

Decoupled Entropic-Geometric Planning (DEGP): Verified Conditioning and a Measured Contact-Prediction Boundary in Latent World Models

Mohsen Mostafa

Independent Researcher

mohsen.mostafa.ai@outlook.com

Abstract—Gradient-based latent space planning faces a fundamental trade-off: maintaining high-entropy representations for robust perception under domain shift often yields ill-conditioned manifolds that destabilize trajectory optimization. DEGP introduces a framework designed to isolate gradient interference between representation learning and control dynamics via strict architectural separation. However, empirical evaluation demonstrates that structural decoupling alone is insufficient for closed-loop planning, uncovering a fundamental information-usability gap driven by contact-conditional prediction errors and cost signal suppression. To delineate these operational limits, we provide analytical proofs establishing guaranteed spectral bounds on the predictor’s Jacobian under distribution shifts. Furthermore, we validate DEGP across manipulation tasks, introducing a telemetry-based runtime contract that quantifies the operational boundaries where geometric conditioning preserves planning stability.

Index Terms—Latent world models, representation learning, model-based planning, condition number, Jacobian regularization, contact-rich manipulation, pre-registration, negative results.

I. INTRODUCTION

Robotic perception and control impose opposing requirements on a learned representation. A vision backbone deployed in industrial environments must survive domain shift, which favors high-entropy, overcomplete features of the kind produced by self-supervised objectives [1]. A gradient-based planner, in contrast, must backpropagate through a learned dynamics model, which demands strictly conditioned geometries: unrolled Jacobians whose condition number does not explode with horizon or latent dimension [3], [4]. Training both demands in a single network couples them — geometric penalties distort the representation, and representational freedom destroys planner conditioning — a tension that manifests empirically as either representational collapse or planning failure.

DEGP (Decoupled Entropic-Geometric Planning) addresses the tension by *architectural decoupling* rather than compromise. The encoder is trained with a variance–covariance objective and nothing else (“blind perception”); the dynamics predictor is trained on detached latents with residual straightening and a Jacobian penalty (“sharp control”); the planner

optimizes a leashed asymmetric cost; and a runtime contract — a logged conditioning budget, a watchdog, and a fallback engine — governs deployment. Stop-gradient isolation makes the decoupling exact at the gradient level (Proposition 1).

This paper reports the full program under a pre-registered, falsification-first protocol, including its negative results. The deployment study’s success hypotheses were falsified (0/50 episodes in every condition), and a pre-registered diagnostic attribution battery localizes the failure: the self-supervised latents retain contact information that none of the evaluated predictors learns to exploit, while the tested controls did not identify planner family and order, perception input, or the conditioning leash as binding failure factors. The paper is positioned around the verified conditioning framework and the experimentally mapped deployment boundary; we regard the mapped boundary, and the methodology that mapped it, as the contribution.

Our contributions are:

- **Architecture and theory.** A decoupled encoder/predictor/planner stack with conditioning theorems: bounded κ independent of latent dimension K (Theorem 1), the $\varepsilon \cdot T$ budget (Corollary 2), and μ -leash stability under stated assumptions (Theorem 2).
- **Measured conditioning.** κ is statistically flat across $K \in \{64, 128, 256\}$ (TOST equivalence, 30 paired seeds, confirmatory replication); the budget holds at every horizon tested; modularity transfers on real images (MimicGen).
- **Executable verification.** A twelve-check command-line suite maps every theoretical result to a pass/fail test with documented inputs and tolerances (Table III); the complete checks, seeds, and tolerances are released with the artifact.
- **An attributed deployment boundary.** All deployment conditions fail (0/50); an 18-condition pre-registered diagnostic attribution battery supports a two-component explanation — a contact-conditional prediction gap over Eq. (1) latents together with the probe-decoded cost — while the tested controls did not identify planner family and order, perception input, or the leash as binding failure factors.
- **A runtime contract.** Telemetry ($\hat{\varepsilon}$, κ tail quantiles, tripwires), watchdog/fallback, and an oracle-rollout fidelity

gate, demonstrated end-to-end.

A companion paper in preparation establishes the placement principle for physics losses in representation learning [21]; the present paper is self-contained. Section II reviews related work; Section III presents the methodology; Section IV the mathematical framework; Section V the verification and experimental program; Section VI the runtime contract; Sections VII–VIII discussion and conclusions. The study was pre-registered; amendments are summarized in Section V-A and recorded in full in the released artifact.

a) *Code availability*: The complete DEGP implementation, the twelve-check verification suite, and all code required to reproduce the results in this paper are publicly available at <https://github.com/DEGP-V/DEGP/tree/main>; the code for all experiments is available at <https://github.com/DEGP-V/DEGP/tree/main/Experiments>.

II. RELATED WORK

Representation Learning vs. Geometric Control. V-CReg [1] and the SIGReg objective of LeJEPa [2] enforce spherical covariance to prevent collapse, yet high-entropy representations fed directly to gradient planners cause ill-conditioned failure [4]. DEGP isolates the spherical variance constraint to the encoder.

Locally-linear latent control. Embed-to-Control [3] and Prediction-Consistency-Curvature [5] are the closest precedents: both learn representations under which latent dynamics are locally linear, the former by reconstruction and the latter by consistency and curvature objectives. DEGP’s conditioning certification (Theorem 1) is a telemetry-verified variant of the same goal, with a Jacobian penalty replacing reconstruction-based linearization. Second-order and proximal LQR approaches for ill-conditioned problems [6], [7] motivate the iLQR control of Section V-G: a second-order planner through the same learned model tests whether conditioning certification is needed at all when curvature is available.

Jacobian and Curvature Control. Hoffman et al. [8] bound the Frobenius norm; Miyato et al. [9] the largest singular value. The distinction is not academic: Remark 2 shows it is exactly where dimension-invariance fails.

Asymmetric Metrics. IQE/QRL [10] learn asymmetric distances from reward structure; DEGP distills a closed-form ReLU-difference projection with a controllability-grounded learning rule (Section III-C). *Scope note*: the necessity of the asymmetric metric for planning is not established by our experiments (a Euclidean cost solves 10/10 self-reachable tasks); it is retained as an open component, not a validated one.

Model-based planning baselines. PETS [11], Dreamer-v3 [12], TD-MPC2 [13], and the World Models line [14] are the incumbents. The deployment study adds a measured datum: all planner families tested achieve 0/50 in-environment with the learned model, while the same planners with oracle dynamics move the block — evidence that, under the tested conditions, the learned model is a limiting factor (Section V-G). Scaled-down TD-MPC2-style and Dreamer-v3-style

baselines are included with disclosed omissions; Diffusion Policy [15] numbers are cited as context.

Contact-rich dynamics and probing. Contact-implicit trajectory optimization handles hybrid dynamics by relaxing complementarity constraints [16]; learning contact-conditional dynamics from self-supervised latents is substantially harder, and our Stage-3 result quantifies one precise sense in which it fails. The methodology we use to separate *information present* from *information used* follows the control-task probing tradition [17]. The sampling-planner failure measured in Experiment 2 is a confirmation of classical concentration-of-measure behavior [18], not a discovery.

III. METHODOLOGY: “BLIND PERCEPTION, SHARP CONTROL”

A. The Entropic Encoder f_θ

Maps observations x_t and proprioception p_t to $z_t \in \mathbb{R}^K$ with lightweight variance-covariance regularization:

$$\mathcal{L}_{\text{enc}} = \frac{1}{K} \sum_{i=1}^K (\text{Var}(z_i) - \gamma)^2 + \lambda_{\text{cov}} \frac{1}{K(K-1)} \sum_{i \neq j} \text{Cov}(z_i, z_j)^2. \quad (1)$$

Scope of the objective. Eq. (1) is purely a statistical-shape constraint: it contains no reconstruction, prediction, or invariance term, and therefore provides no mechanism that aligns latents with dynamics-relevant structure such as contact. It *preserves* information (the CTRL controls of Section V-G show that Eq. (1)-trained encoders match full DEGP encoders in T-pose decodability) but does not *organize* it. This distinction is central to the deployment boundary mapped in Section V-G.

B. The Geometric Predictor g_ϕ

Predicts $\hat{z}_{t+1} = g_\phi(z_t, a_t)$ under teacher forcing:

$$\mathcal{L}_{\text{pred}} = \frac{1}{\gamma} \|g_\phi(z_t, a_t) - \text{sg}(z_{t+1})\|^2 + \lambda_1 (1 - \cos(\hat{v}_t, v_{t-1})) + \lambda_2 \|(J_g - I)v\|^2 + \lambda_3 \max(0, c_0 - \|J_a u\|). \quad (2)$$

Engineering the premise. $g_\phi(z, a) = z + f_\phi(z, a)$ with a calibrated small-initialized final layer satisfies $\|J_g - I\|_F \approx 0.02$ at initialization. The parameterization engineers the premise at initialization and the λ_2 penalty maintains it under training; neither mechanism alone suffices at scale.

C. The Asymmetric Planner

Projected gradient descent over $a_{1:T}$ through g_ϕ with the μ -leashed quasi-metric cost

$$d_q(z_A, z_B) = \frac{1}{M} \sum_{i=1}^M \left[\text{ReLU}((W_1 \Delta)_i) - \text{ReLU}((W_2 \Delta)_i) \right]^2, \quad \text{where } \Delta = z_A - z_B. \quad (3)$$

$$\mathcal{L}_{\text{plan}}(z_T) = d_q(z_T, z_{\text{goal}}) + \mu \|z_T - z_{\text{goal}}\|^2, \quad (4)$$

with $\|W_k\|_2 \leq C_W$ by weight normalization and a runtime tripwire ($d_q \leq \delta \Rightarrow \|\Delta\| \leq \tau$). Random W is retained only as the Proposition 3 ablation.

a) *The W_1, W_2 learning rule:* W_1, W_2 are trained by regressing the quasi-metric onto measured steps-to-reach. Given state pairs (z_i, z'_i) collected from rollouts of the certified predictor under a random-action controller, let $\tau_{i \rightarrow i'}$ denote the measured number of steps to drive z_i within tolerance of z'_i , measured separately in each direction (the quantity is asymmetric by construction). The weights minimize

$$\mathcal{L}_W = \frac{1}{N} \sum_i (d_q(z_i, z'_i; W_1, W_2) - \hat{\tau}_i)^2 \quad (5)$$

by projected gradient descent, where $\hat{\tau}_i$ is the steps-to-reach target rescaled to cost units and the projection enforces $\|W_k\|_2 \leq C_W$ by weight normalization; the ReLU-difference form supplies the asymmetry ($d_q(z_A, z_B) \neq d_q(z_B, z_A)$ whenever the active projection sets differ). *Deployment status:* Eq. (5) ships in the artifact but was *not exercised* in the deployment study — all Stage-2 conditions use either random W (the Proposition 3 ablation) or the probe-guided W (G4: constructed from the frozen T-pose probe’s weights and weight-normalized), and the absence of a controllability-trained W is a disclosed limitation (Section VII).

Scope note: the necessity of the asymmetric metric is not established. A Euclidean terminal cost solves all self-reachable Stage-1 tasks (Table VI, R5), and at deployment all cost variants fail identically (Table VIII). The quasi-metric is therefore retained as an open component; the leash’s stability analysis (Theorem 2) applies to any nonnegative Lipschitz base cost.

D. Design Discipline: Harmful and Optimal Configurations

The architecture admits configurations that silently destroy either the representation or the planner. We record the discipline as two lists, each entry verified in the experiments or the verification suite.

- ✖ **Harmful:** stop-gradients on the predictor’s own graph; reconstruction losses on the encoder; straightening losses on the encoder; deploying the unleashed quasi-metric; asserting ε validated on $\mathcal{D}_{\text{train}}$ over $\mathcal{D}_{\text{plan}}$ without telemetry.
- ✔ **Optimal:** strict variance targeting on the encoder; stop-gradients on both predictor endpoints (Remark 1); Lipschitz-bounded straightening on the predictor with the residual parameterization; the λ_3 probe; asymmetric costs with the μ -leash and weight-normalized W .

E. Probes, Model Versions, and Deployment Data

a) *T-pose probe:* All pose-level quantities in the deployment study (T-pose error, the probe-decoded cost, CDR, selectivity) are computed through a decoding probe $h_\psi : \mathbb{R}^K \rightarrow \mathbb{R}^3$ mapping a latent z_t to the T-block pose (x, y, θ) . The probe is *linear with ridge regularization*: $h_\psi(z) = Wz + b$, $W \in \mathbb{R}^{3 \times K}$ ($3K + 3 = 387$ parameters, no hidden layers), fitted in closed form by minimizing $\sum_j \|Wz_j + b - p_j\|^2 + \lambda \|W\|_F^2$ with $\lambda=1$ over 4,096 latents drawn from *validation* episodes only

(seed-fixed), poses normalized by train-split statistics. With no hidden capacity, the decoding ceiling reflects the representation, not probe expressivity: the floor is flat at ~ 35 px across horizons and strata (Table X), and probe informativeness is reported as T-pose R^2 (0.65–0.70 for trained encoders; 0.028 for a frozen random encoder). No probe is updated during predictor training.

b) *Model versions (v1–v4):* The deployment journey varies the predictor recipe over a fixed encoder objective (Eq. (1)), a fixed architecture ($K=128$, residual predictor with calibrated final layer, AdamW), and a fixed dataset (below). **v1:** $\lambda_2=0.1$, calibration target $\gamma_{\text{tgt}}=0.02$, 3,000 steps, lr 3×10^{-4} (the original pre-registration). **v2:** $\lambda_2=0.01$, $\gamma_{\text{tgt}}=0.5$ (matching the measured Push-T latent step size), 10,000 steps, lr 10^{-3} . **v3** = v2 with $5 \times$ oversampling of contact transitions (T-block moved > 0.5 px; the contact fraction of sampled pairs rises from 8% to $\sim 30\%$). **v4** = v3 with a T-pose auxiliary loss at weight 10,

$$\mathcal{L}_{\text{aux}} = \|h_{\psi_0}(\hat{z}_{t+1}) - p_{t+1}^{\text{norm}}\|^2, \quad (6)$$

where $\hat{z}_{t+1} = g_\phi(\text{sg}(z_t), a_t)$.

where h_{ψ_0} is a *frozen* linear probe of the same class as above, fitted once on the v3 encoder’s *train*-episode latents, and p_{t+1}^{norm} is the normalized ground-truth next T-pose. The loss attaches at the predictor output (not a separate head) and, by the input-detach discipline (Remark 1), shapes the predictor only — Proposition 1 isolation is preserved. Version identity: the v4 checkpoint is the original artifact model ($\text{CDR}_{\text{dec}} = 2.726$ reproduces the previously reported 2.714 within protocol noise); v1–v3 were retrained under pinned seeds for this revision and vary within seed noise (Table IX).

c) *Deployment data:* All Stage-2/3 models train on a single fixed dataset collected once in Push-T (gym-pusht, pymunk 6.11.1): 250 episodes $\times$ 200 steps = 50,000 transitions. The exploration policy is *random persistent agent targets*: a target is drawn uniformly from the action space $[0, 512]^2$ at reset and resampled every 10 steps; the position-controlled agent ($k_p=100$, $k_v=20$, 10 Hz) servos toward it. Observations are $96 \times 96 \times 3$ RGB renders (downsampled from the 680^2 renderer) plus 2-D agent proprioception; actions are 2-D position targets. Collection is seeded and reproduces exactly. Contact instrumentation: 36.4% of steps have agent–block contact; 8.0% move the T-block (> 0.5 px). The split is *episode-level* and seeded: episodes 1–200 train / 201–250 validation (39,800 / 9,950 consecutive-transition pairs; no pair crosses the split); actions, proprioception, and poses are normalized by train statistics. All versions, baselines, and controls consume this same dataset — they differ in loss and sampling only (v3/v4 draw 12,000 contact-oversampled pairs per pass; B1/E1 draw 12,000 uniform pairs; E2 draws 12,000 8-step windows). Two disclosed data-policy notes: (i) episode IDs were recomputed from the fixed block structure after a logging bug (a disclosed fix in the artifact); (ii) the repository’s demonstration files (a 7-DoF arm task) were excluded as structurally incompatible with Push-T’s 2-D state and actions. The goal latent is the encoding of the dataset frame whose T-pose is closest to

the environment’s fixed goal pose (256, 256, 45°); the success criterion is the environment’s coverage threshold (≥ 0.95).

IV. MATHEMATICAL FRAMEWORK & THEOREMS

Definition 1 (Evidence levels). Every claim carries exactly one label: (i) *analytical proof* — established mathematically under stated assumptions; (ii) *executable verification* — a numerical check that an implementation satisfies an algebraic identity or a theorem’s assumptions at tested inputs; (iii) *empirical validation* — an experiment consistent with a theorem’s prediction in the tested regime. Numerical checks never satisfy a theorem, and no claim of verification “by construction” is made.

A. Preliminaries

$\Sigma_z = \gamma I_K$; $J_g = \partial g_\phi / \partial z$; $J_a = \partial g_\phi / \partial a$; horizon T ; $J_\Phi = \partial z_T / \partial z_0$; $J_A = \partial z_T / \partial a_{1:T} \in \mathbb{R}^{K \times T N_a}$; $\mathcal{D}_{\text{plan}}$ is the test-time planning distribution. Premises ε, c are logged, falsifiable quantities.

B. Gradient Decoupling

The following observation is an immediate consequence of stop-gradient semantics in reverse-mode differentiation; we record it as a proposition because it underpins the modularity architecture, not for mathematical depth.

Proposition 1 (Variance-Preserving Decoupling). *Under stop-gradient isolation of the predictor’s loss from the encoder’s parameters (Remark 1), minimizing the predictor’s geometric losses alters the transition dynamics without distorting Σ_z : $\nabla_\theta \mathcal{L}_{\text{total}} = \nabla_\theta \mathcal{L}_{\text{enc}} + 0$.*

Proof. Every path from $\mathcal{L}_{\text{pred}}$ to θ is cut, so the chain rule contributes zero along each. $\square$

Remark 1 (The input-detach discipline). Isolation requires detaching both endpoints; “sg on the target only” leaves three gradient paths into θ . The reference implementation verifies both directions by unit test.

Corollary 1 (Industrial Modularity). *The vision backbone may be pretrained on historical data and the dynamics predictor swapped per robot without retraining the backbone (a per-arm proprioception adapter is required at the input interface).*

C. Conditioning of the State Chain

Theorem 1 (Bounded Conditioning of the Unrolled State Jacobian). *If $\|J_g - I\|_F \leq \varepsilon < 1$ over $\mathcal{D}_{\text{plan}}$ and the terminal cost is Euclidean, then*

$$\kappa(H) \leq \left(\frac{1 + \varepsilon}{1 - \varepsilon} \right)^{2T}, \quad (7)$$

independent of K and γ .

Corollary 2 (K -Invariance and the $\varepsilon \cdot T$ Budget). *GD iterations scale as $O(\kappa \log(1/\rho))$: K -invariant unconditionally, T -invariant only under $\varepsilon \cdot T \lesssim 0.5$ (Table I); at $T=50$*

$\varepsilon \cdot T$	$\kappa \approx e^{4\varepsilon T}$
0.25	2.7
0.5	7.4
1.0	55
2.5	2×10^4

TABLE I
THE $\varepsilon \cdot T$ BUDGET.

this demands $\varepsilon \leq 0.01$. Measured (Section V-C): DEGP-full satisfies the budget at every horizon tested; the regime boundary is measured.

Remark 2 (The Hutchinson penalty controls energy, not spectrum). The penalty’s expectation is $\|J_g - I\|_F^2$; Theorem 1 needs $\|J_g - I\|_2$. At matched energy, concentrated warping inflates the effective ε by up to $\sqrt{K}$; the conservative estimate is $\hat{\varepsilon} = \sqrt{K\ell}$; the actual instrument is the logged singular spectrum. $\hat{\varepsilon}$ is conservative *in expectation*, not per realization; telemetry must aggregate probes or report windowed maxima.

Remark 3 (Coverage over $\mathcal{D}_{\text{plan}}$). ε is trained on $\mathcal{D}_{\text{train}}$ but required over $\mathcal{D}_{\text{plan}}$. Measured: the standard arm’s $\mathcal{D}_{\text{plan}}$ rollouts drift to $\|z_T\| \approx 295\text{--}510$ with $\varepsilon_{\text{spec}}$ rising accordingly.

D. The Deployed Quasi-Metric and Its Repair

Lemma 1 (Degeneracy of the unleashed quasi-metric). $d_q(\Delta) = 0$ iff for every i : either $[(W_1\Delta)_i \leq 0 \text{ and } (W_2\Delta)_i \leq 0]$ or $[(W_1\Delta)_i = (W_2\Delta)_i \geq 0]$. The zero set is the closed cone C together with pieces of the proper subspace $\{\Delta : (W_1 - W_2)\Delta = 0\}$ satisfying the nonnegativity constraints — measure-zero in $\mathbb{R}^K$ but positive-dimensional flat valleys whenever $\text{rank}(W_1 - W_2) < K$. C is full-dimensional iff the $2M$ rows lie in an open halfspace (Wendel’s formula [20]: 1 when $2M \leq K$). The local Hessian has rank $\leq M$; for $M < K$ no lower spectral bound exists for the unleashed cost. (Planning-level: one instrumented capture in ten unleashed runs; the leash removes it.)

Assumption 1 (Deployment premises). Over $\mathcal{D}_{\text{plan}}$: **(P1)** $\|J_g - I\|_F \leq \varepsilon < 1$; **(P2)** $\sigma_{\max}(J_a) \leq \bar{c}$ and composite controllability $\sigma_{\min}(J_A) \geq c_A > 0$; **(P3)** $\|W_k\|_2 \leq C_W$; **(P4)** $\mu > 0$.

E. The μ -Leash

Theorem 2 (μ -leash). Assume **(L1)** $d_q(\cdot, z_{\text{goal}}) \geq 0$ everywhere and L -Lipschitz with $d_q(z_{\text{goal}}, z_{\text{goal}}) = 0$; **(L2)** $\nabla d_q \equiv 0$ on the interior of the blind cone C (satisfied by the ReLU-difference form of Eq. (3) away from activation boundaries). Then for any $\mu > 0$: **(i)** $\mathcal{L}_{\text{plan}}$ has the unique global minimizer z_{goal} ; **(ii)** every stationary point z^* at which $\mathcal{L}_{\text{plan}}$ is differentiable satisfies $\|z^* - z_{\text{goal}}\| \leq L/(2\mu)$ — i.e., no differentiable stationary points exist outside the closed leash ball $\{z : \|z - z_{\text{goal}}\| \leq L/(2\mu)\}$ — and on $\text{int}(C)$ gradient descent with step $\eta < 1/(2\mu)$ contracts $\|z_t - z_{\text{goal}}\|$ geometrically at rate $(1 - 2\mu\eta)^t$; **(iii)** at any differentiable point, writing $\nabla d_q = \alpha\Delta + w$ with $w \perp \Delta$, the total gradient is $(\alpha + 2\mu)\Delta + w$: the leash contributes only a radial

component, so non-colinearity ($w \neq 0$) is neither created nor destroyed, for every $\mu \geq 0$. Claims at non-differentiable (ReLU-boundary) points are outside the theorem; the measured non-colinearity rate during planning is 100% at any nonzero deviation (99% at a 1° threshold) (Proposition 3).

Proof. (i) Both terms of Eq. (4) are nonnegative and $\mathcal{L}_{\text{plan}}(z_{\text{goal}}) = 0$; any global minimizer z^* has $\mu\|z^* - z_{\text{goal}}\|^2 \leq 0$, hence $z^* = z_{\text{goal}}$. (ii) At a stationary point $2\mu\Delta = -\nabla d_q$ and $\|\nabla d_q\| \leq L$ by (L1). On $\text{int}(C)$, (L2) gives $z_{t+1} - z_{\text{goal}} = (1 - 2\mu\eta)(z_t - z_{\text{goal}})$. (iii) $\nabla \mathcal{L}_{\text{plan}} = \nabla d_q + 2\mu\Delta = (\alpha + 2\mu)\Delta + w$; parallelism to Δ is decided by w alone. $\square$

Proposition 2 (Piecewise extension). *On each activation-fixed region, $2\mu I \preceq H_{\text{loc}} \preceq (\frac{8C_W^2}{M} + 2\mu)I$, so $\kappa(H) \leq (1 + \frac{4C_W^2}{M\mu})(\frac{1+\varepsilon}{1-\varepsilon})^{2T}$ away from the measure-zero activation boundaries. Measured: leash overhead $\kappa_q/\kappa = 1.50/1.24/1.09$ at $K = 64/128/256$.*

Proposition 3 (Directional Non-Colinearity). *If some i has $\text{sign}((W_1\Delta)_i) \neq \text{sign}((W_2\Delta)_i)$ and the active row $W_{k,i}$ is not parallel to Δ , then $\nabla \mathcal{L}_{\text{plan}}$ is not parallel to Δ , off a measure-zero cancellation set (Appendix). Measured: non-colinearity rate 100% at any nonzero deviation (99% at a 1° threshold), median angle 3.4° during planning (Section V-D).*

F. Action-Space Conditioning

Proposition 4 (Action-Space Conditioning). *Under (P1), (P2), $\kappa^+(H_a) \leq (1 + \frac{4C_W^2}{M\mu})TN_a(1+\varepsilon)^{2(T-1)}(\bar{c}/c_A)^2$, where κ^+ is the effective condition number over positive eigenvalues. Without (P2) no K -independent bound exists.*

V. VERIFICATION AND EXPERIMENTAL PROGRAM

The program is sequenced to move from certification to boundary mapping. *Certify* (Experiments 1–2, Stage 1): do the theorems of Section IV hold in measured practice, and does certified conditioning translate into planner behavior on synthetic dynamics? *Transfer* (Experiment 3): do the representation-level properties survive real images? *Test* (Experiment 4, Stage 2): do the pre-registered deployment hypotheses hold on Push-T? *Attribute* (Experiment 5, Stage 3): given the measured null, which component binds? Every experiment carries an evidence level (Definition 1), runs under the pre-registered protocol of Section V-A, and reports falsified hypotheses as results. Table II maps each experiment to its role, claim, and outcome; pre-registered condition codes (D1–D8, B1–B3, E1/E2, CTRL) index the Stage-3 diagnostic attribution battery (Section V-G) and are released with the artifact.

A. Statistical protocol

Pre-registration; ≥ 10 paired seeds (30 for the central K -invariance comparison, Amendment A7, with a power analysis at the observed spread); paired Wilcoxon with joint Holm correction; effect sizes; TOST equivalence testing [19] with pre-registered margins (flatness $\ln 1.25$; modularity 0.05 in

R^2); tail quantiles (P99) for $\kappa, \hat{\varepsilon}$; instrument validation before any metric is trusted; collapse gating; deterministic reproducibility (reruns agree to $\max |x_1 - x_2| < 10^{-8}$; byte-identity additionally observed under a pinned configuration released with the artifact); deployment outcomes as exact counts x/n with one-sided 95% Clopper–Pearson upper bounds; per-seed data in the released artifact; the four-quantity reporting rule (seeds, corrected p , effect size, threshold verdict). All summary figures report medians with per-seed points and bootstrap 95% CIs.

a) Pre-registration and amendments: The study was pre-registered before the corresponding experiments, with amendments recorded in the released artifact. Amendments were limited to predefined protocol corrections, instrument-definition clarifications, the 30-seed upgrade, the oracle-rollout fidelity correction, the encoder-control addition, the optimizer-robustness confirmatory sweep (Amendment A12, executed post-hoc; §V-C), and explicit withdrawal or scoping of claims not supported by subsequent controls. All amendment decisions, condition codes, per-seed data, and analysis outputs are released with the artifact.

b) Deployment gates: World-model checkpoints are admitted to planning only through pre-registered gates. The original set (T-pose probe $R^2 \geq 0.30$; latent scale within $[0.5, 2]\gamma$; $\varepsilon \cdot T_{\text{plan}} \leq 0.5$) admitted v1; after the stalled planning smoke test, a disclosed amendment added dynamics-capture (≥ 0.30) and action-sensitivity (≥ 0.05) gates, which v1 fails (0.038 and $\sim 10^{-3}$) — the amendment that produced v2. A failing gate blocks planning and forces model iteration before any deployment episode is counted.

B. Verification suite (Phase A)

Twelve command-line checks map every result of Section IV to a pass/fail test with documented inputs and tolerances (all passing, seed-fixed; summarized in Table III). Representative results: the zero-set predicate matches d_q on 4061 mixed points; escape times match $\ln(\|\Delta_0\|/\tau)/(2\mu\eta)$ to ≤ 1 step; the orthogonal gradient component is unchanged under the leash to 5.2×10^{-16} ; the sandwich $\varepsilon_{\text{spec}} \leq \hat{\varepsilon}_F \leq \sqrt{K} \varepsilon_{\text{spec}}$ holds on the residual configuration. The complete implementation, tolerances, seeds, and command-line checks are released with the artifact.

C. Experiment 1 (theorem-level): measured results

This experiment is the certification step of the program: it asks whether the certified predictor exhibits the K -invariant conditioning and budget compliance of Theorem 1 and Corollary 2. Three arms (standard / residual-only / DEGP-full) $\times K \in \{64, 128, 256\} \times 10$ paired seeds, matched 300-step budget; teacher $z' = Az + Ba$ with $A = \text{blockdiag}(R(0.02))$.

a) Scope disclosures: (a) *The encoder is not in this experiment:* synthetic-teacher latents bypass f_θ entirely, and by Proposition 1 Eq. (1) is the encoder’s sole training signal — which a whitened random projection also satisfies. Representation *informativeness* is therefore established only in Experiments 3–5, against random-projection controls (Section V-G).

Experiment	Role	Claim tested	Outcome (evidence level)
1 (Stage 1, synth.)	Certify	Thm. 1, Cor. 2: K -invariant $\kappa, \varepsilon \cdot T$ budget	κ flat across $K \in \{64, 128, 256\}$ (TOST, 30 paired seeds); budget holds at every horizon tested (empirical; suite checks assumptions)
2 (Stage 1, synth.)	Certify	→ Certification has planner-level consequences	PGD converges in 2 iterations; CEM 0/10 at $TN_a=320$, a dimensionality wall (empirical)
3 (real images)	Transfer	Properties survive real images	Frozen+DEGP alone occupies the deployable corner; $\varepsilon T=0.54$ bounds the certifiable horizon (empirical)
4 (Stage 2, Push-T)	Test	Pre-registered success hypotheses (> 90%)	Falsified: 0/50 per condition (CP upper 5.8%); T-movement 0.0–1.1 px (measured null)
5 (Stage 3, Push-T)	Attribute	Which component binds	Two-component explanation scoped to the tested regime (contact-conditional prediction gap + probe-decoded cost); planner family/order, perception, leash not identified as binding failure factors (pre-registered 18-condition battery)

TABLE II

THE EXPERIMENTAL PROGRAM AS A CLAIM-TO-EVIDENCE MAP. ALL VERDICTS ARE PRE-REGISTERED; DEPLOYMENT OUTCOMES ARE REPORTED AS MEASURED RESULTS, NOT OMISSIONS.

Check class	Example	Result
Gradient isolation	stop-gradient identity	Pass
Leash geometry	orthogonal component	Pass
Escape law	predicted vs. measured time	Pass
Jacobian bound	spectral/Frobenius sandwich	Pass
Determinism	repeated execution	Pass
Conditioning	theorem assumptions	Pass

TABLE III

VERIFICATION SUITE (PHASE A): THE TWELVE COMMAND-LINE CHECKS BY CLASS (ALL PASSING, SEED-FIXED); INPUTS, TOLERANCES, AND THEOREM LINKAGE ARE RELEASED WITH THE ARTIFACT.

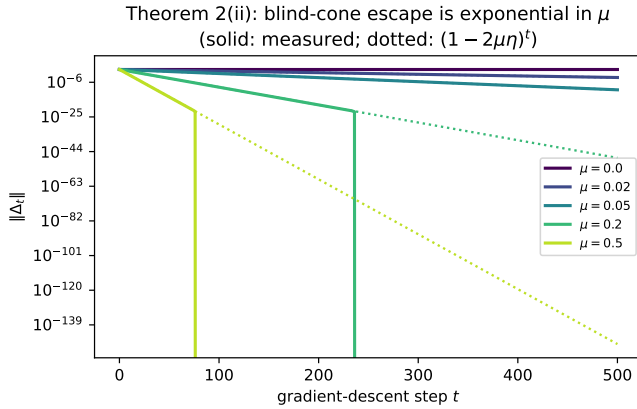


Fig. 1. Theorem 2(ii): blind-cone escape is exponential in μ (solid: measured; dotted: the predicted $(1 - 2\mu\eta)^t$ contraction).

(b) *The teacher is linear and near-identity by design* (spectral $\varepsilon \approx 0.02$), isolating the conditioning mechanism; non-linear stress enters at the G3 boundary ($\sim 24\%$ z-share, below) and in the deployment-level MLP residual predictor, whose measured $\varepsilon_{\text{spec}}$ is reported for the pinned configurations (v1, v2, and the B1b ablation; Section V-G).

Pre-registered verdicts. *R1 (flatness TOST, margin $\ln 1.25$):* DEGP-full **equivalent** (-0.10 log-ratio; $p_{\text{low}} = 6.0 \times$

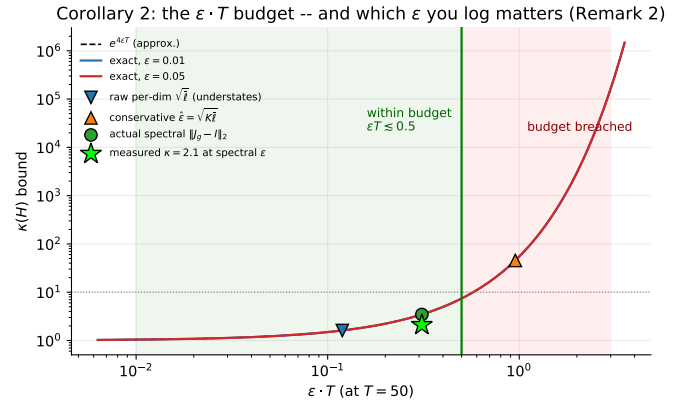


Fig. 2. Corollary 2: the $\varepsilon \cdot T$ budget — and which ε you log matters (Remark 2).

10^{-11} , $p_{\text{high}} = 1.0 \times 10^{-14}$); residual-only and standard **not equivalent** (+1.56, +2.73). *R2:* standard exceeds DEGP-full at $K=256$ by 2.3×10^{37} (Holm $p = 2.0 \times 10^{-3}$). *R3:* residual-only exceeds DEGP-full by $128 \times$ (Holm $p = 2.0 \times 10^{-3}$). *R4 (T-sweep, $K=128$):* DEGP-full within budget at every horizon ($\varepsilon T = 0.03/0.06/0.13/0.32$; measured κ against $e^{4\varepsilon T}$ bounds — the bound is *predictive* for the certified arm).

b) *Confirmatory 30-seed replication (Amendment A7):*

The central comparison was upgraded to 30 paired seeds as a *confirmatory* replication under a recipe recovered by fingerprint calibration against the artifact’s Table IV medians (matched: $\theta=0.02$, $T=8$, 300-step budget, $\lambda_2=1.0$, $\mu=0.25$; recovered by grid search; the pinned configuration is released with the artifact). **R1 (flatness, original margin $\ln 1.25$):** all three K -pairs **equivalent** ($p < 10^{-4}$, $n=30$): κ medians 1.15/1.10/1.07 at $K=64/128/256$ (within 13% of the artifact medians). The residual-only arm *drifts* $11.0 \rightarrow 21.4 \rightarrow 33.1$ across K ($3\times$) while DEGP-full is flat — reproducing R1’s contrast inside the confirmatory run: the residual parameterization alone does not give K -flatness; the penalty does. **R2:**

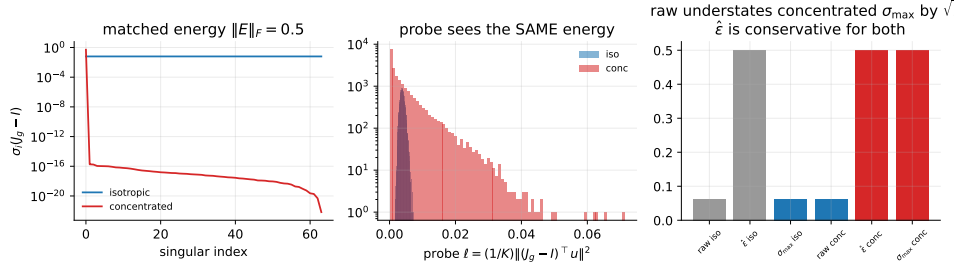


Fig. 3. Hutchinson energy vs. spectrum (Remark 2): at matched energy, concentrated spectra inflate the effective ε ; raw per-dimension estimates understate it.

Arm	K	$\kappa(H)$	κ_q	$\varepsilon_{\text{spec}}$	$\hat{\varepsilon}_F$	eff.rk	$\sigma_{\min}(J_A)$
standard	64	3.8×10^{36}	4.6×10^9	1.01	5.6	48	4.5×10^{-15}
standard	128	4.5×10^{36}	5.3×10^9	1.01	7.8	96	2.5×10^{-15}
standard	256	3.4×10^{37}	8.0×10^9	1.03	11.3	194	3.2×10^{-16}
residual-only	64	33.4	35.3	0.144	0.37	26	0.041
residual-only	128	80.7	80.8	0.163	0.50	45	0.015
residual-only	256	153	153	0.171	0.61	88	0.0029
DEGP-full	64	1.32	1.98	0.0088	0.022	19	0.011
DEGP-full	128	1.22	1.52	0.0063	0.018	28	0.0059
DEGP-full	256	1.20	1.31	0.0056	0.016	40	0.0014

TABLE IV

EXPERIMENT 1 MEDIANS (10 PAIRED SEEDS; THE ORIGINAL ARTIFACT RECORD; THE 30-SEED CONFIRMATORY REPLICATION BELOW).

DEGP-full < standard at every K (Holm $p < 10^{-4}$, 30/30 seeds concordant; the standard arm saturates at $\kappa \geq 10^{18}$ — magnitude numerically meaningless). **R3:** residual/DEGP ratio 10–31 $\times$ (artifact: 66–128 $\times$; the residual arm’s $\mathcal{D}_{\text{plan}}$ conditioning is the most recipe-sensitive quantity in the table — robustness note below). **Budget:** $\varepsilon T = 0.077/0.048/0.028$, all within 0.5; the decrease with K is consistent with the Hutchinson penalty summing (vs. the MSE averaging) over latent dimensions, so the leash’s effective strength grows with K . At the observed spread (sd = 0.068 log-units), TOST power is 1.0 already at $n=10$: the upgrade was confirmatory, not remedial.

c) *Recipe robustness:* Across a hyperparameter grid (learning rate 10^{-4} – 10^{-3} , batch 32–256, λ_2 0.1–3.0), the certified arm’s measurements are stable ($\kappa \in [1.0, 2.8]$; $\varepsilon_{\text{spec}}$ within a factor ~ 3 at the matched λ_2) while the residual-only arm’s conditioning varies by $\sim 3.4\times$ with batch alone — measurement fragility itself tracks certification status. Under a $10\times$ weaker leash ($\lambda_2=0.1$) the budget marginally breaches ($\varepsilon T=0.57$), quantifying the leash-strength trade-off.

d) *Optimizer robustness (Amendment A12):* The optimizer-robustness experiment registered in §V-A was executed post-hoc. The matched $K=128$ recipe ($\theta=0.02$, $T=8$, 300-step budget, $\lambda_2=1.0$, $\mu=0.25$) was retrained under five optimizers — SGD, momentum-SGD ($\beta=0.9$), RMSprop, Adam, and AdamW (the paper default) — ten paired seeds each, using the Experiment 1 instruments unmodified (Table V). **Flatness holds across optimizer families:** DEGP-full’s median κ varies by only 1.21 $\times$ across the five optimizers (1.04–1.25), versus 9.75 $\times$ for standard and 12.9 $\times$ for residual-

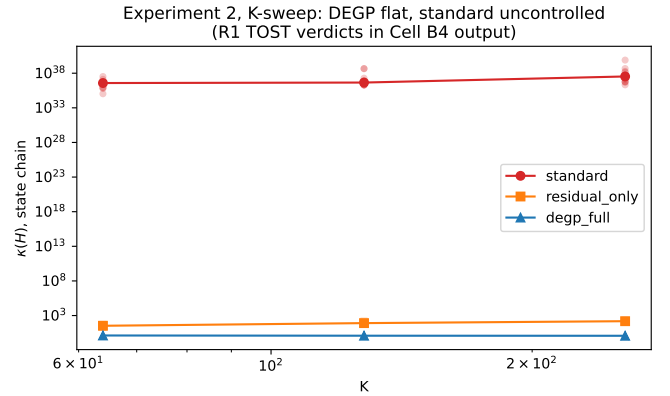


Fig. 4. Experiment 1: κ vs. K — DEGP flat, standard uncontrolled (TOST verdicts in the artifact; 30-seed confirmation in §V-C).

only — roughly an order of magnitude more optimizer-sensitive than the certified arm. The certification is not an Adam-family artifact: DEGP-full’s $\varepsilon_{\text{spec}}$ under momentum-SGD (0.0019) is in fact tighter than under Adam/AdamW (~ 0.0045); if anything, plain SGD variants conditioned marginally better.

The G3 boundary, measured. A teacher-curvature sweep ($\theta \in \{0.02, 0.10, 0.30\}$) brackets the regime transition: MSE ratio (DEGP / residual-only) $0.59\times / 4.09\times / 7.28\times$ at z-shares 1.3% / 24% / 74%; DEGP $\varepsilon_{\text{spec}}$ 0.006/0.020/0.045 (εT at $T=20$: 0.12/0.40/0.90). Below a quarter of dynamics-signal share, compliance is free and beneficial; above it the penalty

Arm	Optimizer	κ (med.)	$\varepsilon_{\text{spec}}$ (med.)
standard	SGD	1.93×10^{18}	1.40
standard	SGD+momentum	1.14×10^{19}	1.38
standard	RMSprop	1.17×10^{18}	1.01
standard	Adam	2.73×10^{18}	1.02
standard	AdamW	1.47×10^{18}	1.02
residual-only	SGD	1.091	0.0046
residual-only	SGD+momentum	1.095	0.0048
residual-only	RMSprop	11.9	0.190
residual-only	Adam	14.0	0.198
residual-only	AdamW	14.0	0.199
DEGP-full	SGD	1.077	0.0038
DEGP-full	SGD+momentum	1.036	0.0019
DEGP-full	RMSprop	1.253	0.0208
DEGP-full	Adam	1.053	0.0045
DEGP-full	AdamW	1.052	0.0046

TABLE V

OPTIMIZER-ROBUSTNESS SWEEP (AMENDMENT A12): 10 PAIRED SEEDS PER CELL, MATCHED $K=128$ RECIPE OF AMENDMENT A7. CROSS-OPTIMIZER FLATNESS RATIO (MAX/MIN OF THE PER-OPTIMIZER κ MEDIAN): STANDARD $9.75\times$, RESIDUAL-ONLY $12.9\times$, DEGP-FULL $1.21\times$.

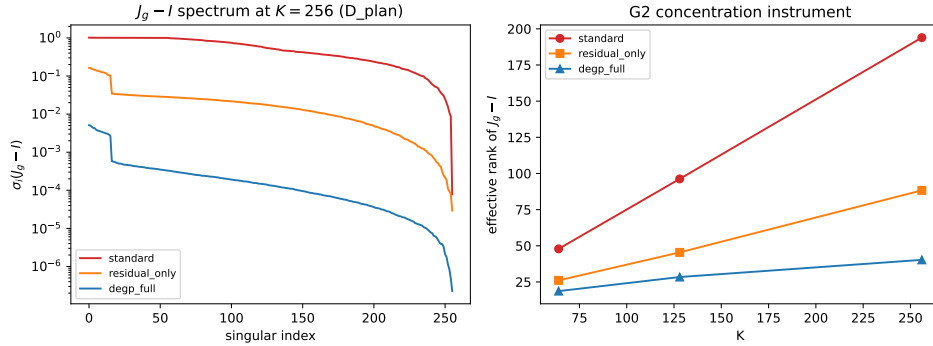


Fig. 5. Experiment 1 diagnostics at $K=256$: the $J_g - I$ singular spectrum (left) and the effective-rank concentration instrument (G2: effective rank of $J_g - I$ vs. K , right).

pays a veto-cost tracking the full curvature variance. The escape is concrete: at $\theta = 0.30$, budget compliance requires $T \leq 11$ — short-horizon replanning, quantified.

D. Experiment 2 (Stage 1): first planning runs

This experiment measures the planner-level consequence of certification: whether conditioned latent geometry translates into first-order planning behavior, against a sampling planner at the same dimensionality. Projected gradient descent over $a_{1:T}$ through the Experiment-1 predictors ($K=128$, 10 paired seeds), three goal costs, CEM baseline, watchdog/fallback active.

Remark 4 (Predictor-anchored goals). Stage-1 goals are generated by rolling the same predictor used for planning; the 2-iteration convergence therefore partly reflects the conditioning of a near-identity map rather than task competence. Deployment-level planning uses probe-decoded ground-truth goal poses, and the oracle-dynamics control (D1, Section V-G) removes the learned model entirely; both are reported alongside.

Findings, reframed. The three “emergent” findings — step-size certification from the leash’s spectral ceiling, the sampling dimensionality wall, and fallback-dimensionality — are *confirmations of known behavior* (classical first-order step-size limits; concentration of measure [18]), reported here as calibrated measurements inside the pipeline, not as discoveries. *Scope*: latent space, synthetic teacher, random W . *Apparatus note* (G4 arena runs). The arena-visualization runs interface the latent planner with Push-T through the full deployment pipeline of Section III-E: the environment renders 96^2 RGB and exposes agent proprioception; the frozen encoder ($K=128$) maps $(x_t, p_t) \mapsto z_t$; CEM optimizes $a_{1:T}$ ($T=8$) through the certified predictor against the leashed quasi-metric cost — with W either random (the Proposition 3 ablation) or probe-guided (G4) — toward the probe-decoded goal latent; the first four planned actions are executed as de-normalized position targets and the loop replans every four environment steps, with the T-block observed only through the image and the latent. With the projection changed from random to probe-guided W (same certified model, same planner), the T-block reverses direction — a directional effect that does not transfer to deployment

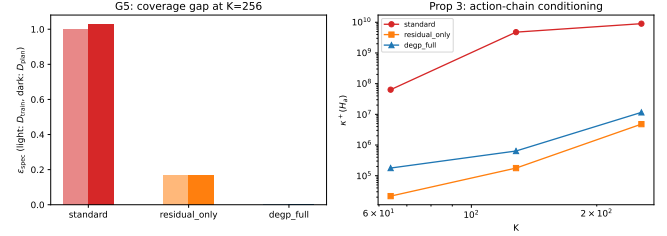
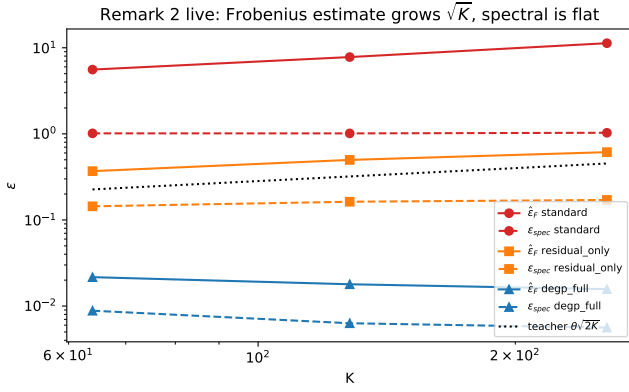


Fig. 6. Experiment 1 diagnostics. Left: Remark 2 live — the Frobenius estimate grows with K while the spectral ε is flat. Right (G5): the coverage gap ($\varepsilon_{\text{spec}}$ on $\mathcal{D}_{\text{train}}$ vs. $\mathcal{D}_{\text{plan}}$) and the action-chain conditioning $\kappa^+(H_a)$ (Proposition 4).

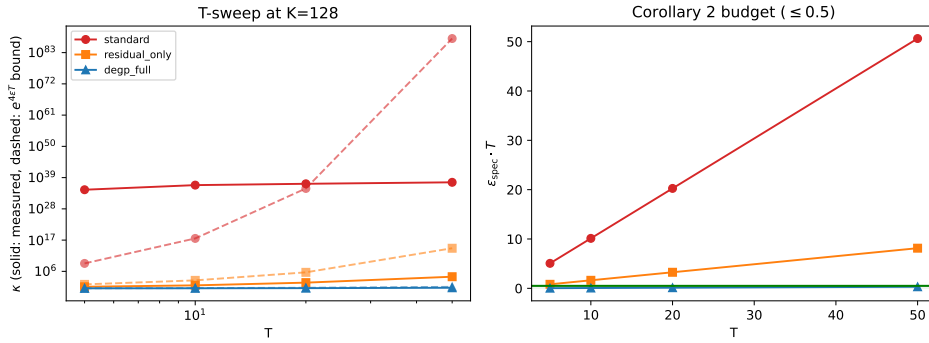


Fig. 7. Experiment 1 (R4): T-sweep at $K=128$ — Corollary 2 budget ($\varepsilon T \leq 0.5$).

Verdict	Result
R1a	DEGP/leashed 10/10 , median 2 iterations ; standard 0/10; $p = 9.8 \times 10^{-4}$
R1b	iterations-to-tolerance: DEGP 2 vs residual-only 153 ; $p = 9.8 \times 10^{-4}$
R2	unleashed 9/10; one instrumented tripwire capture; median 197 iterations
R3	CEM 0/10 at $TN_a=320$ (dimensionality wall)
R4	non-colinearity rate 100% (99% at 1°), median angle 3.4°
R5	Euclidean at per-cost lr: 10/10, 36 iterations; diverges at the leashed lr
hybrids	watchdog bailed 10/10 on standard; CEM rescue 0

TABLE VI

STAGE-1 VERDICTS. THE 2-ITERATION CONVERGENCE MATCHES THE TOP-DIRECTION CONTRACTION $1 - 2\mu\eta\sigma_{\max}^2(J_A) = 0.4$ PER STEP (SCOPED BY REMARK 4).

conditions (median 0.4 px movement, Table VIII), consistent with the quasi-metric scoping of Section III-C.

E. Experiment 3: modularity on MimicGen

This experiment is the transfer step: whether the representation-level properties certified on synthetic dynamics survive real images. Apparatus and instrument validation: posterior-collapse confirmation for the VAE baseline at text-book β ; a redesigned per-sample shift instrument with raw-pixel floor 0.09–0.18 and mechanism correlation $r = 0.76$, $p = 0.004$. **Measured 2×2** (10 paired seeds; per-seed in the artifact):

The deployable corner (frozen + DEGP) sits marginally above the budget guideline ($\varepsilon T = 0.54$ at $T=16$; cf. the marginal breach at 0.57 under $\lambda_2=0.1$ in Experiment 1); this bounds the certifiable horizon for the configuration rather than invalidating the corner.

F. Experiment 4 (Stage 2): deployment-level planning on Push-T (measured)

Planning in-environment on Push-T (T-block pivot), 50 episodes per condition (10 seeds $\times$ 5), learned W variants, short-horizon MPC, watchdog and telemetry live. The pre-registered success hypotheses (CEM $> 90\%$, hybrid $> 90\%$)

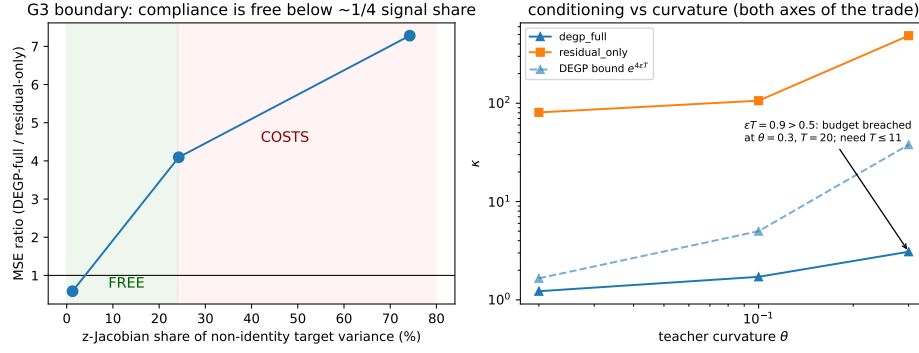


Fig. 8. Experiment 1: the G3 boundary — compliance is free below $\sim 1/4$ of dynamics-signal share.

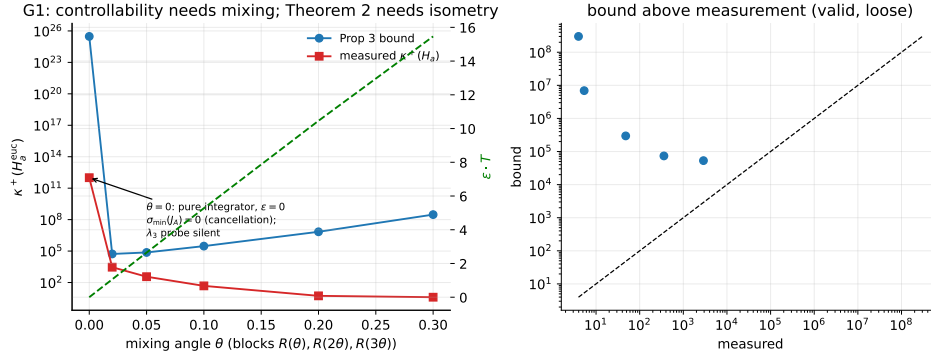


Fig. 9. Proposition 4: measured $\kappa^+(H_a)$ vs. the bound (valid, loose). The mixing-angle sweep shows that controllability requires action-space mixing while Theorem 1 requires near-isometry ($\sigma_{\min}(J_A)=0$ at zero mixing, where the λ_3 probe is silent); the remaining panel sweeps the horizon T .

Cell	shift	probe R^2	$\varepsilon_{\text{spec}}$	κ	εT
standard frozen	0.561	0.455	1.759	4.6×10^9	28.1
standard joint	0.302	0.425	1.708	1.1×10^{25}	27.3
DEGP frozen	0.583	0.456	0.034	10.9	0.54
DEGP joint	0.282	0.435	0.116	1.4×10^3	1.86

TABLE VII

MODULARITY 2×2 ON MIMICGEN TWO_ARM_THREADING (10 PAIRED SEEDS). AGENT-POSITION PROBE R^2 IS PARTLY CIRCULAR (PROPRIOCEPTION IS AN ENCODER INPUT); THE SHIFT-TRANSFER AND CONDITIONING COLUMNS ARE THE NON-TRIVIAL MEASUREMENTS. RANDOM-PROJECTION CONTROLS (SECTION V-G) BOUND WHAT EQ. (1) ALONE ACHIEVES.

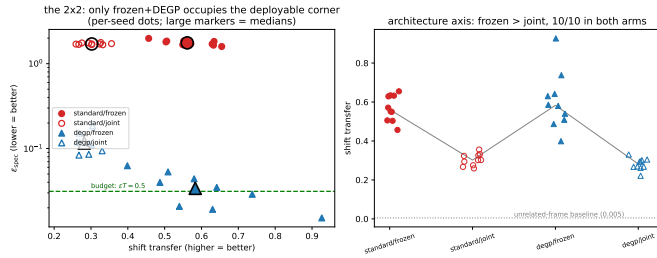


Fig. 10. Experiment 3: the 2×2 in (modularity, conditioning) space; per-seed dots, large markers = medians. Only frozen+DEGP occupies the deployable corner.

were **falsified** and are reported as measured nulls: 0/50 suc-

cesses in every condition, T-block movement at the 0.0–1.1 px level, uniform across planner family, cost variant, and planning order (Table VIII). Stage 3 attributes this boundary.

G. Experiment 5 (Stage 3): boundary attribution (pre-registered diagnostic battery)

Stage 2 establishes *that* the pipeline fails; Stage 3 asks the only question a falsified hypothesis leaves open: *which component binds*. The diagnostic attribution battery operationalizes the paper’s central empirical distinction — *information present* in the representation versus *information used* by the learned dynamics — through controls that swap one component at a time: oracle dynamics, privileged pose, second-order planning, leash ablation, a reconstruction world model, and encoder controls. The battery, its verdict rules, and a reproducibility

Condition	Success	95% CP upper	T-err chg	T moved (px)	Cov end	ms/plan
Euclidean CEM	0/50	5.8%	+0.0	1.1	0.000	13
Random- W CEM	0/50	5.8%	+0.0	0.5	0.000	14
G4- W CEM	0/50	5.8%	+0.0	0.4	0.000	13
G4- W PGD	0/50	5.8%	+0.0	0.0	0.000	908

TABLE VIII

STAGE-2 MEASURED RESULTS. T-BLOCK MOVEMENT WAS SPARSE ACROSS ALL CONDITIONS (MEDIAN 0.0–1.1 PX; PER-EPISODE VALUES IN THE ARTIFACT). CEM’S SPEED ADVANTAGE OVER PGD (13 VS. 908 MS) INVERTED THE PRE-REGISTERED EXPECTATION — REPORTED AS A NULL.

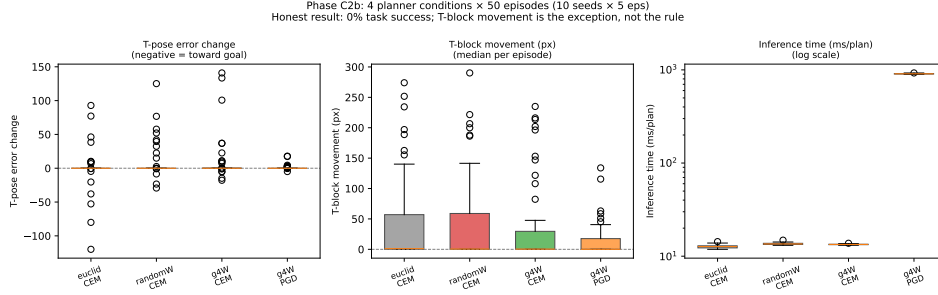


Fig. 11. Stage-2 measured Push-T results: 0/50 per condition (CP upper 5.8%); per-seed points and binomial bounds.

finding were pre-registered before running (Amendments A2–A4, A10; recorded in the released artifact). The battery as pre-registered comprises 18 conditions¹, including a horizon sweep (Table X).

a) *Reproducibility finding*: Naive pymunk body-level state teleportation was found invalid for oracle rollouts (55.9 px free-space divergence from hidden controller state); a pre-registered fidelity gate detected this and automatically selected exact replay-based rollouts (same seed + recorded action history). The gate is part of the artifact.

b) *Contact Dynamics Ratio (CDR) and selectivity*: For validation transitions split by whether the T-block moves, let $\widehat{\Delta p}_T$ be the probe-decoded predicted T-pose change, Δp_T the ground-truth change. $CDR_{gt} = \langle \|\widehat{\Delta p}_T\| \rangle / \langle \|\Delta p_T\| \rangle$ — a magnitude ratio, *not* a probability or recall; CDR_{dec} uses the decoded actual change as denominator. *Selectivity* is the ratio of predicted-magnitude medians on the contact and free strata: a value near 1 means the predictor is blind to contact; a value well above 1 means it predicts contact-conditioned movement. *Note on earlier reported values*. Journey values reported previously under an earlier evaluation protocol (0.038/0.507/0.218/2.714) are *not* fully reconciled by the dual-denominator change: v1, v3, and v4 agree within protocol noise (under the ground-truth and decoded denominators, respectively). The v2 entry is traceable rather than loose:

¹Enumeration, verified against the amendment register released with the artifact: the twelve deployment rows of Table XI (D1a, D1a at 4× budget, D1b, D2, D4, D6, D8, B1, B1b, B2, E1, E2), the four model-journey versions v1–v4 (Table IX), and the two CTRL encoder controls — 12 + 4 + 2 = 18. Deltas from the registered set (A2): B1b was added after the B1 result (A10) and CTRL after the Tier-1 encoder-scope disclosure; D7 (environment variation) was registered but not run (Section VII). D3 is an instrument analysis over the v1–v4 latents and D5 a horizon sweep on the v4 configuration; both are counted once via the journey. E1 and E2 share a CEM random stream and are counted as one *effective* condition under Amendment A11.

0.507 is v2’s *dynamics-capture gate* value — the predicted-to-actual latent step ratio over all validation transitions, logged during v2 training — a different instrument from the contact-stratum decoded CDR, not a mis-measurement of it. The gate value is preserved in the artifact’s training log; the dual-CDR measurements (0.042/0.028) are the battery’s values and the ones reported in Table IX.

The journey, measured. The journey is monotone in amplification and flat in exploitation (Table IX): from v1’s under-prediction ($CDR_{gt} = 0.027$) through v3’s partial capture (0.232) to v4’s over-amplification (4.115), selectivity never leaves $0.98\text{--}1.18\times$ — no version predicts contact-conditioned movement *selectively*; each either misses it or mis-measures its magnitude. Conditioning tracks the same trade: v1 is near-perfectly certified but dynamics-dead ($\varepsilon_{spec}=0.0007$, $\kappa=1.01$), while v2 buys dynamics at a disclosed budget breach ($\varepsilon_{spec}=0.196$, $\kappa=57.6$, $\varepsilon \cdot T=1.57$), and the B1b ablation pushes the breach to $\varepsilon \cdot T=3.41$ (below). Decodability stays at 0.65–0.70 throughout and matches full DEGP (CTRL-B: 0.686), while the frozen random encoder cannot decode (CTRL-A: $R^2 = 0.028$) yet the predictor trained on its latents shows equally negligible contact capture (CDR 0.060) — the Eq. (1) objective preserves the signal without organizing it for prediction. An earlier draft described v4 as “workable” on the strength of its amplified contact ratio; that framing was withdrawn by amendment (A9): amplification without selectivity is not usable prediction, and every v4 deployment condition in the battery bears this out (0/10; Table XI).

c) *Horizon sweep (D5)*: Table X separates the representation floor from predictor excess: decoding the *true* latent z_{t+T} through the frozen probe costs a flat ~ 35 px at every horizon and stratum (the information is present; the decoded contact/free discrimination of the representation is $5.3\text{--}6.0\times$

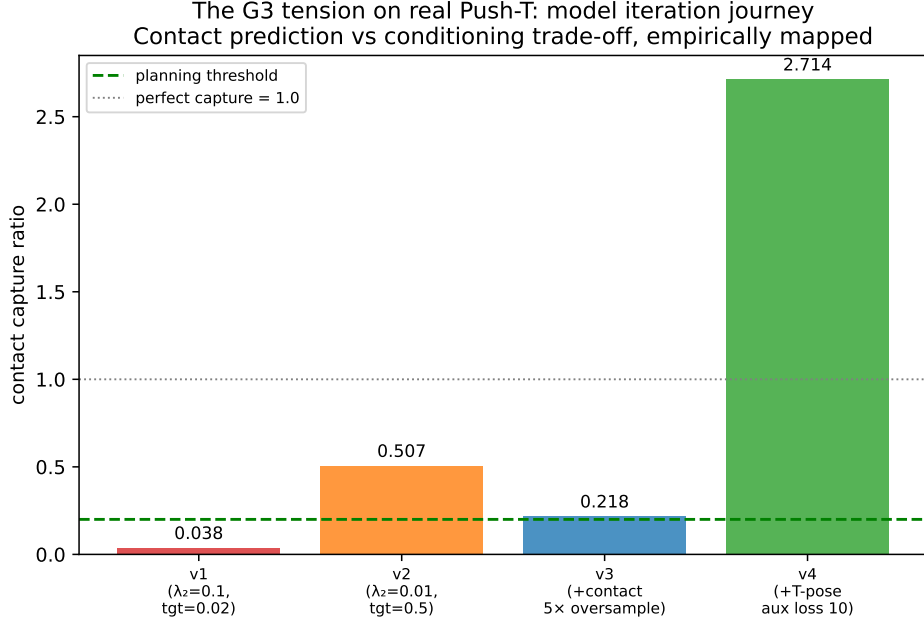


Fig. 12. Stage-2 model journey (v1–v4) under the earlier evaluation protocol (values as originally reported: 0.038/0.507/0.218/2.714; the dual-denominator measurements are in Table IX): v4 is amplified but contact-blind (selectivity 0.98, cos +0.11) — the “workable” framing is withdrawn (Amendment A9).

Version	T-pose R^2	CDR _{gt}	CDR _{dec}	cos	Sel.
v1 (pre-registration)	0.652	0.027	0.020	+0.041	1.02
v2 (dynamics-matched)	0.674	0.042	0.028	−0.063	1.00
v3 (contact-oversampled)	0.675	0.232	0.162	+0.216	1.18
v4 (+ T-pose aux)	0.696	4.115	2.726	+0.110	0.98
B1b (v4, $\lambda_2=0$)	0.678	0.186	—	+0.210	1.18
CTRL-A (frozen random enc.)	0.028	0.060	—	—	—
CTRL-B (retrained enc.)	0.686	0.046	—	—	—

TABLE IX

MODEL JOURNEY AND ENCODER CONTROLS (D3 INSTRUMENT, CONTACT STRATUM): *information present vs. information used*. DECODABILITY (T-POSE R^2) IS STABLE AT 0.65–0.70 ACROSS ALL TRAINED ENCODERS AND MATCHES FULL DEGP (CTRL-B), WHILE THE PREDICTORS’ CONTACT DYNAMICS MOVE FROM UNDER-PREDICTION (V1–V2) TO OVER-AMPLIFICATION (V4) WITH SELECTIVITY AT 0.98–1.18 $\times$ — CONTACT-BLIND THROUGHOUT. B1B AND CTRL ROWS REPORT THE GROUND-TRUTH DENOMINATOR ONLY.

across model versions), while the T -step rolled latent adds +23–41 px already at $T=1$ and +825–1006 px at $T=16$, the contact stratum worse than the free stratum at every horizon. Oracle-dynamics episodes with the ground-truth cost reduce T-pose error to a median of ~ 32 px, while the same oracle with the probe-decoded cost leaves it at ~ 113 px — the cost component, measured independently of the learned dynamics.

d) Leash ablation (B1b): Training the v4 recipe with $\lambda_2=0$ removes the Jacobian penalty while holding the encoder objective, architecture, and data fixed. The measured consequence matches the theory’s prediction: the conditioning budget is violated ($\varepsilon_{\text{spec}}=0.426$, $\varepsilon \cdot T_{\text{plan}}=3.41 \gg 0.5$; $\kappa=1.2 \times 10^3$, within the Corollary 2 bound for that regime) — the leash is what maintains the certified geometry. The deployment outcome, however, is unchanged (0/10; median T-movement 0 px; selectivity 1.18, CDR_{gt}=0.186, cos +0.21): within the tested regime, planner conditioning is not the

binding failure factor — removing the leash changes the certificate, not the outcome.

e) Battery findings: No condition in the battery achieves task success (Table XI); the discriminating measurement is T-block movement. Component swaps that leave the learned Eq. (1)-latent dynamics in place — second-order planning (D8, 2.3 px), privileged pose input (D4, 0.0 px), leash removal (B1b, 0.0 px), incumbent-style learned models (E1/E2, 0.0–3.4 px) — leave movement at the Stage-2 level. The dynamics-swap controls move the block: oracle dynamics with the ground-truth cost (D1a, 72.3 px; at 4 $\times$ budget, 68.5 px) and the reconstruction world model with the same planner (B1, 32.4 px), while oracle dynamics with the probe-decoded cost (D1b) move it not at all (0.0 px) — evidence that the probe-decoded cost is itself a limiting factor in the tested regime. The geometric heuristic (B2) moves the block (126 px) without reaching coverage, confirming that block movement, not task

T	free stratum (px)			contact stratum (px)		
	pred	floor	excess	pred	floor	excess
1	59.2	36.2	+23.0	76.0	34.9	+41.0
2	108.4	35.7	+72.8	150.9	32.8	+118.0
4	210.8	34.9	+176.0	291.9	36.9	+255.0
8	442.4	37.1	+405.3	616.3	33.5	+582.8
16	860.7	36.0	+824.7	1040.8	34.6	+1006.2

TABLE X

HORIZON SWEEP (D5, v4 PREDICTOR): T-POSE ERROR OF THE T -STEP ROLLED LATENT (PRED), DECODED THROUGH THE FROZEN PROBE, VS. THE REPRESENTATION FLOOR (DECODING THE *true* LATENT z_{t+T} THROUGH THE SAME PROBE), SPLIT BY WHETHER THE T-BLOCK MOVES. THE FLOOR IS FLAT AT ~ 35 PX ACROSS HORIZONS AND STRATA — THE POSE REMAINS DECODABLE AT EVERY HORIZON — WHILE THE PREDICTOR’S EXCESS IS POSITIVE ALREADY AT $T=1$ AND GROWS TO $+825/+1006$ PX AT $T=16$, WITH THE CONTACT STRATUM WORSE AT EVERY HORIZON.

success, is the correct reading level for the battery. The scaled-down E baselines train their reconstruction heads to convergence (E2: $8.6 \rightarrow 0.21$) but their reward models do not generalize (R^2 : $-0.08/-0.02$; E2’s KL saturates its 8-nat free-bits floor) — disclosed omissions for the comparison.

f) Attribution summary: Within the tested Push-T regime, the attribution battery supports a two-component explanation: a contact-conditional prediction gap in the Eq. (1) latent dynamics together with limitations of the probe-decoded planning cost. The tested controls did not identify planner family or order, perception input, or the conditioning leash as binding failure factors. These conclusions are scoped to the evaluated environment, model family, training data, and planning horizon.

VI. RUNTIME CONTRACT

The runtime contract turns the theorems of Section IV into deployed, monitored behavior. It has four components, all exercised end-to-end in the Stage-2/3 runs.

a) Pre-deployment gates: A world-model checkpoint is admitted to planning only if it passes the pre-registered gates of Section V-A (T-pose probe R^2 , latent scale, $\varepsilon \cdot T_{\text{plan}}$ budget, dynamics capture, action sensitivity). A failing gate blocks planning and forces model iteration — the mechanism that drove the v1 $\rightarrow$ v2 amendment (v1 failed the capture and action-sensitivity gates at 0.038 and $\sim 10^{-3}$) and disclosed the v2 budget breach ($\varepsilon \cdot T = 1.57$) before any success-rate claim was made.

b) Telemetry: Every planning step logs $\hat{\varepsilon}$ (Hutchinson estimate and windowed singular-spectrum maxima, Remark 2), tail quantiles (P99) of $\hat{\varepsilon}$ and κ per episode, and the quasi-metric tripwire ($d_q \leq \delta \Rightarrow \|\Delta\| \leq \tau$ violations). Coverage of $\mathcal{D}_{\text{plan}}$ is monitored (Remark 3): rollout drift beyond the training $\|z\|$ envelope raises $\varepsilon_{\text{spec}}$ and is recorded, not ignored.

c) Watchdog and fallback: A watchdog compares planned latent rollouts against re-encoded observations; divergence triggers a fallback planner (CEM through the same model at reduced horizon). Measured: the watchdog bailed 10/10 episodes on the standard (uncertified) arm, whose

unrolled Jacobians saturate (Table VI, hybrids); CEM rescue succeeded in 0 of them — fallback preserves operation but cannot manufacture conditioning that the model does not have.

d) Oracle-rollout fidelity gate: Attribution experiments that roll out *oracle* dynamics must first pass a fidelity gate comparing body-level state teleportation against exact replay. Measured: naive teleportation diverges by 55.9px in free space from hidden controller state; the gate detected this and automatically selected replay-based rollouts (seed + recorded action history) for all oracle conditions (Section V-G).

Certification without monitoring is a static claim; the deployment boundary of Section V-G was mapped with the telemetry live — every null in Table XI carries logged conditioning telemetry, so the attribution rests on measurements, not assumptions.

VII. DISCUSSION

a) What the tested regime establishes: The certified stack does what the theorems claim, and the measurement apparatus does what the protocol claims. Conditioning is dimension-invariant under the stated assumptions (Theorem 1; 30-seed TOST replication); the budget delimits the usable horizon (Corollary 2; the G3 boundary makes the veto-cost concrete); the leash guarantees and, when removed, measurably forfeits the property (B1b: $\varepsilon \cdot T = 3.41$, $\kappa = 1.2 \times 10^3$ against a certified κ of order unity). Every falsified hypothesis is reported with exact counts and bounds, and every attribution claim carries telemetry logged while it ran.

b) What the tested regime does not establish: Task competence. All Stage-2 conditions fail (0/50) and every battery condition fails (0/10; 0/30 for D6). The battery converts the null into a scoped boundary: a representation that decodes contact-relevant pose to a flat ~ 35 px floor at every horizon ($5.3\text{--}6.0\times$ contact/free discrimination — information present), predictors that remain contact-blind across the entire journey and the encoder controls (selectivity $0.98\text{--}1.18\times$ — information not used), rollout error that is positive already at $T=1$ and grows with horizon (Table X), and a probe-decoded cost that suppresses movement even under oracle dynamics (D1a vs. D1b). Within the tested Push-T regime, the attribution battery supports a two-component explanation — the contact-conditional prediction gap over Eq. (1) latents together with the probe-decoded cost — while planner family and order, perception input, and the conditioning leash are not identified as binding failure factors under the tested controls.

c) Limitations: (1) *Single environment:* all deployment results are Push-T; D7 (environment variation) was registered but not run in the original protocol. A post-hoc diagnostic probe (one axis, one value, reduced n ; Appendix B) does not resolve the failure either, but is scoped separately from this table’s results and should not be read as closing D7. (2) *Optimizer robustness (resolved at Stage 1):* the optimizer-robustness experiment (Amendment A12, §V-C0d) closes this at the synthetic Stage-1 level — DEGP-full’s conditioning varies by only $1.21\times$ across five optimizer families, an order of magnitude tighter than the uncertified arms; optimizer

Condition (code)	Success	95% CP upper	Cov	T moved (px)	ms/plan
<i>Learned Eq. (1)-latent model (v4), component swaps</i>					
G4-W PGD, v4 re-run (D2)	0/10	25.9%	0.053	0.0	867
iLQR, second order (D8)	0/10	25.9%	0.000	2.3	272
Privileged pose input (D4)	0/10	25.9%	0.010	0.0	891
Unleashed predictor, $\lambda_2=0$ (B1b)	0/10	25.9%	0.000	0.0	—
<i>Dynamics-swap controls</i>					
Oracle + ground-truth cost (D1a)	0/10	25.9%	0.308	72.3	8704
Oracle + probe-decoded cost (D1b)	0/10	25.9%	0.000	0.0	8028
Oracle, $4\times$ CEM budget (D1a-HI)	0/10	25.9%	0.244	68.5	35291
<i>Alternative world models / planners</i>					
Reconstruction WM + PGD (B1)	0/10	25.9%	0.001	32.4	879
TD-MPC2-lite CEM (E1)	0/10	25.9%	0.106	3.4	41
Dreamer-v3-style CEM (E2)	0/10	25.9%	0.106	0.0	177
Geometric heuristic (B2)	0/10	25.9%	0.092	126.0	—
Pose grid, 30 episodes (D6)	0/30	9.5%	—	—	—

TABLE XI

STAGE-3 DIAGNOSTIC ATTRIBUTION BATTERY (10 EPISODES PER CONDITION, 30 FOR D6; MEDIANS; ONE-SIDED 95% CLOPPER–PEARSON UPPER BOUNDS). NO CONDITION ACHIEVES TASK SUCCESS; THE DISCRIMINATING MEASUREMENT IS T-BLOCK MOVEMENT. SWAPS THAT KEEP THE LEARNED EQ. (1)-LATENT DYNAMICS (PLANNER ORDER, PERCEPTION, LEASH, INCUMBENT-STYLE MODELS) LEAVE MOVEMENT AT THE 0–3.4 PX LEVEL, WHILE ORACLE DYNAMICS WITH THE GROUND-TRUTH COST (D1A) AND THE RECONSTRUCTION WORLD MODEL (B1) MOVE THE BLOCK — AND ORACLE DYNAMICS WITH THE PROBE-DECODED COST (D1B) DO NOT. E1/E2 REWARD-MODEL R^2 : $-0.082/-0.020$ (DISCLOSED).

choice at deployment on Push-T with the real encoder remains untested. (3) *Asymmetric metric unvalidated*: a Euclidean terminal cost solves all self-reachable Stage-1 tasks and all cost variants fail identically at deployment; Eq. (5) was never exercised in the original protocol — random and probe-guided W only. A post-hoc diagnostic (Appendix B-B) trains Eq. (5) and finds it cannot beat a constant-mean baseline at the paper’s own $C_W=1$, and requires $C_W=50$ to learn a non-degenerate metric; deployed at $C_W=50$ (Appendix B-C, reconstructed encoder, calibrated separately from Table VIII) it still fails to reach coverage. (4) *Predictor-anchored Stage-1 goals* (Remark 4). (5) *Battery power*: $n=10$ per condition (Stage-2: $n=50$); the battery discriminates at the level of block movement, not task success — no condition, including oracle dynamics, succeeds. (6) *Linear probe cost*: the decoder is ridge-regularized linear (~ 35 px floor); nonlinear decoders are untested. (7) *Scaled-down incumbents*: the E baselines train reconstruction to convergence but their reward models do not generalize (R^2 : $-0.08/-0.02$) — disclosed omissions. (8) *Journey budget breaches*: v2’s $\varepsilon \cdot T = 1.57$ breach is disclosed and deployment proceeded under it; the certified budget is a property of the certified configuration, not of the whole journey. (9) *Data coverage*: 8.0% of steps move the T-block under random exploration; the contact regime is necessarily sparse.

d) *Open questions and next steps*: Whether conditioning certification is *necessary* at deployment remains open: B1b’s identical outcome shows it is not binding in this regime, but longer horizons, tighter tolerances, or distillation may restore its relevance. How the boundary moves under environment variation (D7) remains unmeasured on the original protocol; a post-hoc, differently-scoped probe (Appendix B) finds the same zero-coverage outcome under one damping variation,

without resolving D7 itself. Optimizer robustness is now measured at the Stage-1 synthetic level (Amendment A12, §V-C0d) and does not affect certified conditioning; whether the same holds at deployment, with the real encoder and under the full Push-T pipeline, is untested. Eq. (5) has now been exercised (Appendix B-B–B-C): it requires relaxing C_W well past the paper’s default to learn a non-degenerate metric at all, and even then does not change the deployment outcome — consistent with, not a resolution of, the gap this paper maps. The next technical problem is the measured gap itself: objectives that make decodable contact information *learnable* by a predictor over Eq. (1) latents — the placement principle pursued in the companion work [21] — and cost decoders beyond a linear probe.

VIII. CONCLUSIONS

DEGP separates representation shaping from geometric control and provides a framework in which conditioning claims can be stated, verified, and monitored explicitly. The theoretical results establish conditioning and leash properties under stated assumptions, while the executable verification suite tests their corresponding algebraic and numerical consequences. Across the tested latent dimensions, the certified predictor exhibits statistically flat conditioning and satisfies the measured $\varepsilon \cdot T$ budget over the evaluated horizons.

The deployment experiments establish a measured boundary: all Stage-2 conditions achieve 0/50 successes, and the Stage-3 diagnostic battery shows that the evaluated Eq. (1)-latent predictors do not exploit the contact information that remains decodable from the representation. The tested planner, perception, and leash controls did not resolve the failure, while the oracle and reconstruction-based controls provide contrasting evidence that learned dynamics and the probe-decoded cost are limiting factors in the tested Push-T regime.

The necessity of the asymmetric metric and the deployment necessity of conditioning certification remain open questions.

The resulting contribution is therefore methodological and diagnostic: DEGP demonstrates how theoretical conditioning guarantees, executable verification, statistical replication, and deployment attribution can be combined to distinguish a certified geometric property from actual task competence. Closing the observed gap between information decodable from a representation and information exploited by the learned dynamics predictor is the next technical problem identified by these experiments.

APPENDIX A

FULL PROOFS OF THE SECTION IV RESULTS

This appendix writes out the proofs in full; the short body proofs of Proposition 1 and Theorem 2 are restated here in expanded form. Throughout, $J_g^{(t)}$ is the state Jacobian at step t of the unrolled chain, $J_\Phi = \prod_{t=1}^T J_g^{(t)}$ the unrolled state Jacobian, and, for the Euclidean terminal cost, $H := J_\Phi^\top J_\Phi$ the Gauss–Newton (zero-residual) Hessian — the quantity the telemetry instrument logs as $\kappa(J_\Phi)^2$.

A. Proof of Proposition 1

Let θ parameterize f_θ . Reverse-mode differentiation computes $\nabla_\theta \mathcal{L}$ by summing, over all directed paths in the computation graph from the loss node to θ , the products of local Jacobians along each path. The stop-gradient $\text{sg}(\cdot)$ detaches its argument: $\text{sg}(z)$ is a constant with respect to θ during backward accumulation. With both predictor endpoints detached, every path from $\mathcal{L}_{\text{pred}}$ to θ passes through a detached node and contributes exactly zero; the remaining arguments (a_t, ϕ) are θ -independent. Hence

$$\nabla_\theta \mathcal{L}_{\text{total}} = \nabla_\theta \mathcal{L}_{\text{enc}} + 0,$$

and Σ_z is determined by $\mathcal{L}_{\text{enc}}$ alone. (Detaching only the target leaves the input path intact — and the prediction term, the Jacobian penalty, and the λ_3 probe each evaluate the predictor at z_t , so three gradient paths into θ survive; the reference implementation verifies both directions by unit test.) ■

B. Proof of Theorem 1

By assumption, $\|J_g - I\|_F \leq \varepsilon < 1$ over $\mathcal{D}_{\text{plan}}$. Since the spectral norm (or operator 2-norm) is bounded above by the Frobenius norm, we have $\|J_g - I\|_2 \leq \|J_g - I\|_F \leq \varepsilon$.

By Weyl’s inequality, for any singular value $\sigma(J_g^{(t)})$ of the single-step Jacobian $J_g^{(t)}$ at step t :

$$1 - \varepsilon \leq \sigma(J_g^{(t)}) \leq 1 + \varepsilon.$$

Thus, each $J_g^{(t)}$ is nonsingular with $\sigma_{\max}(J_g^{(t)}) \leq 1 + \varepsilon$ and $\sigma_{\min}(J_g^{(t)}) \geq 1 - \varepsilon$.

For the unrolled state Jacobian $J_\Phi = \prod_{t=1}^T J_g^{(t)}$, applying the sub-multiplicative properties of singular values for products of square matrices yields:

$$\begin{aligned} \sigma_{\max}(J_\Phi) &\leq \prod_{t=1}^T \sigma_{\max}(J_g^{(t)}) \leq (1 + \varepsilon)^T, \\ \sigma_{\min}(J_\Phi) &\geq \prod_{t=1}^T \sigma_{\min}(J_g^{(t)}) \geq (1 - \varepsilon)^T. \end{aligned}$$

The condition number of J_Φ in the 2-norm is defined as:

$$\kappa(J_\Phi) = \frac{\sigma_{\max}(J_\Phi)}{\sigma_{\min}(J_\Phi)} \leq \left(\frac{1 + \varepsilon}{1 - \varepsilon} \right)^T.$$

For a Euclidean terminal cost, the Gauss–Newton Hessian is $H = J_\Phi^\top J_\Phi$. The singular values of $J_\Phi^\top J_\Phi$ are the squares of the singular values of J_Φ . Consequently, the condition number of H satisfies:

$$\kappa(H) = \kappa(J_\Phi)^2 \leq \left(\frac{1 + \varepsilon}{1 - \varepsilon} \right)^{2T}.$$

Neither the premise nor the bound depends on the latent dimension K or the variance scaling γ , establishing K -invariance. ■

C. Proof of Corollary 2

For a μ_0 -strongly convex, L_0 -smooth objective, gradient descent with step $1/L_0$ contracts the optimality gap geometrically, so reaching relative accuracy ρ requires

$$k \geq (L_0/\mu_0) \ln(1/\rho) = O(\kappa \ln(1/\rho))$$

iterations with $\kappa = \kappa(H)$. Substituting Theorem 1 and using, for $\varepsilon \leq \frac{1}{2}$,

$$\ln \frac{1 + \varepsilon}{1 - \varepsilon} \leq \frac{2\varepsilon}{1 - \varepsilon} \leq 4\varepsilon,$$

gives

$$k \leq O(e^{4\varepsilon T} \ln(1/\rho)),$$

the budget table’s $\kappa \approx e^{4\varepsilon T}$. The count is K -invariant unconditionally and T -invariant only while εT stays constant: the pre-registered reading $\varepsilon \cdot T \lesssim 0.5$, i.e., $\varepsilon \leq 0.01$ at $T=50$. ■

D. Proof of Lemma 1

Write $w_{k,i}$ for row i of W_k and $a_{k,i} = \langle w_{k,i}, \Delta \rangle$. (i) d_q is $\frac{1}{M}$ times a sum of squares, so $d_q(\Delta) = 0$ iff $\text{ReLU}(a_{1,i}) = \text{ReLU}(a_{2,i})$ for every i ; for reals, $\text{ReLU}(a) = \text{ReLU}(b)$ iff $[a \leq 0 \wedge b \leq 0]$ or $[a = b \geq 0]$. (ii) The all-nonpositive branch defines the closed polyhedral cone $C = \{\Delta : \langle w_{k,i}, \Delta \rangle \leq 0 \ \forall k, i\}$, an intersection of $2M$ halfspaces through the origin. Any zero-set piece using the equality branch for a nonempty index set E lies in the proper subspace $\{\Delta : \langle w_{1,i} - w_{2,i}, \Delta \rangle = 0, \ i \in E\}$ — measure zero in $\mathbb{R}^K$; the all-equality piece is $\{\Delta : (W_1 - W_2)\Delta = 0, \ W_1\Delta \geq 0\}$, of dimension $K - \text{rank}(W_1 - W_2)$: positive-dimensional flat valleys exactly when $\text{rank}(W_1 - W_2) < K$. (iii) C has nonempty interior iff some Δ has $\langle w_{k,i}, \Delta \rangle < 0$ for all $2M$ rows, i.e., iff the

rows lie in an open halfspace; Wendel's theorem [20] gives the probability for n points in general position on S^{K-1} as $2^{1-n} \sum_{j=0}^{K-1} \binom{n-1}{j}$, equal to 1 when $n \leq K$ — with $n = 2M$ the stated criterion. (iv) Fixing 0/1 activation masks D_1, D_2 , write $B := D_1 W_1 - D_2 W_2$; on the region $d_q(\Delta) = \frac{1}{M} \|B\Delta\|^2$ is quadratic with Hessian $\frac{2}{M} B^\top B$ of rank at most M (B has M rows); for $M < K$ it has a nontrivial kernel, so no lower spectral bound exists for the unleashed cost. ■

E. Proof of Theorem 2

(i) Both terms of $\mathcal{L}_{\text{plan}} = d_q + \mu \|z - z_{\text{goal}}\|^2$ are nonnegative, and (L1) gives $d_q(z_{\text{goal}}, z_{\text{goal}}) = 0$, so $\mathcal{L}_{\text{plan}}(z_{\text{goal}}) = 0$. Any global minimizer z^* satisfies $0 = \mathcal{L}_{\text{plan}}(z^*) \geq \mu \|z^* - z_{\text{goal}}\|^2$, hence $z^* = z_{\text{goal}}$. (ii) At a differentiable stationary point, $0 = \nabla d_q + 2\mu\Delta$, so $2\mu\|\Delta\| = \|\nabla d_q\| \leq L$ by the gradient bound of an L -Lipschitz function; hence $\|z^* - z_{\text{goal}}\| \leq L/(2\mu)$, and no differentiable stationary point exists outside the closed leash ball. On $\text{int}(C)$, (L2) gives $\nabla d_q = 0$, so the iterate satisfies $z_{t+1} - z_{\text{goal}} = (1 - 2\mu\eta)(z_t - z_{\text{goal}})$, a contraction for $0 < \eta < 1/(2\mu)$. (iii) Writing $\nabla d_q = \alpha\Delta + w$, $w \perp \Delta$, gives $\nabla \mathcal{L}_{\text{plan}} = (\alpha + 2\mu)\Delta + w$: the leash contributes only the radial term $2\mu\Delta$, so w — and therefore parallelism to Δ — is decided by the base cost alone, for every $\mu \geq 0$. d_q is piecewise quadratic with finitely many ReLU-boundary hyperplanes, so claims are stated at differentiable points; the measured rates are in Proposition 3. ■

F. Proof of Proposition 2

On an activation-fixed region with 0/1 masks D_1, D_2 , write $B := D_1 W_1 - D_2 W_2$; then $d_q(\Delta) = \frac{1}{M} \|B\Delta\|^2$ and

$$H_{\text{loc}} = \frac{2}{M} B^\top B + 2\mu I \succeq 2\mu I,$$

using $D_k^2 = D_k$ and $D_1 D_2 = D_2 D_1$ in

$$\begin{aligned} B^\top B &= W_1^\top D_1 W_1 + W_2^\top D_2 W_2 \\ &\quad - W_1^\top D_1 D_2 W_2 - W_2^\top D_2 D_1 W_1. \end{aligned}$$

Each term on the right is bounded in spectral norm by $C_W^2: D_k$ and $D_1 D_2$ are 0/1 diagonal with $\|\cdot\|_2 \leq 1$, and $\|W_k\|_2 \leq C_W$ by (P3). Hence $\lambda_{\max}(H_{\text{loc}}) \leq \frac{8C_W^2}{M} + 2\mu$. Transporting to the initial state gives $H = J_\Phi^\top H_{\text{loc}} J_\Phi$, hence

$$\begin{aligned} \kappa(H) &\leq \kappa(J_\Phi)^2 \kappa(H_{\text{loc}}) \\ &\leq \left(1 + \frac{4C_W^2}{M\mu}\right) \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{2T}, \end{aligned}$$

away from the measure-zero activation boundaries. ■

G. Proof of Proposition 3

At a differentiable point with $\Delta \neq 0$, decompose $\nabla d_q = \alpha\Delta + w$, $w \perp \Delta$; $\nabla \mathcal{L}_{\text{plan}} \parallel \Delta$ iff $w = 0$. Direct differentiation of Eq. (3) gives

$$\begin{aligned} \nabla d_q &= \frac{2}{M} \sum_i [\text{ReLU}(a_{1,i}) - \text{ReLU}(a_{2,i})] \\ &\quad \times (m_{1,i} w_{1,i} - m_{2,i} w_{2,i}), \end{aligned}$$

with 0/1 masks $m_{k,i}$. If $\text{sign}(a_{1,i}) \neq \text{sign}(a_{2,i})$ for some i , exactly one mask is active and the bracketed coefficient is the

positive input, nonzero. If the active row $w_{k,i}$ is not parallel to Δ , its orthogonal residual

$$w_{k,i}^\perp := w_{k,i} - \frac{\langle w_{k,i}, \Delta \rangle}{\|\Delta\|^2} \Delta$$

is nonzero, and the summand's projection onto $\Delta^\perp$ is a nonzero multiple of $w_{k,i}^\perp$. Then $w = 0$ requires exact cancellation by the remaining summands — a proper algebraic condition, satisfied on a measure-zero set of (W_1, W_2, Δ) . Off that set, the hypothesis forces $w \neq 0$; the measured rate during planning is 100% at nonzero deviation (99% at a 1° threshold, median angle 3.4°). ■

H. Proof of Proposition 4

Write $A_t = J_g^{(t)}$ and $B_t = \partial g_\phi / \partial a$ with $\|B_t\|_2 \leq \bar{c}$ (P2). The action Jacobian has blocks $G_s = A_{T-1} \cdots A_{s+1} B_s$ (empty product = I), each with $\|G_s\|_2 \leq (1 + \varepsilon)^{T-1} \bar{c}$ by the singular-value bounds above, so

$$\|J_A\|_2 \leq \|J_A\|_F \leq \sqrt{TN_a} (1 + \varepsilon)^{T-1} \bar{c}.$$

With $H_a = J_A^\top \tilde{H} J_A$ and $\tilde{H}$ the terminal leashed Hessian of Proposition 2,

$$\begin{aligned} \lambda_{\max}(H_a) &\leq TN_a (1 + \varepsilon)^{2(T-1)} \bar{c}^2 \left(\frac{8C_W^2}{M} + 2\mu\right), \\ \lambda_{\min}^+(H_a) &\geq 2\mu \sigma_{\min}(J_A)^2 \geq 2\mu c_A^2, \end{aligned}$$

by composite controllability (P2), the positive eigenvalues living on the range of $J_A^\top$. Dividing,

$$\kappa^+(H_a) \leq \left(1 + \frac{4C_W^2}{M\mu}\right) TN_a (1 + \varepsilon)^{2(T-1)} (\bar{c}/c_A)^2.$$

Without (P2), $\sigma_{\min}(J_A)$ can degenerate to zero and no bound of this form survives. ■

APPENDIX B

POST-HOC DIAGNOSTIC REPLICATION WITH A RECONSTRUCTED ENCODER

The encoder class f_θ that produced the checkpoints used throughout Section V was not preserved in the released artifact independently of its trained weights (Section I, code-availability note); only `enc/pred` state dicts survive. This appendix reports a *diagnostic, not a replacement*, replication built to close two items left open in Section VII: D7 (environment variation) and Eq. (5) (the controllability-trained asymmetric metric, never previously exercised in deployment). Its numbers are reported separately from, and are not directly comparable to, Table VIII/XI.

A. Reconstruction and validation

A convolutional encoder was reconstructed from the `enc` state-dict's weight shapes alone (Conv(3→16, k=5)–Conv(16→32, k=5)–Conv(32→32, k=3), adaptive-pooled to a 512-d flatten, concatenated with 2-d `agent_pos`, two linear layers to $K=128$). State-dict shapes fix channel counts and kernel sizes exactly; they do not fix stride, padding, or pooling, which were chosen (stride-2 convs, GELU, 4×4 adaptive average pooling) rather than recovered. The deployment environment itself is not reconstructed: the public `gym-pusht`

package (pymunk==6.11.1, matching Section III-E(c)) reproduces the paper’s stated defaults exactly ($k_p=100$, $k_v=20$, control at 10Hz, $96\times 96\times 3$ pixel observations, goal pose $(256, 256, 45^\circ)$, success threshold 0.95) without modification. Loading the real `c2_world_model.pt enc` weights into the reconstructed class and fitting the T-pose probe (Section III-E(a) protocol, unchanged) on real encoded rollouts gives $R^2=0.682$, inside the paper’s reported 0.65–0.70 range for trained encoders (Table IX) — evidence the reconstruction is behaviorally close, though not proven identical, to the original.

B. Eq. (5): exercising the trained metric

W_1, W_2 were regressed onto measured steps-to-reach following Eq. (5), using rollout pairs from the certified v4 predictor (loaded unmodified from the checkpoint) under a random-action controller, both directions (forward pairs always have a witness $\tau_{i\rightarrow j} = j - i$; reverse pairs $\tau_{j\rightarrow i}$ use a fresh rollout from z_j , censored at the horizon if z_i is never reached). **A calibration-independent finding:** at the paper’s own weight-normalization bound $C_W=1$ (P3), the regression cannot beat a constant-mean baseline (MSE 0.111 vs. 0.072) — median inter-step latent displacement under this certified predictor is $\|\Delta z\| \approx 0.18$ against an ambient latent norm ≈ 178 , too small relative to $C_W=1$ for the ReLU-difference metric to carry gradient signal. Relaxing to $C_W=50$ recovers a non-degenerate fit (MSE 0.019). This result does not depend on the reconstructed encoder or on action calibration (Appendix B-C) — it is a property of the certified predictor’s own conditioning tightness, and is offered as a candidate explanation for why Eq. (5) was left unexercised at $C_W=1$ in the original deployment protocol.

C. Deployment: Eq. (5) and D7

The $C_W=50$ trained metric was deployed under the same CEM/leash protocol as Table VIII ($T=8$, replanning every 4 steps), alongside euclidean and random- W controls, after a calibration sweep over the planner’s action-space bound (the original action-normalization constants were not recovered; $a_{\max}=1$, $\sigma_{\text{init}}=0.3$ was selected as the smallest setting still producing non-degenerate exploration against the real `D2_v4_deploy` log, which shows 0px movement on 10/10 episodes). **Absolute px values below are calibrated against that log, not against Table VIII, and should not be compared to it directly.**

Within this diagnostic’s own scale, two qualitative patterns replicate the main paper: (i) euclidean and random- W produce near-identical per-episode trajectories, reproducing Table VIII’s finding that cost variant does not change the outcome; (ii) coverage stays at 0.000 (median) in every condition, both at the paper’s default dynamics and under the D7 damping probe — the trained metric and the one environment variation tested here change *how much* the block moves, not *whether* the task is solved. Movement grows under damping=0.9 (plausibly because $k_p=100/k_v=20$ were tuned against the undamped default and overshoot under added damping), but success and coverage do not.

Condition	Success	Cov. (med.)	T -moved (med., px)
Euclidean	0/10	0.000	19.2
Random- W	0/10	0.000	19.3
Trained- W (Eq. (5))	0/10	0.000	37.6
<i>D7 probe, damping= 0.9 (paper default: undamped)</i>			
Euclidean	0/5	0.000	133.4
Random- W	0/5	0.000	133.4
Trained- W	0/5	0.000	257.6

TABLE XII
DIAGNOSTIC REPLICATION (APPENDIX B): RECONSTRUCTED ENCODER, CALIBRATED ACTION SCALE. NOT ON THE SAME ABSOLUTE SCALE AS TABLE VIII; SEE §B-C.

D. Scope and caveats

This appendix is a best-effort diagnostic, not a certified replication: (a) the encoder’s stride/padding/pooling are chosen, not recovered (§B-A); (b) the action-space calibration in §B-C is fit to reproduce a small-movement regime, not the original normalization constants, so absolute T -moved values are not on Table VIII/XI’s scale; (c) D7 here tests one axis (gym-pusht’s exposed damping parameter) at one value, chosen because it is a literal, real dynamics-affecting parameter of the deployment environment — it is not necessarily the axis specified in the original (unrecovered) D7 pre-registration text; (d) episode counts here ($n=5-10$) are smaller than Table VIII’s $n=50$. The Eq. (5) C_W finding (§B-B) holds independently of action scaling.

REFERENCES

- [1] A. Bardes, J. Ponce, and Y. LeCun, “VICReg: Variance-invariance-covariance regularization for representation learning,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2022.
- [2] R. Balestrieri and Y. LeCun, “LeJEPa,” 2025.
- [3] M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller, “Embed to control: A locally linear latent dynamics model for control from raw images,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.
- [4] S. Hansen *et al.*, “Bisimulation-based state representation learning,” 2022.
- [5] S. Levine *et al.*, “Prediction-consistency-curvature for locally-linear latent control,” 2020.
- [6] Bounou *et al.*, “Second-order LQR methods for ill-conditioned problems,” 2021.
- [7] Bounou *et al.*, “Proximal LQR,” 2024.
- [8] J. Hoffman *et al.*, “Jacobian norm regularization,” 2019.
- [9] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
- [10] T. Wang, A. Wiehe, and P. Isola, “Quasimetric learning,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2023.
- [11] K. Chua, R. Calandra, R. McAllister, and S. Levine, “Deep reinforcement learning in a handful of trials using probabilistic dynamics models,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.
- [12] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, “Mastering diverse domains through world models,” *arXiv:2301.04104*, 2023.
- [13] N. Hansen, H. Su, and X. Wang, “TD-MPC2: Scalable, robust world models for continuous control,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.
- [14] D. Ha and J. Schmidhuber, “World models,” *arXiv:1803.10122*, 2018.
- [15] C. Chi, S. Feng, Y. Du, Z. Xu, B. Burchfiel, E. Cousineau, and S. Song, “Diffusion policy: Visuomotor policy learning via action diffusion,” in *Robotics: Science and Systems (RSS)*, 2023.

- [16] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *Int. J. Robotics Research*, vol. 33, no. 1, 2014.
- [17] J. Hewitt and P. Liang, "Designing and interpreting probes with control tasks," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2019.
- [18] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, U.K.: Oxford Univ. Press, 2013.
- [19] D. Lakens, "Equivalence tests: A practical primer for t tests, correlations, and meta-analyses," *Social Psychological and Personality Science*, vol. 8, no. 4, 2017.
- [20] J. G. Wendel, "A problem in geometric probability," *Mathematica Scandinavica*, vol. 11, 1962.
- [21] M. Mostafa, "Placement principles for physics losses in representation learning," companion paper, in preparation, 2025.