$ small. That objective is a partially observed dynamic program and is intractable to solve exactly, so Section 4 adopts a greedy, cost-normalized surrogate.

04 Method

4.1 Posterior with hierarchical shrinkage

The per-axis gains share actuation: the same legs and the same gait scheduler realize longitudinal, lateral, and yaw motion, so a firmware scaling or a change in foot friction tends to shift all three gains together. This is encoded with a hierarchical prior,

$$\log g_i = \mu_g + \eta_i, \quad \mu_g \sim \mathcal{N}(m_g, s_g^2), \quad \eta_i \sim \mathcal{N}(0, \lambda^2), \quad (7)$$

for $i \in \{x, y, \omega\}$, with independent priors $\log \bar{v} \sim \mathcal{N}(m_v, s_v^2)$, $\log \bar{\omega} \sim \mathcal{N}(m_\omega, s_\omega^2)$, and $d \sim \mathcal{U}[d_{\min}, d_{\max}]$. With $\lambda < s_g$, evidence about one gain shrinks the others toward a common value, so directions that have not yet been probed borrow strength from those that have. In the ablation without shrinkage, the gains receive independent priors with the same marginal variance $s_g^2 + \lambda^2$, which isolates the effect of the coupling from that of the prior width.

The posterior is represented by $N = 512$ weighted particles $\{(\theta^{(n)}, \mu_g^{(n)}, w^{(n)})\}_{n=1}^N$ [41]. A particle filter is preferred to an unscented Kalman filter [42] because saturation induces non-Gaussian posteriors: when a probe saturates, the gain and the saturation limit trade off along a curved ridge in parameter space, which sigma-point linearization of the clipping function represents poorly. Body-velocity measurements oscillate at the stride frequency, so the raw 100 Hz signal is averaged over windows of length $T_w = 0.25$ s before it enters the likelihood. With $\bar{\mathbf{y}}_m^{(k)}$ denoting the m -th windowed measurement of probe k and $h_m(\theta, a_k)$ the corresponding windowed prediction obtained by simulating Eqs. (1)–(3) from the measured twist at probe start, the weights are updated as

$$w^{(n)} \propto w^{(n)} \exp\left(-\frac{1}{2} \sum_m \|\bar{\mathbf{y}}_m^{(k)} - h_m(\theta^{(n)}, a_k)\|_{R_m^{-1}}^2\right), \quad (8)$$

where R_m is the windowed noise covariance. When the effective sample size $1/\sum_n (w^{(n)})^2$ falls below $N/2$, particles are resampled systematically and perturbed with Gaussian jitter whose covariance is 0.01 times the weighted particle covariance, which prevents collapse when the prior is poorly placed. The point estimate $\hat{\theta}$ is the weighted posterior mean.

4.2 Saturation-aware information gain

For a candidate probe a , let $\mathbf{h}(\theta, a) \in \mathbb{R}^{3M_a}$ stack the M_a windowed predictions of the probe. The particle predictions define a predictive mean

$$\bar{\mathbf{h}}_a = \sum_n w^{(n)} \mathbf{h}(\theta^{(n)}, a) \quad (9)$$

and, with the deviations $\tilde{\mathbf{h}}_a^{(n)} = \mathbf{h}(\theta^{(n)}, a) - \bar{\mathbf{h}}_a$, a predictive covariance

$$C_a = \sum_n w^{(n)} \tilde{\mathbf{h}}_a^{(n)} (\tilde{\mathbf{h}}_a^{(n)})^\top, \quad (10)$$

and the expected information gain about θ is estimated as

$$\hat{I}(a) = \frac{1}{2} \log \det(I + R_a^{-1} C_a). \quad (11)$$

With Gaussian noise of covariance R_a , the conditional entropy of the observation given θ is $\frac{1}{2} \log \det(2\pi e R_a)$ exactly, and the Gaussian maximizes entropy for a given covariance, so Eq. (11) upper-bounds the true mutual information. It is cheap, since

$3M_a$ is at most a few tens, and it reduces to the D-optimal criterion when the model is linear in θ [8]. Following the heteroscedastic treatment of Kirschner and Krause [16], the noise covariance grows with commanded speed, $R_a = \text{diag}(\rho_0^2 + \rho_1^2 \|W \mathbf{u}_m^a\|^2)$ per window, reflecting larger gait-induced fluctuations at higher speeds.

The estimate is saturation-aware because $\mathbf{h}$ is computed through σ_θ in Eq. (2). Particles that differ only in gain produce identical predictions once the command is saturated, so a probe that drives deep into saturation contributes little about the gains, while a probe that straddles the boundary is informative about $\bar{v}$ or $\bar{\omega}$. The ablation without saturation awareness scores probes with a linear surrogate in which σ_θ is replaced by the identity, while the posterior update still uses the full model. The surrogate's predicted spread grows in proportion to amplitude, so it systematically over-rewards large probes.

4.3 Information per unit of commanded motion

The next probe is selected as

$$a_{k+1} = \arg \max_{a \in \mathcal{A}_{\text{safe}}(k)} \frac{\hat{I}(a)}{c(a) + c_0}, \quad (12)$$

where $\mathcal{A}_{\text{safe}}(k)$ is the safe candidate set defined below and $c_0 = 0.1$ m is a small regularizer that prevents near-zero-motion probes, whose fixed time overhead is not reflected in Eq. (6), from dominating. Classical IDS [10] minimizes $\Delta(a)^2/I(a)$, where Δ is the expected regret. Substituting cost for regret would give $c(a)^2/\hat{I}(a)$, which penalizes long probes quadratically. The first-power ratio in Eq. (12) is preferred because it measures information per meter and is invariant to splitting a probe into two identical halves when their information is additive, so the criterion does not favor short probes for their own sake. The ablation without per-unit-cost weighting maximizes $\hat{I}(a)$ alone.

4.4 Safety filtering and execution

The candidate set combines the 26 directions of $\mathcal{D}$, four amplitudes $\alpha \in \{0.25, 0.5, 0.75, 1.0\}$, and two hold durations $T_h \in \{0.6, 1.2\}$ s, giving 208 candidates. A candidate belongs to $\mathcal{A}_{\text{safe}}(k)$ if its predicted footprint, simulated under the 95th-percentile posterior gains, stays inside the free workspace polygon shrunk by a margin of 0.5 m. If no candidate is admissible, a recentering segment drives the robot toward the workspace center under a proportional controller; its data are used in the posterior update and its path length counts toward the motion budget. The vendor's own safety layer remains active throughout. If it interrupts a probe, the probe is truncated and flagged, and the flagged segment is excluded from the update because the commanded profile no longer describes the input that was actually applied.

Probe commands are streamed at 100 Hz. Between probes the robot holds a zero command for $T_s = 0.3$ s, which restores the near-zero initial twist assumed by the predictor and gives time for the posterior update and the next selection. Each selection requires $N \times 208$ rollouts of the three-state model in Eq. (3), which are vectorized across particles and candidates.

4.5 Procedure

Each calibration session proceeds as follows.

1. Draw N particles from the prior in Eq. (7) and set uniform weights.

2. Build $\mathcal{A}_{\text{safe}}(k)$ from the current pose and the posterior; if it is empty, execute a recentering segment and go to step 5.
3. For every admissible candidate, roll out the particles, compute C_a and $\hat{I}(a)$ with Eqs. (9)–(11), and select a_{k+1} with Eq. (12).
4. Execute a_{k+1} at 100 Hz, logging the measured twist; flag the probe if the safety layer intervenes.
5. Unless the segment is flagged, update the weights with Eq. (8) and resample if the effective sample size falls below $N/2$.
6. Hold zero command for T_s and return to step 2 until the session ends.

Table 1 lists the settings used in all experiments.

05 Experiments

5.1 Platform and protocol

The testbed is a Unitree B2 quadruped operating on a lab floor and a stair course. The main evaluation consists of 200 calibration sessions of 240 s each in simulation, using Isaac Lab with the B2 asset. In each simulated session the interface parameters are drawn from a six-dimensional prior over the parameters in Eq. (4), so every session presents a different unknown interface. Real-robot validation comprises five sessions on the physical B2. In all sessions probe commands are executed at 100 Hz, and the motion budget is measured as commanded path length, Eq. (6). The worst-direction error, median-direction error, and motion budget are recorded at the navigation-grade time T_{nav} , so all methods are compared at the same accuracy criterion; sessions continue to the 240 s limit so that convergence beyond T_{nav} can be observed.

To isolate the effect of probe selection, all methods share the interface model, the particle estimator, the reachability filter, and the settling hold, and differ only in how the next probe is chosen. For GAUGE this means that its planner-conditioned probe schedule is paired with the estimator of Section 4 rather than with its own model-fitting stage.

5.2 Baselines

Uniform sweep excitation cycles through the 26 directions and the amplitude levels of the candidate set, skipping candidates that fail the reachability filter. **GAUGE planner-conditioned probes** [7] draw probe commands from the command distribution of the navigation planner on the same lab floor and stair course, so probing follows the commands the planner actually issues. **Random probes with motion budget** samples candidates uniformly from the safe set, subject to the same motion-budget accounting as the other methods. **Ours** applies the selection rule of Eq. (12).

5.3 Main comparison

Table 2 reports the main comparison. Information-directed probing attains the lowest worst-direction error, 8.9%, against 9.5% for uniform sweep, 10.1% for GAUGE, and 10.3% for random probing; these are reductions of 0.6, 1.2, and 1.4 percentage points, or about 6%, 12%, and 14% in relative terms. It reaches navigation-grade accuracy in 38 s, compared with 40 s for uniform sweep and random probing and 42 s for GAUGE, and it uses 11.4 m of commanded motion, compared with 11.8 m, 13.0 m, and 12.3 m. Its median-direction error of 3.2%

is also the lowest, although the spread across methods on this metric is narrow (3.2–3.6%).

The pattern across baselines is informative. GAUGE has the best median-direction error among the baselines (3.4%) but the second-highest worst-direction error (10.1%), and it uses the most motion (13.0 m) and time (42 s) to reach navigation-grade accuracy. This is consistent with planner conditioning: probes concentrate on the commands the planner issues most often, which calibrates the typical direction well while leaving the least-used directions to be resolved last. Its ratio of worst-direction to median-direction error, about 2.97, is the largest of the four methods. Uniform sweep has the smallest ratio, about 2.71, because it excites every direction equally, but it pays for that uniformity in motion spent on directions that are already resolved. The proposed method’s ratio is about 2.78, and both of its error figures are lower than those of uniform sweep.

The motion budgets and times also imply nearly the same average commanded rate for all methods: 11.4 m in 38 s, 11.8 m in 40 s, 13.0 m in 42 s, and 12.3 m in 40 s each correspond to roughly 0.30 m/s. The methods therefore do not differ in how fast they move, but in where they spend that motion. Information-directed selection reaches the accuracy criterion sooner because each meter of probing removes more posterior uncertainty, not because it probes more aggressively.

The differences in Table 2 are modest in absolute terms: at most 1.4 percentage points in worst-direction error, 4 s in time, and 1.6 m in motion. The tabulated record contains aggregate values without per-session dispersion, so no claim of statistical significance is made, and the comparison should be read as a consistent ordering across four metrics rather than as a large effect.

5.4 Ablation

Table 3 removes one component at a time. Without per-unit-cost weighting, the worst-direction error rises from 8.9% to 10.1%; without the saturation-aware information gain it rises to 10.3%; and without posterior shrinkage it rises to 9.9%. The saturation-aware gain has the largest effect (1.4 percentage points), followed by cost weighting (1.2) and shrinkage (1.0).

Each ablated variant has a higher worst-direction error than uniform sweep excitation (9.5%). Information-directed selection is therefore not beneficial by default: a misspecified information estimate or an unweighted objective concentrates probes in ways that are worse than spreading them evenly. A plausible explanation is as follows. Without cost weighting, the criterion favors long, high-amplitude probes, which carry the most information per probe but the least per meter, so fewer probes fit before the accuracy criterion is met and some directions remain unvisited. Without saturation awareness, the linear surrogate over-rewards probes deep in saturation, which reveal little about the gains, so worst-direction error is highest of all. Without shrinkage, each gain must be identified from its own axis alone, and the least-excited directions remain uncertain for longer.

5.5 Failure cases

Degenerate prior. Two simulated sessions were initialized with a near-degenerate prior whose gain was misplaced by a factor of three. These sessions took 1.7 times longer than usual to reach navigation-grade accuracy, but the worst-direction error still converged. A likely cause is that nearly all particles initially lie far from the true gains, so the first probes are spent

Table 1: Method and interface-model settings used in all experiments.

Setting	Value	Role
Control and probe period Δt	10 ms (100 Hz)	Command streaming and model integration step
Command box $(\bar{u}_x, \bar{u}_y, \bar{u}_\omega)$	1.8 m/s, 0.8 m/s, 1.5 rad/s	Admissible probe commands; exceeds the prior saturation range so the limits are observable
Characteristic radius ℓ	0.6 m	Converts yaw rate to tangential speed in Eqs. (5) and (6)
Particles N	512	Posterior representation
Measurement window T_w	0.25 s	Averages out stride-frequency oscillation
Candidate set	26 directions $\times$ 4 amplitudes $\times$ 2 holds	208 probes; ramp $T_r = 0.2$ s, holds 0.6 s and 1.2 s
Cost regularizer c_0	0.1 m	Prevents near-zero-motion probes from dominating Eq. (12)
Workspace margin	0.5 m	Shrinks the free-space polygon for the reachability filter
Settling hold T_s	0.3 s	Restores the zero-twist initial condition between probes

Table 2: Main comparison. Errors and motion budget are recorded at the navigation-grade time; lower is better in every column.

Method	Worst-direction calibration error (%)	Calibration time to nav-grade (s)	Motion budget (m)	Median-direction error (%)
Uniform sweep excitation	9.5	40	11.8	3.5
GAUGE planner-conditioned probes [7]	10.1	42	13.0	3.4
Random probes with motion budget	10.3	40	12.3	3.6
Ours	8.9	38	11.4	3.2

Table 3: Ablation of the information-directed selection rule.

Variant	Worst-direction calibration error (%)
Ours (information-directed)	8.9
w/o per-unit-cost weighting	10.1
w/o saturation-aware gain	10.3
w/o posterior shrinkage	9.9

moving the particle cloud and the resampling jitter must carry the posterior to the correct region; the information estimate is computed from a posterior that is confidently wrong, and it underestimates the value of probes that would quickly expose the mismatch. A defensive mixture component in the prior or tempering of early likelihoods would address this.

Saturation boundary on hardware. On the real B2, one probe near the saturation boundary triggered the gait manager’s safety brake. The safety layer recovered, and the probe was truncated and flagged. The event exposes a structural tension in the method: the saturation-aware criterion is drawn toward exactly the commands that are most informative about $\bar{v}$ and $\bar{\omega}$, which lie close to the region where the vendor’s protective logic engages. Excluding flagged segments keeps the posterior consistent, but the safe set should also account for the vendor’s limits, for instance through a chance constraint on the predicted command magnitude relative to documented hard limits.

Wasted motion under random probing. To reach the same median-direction error as the proposed method, random probing required 2.4 times the motion budget, with the excess concentrated in directions the planner already uses well. This ratio refers to a matched median-error comparison and is distinct from the navigation-grade operating point of Table 2, at which random probing had a higher median error (3.6% against 3.2%) and used 12.3 m against 11.4 m. The observation illustrates the central argument: undirected probing keeps revisiting directions whose response is already well determined.

06 Limitations

The improvements are consistent but small, and the compiled record contains aggregate values without per-session disper-

sion; a larger study with confidence intervals is needed to establish the size of the effect. Real-robot validation covers five sessions, and the tables do not break down simulated and physical results separately.

The interface model is deliberately simple. It has diagonal gains and therefore cannot represent cross-coupling such as yaw drift induced by lateral commands; it uses a fixed lag time constant; and it assumes that the parameters do not depend on terrain, although the stair course and the lab floor may in practice present different interfaces. Extending θ with coupling terms and terrain-conditioned parameters is straightforward in principle but increases the dimension that the particle posterior must cover.

Selection is greedy over one probe. Non-myopic design, for instance amortized as in deep adaptive design [14], may improve worst-direction error further. The moment-matched estimate in Eq. (11) is an upper bound and can overvalue probes whose predictive distribution is multimodal. The navigation-grade criterion is a single scalar tolerance on direction-averaged error, and the results summarize each session at that operating point rather than over the full curve of calibration error against probe budget. Finally, the method ignores planner relevance entirely. A natural combination with GAUGE [7] would weight the information objective by the planner’s command distribution while retaining a floor on worst-direction uncertainty.

07 Conclusion

This paper formulated calibration of an opaque legged-robot velocity interface as information-directed sampling over a six-parameter state-delay model and selected probes by saturation-aware expected information gain per unit of commanded path length. On a Unitree B2 testbed with 200 simulated sessions and five real-robot sessions, the method reached navigation-grade accuracy in 38 s with 11.4 m of commanded motion and a worst-direction error of 8.9%, ahead of uniform sweep excitation, random probing, and GAUGE’s planner-conditioned probes on all reported metrics. The ablation shows that the benefit depends on all three components: without cost weighting, saturation awareness, or gain shrinkage, information-directed probing performs worse than uniform excitation. The failure cases point to the next steps: robustness to a misplaced prior

and explicit modeling of the vendor's safety limits near the saturation boundary.

References

- [1] D. Fox, W. Burgard, and S. Thrun. The Dynamic Window Approach to Collision Avoidance. *IEEE Robotics and Automation Magazine*, 4(1):23–33, 1997.
- [2] G. Williams, A. Aldrich, and E. Theodorou. Model Predictive Path Integral Control using Covariance Variable Importance Sampling. *arXiv:1509.01149*, 2015.
- [3] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter. Advanced Skills by Learning Locomotion and Local Navigation End-to-End. *arXiv:2209.12827*, 2022.
- [4] 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. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [5] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter. Learning Agile and Dynamic Motor Skills for Legged Robots. *arXiv:1901.08652*, 2019.
- [6] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning Robust Perceptive Locomotion for Quadrupedal Robots in the Wild. *arXiv:2201.08117*, 2022.
- [7] T. Zang, H. Huang, S. Cheng, and W. Zhang. GAUGE: Planner-Conditioned Active Calibration of Opaque Quadruped Velocity Interfaces. *arXiv:2607.02841*, 2026.
- [8] K. Chaloner and I. Verdine. Bayesian Experimental Design: A Review. *Statistical Science*, 10(3):273–304, 1995.
- [9] T. Rainforth, A. Foster, D. R. Ivanova, and F. Bickford Smith. Modern Bayesian Experimental Design. *arXiv:2302.14545*, 2023.
- [10] D. Russo and B. Van Roy. Learning to Optimize via Information-Directed Sampling. *arXiv:1403.5556*, 2014.
- [11] J. Borenstein and L. Feng. Measurement and Correction of Systematic Odometry Errors in Mobile Robots. *IEEE Transactions on Robotics and Automation*, 12(6):869–880, 1996.
- [12] A. Censi, A. Franchi, L. Marchionni, and G. Oriolo. Simultaneous Calibration of Odometry and Sensor Parameters for Mobile Robots. *IEEE Transactions on Robotics*, 29(2):475–492, 2013.
- [13] A. Foster, M. Jankowiak, E. Bingham, P. Horsfall, Y. W. Teh, T. Rainforth, and N. Goodman. Variational Bayesian Optimal Experimental Design. *arXiv:1903.05480*, 2019.
- [14] A. Foster, D. R. Ivanova, I. Malik, and T. Rainforth. Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design. *arXiv:2103.02438*, 2021.
- [15] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel. Bayesian Active Learning for Classification and Preference Learning. *arXiv:1112.5745*, 2011.
- [16] J. Kirschner and A. Krause. Information Directed Sampling and Bandits with Heteroscedastic Noise. *arXiv:1801.09667*, 2018.
- [17] J. Kirschner, T. Lattimore, and A. Krause. Information Directed Sampling for Linear Partial Monitoring. *arXiv:2002.11182*, 2020.
- [18] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. *arXiv:0912.3995*, 2009.
- [19] R. K. Mehra. Optimal Input Signals for Parameter Estimation in Dynamic Systems: Survey and New Results. *IEEE Transactions on Automatic Control*, 19(6):753–768, 1974.
- [20] A. Wagenmaker and K. Jamieson. Active Learning for Identification of Linear Dynamical Systems. *arXiv:2002.00495*, 2020.
- [21] H. Mania, M. I. Jordan, and B. Recht. Active Learning for Non-linear System Identification with Guarantees. *arXiv:2006.10277*, 2020.
- [22] M. Buisson-Fenet, F. Solowjow, and S. Trimpe. Actively Learning Gaussian Process Dynamics. *arXiv:1911.09946*, 2019.
- [23] P. Shyam, W. Jaśkowski, and F. Gomez. Model-Based Active Exploration. *arXiv:1810.12162*, 2018.
- [24] D. Pathak, D. Gandhi, and A. Gupta. Self-Supervised Exploration via Disagreement. *arXiv:1906.04161*, 2019.
- [25] B. Sukhija, L. Treven, C. Sankar, S. Blaes, S. Coros, and A. Krause. Optimistic Active Exploration of Dynamical Systems. *arXiv:2306.12371*, 2023.
- [26] M. Memmel, A. Wagenmaker, C. Zhu, P. Yin, D. Fox, and A. Gupta. ASID: Active Exploration for System Identification in Robotic Manipulation. *arXiv:2404.12308*, 2024.
- [27] T. Lew, A. Sharma, J. Harrison, A. Bylard, and M. Pavone. Safe Active Dynamics Learning and Control: A Sequential Exploration-Exploitation Framework. *arXiv:2008.11700*, 2020.
- [28] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning Quadrupedal Locomotion over Challenging Terrain. *arXiv:2010.11251*, 2020.
- [29] N. Rudin, D. Hoeller, P. Reist, and M. Hutter. Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning. *arXiv:2109.11978*, 2021.
- [30] V. Makoviychuk et al.. Isaac Gym: High Performance GPU-Based Physics Simulation for Robot Learning. *arXiv:2108.10470*, 2021.
- [31] M. Mittal et al.. Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments. *arXiv:2301.04195*, 2023.
- [32] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *arXiv:1703.06907*, 2017.
- [33] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. *arXiv:1710.06537*, 2017.
- [34] W. Yu, J. Tan, C. K. Liu, and G. Turk. Preparing for the Unknown: Learning a Universal Policy with Online System Identification. *arXiv:1702.02453*, 2017.
- [35] A. Kumar, Z. Fu, D. Pathak, and J. Malik. RMA: Rapid Motor Adaptation for Legged Robots. *arXiv:2107.04034*, 2021.
- [36] F. Ramos, R. C. Possas, and D. Fox. BayesSim: Adaptive Domain Randomization via Probabilistic Inference for Robotics Simulators. *arXiv:1906.01728*, 2019.
- [37] Y. Chebotar, A. Handa, V. Makoviychuk, M. Macklin, J. Isaac, N. Ratliff, and D. Fox. Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience. *arXiv:1810.05687*, 2018.
- [38] G. Kahn, P. Abbeel, and S. Levine. BADGR: An Autonomous Self-Supervised Learning-Based Navigation System. *arXiv:2002.05700*, 2020.
- [39] J. Frey, M. Mattamala, N. Chebrolu, C. Cadena, M. Fallon, and M. Hutter. Fast Traversability Estimation for Wild Visual Navigation. *arXiv:2305.08510*, 2023.
- [40] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada. Control Barrier Functions: Theory and Applications. *arXiv:1903.11199*, 2019.
- [41] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

- [42] E. A. Wan and R. van der Merwe. The Unscented Kalman Filter for Nonlinear Estimation. Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium, 153–158, 2000.