

Physics Guards for Residual Policies at Breakaway: When Learned Corrections Should Defer to a Friction Model

Clement Aidoo

CANLABS, Ghana clement.aidoo@canlabs.org

Code, data and protocol: <https://github.com/CAN-Aidoo/breakaway-shield> Archive: doi:10.5281/zenodo.23503813

Abstract

A residual policy learned on top of a physics controller can cut the number of costly measurements a manipulation task needs, but in friction-limited tasks its worst errors happen at breakaway, when friction drops from its static to its kinetic level. We ask when such a learned correction should defer to a friction model. We derive two breakaway shields. For gravitational pivoting, a closed-form grip floor certifies that post-slip travel stays within the target band whenever the static-to-kinetic ratio r is at most an assumed $r_{\max}$. For controlled sliding, the shield bounds the time spent below the arrest force. For pivoting, a certified release from rest exists only if $r < \phi_{\text{lim}}/\sin \phi_{\text{lim}}$, so shallow targets are infeasible for realistic r . In a pre-registered MuJoCo study with emulated stick-slip, the two tasks split. In sliding, a residual trained under the shield cut overshoot from 30% to 2% at $r = 1.2$ and ended within 5 mm on 96–98% of episodes, but kept only part of the unshielded residual’s measurement saving; a confirmation set replicated this. In pivoting, the shield found no certified release for exactly the half of the test targets that the feasibility condition predicts. Deferring to the model then left overshoot at 63% at $r = 1.1$. Declining prevented overshoot, but only by stopping short. Within the assumed range, certificates were violated on at most 0.5% of steps, mostly through soft-contact effects of the simulator. A breakaway shield helps where the physics leaves room for a certified action and, where it does not, makes that limit explicit.

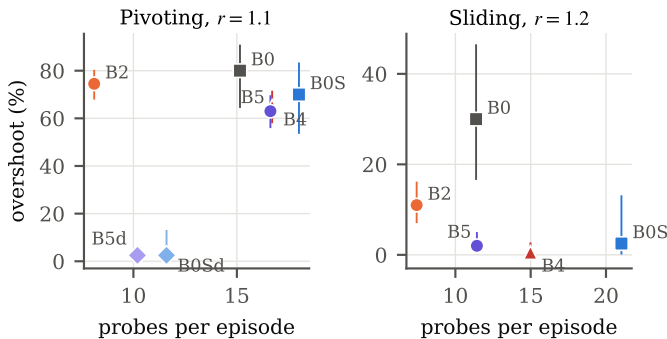


Figure 1. Measurements against safety at the largest static/kinetic ratio in the training range (40 test objects; bars: 95% Clopper–Pearson intervals). In sliding (right) the shield-trained residual (B5) keeps overshoot near zero at about the model controller’s measurement cost. In pivoting (left) half of the targets admit no certified release: with the defer fallback the shield cannot stop overshoot, and only declining (B0Sd, B5d) does so, by stopping short.

I. INTRODUCTION

Residual learning keeps a hand-designed controller and learns a correction to its output [1, 2]. In earlier work we applied it to gravitational pivoting with a parallel-jaw gripper that senses the object only through wrist motor current [3]. The residual roughly halved the number of slow angle measurements per pivot, from 18.0 to 9.3, but it overshoot targets more often. Almost every overshoot happened on one step, the step on which the rod first slipped [4]. A hand-made guard repaired most of the damage, but it was designed after we had seen the test results and carried no guarantee.

That failure has a physical cause. When a grasped object breaks away, friction falls from its static to its kinetic level, and the surplus of gravity over kinetic friction becomes kinetic energy. A quasi-static friction model, which both the base controller and our earlier guard rely on, ignores this energy. A learned correction that makes releases larger or faster than the model’s falls exactly into the region where the model is least reliable.

This paper asks when a learned correction should defer to a friction model. We turn the first-slip guard into a *shield* [5]: a closed-form set of grip actions that keeps the post-breakaway travel within the target band whenever the true static-to-kinetic ratio $r = \mu_s/\mu_k$ does not exceed an assumed bound $r_{\max}$. Outside that set, the learned correction is dropped. We study two tasks, gravitational pivoting and controlled sliding of a rod through the fingers, in MuJoCo simulation with explicit stick-slip. Our contributions are:

- A breakaway shield for pivoting with a one-line closed form (Proposition 1), together with the exact conditions under which it reduces to our earlier quasi-static guard, and a timing shield for sliding (Proposition 2). Sliding needs a bound on time below the arrest force, not a force floor.
- A feasibility analysis showing that, from rest, a certified release toward a target θ^* exists only if $r < \phi_{\text{lim}}/\sin \phi_{\text{lim}}$. Shallow pivots are therefore infeasible for realistic r , and the shield’s own infeasible region coincides with this physical one.
- A pre-registered simulation study (Q1–Q4) comparing the model controller, unshielded and shielded residual policies, and shields applied at training or only at test time, across r and under $r_{\max}$ mismatch, with certificate-

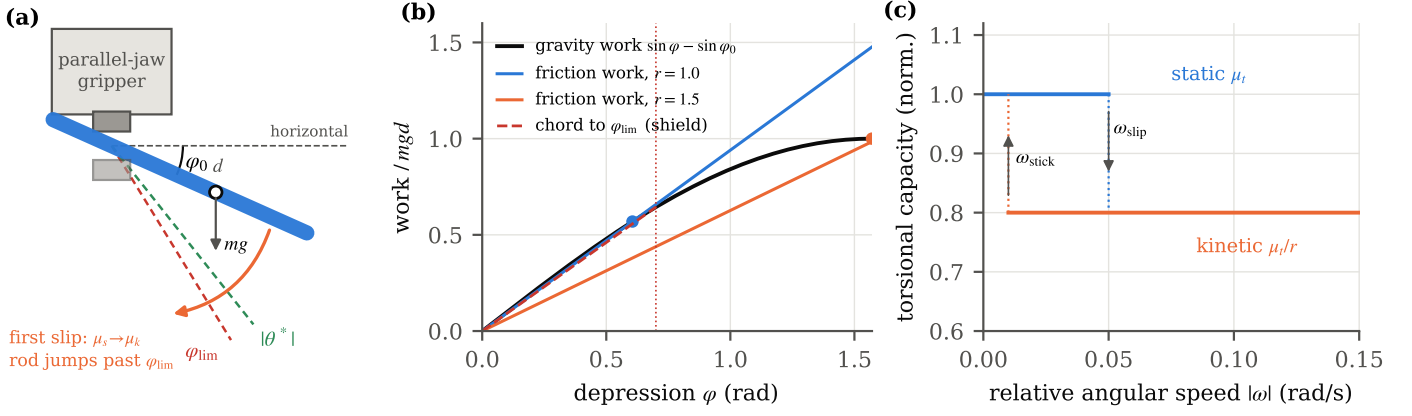


Figure 2. The breakaway problem. (a) A rod held off its centre of mass pivots under gravity; at the first slip friction drops from static to kinetic and the rod can run past the limit φ_{lim} . (b) Energy balance from rest at $\varphi_0 = 0$: the rod stops where friction work (lines of slope $k = F/(rF_h)$) catches up with gravity work. With the same grip, $r = 1.5$ stops far later than $r = 1$. The shield requires the friction line to lie above the chord to φ_{lim} . (c) The emulated friction law: static capacity until the relative speed exceeds ω_{slip} , kinetic capacity μ_k/r until it falls below ω_{stick} .

violation rates reported alongside task metrics.

II. RELATED WORK

Residual learning and physics priors. Residual RL learns an additive correction on top of a controller [1, 2] or a learned base policy [6]. Corrections are usually limited by a fixed magnitude or a penalty: ForceRFT caps a wrist-force-conditioned residual at fixed per-axis bounds [7]. Chen, Lu, Colgate and Lynch embed grasp-quality and fingertip-geometry priors into the reward and the hardware for in-hand rotation [8]. Our bound is different in kind: it is *state-dependent*, derived from contact mechanics, and enforced at run time.

Shields and safety filters. Shielding overrides unsafe actions using a model of the system [5]. Control-barrier-function filters [9, 10], Hamilton–Jacobi reachability [11] and predictive safety filters [12] generalise the idea; Brunke et al. review the field [13]. Adaptive shielding infers hidden parameters such as friction online [14], and ResSafe learns the safety filter itself as a residual [15]. Closest to us on the friction side, Johansson et al. filter bimanual manipulation actions with a Coulomb factor of safety evaluated by rollouts [16]. Closest in spirit, Shuto and Vargas derive a certified, sensorless grasp-damage bound from a stiffness lower bound and propose it as a safe-action parameter for learned grasping [17]. We make two changes. Our certificate rests on the static-to-kinetic *transition*, which a single- μ Coulomb model cannot express. And we actually train and test learners under it.

Pivoting and controlled sliding. Gravitational pivoting has been planned open loop [18], controlled adaptively [19] and learned with tactile sensing [20]. Dynamic in-hand sliding has been planned with soft-finger limit surfaces [21], and prehensile pushing uses Coulomb motion cones [22]. Soft-finger friction scales nonlinearly with load [23, 24]. Kolbert et al. found that MuJoCo under-predicted a measured pivot [25]. None of these works isolates breakaway as the failure mode of a learned correction.

III. BREAKAWAY SHIELDS

A. Pivoting

A rod of mass m is held at offset d from its centre of mass by two soft fingers with grip force F . Let $\varphi \in [0, \pi/2]$ be the depression of the centre of mass below horizontal. We assume (A1) a rigid rod rotating about the grip axis, (A2) torsional capacity linear in grip, $\tau_s = c_s F$ static and $\tau_k = c_s F/r$ kinetic, (A3) a commanded grip that is tracked and non-increasing during a release, (A4) no stored pad compliance and no damping, and (A5) gravity torque $mgd \cos \varphi$. The horizontal breakaway grip is $F_h = mgd/c_s$.

Proposition 1. Suppose the rod is stuck at φ_0 under grip F_{last} , and the true ratio satisfies $r \leq r_{\text{max}}$. Then any release to

$$F \geq F_{\text{last}} \frac{r_{\text{max}}}{\cos \varphi_0} S(\varphi_0, \varphi_{\text{lim}}), \quad S(a, b) = \frac{\sin b - \sin a}{b - a}, \quad (1)$$

leaves the rod stuck at some $\varphi \leq \varphi_{\text{lim}}$.

Proof. Sticking at φ_0 implies $c_s F_{\text{last}} \geq mgd \cos \varphi_0$, so $F_h \leq F_{\text{last}}/\cos \varphi_0$. If the rod breaks away under final grip F , its kinetic energy obeys $\frac{1}{2} I \dot{\varphi}^2 \leq mgd g(\varphi)$ with $g(\varphi) = \sin \varphi - \sin \varphi_0 - k(\varphi - \varphi_0)$ and $k = F/(rF_h)$. By A3, the grip during sliding is at least F ; by A4, no energy enters from compliance. The rod stops at the first root of g after φ_0 . Because $\sin$ is concave on $[0, \pi/2]$, the chord slope $S(\varphi_0, \varphi)$ decreases in φ , so that root is at most φ_{lim} iff $k \geq S(\varphi_0, \varphi_{\text{lim}})$. At the root, $\cos \varphi < k$, so even kinetic friction exceeds gravity and the rod re-sticks. Substituting the bounds on F_h and r gives (1). $\square$

Relation to the earlier guard. With $r_{\text{max}} = 1$ and $\varphi_0 = 0$, (1) gives $F \geq F_{\text{last}} \sin \varphi_{\text{lim}}/\varphi_{\text{lim}}$. Our earlier guard, $F \geq F_{\text{last}} \cos \varphi_{\text{lim}}$ [4], follows only if the energy balance is replaced by quasi-static equilibrium ($\cos \varphi = k$). The two floors differ by $\tan \varphi_{\text{lim}}/\varphi_{\text{lim}}$: about 1.03 at 0.3 rad, 1.20 at 0.7 rad and 1.56 at 1.0 rad. The quasi-static guard therefore ignores the inertia of the first slip even when $r = 1$.

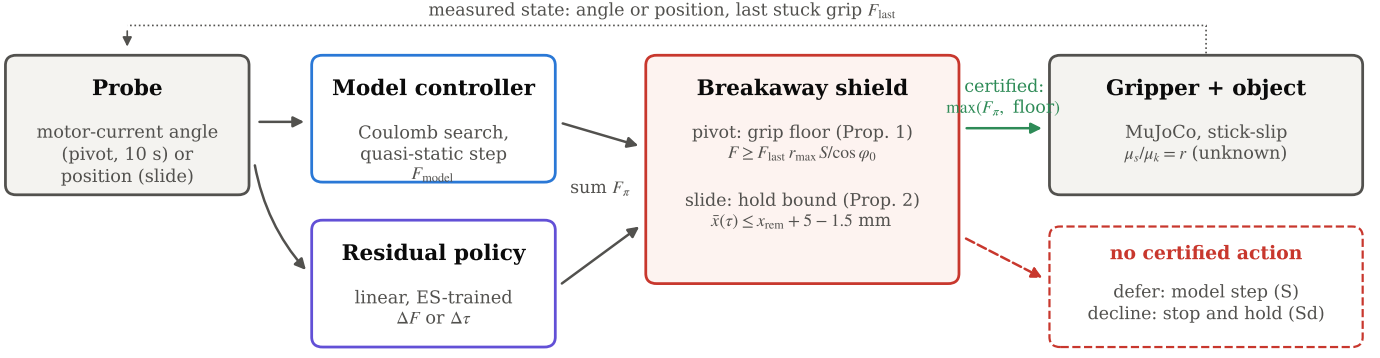


Figure 3. Shielded residual control. A probe returns the measured state and the last grip at which the object was stuck. The model controller proposes an action and the residual policy corrects it. The shield passes certified actions (clipping grip up, or the hold down, to the bound) and falls back to the model step (defer) or stops (decline) when no certified action exists.

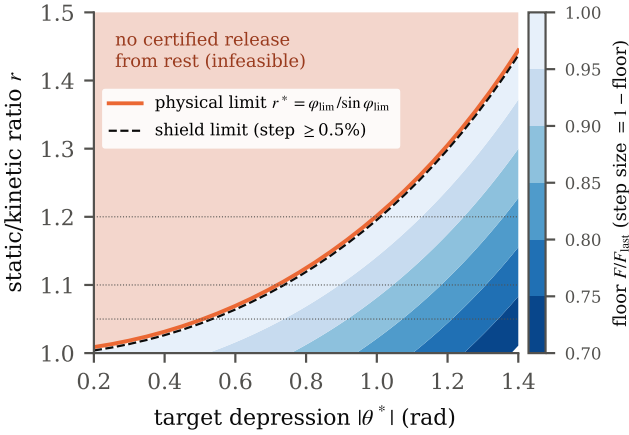


Figure 4. Where a certified pivot from rest exists. Shaded blue: the shield’s grip floor as a fraction of F_{last} (smaller means larger steps). Pink: no certified release. The shield’s limit (dashed) lies on the physical limit r^* (orange). For $r = 1.1$, targets shallower than about 0.75 rad are infeasible.

Feasibility. A release that causes motion needs $F < F_h \cos \phi_0$, while the shield needs $F \geq r F_h S$. A certified release from rest therefore exists only if

$$r < r^*(\phi_{\text{lim}}) = \phi_{\text{lim}} / \sin \phi_{\text{lim}}, \quad (2)$$

which is 1.04 at 0.5 rad, 1.09 at 0.7 rad and 1.19 at 1.0 rad. Because the shield computes its floor from F_{last} rather than the unknown F_h , it can only shrink the step toward breakaway, by the factor $r_{\text{max}} S / \cos \phi_0$. Whenever that factor is below one, the search still reaches breakaway, so the shield’s infeasible region is the physical one (Fig. 4). The price is smaller steps, and so more measurements.

Deferral rule. At each step we compute the floor. If the floor is below $0.995 F_{\text{last}}$, the step is *certified* and the commanded grip is $\max(F_{\text{policy}}, \text{floor})$, a clip that never lowers grip. Otherwise no certified release exists. We pre-specified two fallbacks: *defer*, which drops the learned residual and takes the model controller’s step (uncertified); and *decline*, which stops and holds.

We set $\phi_{\text{lim}} = |\theta^*| + 0.05 - 0.02$ rad, the overshoot threshold minus a 2.5σ sensing margin.

B. Controlled sliding

A vertical rod held at its centre of mass slides along its axis under gravity. Static breakaway occurs at $F_b = mg / (2\mu_s)$, but stopping a sliding rod needs $F > r F_b$. Gravity does not weaken as the rod slides, so for $r > 1$ any release that causes breakaway accelerates the rod until the grip rises above $r F_b$. A force floor cannot bound this travel. What bounds it is the time spent below the arrest force.

Proposition 2. Suppose the rod was last observed stuck at grip F_{stuck} and $r \leq r_{\text{max}}$. For a planned grip profile $F(t)$ starting from rest, the travel is at most $\bar{x} = \int_0^T \bar{v} dt$, where $\bar{v} = g(1 - F(t)/(r_{\text{max}} F_{\text{stuck}}))$ and $\bar{v}$ is clipped at zero.

Proof. While sliding, the acceleration is $g(1 - F/(r F_b))$, and $r F_b \leq r_{\text{max}} F_{\text{stuck}}$ because the rod was stuck at F_{stuck} . So the true acceleration never exceeds $\bar{v}$, whether positive or negative, and the true velocity never exceeds $\bar{v}$. $\square$

The shield shortens the hold of each pulse, by bisection, until $\bar{x}$ is within the remaining distance plus 5 mm minus a 1.5 mm sensing margin. Unlike pivoting, sliding has no infeasible region caused by r . Infeasibility arises only while F_{stuck} is still loose, early in the search.

IV. EXPERIMENTAL SETUP

Simulator. We use MuJoCo 3.15.0 [26] with the rig of [3, 4]: soft-finger contacts (condim 4), an elliptic cone, impratio 10, 20 noslip iterations and a 1 ms step. MuJoCo has a single friction coefficient per geom and no true stick state [27]. We therefore emulate stick-slip by switching the torsional coefficient (pivoting) or the sliding coefficient (sliding) of all contact geoms between μ and μ/r , with hysteresis on the relative speed: switching to kinetic friction above 0.05 rad/s or 5 mm/s, and back to static below 0.01 rad/s or 1 mm/s (Fig. 2c). With $r = 1$ the rig reproduces our earlier results exactly, to 10^{-6} rad. Section A reports the validation checks.

Tasks and controllers. *Pivoting* uses the object distribution and model controller of [4]: a $\times 0.94$ grip search until motion, then quasi-static steps toward the target; angle measurements have $\sigma = 0.008$ rad. *Sliding* lowers the rod by $x^* \in [10, 50]$ mm using release pulses. Each pulse releases in 0.1 s, holds for τ and re-clamps in 0.3 s, followed by a position measurement ($\sigma = 0.5$ mm). The model controller searches with $\times 0.9$ steps and then chooses the hold from a Coulomb ($r = 1$) prediction, corrected by the ratio of observed to predicted travel. In both tasks, each measurement is one *probe*, the costly sensing action.

Learning. Linear residual policies, as in [4], are trained with antithetic evolution strategies (1560 episodes per run). Training randomises $r \sim U[1, 1.1]$ for pivoting, with shield $r_{\max} = 1.1$, and $r \sim U[1, 1.2]$ for sliding, with $r_{\max} = 1.2$. For each task we train two methods with 5 seeds each: an unshielded residual, and a residual with the shield in the training loop (defer fallback).

Methods. B0 is the model controller; B1 adds the earlier quasi-static guard (pivoting only); B0S is the model controller with the shield. B2 is the unshielded residual; B3 adds the earlier guard (pivoting only). B4 applies the shield to the unshielded residual at test time only. B5 is the shield-trained residual. A suffix “d” denotes the decline fallback.

Protocol and statistics. The protocol, including hypotheses, sets, metrics and tests, was committed before any test evaluation; the sliding task was added as a dated deviation before any sliding test data existed. Each test set has 40 held-out objects evaluated at fixed r values, and a confirmation set of 40 fresh objects was run once after the analysis. Learned methods are averaged over seeds per object. We use paired Wilcoxon tests against B0, hierarchical bootstrap 95% CIs (seeds, then objects), Holm correction within each task’s B5-vs-B0 family, and Clopper–Pearson intervals for rates.

V. RESULTS

A. Simulator validation

Before any test run we checked the stick–slip emulation (protocol checks V1–V6, Table 1). Slow-ramp breakaway occurred within 0.3–2.7% of the nominal grip, for both $r = 1$ and $r = 1.2$. The contact itself has a small static/kinetic gap: breakaway at 0.988 and sliding at 0.980 of nominal capacity, an intrinsic ratio of 1.008, so an emulated $r = 1.2$ measured 1.206. Creep in the stuck phase was below 10^{-5} rad/s. Halving the time step changed first-slip travel by less than 0.025 rad, and scaling both switching thresholds by 0.5 or 2 changed it by less than 0.01 rad.

The direct test of Proposition 1 held the rod at 1.01 times its measured breakaway grip and then released it to qF_{last} (Fig. 5). Over 128 releases, simulated travel averaged 0.90–0.97 of the bound, so the bound is tight. However, 24 releases exceeded it, by at most 0.048 rad, more than our 0.02 rad sensing margin. We recorded this before testing (deviation D2) and expected certificate violations in the test runs.

Table 1. Simulator validation (validation objects only; recorded before any test data).

Check	Quantity	Result
V1	slow-ramp breakaway / nominal	1.003–1.027
V2	intrinsic static/kinetic ratio	1.008 (1.006–1.010)
V2	emulated $r = 1.2$, measured	1.206
V3	creep at $1.05 F_h$	$< 10^{-5}$ rad/s
V4	first-slip travel, dt halved	$ \Delta < 0.025$ rad
V5	thresholds $\times 0.5$, $\times 2$	$ \Delta < 0.01$ rad
V6	travel / Prop. 1 bound	0.90–0.97 (mean)
V6	releases over bound	24/128, max 0.048 rad

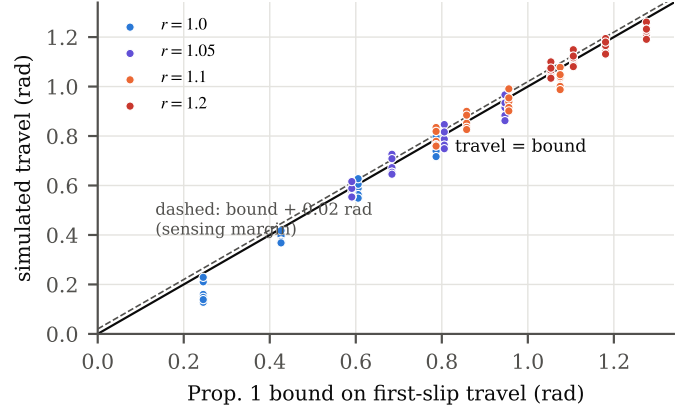


Figure 5. Proposition 1 against simulation (V6). Each point is one release from a measured stuck grip. Points above the solid line exceed the bound; points above the dashed line exceed it by more than the sensing margin.

B. Q1: Does the shield keep the residual’s savings?

Sliding (Table 2). The unshielded residual B2 needs far fewer measurements than the model controller: 7.4 against 11.4 at $r = 1.2$, and 3.5–5.3 fewer at every r . The shield-trained residual B5 keeps only part of this. It saves -2.5 probes at $r = 1.0$ (95% CI -4.8 to -0.6 , Holm $p = 0.020$) and -1.6 at $r = 1.1$ (not significant), and none at $r = 1.2$ or 1.3 . In sliding, then, the safety bought by the shield costs most of the residual’s measurement saving. Two comparisons show where the cost goes. Shielding the unshielded residual only at test time (B4) needs 13.1–15.0 probes, against 10.8–12.1 for B5: a policy trained under the shield learns to work within it. Shielding the model controller (B0S) nearly doubles its probes, to 21–27. The shield shortens its holds, so the rod stops short and needs more pulses.

Pivoting (Table 3). The unshielded residual again saves measurements: 7.0–7.4 fewer probes than the model at every r (e.g. 8.1 against 15.2 at $r = 1.1$). Under the shield this saving disappears entirely. B5 uses *more* probes than B0, by 2.5 at $r = 1.0$ (CI 0.8 to 4.5) and by 1.0–2.3 at higher r . Training under the shield made no difference here: B4 and B5 are within 0.25 probes and 3.5 points of overshoot of each other at every r . Between 43 and 47% of B5’s steps were deferred to the model, so the policy rarely got to act. At $r = 1$ the best trade-off belongs to the earlier quasi-static guard (B3: 10.5 probes, 6% overshoot, 94% within), but it stops helping as soon as $r > 1$.

Table 2. Sliding test set (40 objects; learned methods: 5 seeds, 200 episodes per cell). Within: final error ≤ 5 mm and held. Over: travel past the target by more than 5 mm. Probes: position measurements per episode. Shield $r_{\max} = 1.2$.

ID	method	$r = 1.0$			$r = 1.1$			$r = 1.2$			$r = 1.3$		
		within	over	probes	within	over	probes	within	over	probes	within	over	probes
B0	model	82.5	15	14.7	90	7.5	13.5	70	30	11.4	72.5	27.5	11.5
B0S	model + shield	77.5	0	24.2	65	2.5	26.6	75	2.5	21.0	67.5	5	21.1
B2	residual	90.5	8	9.3	90	10	8.2	89	11	7.4	86	13	7.9
B4	residual, shield at test	94.5	1.5	13.9	90.5	1.5	14.5	91.5	0.5	15.0	94	1	13.1
B5	shield-trained residual	98	2	12.1	98.5	1	11.9	96.5	2	11.4	96	2.5	10.8

Table 3. Pivoting test set (40 objects; learned methods: 5 seeds, 200 episodes per cell). Within: final error ≤ 0.05 rad and held. Over: overshoot past the target by more than 0.05 rad. Probes: angle measurements per pivot. Shield $r_{\max} = 1.1$.

ID	method	$r = 1.0$			$r = 1.05$			$r = 1.1$			$r = 1.2$		
		within	over	probes	within	over	probes	within	over	probes	within	over	probes
B0	model	87.5	12.5	15.8	42.5	57.5	15.5	20	80	15.2	7.5	92.5	14.4
B1	model + quasi-static guard	87.5	12.5	16.2	42.5	57.5	15.7	20	80	15.4	7.5	92.5	14.6
B0S	model + shield	85	7.5	19.8	42.5	47.5	19.1	22.5	70	18.0	7.5	80	16.9
B0Sd	model + shield (decline)	7.5	0	13.2	12.5	2.5	12.6	17.5	2.5	11.6	7.5	30	11.1
B2	residual	74	26	8.4	48.5	51.5	8.3	25.5	74.5	8.1	8	92	7.4
B3	residual + quasi-static guard	94	6	10.5	48	52	10.0	26.5	73.5	9.8	8	92	8.9
B4	residual, shield at test	88.5	3.5	18.3	47.5	45	18.0	26.5	65	16.7	9.5	82.5	15.5
B5	shield-trained residual	85	7	18.3	50.5	42	17.8	29	63	16.6	9.5	82.5	15.3
B5d	shield-trained (decline)	10.5	0	11.6	12	1.5	11.3	21.5	2.5	10.2	9.5	32.5	9.9

C. Q2: Does the shield bound overshoot?

Sliding. The model controller overshoots 15, 7.5, 30 and 27.5% of episodes at $r = 1.0, 1.1, 1.2$ and 1.3 . The Coulomb prediction it uses for the hold underestimates the post-breakaway acceleration once $r > 1$. The unshielded residual lowers this to 8–13%. The shield-trained residual overshoots 1–2.5% at every r , with 96–98% of episodes ending within 5 mm. Against B0, the reduction is –28 points at $r = 1.2$ (CI –42 to –15, Holm $p = 0.008$) and –25 at $r = 1.3$ (Holm $p = 0.015$). At $r \leq 1.1$, where B0 overshoots less, the reduction is not significant after correction.

The certificate itself, that a certified pulse travels no further than its limit, failed on 4, 3, 5 and 6 of about 1 900 certified B5 pulses per r (0.15–0.35%). Section E explains these failures.

Pivoting. The model controller’s overshoot grows steeply with r : 12.5, 57.5, 80 and 92.5%. The unshielded residual doubles it at $r = 1$ (26%). With the defer fallback, the shield barely helps: B5 overshoots 7, 42, 63 and 82.5%, and no reduction against B0 survives Holm correction. Deferral hands control back to the model step precisely where the model overshoots. The decline fallback does stop overshoot within the assumed range: B5d overshoots 0, 1.5 and 2.5% for $r \leq r_{\max}$ (all $p < 0.001$ against B0 for $r \geq 1.05$). But it lands within the band on only 10.5–21.5% of episodes, because declined rods stop short, on average 0.35 rad from the target. For $r \geq 1.1$ no method ends within 0.05 rad on more than 30% of episodes (Fig. 6). Certified B5 steps violated their certificate on 0, 4 and 7 of about 1 600 for $r \leq r_{\max}$ (at most 0.5%).

D. Q3: How much of the task admits a certified action?

Pivoting is limited by Eq. (2). With $r_{\max} = 1.1$, the equation predicts that 20 of the 40 test targets have no certified release from rest. The model controller with the decline fallback (B0Sd) declined at its very first step on exactly these 20 objects, both at $r = 1.0$ and at $r = 1.1$. A further 14 (at $r = 1.0$) and 11 (at $r = 1.1$) declined later, after 5–28 steps, once the rod had moved part of the way and the remaining arc could no longer be certified. Overall B0Sd declined on 34, 32, 31 and 22 of 40 objects. With the defer fallback, 53–61% of steps were certified. The infeasible region is therefore not a corner case: for realistic friction ratios it covers most of the task.

Sliding behaves as Proposition 2 predicts. Every shielded episode deferred exactly once, on its first pulse, while F_{stuck} was still unknown. No later pulse was infeasible, at any r (B0S 160 of 160 episodes, B4 and B5 800 of 800 each).

E. Q4: Why do violations and overshoots occur?

Attribution. We replayed every test episode with a certificate violation and classified each violating step (an exploratory analysis, run after the pre-registered tests). For sliding, 10 of the 32 violating steps occurred at $r = 1.3 > r_{\max}$, where no guarantee applies. One followed a recorded stuck grip that was below the true arrest requirement: the rod had slipped by less than the 2.5 mm motion threshold. The remaining 21 occurred with every assumption of Proposition 2 satisfied. All 21 were on three objects, all heavy ($m = 0.178$ – 0.192 kg). Direct probes on these objects found simulated travel up to 1.8 times the bound for short pulses released close to the stuck grip, where the bound is only a few millimetres. On light objects, travel stayed below

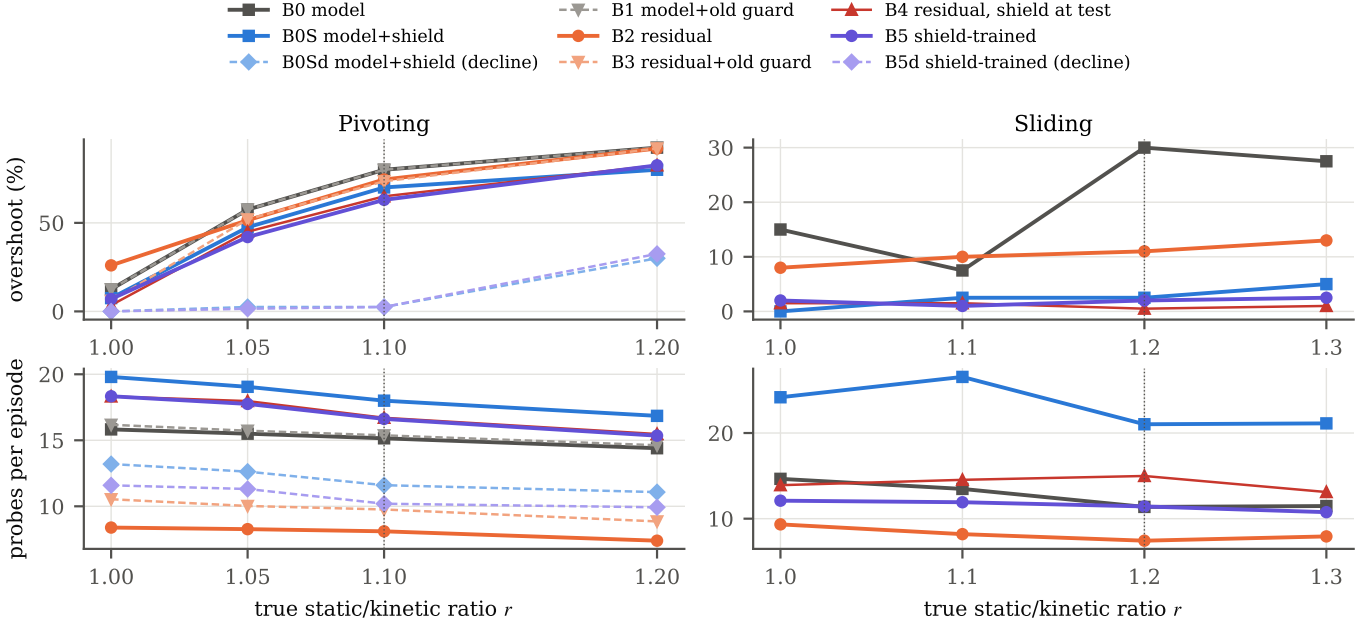


Figure 6. Overshoot rate (top) and probes per episode (bottom) against the true static/kinetic ratio r , on the test sets. Dotted vertical line: the shield’s assumed $r_{\max}$. In sliding the shield-trained residual (B5) stays near zero overshoot across r . In pivoting only the decline fallbacks (B0Sd, B5d) prevent overshoot, by stopping short.

the bound. We attribute these to MuJoCo’s regularised friction, which lets a heavily loaded contact creep [27]; this is a simulator effect and not a sensing error.

For pivoting, violations rise roughly tenfold when the assumption fails: at $r = 1.2 > r_{\max}$, 68 of 1388 certified B5 steps (4.9%) violated the certificate, and decline-mode overshoot rose to 32.5%. At $r \leq r_{\max}$ the violation rate is consistent with the model–simulator gap measured in V6.

Mismatch. Fig. 7 shows what happens when $r_{\max}$ is wrong. In sliding, assuming $r_{\max} = 1.0$ when the true r is 1.1 or 1.2 raises B5’s overshoot to 5.5 and 6.5% and its violation rate to about 1%. Its probes fall to 8.5, the level of the unshielded residual. Assuming $r_{\max} = 1.3$ keeps overshoot at 0.5–1.5% at a cost of 13.3 probes. The bound trades measurements for overshoot monotonically, and an underestimate degrades the shield gradually rather than abruptly. Pivoting behaves differently. With $r_{\max} = 1.0$ every step is certifiable, but the certificates are wrong: 6.6% of B5’s certified steps violated them at $r = 1.1$. Yet B5’s overshoot (58.5%) is *lower* than with $r_{\max} = 1.2$ (73%), where about 80% of steps are deferred and no certificate is violated. A larger $r_{\max}$ hands more steps back to the model, which overshoots. With the defer fallback, a valid certificate on every certified step does not make the episode safe.

Confirmation. The confirmation sets, 40 fresh objects per task with r drawn from the training range, were each run once after the test analysis had been committed. In sliding, B5 overshoot 0.5% of episodes against 17.5% for B0 (Wilcoxon $p = 0.010$), ended within 5 mm in 98.5% against 82.5%, and used 11.9 probes against 14.0 (difference not significant). The test-set conclusions replicate. In pivoting (r drawn from $[1.0, 1.1]$), the confirmation set again showed no safety gain from the defer

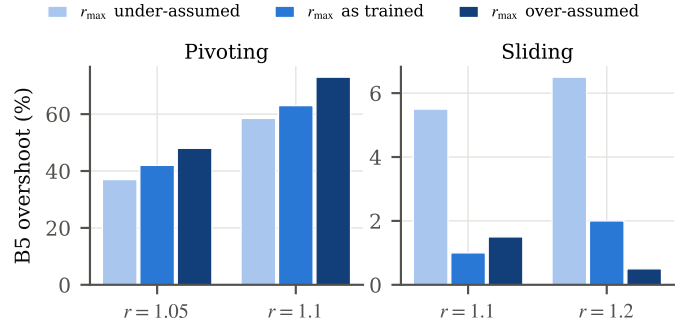


Figure 7. Overshoot of the shield-trained residual (B5, defer fallback) when the assumed $r_{\max}$ is wrong. Pivoting: $r_{\max} = 1.0, 1.1$ (as trained) and 1.2. Sliding: 1.0, 1.2 (as trained) and 1.3. In sliding an underestimate raises overshoot and an overestimate lowers it. In pivoting the order reverses, because a larger $r_{\max}$ defers more steps to the overshooting model controller.

fallback: B5 overshoot 36% against 37.5% for B0, while using 17.8 probes against 15.2 (Wilcoxon $p < 0.001$). The decline fallback again removed overshoot (B0Sd 0%) but declined on 31 of 40 objects and ended within the band on 15%. No certified B5 step violated its certificate (0 of 1709).

VI. LIMITATIONS

Infeasible region. The main limitation is physical rather than algorithmic. By (2), no quasi-static release from rest can take the rod to a shallow target without passing it once r exceeds $\phi_{\lim}/\sin \phi_{\lim}$. For $r = 1.1$ this rules out every target shallower than about 0.75 rad. In this region the shield can only defer, which reproduces the model controller’s overshoot, or decline, which stops short. A fast slip-triggered re-clamp (a “catch”)

could reach these targets: in a separate probe, a 50 ms detection-plus-clamp delay limited the first-slip jump to 0.03–0.07 rad even at $r = 1.5$. However, it needs slip detection far faster than our 10 s current-based angle measurement, and we leave it to future work.

Certified under a model, not under the world. Proposition 1 assumes a rigid rod, torsional capacity linear in grip, constant kinetic friction and no stored pad energy. In simulation the bound held on average, with simulated travel at 0.90–0.97 of the bound, but 24 of 128 direct test releases exceeded it by up to 0.048 rad. We attribute this to soft-contact geometry, and it exceeds our 0.02 rad sensing margin. On hardware, three effects could make the bound optimistic: soft-finger friction grows sub-linearly with load [23], so capacity falls faster than grip when grip is reduced; elastomer pads store torsional energy before breakaway; and static friction rises with dwell time [28]. A practical value of $r_{\max}$ should therefore be measured on the actual pad–object pair, at the hold times the controller uses, and taken from the upper tail of the measured ratios.

Simulator. MuJoCo has no true stick state [27]. We emulated static and kinetic friction by switching coefficients with hysteresis. The emulation reproduced the intended ratio (an emulated 1.2 measured 1.206) and an intrinsic ratio of 1.008, and first-slip travel was insensitive to the time step and switching thresholds. It is nonetheless a step-function Stribeck curve without rate-and-state effects [29]. Kolbert et al. found that MuJoCo under-predicted a real pivot [25], so absolute travel values should not be transferred to hardware.

Learning and scope. We used linear policies, evolution strategies, 1560 episodes per run and 5 seeds per method. Larger policies or other optimisers could change the comparison between shield-in-training and shield-at-test. Sensing during training was a noise model, not the motor-current pipeline. Lift-with-slip, the third task we originally planned, is not studied.

VII. CONCLUSION

We asked when a learned correction should defer to a friction model at breakaway, and turned the answer into shields with closed-form certificates for pivoting and sliding. The answer depends on the task’s physics more than on the learner. In sliding, a certified action almost always exists. A residual trained under the shield then keeps overshoot near zero across friction ratios, and degrades gradually when $r_{\max}$ is underestimated, although it gives up most of the unshielded residual’s measurement saving. In pivoting, the stick–slip energy budget leaves no certified release for most shallow targets. There a shield cannot make a learned correction both efficient and safe: it can only choose between the model’s overshoot and stopping short, and the feasibility condition $r < \phi_{\lim}/\sin \phi_{\lim}$ predicts in advance which targets force that choice. We see two next steps: a fast slip-triggered catch, which would enlarge the feasible region for pivoting, and measuring $r_{\max}$ on real pad–object pairs to test whether these certificates survive contact with hardware.

DATA AVAILABILITY

All code, the pre-registered protocol with its dated deviations, every evaluated episode and the analysis scripts are available at <https://github.com/CAN-Aidoo/breakaway-shield> and archived at Zenodo, doi:10.5281/zenodo.23503813. Every number in Section V is generated from these files.

ACKNOWLEDGEMENTS AND DISCLOSURE

Simulation code, analysis and drafting were carried out with the assistance of an AI model (Claude, Anthropic) under the author’s direction. The author takes responsibility for the content. All results are from simulation.

REFERENCES

- [1] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine, “Residual reinforcement learning for robot control,” in *Proc. IEEE ICRA*, 2019, arXiv:1812.03201.
- [2] T. Silver, K. Allen, J. Tenenbaum, and L. Kaelbling, “Residual policy learning,” arXiv:1812.06298, 2018.
- [3] C. Aidoo, “Gravity pivoting from motor current: calibration-free angle estimation and the sensing budget it requires,” *enrXiv*, doi:10.31224/8492, 2026.
- [4] C. Aidoo, “Residual learning on a physics pivoting controller: faster search, and where the quasi-static model breaks,” *enrXiv*, doi:10.31224/8542, 2026.
- [5] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, “Safe reinforcement learning via shielding,” in *Proc. AAAI*, 2018, arXiv:1708.08611.
- [6] L. Ankile, A. Simeonov, I. Shenfeld, M. Torne, and P. Agrawal, “From imitation to refinement – residual RL for precise assembly,” arXiv:2407.16677, 2024.
- [7] Y. Wang, C. Zhang, X. Su, J. Ma, H. Zhu, and X. Li, “ForceRFT: Refining VLA actions through force-guided residual reinforcement learning,” arXiv:2609.22840, 2026.
- [8] Y. Chen, S. Lu, E. Colgate, and K. Lynch, “Robust in-hand manipulation via priors in reinforcement learning and mechanical design,” arXiv:2607.12105, 2026.
- [9] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick, “End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks,” in *Proc. AAAI*, 2019, arXiv:1903.08792.
- [10] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: theory and applications,” in *Proc. ECC*, 2019, arXiv:1903.11199.
- [11] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, “A general safety framework for learning-based control in uncertain robotic systems,” *IEEE Trans. Autom. Control*, 2019, arXiv:1705.01292.
- [12] K. P. Wabersich and M. N. Zeilinger, “A predictive safety filter for learning-based control of constrained nonlinear dynamical systems,” *Automatica*, 2021, arXiv:1812.05506.
- [13] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, “Safe learning in robotics: from learning-based control to safe reinforcement learning,” *Annu. Rev. Control Robot. Auton. Syst.*, 2022, arXiv:2108.06266.

- [14] M. Kwon, T. Ingebrand, U. Topcu, and L. Feng, “Adaptive shielding for safe reinforcement learning under hidden-parameter dynamics shifts,” arXiv:2506.11033, 2025.
- [15] G. Qu *et al.*, “ResSafe: learning safety filtering with residual reinforcement learning for humanoids,” arXiv:2609.15988, 2026.
- [16] A. Johansson, D. Lindmark, V. Wiberg, and M. Servin, “Safety filtering of robotic manipulation under environment uncertainty: a computational approach,” arXiv:2509.12674, 2025.
- [17] Y. Shuto and D. V. Vargas, “Sensorless damage-safe grasping,” arXiv:2608.23983, 2026.
- [18] A. Holladay, R. Paolini, and M. T. Mason, “A general framework for open-loop pivoting,” in *Proc. IEEE ICRA*, 2015, pp. 3675–3681.
- [19] F. E. Viña B., Y. Karayiannidis, C. Smith, and D. Kragic, “Adaptive control for pivoting with visual and tactile feedback,” in *Proc. IEEE ICRA*, 2016, pp. 399–406.
- [20] J. Toskov, R. Newbury, M. Mukadam, D. Kulić, and A. Cosgun, “In-hand gravitational pivoting using tactile sensing,” in *Proc. CoRL*, 2022, arXiv:2210.05068.
- [21] J. Shi, J. Z. Woodruff, P. B. Umbanhowar, and K. M. Lynch, “Dynamic in-hand sliding manipulation,” *IEEE Trans. Robot.*, 33(4):778–795, 2017.
- [22] N. Chavan-Dafle and A. Rodriguez, “Prehensile pushing: in-hand manipulation with push-primitives,” in *Proc. IEEE/RSJ IROS*, 2015.
- [23] N. Xydias and I. Kao, “Modeling of contact mechanics and friction limit surfaces for soft fingers in robotics, with experimental results,” *Int. J. Robot. Res.*, 18(9):941–950, 1999.
- [24] R. D. Howe and M. R. Cutkosky, “Practical force-motion models for sliding manipulation,” *Int. J. Robot. Res.*, 15(6):557–572, 1996.
- [25] R. Kolbert, N. Chavan-Dafle, and A. Rodriguez, “Experimental validation of contact dynamics for in-hand manipulation,” in *Proc. ISER*, 2016, arXiv:1702.07252.
- [26] E. Todorov, T. Erez, and Y. Tassa, “MuJoCo: a physics engine for model-based control,” in *Proc. IEEE/RSJ IROS*, 2012, pp. 5026–5033.
- [27] MuJoCo developers, “Regularized friction never sticks,” GitHub issue #3328, google-deepmind/mujoco, 2026; and MuJoCo documentation, *Computation: soft contacts*.
- [28] J. H. Dieterich, “Time-dependent friction in rocks,” *J. Geophys. Res.*, 77(20):3690–3697, 1972.
- [29] C. Canudas de Wit, H. Olsson, K. J. Åström, and P. Lischinsky, “A new model for control of systems with friction,” *IEEE Trans. Autom. Control*, 40(3):419–425, 1995.