# ECA: EFFERENCE-COPY ADAPTATION FOR FROZEN CONTROL POLICIES

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Learned policies are widely used to control autonomous systems, but actuator ordering, direction, or gain errors can exceed the tolerance of their feedback. Identifying and inverting the actuator map restores performance; applying an estimated inverse when no such map is present can instead degrade working feedback. We propose Efference-Copy Adaptation (ECA), which augments an identification-and-inversion core with static-map and velocity-coupled hypotheses, a rejection gate, a unit amplification bound, and an identifier hold. ECA acts under the best-supported hypothesis, withdraws when the identity or a delayed identity explains recent transitions better, decides only on informative data, and never amplifies the command it corrects. On two MuJoCo manipulators, it turns the seven-actuator permutation and the orthogonal mix, on which identification alone loses $24.0$ and $2.8$ return, into gains of $42.7$ and $65.1$. It removes the delay losses of $44.6$ and $113.2$ return that identification causes, and converts a velocity-coupled disturbance that a map-only gate mis-fits by $4.3$ return into a gain of $15.9$. No conservative identification control in a full grid attains ECA's recovery with a smaller worst mean loss over the four recorded preservation conditions. On a cohort trained after the method was frozen and evaluated once, recovery is met on both plants; a $1.2$ return loss under gain loss leaves preservation unmet on one. On a contact-rich third plant, recovery transfers but preservation does not, and a healthy-body measurement predicts the difference.

## 1 INTRODUCTION

Learned control policies support autonomous systems in locomotion and manipulation (Tan et al., 2018; Hwangbo et al., 2019; Zhao et al., 2020). Their feedback tolerates moderate changes in the body or environment, and training-time randomisation and adaptation extend this tolerance (Peng et al., 2018; Kumar et al., 2021). After deployment, however, robustness remains limited by the conditions represented in training and by the correspondence between policy commands and physical actuators. A loss of actuator effectiveness rescales a command, a standard fault model in control (Tao et al., 2004; Blanke et al., 2006; Zhang & Jiang, 2008). Interface errors can instead permute, reverse, or mix action channels when actuator ordering, sign conventions, or normalisation differ across simulators, drivers, and robot configurations. Related alignment problems arise when data are combined across embodiments (O'Neill et al., 2024). These examples motivate a fixed linear input map: the policy requests $a$, but the body executes $Ma$. We study this fault class without assuming that it accounts for most deployment failures.

Frozen policies on two simulated manipulators tolerate a $30^\circ$ rotation of an actuator pair, but lose substantial return at $90^\circ$ and beyond (Figure 1). Adaptive correction does not automatically solve the problem: a residual update based on nominal actuator directions reinforces the error after a large rotation, and an unconstrained identifier interprets a delay as a static actuator map and destroys the feedback that was working. When the fault class is unknown, a compensator needs a correction mechanism and a rule for when and under which hypothesis to apply it. Since ungated identification is already strong when a map fault is known to be present, that rule must be judged by whether it costs recovery.

We propose *Efference-Copy Adaptation (ECA)*, which supplies that rule. ECA wraps an identification and inversion core (ID) in a test of which fault hypothesis, if any, explains what the body is



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

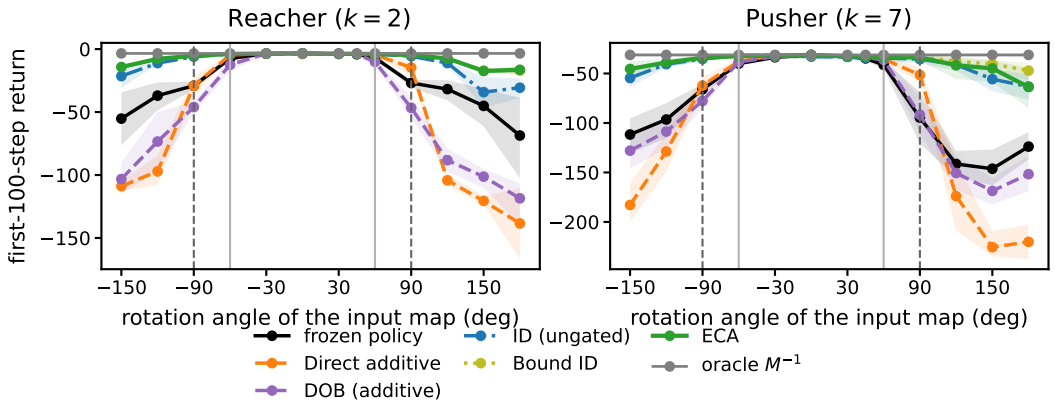



Figure 1: Return under actuator rotations on the development policies. Curves show means and 95% bootstrap intervals over five policy seeds; Bound ID is the ungated inverse under ECA's unit bound. Vertical guides mark $\pm 60^\circ$ and $\pm 90^\circ$; they are not certified stability boundaries.

doing. A nominal consequence model predicts the response to the issued command. Its residual drives recursive estimates of two fault hypotheses: a static input map and a map with a velocity-coupled term for disturbances such as damping changes and cross-joint coupling. ECA applies a regularised inverse from the first step, exactly as ID does, and scores each hypothesis on recent transitions. The references are two zero-parameter alternatives: the identity map and a delayed identity. ECA withdraws compensation when an alternative explains the data better. The velocity-coupled hypothesis acts only when it also beats the map hypothesis, and neither decision changes on data that cannot discriminate. A unit bound restricts the inverse to the norm of the command it corrects (the policy's own command under the map hypothesis), so ECA redirects rather than amplifies. The identifiers are held while the data are uninformative. Policy and nominal model remain fixed, and online adaptation uses neither reward nor fault labels. A local analysis motivates the design. Frozen feedback has a small-gain margin; an idealised additive residual update requires input-map eigenvalues with positive real parts; identification avoids this angular restriction under an exact reference and sufficient excitation. None of these results certifies the nonlinear algorithm.

Our contributions are: 1) a local analysis that distinguishes the limits of frozen feedback, additive correction, and identification under fixed input maps; 2) one algorithm combining a hypothesis library, rejection, selection, and bounded compensation, with ablations that connect each component to a specific failure; 3) an evaluation that separates recovery from interference, on policies trained after the method was frozen and against conservative variants of the same inverse; and 4) a map of where the result stops, including alternatives that retrain the policy or use reward, faults that begin mid-episode, and a plant on which the method's preservation property fails for a reason that can be measured in advance.

## 2 Related Work

**Actuator faults and adaptive control.** Multiplicative actuator faults and loss of effectiveness are established topics in fault-tolerant control (Tao et al., 2004; Blanke et al., 2006; Zhang & Jiang, 2008), and indirect adaptive control identifies plant parameters before updating a controller (Åström & Wittenmark, 2008; Ljung, 1999). ECA estimates a full input matrix around a frozen learned policy, and its rejection test compares consequence predictions of the estimated map and of zero-parameter alternatives rather than assigning a fault label.

**Multiple models and switched adaptive control.** Selecting among a bank of candidate models by their recent prediction error, with hysteresis or a dwell time to prevent chattering, is the core of multiple-model adaptive control and of supervisory switching control (Narendra & Balakrishnan, 1997; Morse, 1996; Anderson et al., 2001; Hespanha et al., 2003). ECA's hypothesis library and selection rule are an instance of that idea, and we claim no new switching theory. What differs is

what may be designed: the controller is a frozen learned policy that cannot be redesigned or matched to a candidate plant, the reference is a learned nominal model rather than an exact plant family, and two candidates have no parameters, so "do not compensate" competes on the same footing. Where that literature bounds switching cost by dwell-time or hysteresis arguments, the unit amplification bound limits the damage of a wrong selection here, and we verify that empirically.

**Robust policies, residual compensation, and efference copies.** Actuation models matter for transfer from simulation to hardware (Tan et al., 2018; Hwangbo et al., 2019); domain randomisation (Tobin et al., 2017; Peng et al., 2018), learned adaptation modules (Kumar et al., 2021; Liang et al., 2024), policies conditioned on identified parameters (Yu et al., 2017; Chen et al., 2018) and behaviour repertoires searched after damage (Bongard et al., 2006; Cully et al., 2015) all address a changed body. Each needs the variation represented in training, or reward-guided trials. ECA instead uses a nominal model after training, without reward. Disturbance observers (Ohnishi et al., 1996), $\mathcal{L}_1$ augmentation (Hovakimyan & Cao, 2010; Cheng et al., 2022) and residual policies (Silver et al., 2018; Johannink et al., 2019) correct an existing controller additively; the closest recent instance learns a cerebellum-inspired additive residual online around a frozen policy (Jayasinghe et al., 2026). ECA instead estimates the actuator map itself and inverts it, so the correction is multiplicative and bounded rather than learned, and our analysis covers only the implementations we compare. The prediction signal is inspired by forward models and efference copies in motor control (Kawato et al., 1987; Wolpert & Kawato, 1998; Manoonpong & Wörgötter, 2009), where altered limb dynamics and visuomotor rotations are standard probes (Shadmehr & Mussa-Ivaldi, 1994; Krakauer et al., 2000); ECA uses it to estimate an actuator map, not an additive correction.

## 3 Problem and local analysis

### 3.1 Adaptation to an input map

A frozen policy produces $a = \pi(s) \in [-a_{\max}, a_{\max}]^k$ from observation $s \in \mathbb{R}^n$. Let $y = \Pi(s' - s) \in \mathbb{R}^m$ denote the observed consequence on selected actuated coordinates. The nominal response is $P(s, a)$. After a fault, a compensator issues $u$ and observes

$$y = P(s, Mu), \qquad M \in \mathbb{R}^{k\times k} \text{ fixed, invertible, and unknown.} \tag{1}$$

Neither $M$ nor a fault label is included in $s$. The ideal command $M^{-1}a$ reproduces the nominal response when it lies within the actuator range; gain loss belongs to this class, but saturation can prevent exact recovery. The healthy body is the identity case $M = I$. We also consider the affine extension $y = P(s, Mu + K\dot{q})$, where $\dot{q}$ are the actuated joint velocities read from $s$ and $K$ is a fixed coupling matrix: it contains damping changes and torques proportional to the velocity of another joint, and the static map is its case $K = 0$; the ideal command is then $M^{-1}(a - K\dot{q})$. Delays admit neither representation.

### 3.2 Feedback, additive correction, and identification

Consider a fixed local linearisation with nominal dynamics $x^+ = (A + BK)x$, policy gain $K$, and $\Delta = M - I$. Proofs and assumptions are detailed in Appendix A.

**Proposition 1** (Frozen feedback)**.** *Assume $A + BK$ is Schur stable. Define $G(z) = K(zI - A - BK)^{-1}B$ and $\gamma = \|G\|_\infty$. The faulted linear loop is stable if $\|\Delta\|_2\gamma < 1$. For a rotation by $\theta \in [-\pi, \pi]$ in one actuator pair, $\|\Delta\|_2 = 2|\sin(\theta/2)|$.*

The bound is a sufficient local margin, and we do not estimate $\gamma$ for the neural policies; their loss of return near 60–90° shows limited tolerance without measuring it.

**Proposition 2** (Ideal additive update)**.** *Fix $s$ and $a$, assume an exact affine consequence model with a full column rank input Jacobian $J$, and ignore clipping and leakage. An update $d^+ = d - \lambda J^+[P(s, M(a + d)) - P(s, a)]$ obeys*

$$d^+ = (I - \lambda M)d - \lambda(M - I)a. \tag{2}$$

*It converges to $d^\star = M^{-1}a - a$ for some $\lambda > 0$ if and only if every eigenvalue of $M$ has positive real part.*

For a rotation, this condition is $\cos\theta > 0$; channel swaps and cyclic permutations violate it. Filtering, leakage, changing actions, and nonlinear dynamics modify the iteration, so the proposition is a diagnostic for our additive baselines, not an impossibility result.

**Proposition 3** (Ideal identification)**.** *Assume an exact affine consequence model with a full column rank input Jacobian. In the noiseless case, the undamped command residual is $\delta = \Delta u$. If the command Gram matrix is full rank, least squares identifies $\Delta$. With persistent excitation, recursive least squares converges to this map. Applying $M^{-1}a$ then restores the nominal response wherever actuator limits permit.*

Identification has no eigenvalue half-plane restriction; its accuracy instead depends on excitation and the nominal model. With noise, forgetting, capped innovations, and a gate, Proposition 3 is an ideal reference rather than a convergence guarantee for the full algorithm.

## 4 Efference-Copy Adaptation

ECA uses the nominal model, recursive estimator, and regularised inverse of the identification core (ID). It adds a rejection gate; a unit amplification bound with a hold on uninformative data; and a velocity-coupled hypothesis selected against the static map (Figure 2 and Algorithm 1). The contribution is selective application of the core, not a new identification rule. ECA-m retains only the map hypothesis; an earlier closed-start gate is reported in Appendix C.

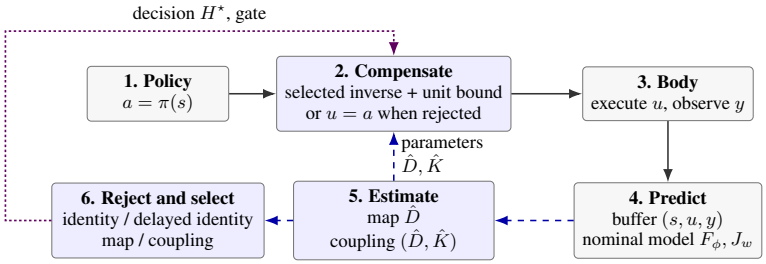



Figure 2: ECA data flow. Solid arrows carry the command being executed now; dashed arrows carry what the compensator learns from it into the next control step.

### 4.1 Nominal model and hypothesis estimation

A model $F_\phi(s, u)$ predicts nominal consequences. We fit it on healthy transitions from an earlier cohort of policies and freeze its parameters. Residual mean $\mu$ and coordinate scales $\sigma$ are calibrated on healthy deterministic rollouts. One model serves all evaluated policies for each body.

For an issued command $u$, define the whitened residual and Jacobian by $r = (y - F_\phi(s, u) - \mu)/\sigma$ and $J_w = \mathrm{diag}(\sigma)^{-1}\partial F_\phi(s, u)/\partial u$. ECA computes

$$\delta = (J_w^\top J_w + \rho_J I)^{-1} J_w^\top r, \qquad \rho_J = 0.01\|J_w\|_F^2/k + 10^{-6}. \tag{3}$$

The approximation $\delta \approx (M - I)u + K\dot{q}$ motivates two hypotheses estimated from the same residual. The *map hypothesis* regresses $\delta$ on $\varphi = u$ and gives $\hat{D} = I + \hat{E}$. The *coupling hypothesis* regresses $\delta$ on the stacked regressor $\varphi = (u, \dot{q})$ and gives $(\hat{E}, \hat{K})$, that is, the executed command $\hat{D}u + \hat{K}\dot{q}$. Each hypothesis keeps its own recursive least-squares estimate $\hat{E}_\varphi$ with covariance $C$, initialised to $10I$ and $\hat{E}_\varphi = 0$, forgetting factor $\beta = 0.995$, an innovation capped in norm at $a_{\max}/2$, and a surprise weight $g \in [0, 1)$ that suppresses the innovation when the intended command already explains the response (eq. (6), Appendix B); $\mathrm{cap}_b(v) = v\min(1, b/\max(\|v\|_2, 10^{-8}))$ throughout. Both updates are skipped while the validation buffer below is uninformative, since regressing on such data lets the estimates drift on model noise.

### 4.2 Rejection gate and bounded compensation

ECA stores the most recent $w = 20$ transitions $(s, u, y)$ in a buffer $\mathcal{B}$, with the commands issued in the three preceding steps. Let $v_H$ be the executed command predicted by hypothesis $H$: $v_D = \hat{D}u$

and $v_K = \hat{D}u + \hat{K}\dot{q}$. ECA scores each hypothesis against the identity and against a delayed identity with lag $L \in \{1, 2, 3\}$:

$$e_0 = \epsilon(u), \qquad e_H = \epsilon(v_H), \qquad e^{(L)} = \epsilon(u_{t-L}),$$
$$\epsilon(v) = \frac{1}{|\mathcal{B}|} \sum_{(s,u,y)\in\mathcal{B}} \left\| \frac{y - F_\phi(s, v) - \mu}{\sigma} \right\|_2^2, \tag{4}$$

where $u_{t-L}$ is the command issued $L$ steps before the transition. The ratio $q_H = e_H/b$, with $b = \min(e_0, e^{(1)}, e^{(2)}, e^{(3)})$ floored at $10^{-9}$, compares hypothesis $H$ with the best zero-parameter alternative. A buffer is *informative* when $e_0 > 10k$: the identity fails by a margin that model noise does not produce on the healthy body.

Each hypothesis has a gate $h_H$, revised only on informative buffers. If an open gate has $q_H > 0.5$, compensation is vetoed ($h_H \leftarrow 0$); if a closed gate has $q_H < 0.25$, it is restored ($h_H \leftarrow 1$). Otherwise the state is unchanged. The map gate starts open and decides from five samples. The coupling gate starts closed and decides from ten samples because a term in $\dot{q}$ can fit many disturbances over a short window.

This design has three immediate consequences. Before the first decision, ECA issues exactly the ID command: the gate can remove compensation but cannot delay its initial use. An earlier closed-start gate first had to be convinced and often failed to open on the seven-actuator arm (Section 5.3). A delay of $L$ steps is explained exactly by the $L$-lagged identity, so the veto fires once the buffer fills. On the healthy body, $\hat{E} \approx 0$ makes an open map gate reproduce the policy command; the identifiers are held because the buffer is nearly always uninformative.

Let $H^\star$ denote the acting hypothesis (Section 4.3). With $h_{H^\star} = 0$, ECA issues $u = a$. With $h_{H^\star} = 1$, it first removes the coupling term, $\bar{a} = a - \mathrm{cap}_{2a_{\max}}(\hat{K}\dot{q})$ under the coupling hypothesis and $\bar{a} = a$ under the map, then computes the regularised inverse

$$\tilde{u} = (\hat{D}^\top \hat{D} + 0.01I)^{-1}(\hat{D}^\top \bar{a} + 0.01\bar{a}), \tag{5}$$

clips $\tilde{u} - \bar{a}$ to $\pm a_{\max}$, adds the result to $\bar{a}$, clips to the actuator range, and finally bounds the command norm by $\|\bar{a}\|_2$. The unit bound makes ECA redirect rather than amplify: it preserves the norm of $\bar{a}$, which equals the policy's command whenever the coupling term is inactive. It is exact for every orthogonal map, and so for the rotations, permutations and mixes studied here. For a gain loss it forgoes the amplifying part of the inverse, a loss the frozen policies already tolerate. The earlier gate's factor-two bound and Gaussian probing are removed, since with the open start the policy's own commands excite the map (Section 5.3).

### 4.3 Hypothesis selection

The map hypothesis acts by default. On every informative buffer of at least ten samples, ECA compares the coupling hypothesis with the best of the map hypothesis and the zero-parameter alternatives, $\varsigma = e_K / \min(e_D, b)$; the coupling hypothesis takes over when $\varsigma < 0.5$ and hands back when $\varsigma > 1$. Comparing with the alternatives, not with the map alone, keeps it from taking over a window on which neither hypothesis beats the identity (Section 5.3). On a genuine input map $\hat{K} \to 0$ and the two hypotheses issue the same command. On a velocity-coupled disturbance the map hypothesis fits a contracting map, whose inverse would amplify and which the bound therefore withholds, whereas the coupling hypothesis cancels the term additively and inverts only the remaining map. The library is deliberately small: it holds the two hypotheses of Section 3, and the delay, which neither covers, is handled by rejection instead of a third parametric hypothesis.

## 5 Experiments

### 5.1 Evaluation protocol

We use Reacher-v5 ($k = 2$, 250 steps) and Pusher-v5 ($k = 7$, 200 steps) from Gymnasium/MuJoCo (Todorov et al., 2012; Towers et al., 2024), with SAC policies (Haarnoja et al., 2018; Raffin et al., 2021) trained on healthy dynamics. Five policies form the development split, on which

every method choice was made. Three further cohorts of five were trained for held-out evaluation, one per stage of the method, each after the corresponding setting had been frozen and each evaluated once against criteria recorded before its training; we call them the first, second and third cohort, the third being ECA's. The first cohort afterwards carried the fresh-episode control grid, on which the later gates are additional paired evaluations, not held-out tests. Input-map faults are rotations of one actuator pair, a cyclic permutation, a fixed orthogonal mix, and gain $0.5$ on half the channels. The *preservation set* is reported separately: the healthy body, gain loss, a velocity field (a torque on each actuator proportional to a neighbouring joint's velocity, outside the static map class but inside the velocity-coupled one), and a two-step actuator delay. Gain loss and the field sit there because the recorded criteria treat them as conditions a compensator must not harm, so the two sets are not a partition by fault class. All perturbations start at the first step. Appendices B and D give implementation details and the seeds of every split.

**Comparators and ablations.** We compare ECA with the frozen policy, a disturbance observer (DOB), a specific $\mathcal{L}_1$ implementation, a direct additive residual update (Add.), the shared identification and inversion core (ID), which applies its estimated inverse at every step, and Bound ID, the same inverse under ECA's unit bound at every step with no gate, hold or second hypothesis. ECA-m, ECA with the map hypothesis alone, is the one ablation carried through the main results; the earlier closed-start gate is reported in Appendices C and E, which also add a clipped inverse of the known $M$ as a privileged diagnostic and report reuse of an estimate across policies. All learned compensators share the nominal model and the evaluation episodes, so the comparison with ID measures gate, bound, hold and library together, and the comparison with Bound ID isolates gate, hold and library.

**Statistics.** Our primary metric is return over the first $100$ steps, the window on which adaptation acts, with five episodes per policy; we take paired differences against the frozen policy, average within policy, and bootstrap the five policy means. A star marks a $95\%$ interval excluding zero. With five clusters these intervals are approximate and uncorrected for the many cells, so the claims rest on replication across the development split, fresh episodes and the fresh cohort rather than on any single interval, and complete episodes are checked in Appendix E. As a check independent of the bootstrap, ECA is ahead of ID in $62$ of the $70$ policy-condition pairs of the input-map class on the development split, and again in $62$ of $70$ on the held-out cohort.

### 5.2 Recovery and interference with existing feedback

Table 1 sets the two objectives against each other: recovery from input-map faults, and leaving working feedback alone. ECA matches or exceeds ID in every cell of the recovery block, where the recorded criteria place gain loss with the preservation set: higher in twelve of the fourteen, significantly in six, and never behind by more than $0.6$ return with an interval containing zero. Two contrasts carry the section. The first is the pair of conditions on which ungated identification fails outright, the seven-actuator permutation and the orthogonal mix ($+42.7$ and $+65.1$ against $-24.0$ and $-2.8$): its inverse of a map identified only in the excited directions amplifies the command along the others, and the unit bound removes that amplification while keeping the redirection. The second is in the summary rows. Over the eight input maps ECA gains $19.7$ and $54.8$ against ID's $15.1$ and $36.5$, while Bound ID, the same inverse under the same bound but always acting, gains $19.7$ and $58.8$. The bound therefore accounts for the whole difference from ID, and never withdrawing buys four more return on Pusher, significantly on three of its cells. That is not free. Bound ID's worst preservation loss is $1.8$ and $29.8$ against ECA's $0.0$ and $1.1$, and it fails the recorded criterion on both plants, as does every other comparator in the table that recovers input maps. Under gain loss, the one input map whose inverse asks for amplification, ID is ahead by $1.0$ on Pusher. ECA-m stays within $0.4$ return of ECA on every input-map cell, as the selection rule intends, and the additive comparators recover no rotation beyond $60^\circ$, consistent with Proposition 2.

On the preservation set, ECA is indistinguishable from the frozen policy on the healthy body and under gain loss, and its delay losses are $0.0$ and $1.1$ return where ID loses $44.6$ and $113.2$. The velocity field separates the library from the map-only gate. ECA-m loses $4.3$ return on Pusher: over the first twenty to thirty steps a static, contracting map explains the velocity-coupled torque better than the identity, so its gate opens, but inverting a contracting map calls for amplification that the unit bound withholds. ECA gains $15.9$ instead, matching the direct additive update and ahead of ID

Table 1: Recovery and interference on development policies, first 100 steps. Frozen gives absolute return and the other columns paired differences. * marks a 95% policy-bootstrap interval excluding zero, and bold the largest positive mean on a recovery condition. The summary rows give the mean over the eight input maps, the largest mean loss on the four preservation conditions, and the recorded rule. Later tables use this notation.

| Plant | Condition | Frozen | Add. | DOB | ID | Bound ID | ECA-m | **ECA** |
|---|---|---|---|---|---|---|---|---|
| Reacher | healthy | -3.4 | -0.4* | +0.0 | +0.0 | +0.0 | +0.0 | +0.0 |
| | gain 0.5 | -3.6 | +0.1 | +0.0 | +0.1* | +0.0 | +0.0 | +0.0 |
| | velocity field | -199.0 | +1.3 | +0.2 | -0.6 | -1.8* | +0.0 | +0.2 |
| | delay 2 | -3.9 | -52.3* | -0.8* | -44.6* | -0.2* | +0.0 | +0.0 |
| | rot 60° | -6.9 | +2.3* | -3.3* | +2.9* | **+3.0*** | +3.0* | +3.0* |
| | rot 90° | -26.9 | +12.3* | -19.8* | +21.2* | **+22.3*** | +22.2* | +22.2* |
| | rot 120° | -31.8 | -72.4* | -56.5* | +20.7* | **+24.6*** | +24.4* | +24.4* |
| | rot 150° | -45.1 | -75.5* | -56.2* | +10.9* | +27.9* | **+28.1*** | +27.7* |
| | rot 180° | -68.6 | -69.8* | -49.9* | +38.0* | +51.7* | +52.2* | **+52.2*** |
| | permutation | -14.9 | -41.0* | -22.9* | +7.7* | +8.5* | +8.6* | **+8.7*** |
| | orthogonal mix | -23.8 | +17.6* | -12.0* | +19.1* | **+19.5*** | +19.4* | +19.4* |
| | *input-map mean* | | -28.3* | -27.5* | +15.1* | +19.7* | +19.7* | +19.7* |
| | *worst loss* | | 52.3 | 0.8 | 44.6 | 1.8 | 0.0 | 0.0 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Pass | Pass |
| Pusher | healthy | -31.2 | -0.3 | -0.3 | -1.8* | -0.5* | +0.0 | -0.1 |
| | gain 0.5 | -33.4 | +0.5* | +0.4 | +1.3* | -0.7 | +0.3* | +0.3* |
| | velocity field | -52.7 | +16.5* | +7.4* | +12.5* | -29.8* | -4.3* | +15.9* |
| | delay 2 | -37.5 | -40.6* | -5.0* | -113.2* | -26.0* | -1.1* | -1.1* |
| | rot 60° | -41.3 | +5.9* | +5.1* | +7.0* | +7.9* | **+8.1*** | +7.9* |
| | rot 90° | -94.6 | +43.1* | +2.9 | +59.1* | **+61.3*** | +60.8* | +60.9* |
| | rot 120° | -141.5 | -32.5* | -9.1* | +100.3* | **+103.8*** | +99.3* | +99.7* |
| | rot 150° | -146.2 | -79.3* | -22.4* | +90.5* | **+105.8*** | +101.5* | +101.4* |
| | rot 180° | -123.9 | -96.2* | -27.9* | +60.6* | **+76.8*** | +60.3* | +60.2* |
| | permutation | -99.1 | -110.3* | -22.1* | -24.0* | **+48.7*** | +42.9* | +42.7* |
| | orthogonal mix | -113.0 | -54.9* | -19.0* | -2.8 | **+67.2*** | +65.1* | +65.1* |
| | *input-map mean* | | -40.4* | -11.5* | +36.5* | +58.8* | +54.8* | +54.8* |
| | *worst loss* | | 40.6 | 5.0 | 113.2 | 29.8 | 4.3 | 1.1 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Fail | Pass |

and DOB, because the coupling hypothesis wins the selection on that window, cancels the term and inverts the remaining near-identity map. On Reacher the field exceeds the actuators' authority, the frozen return falling to $-199$, and no comparator recovers more than 6 of the 196 lost return. The core thus supplies the recovery; ECA's additions keep it, add more where the core amplifies its own estimation error, remove the delay losses, and extend it to the velocity-coupled class.

**Is the gate only a weaker inverse?** A weakened inverse would buy the same preservation more cheaply. Appendix G places a full grid of such settings on the recovery–interference plane: partial and ridge inverses, norm bounds, and a forced-open control, all sharing ECA's estimator and nominal model, run on fresh episodes of the first cohort. No setting attains ECA's recovery with less interference on either plant. Bound ID ($b = 1$) is the strongest of them, reaching ECA's recovery with overlapping intervals, and its interference is what the gate and the library remove: 36.5 return of preservation loss on Pusher and 24.9 on the additional disturbances, against ECA's 1.8 and 0.3. No setting on that plant with a smaller loss than ECA's recovers more than 4.1, so the gate costs four return there and none on Reacher. Over complete episodes the same trade reads 145.0 against 131.6 on the Pusher input maps, with 49.2 against 0.1 return of delay loss (Appendix E): the gate gives up about a tenth of the recovery and removes almost all of the interference.

### 5.3 Component ablations and failure analysis

Table 2 removes each addition in turn; they address different failures, and the gate's design changes the recovery side, not only the interference side. The largest single effect is where compensation starts: the earlier gate, waiting to be convinced, recovers $+16.9$ on Pusher against $+54.8$, and

Table 2: Component ablations on development policies. Gain is the mean over the eight input maps. Field is the velocity field. Worst is the largest preservation loss with the condition that produces it. Rule is the recorded criterion. Indented rows change one component of ECA-m. Notation as in Table 1. Table 6 in Appendix C gives the full grid with bootstrap intervals.

| Configuration | Reacher gain | Field | Worst | Rule | Pusher gain | Field | Worst | Rule |
|---|---|---|---|---|---|---|---|---|
| ID (ungated) | +15.1* | -0.6 | 44.6 delay | Fail | +36.5* | +12.5* | 113.2 delay | Fail |
| closed-start gate | +13.2* | +0.0 | 0.3 gain | Pass | +16.9* | -3.6 | 3.6 field | Pass |
| ECA-m: map hypothesis only | +19.7* | +0.0 | 0.0 delay | Pass | +54.8* | -4.3* | 4.3 field | Fail |
| − open start | +18.7* | +0.0 | 0.0 healthy | Pass | +43.2* | -3.2 | 3.2 field | Pass |
| − lagged hypotheses | +19.7* | -0.1 | 0.1 field | Pass | +55.9* | -5.2* | 5.2 field | Fail |
| bound $b = 2$ | +17.8* | +0.0 | 0.8 delay | Pass | +36.8* | +6.5* | 4.0 delay | Fail |
| coupling hypothesis alone | -0.6 | +2.6* | 1.5 delay | Fail | +22.2* | +19.1* | 3.6 delay | Fail |
| **ECA** | +19.7* | +0.2 | 0.0 delay | Pass | +54.8* | +15.9* | 1.1 delay | Pass |

restoring only its closed start inside ECA-m leaves $+43.2$; Appendix C traces the self-locking failure behind this.

**Delayed-identity hypotheses.** Without them the gate compares the map with the identity alone. A two-step delay violates the identity, and the map fitted to the delayed response is briefly the better explanation, so the gate lets it act: the loss is $3.9$ on Pusher against $1.1$ with the lagged hypotheses, and they cost nothing on the input-map class, since a genuine map is never explained better by a lagged command.

**Unit bound.** The bound is what changes the comparison with ID on the input-map class. With $b = 2$ the Pusher gain falls from $+54.8$ to $+36.8$ and the permutation reverts to a loss ($-6.6$ against $+42.9$). A looser bound also lets the map-only gate act on the field ($+6.5$ at $b = 2$ against $-4.3$ at $b = 1$): the field is undone by amplification and the permutation by redirection, and one bound cannot serve both. Appendix C lists the further gate hypotheses we tried on ECA-m to reject the field; the four that bring its loss below 2 return give up between 2 and 32 of the Pusher input-map gain, and none removes it outright. The coupling hypothesis resolves the conflict without touching the bound.

**Coupling hypothesis and selection.** Alone, the coupling hypothesis is not a compensator for the input-map class: with a term in $\dot{q}$ absorbing part of what the map should explain, it gains 22.2 on Pusher and nothing on Reacher, and fails the preservation rule on both plants. Its value is as a second member selected by evidence. Opening it at the start, deciding from five samples, or selecting it against the map alone each leave the map results unchanged and the Pusher field at $+17.9$, but cost up to $0.9$ on the Reacher field and $2.2$ on the delays, where a velocity term briefly fits a window on which nothing beats the identity. The reported setting removes those losses for 2 return of the Pusher field; twenty samples reduce that gain to $+11.4$ and a covariance prior of $I$ to $+16.3$.

### 5.4 Boundary of the result

Three questions sit outside the protocol above; Appendix F answers each in full. Retraining the policy on the fault class recovers part of the loss (eight-map mean $-27.7$ to $-16.1$ on Reacher, $-99.1$ to $-73.7$ on Pusher) but pays for it on the healthy body, under delay and under gain loss, and still trails the healthy-trained policy with ECA ($-8.0$ and $-44.3$); ECA on the randomised policy gains a further $3.3$ and $20.7$, so the two compose. Node perturbation on a command matrix uses reward in place of a model. It needs a median of $5$ to $23$ episodes on Reacher and $3$ to $82$ on Pusher to reach the return of ECA's single episode, and after $120$ episodes ends $32.9$ return below the frozen policy under delay, because it cannot express not acting. A fault switched on at step $30$ is recovered as well as one present from reset: after the switch ECA recovers $110.2$ of the $121.4$ return lost at Reacher $120°$, and about three quarters of the loss on the Pusher permutation and mix, where the ungated core recovers a third and a fifth.

Table 3: Held-out evaluation of ECA on the third cohort, run once. Notation as in Table 2. The last three columns are paired gains over the frozen policy at $90^\circ$, at $120^\circ$, and over the eight input maps. The ECA−ID rows give the paired difference between ECA and the ungated core, with policy-bootstrap intervals. Table 13 lists every condition and comparator.

| Plant | Method | Worst loss | Rule | $90^\circ$ | $120^\circ$ | Maps |
|---|---|---|---|---|---|---|
| Reacher | ID | 56.4 (delay 2) | Fail | +29.8* | +37.8* | +19.2* |
| | ECA-m | 0.0 (delay 2) | Pass | +31.7* | +44.3* | +26.6* |
| | **ECA** | 0.0 (delay 2) | Pass | +31.7* | +44.3* | +26.6* |
| | ECA−ID | | | +1.8 [+1.3, +2.4] | +6.5 [+3.4, +9.6] | +7.4 [+5.0, +9.5] |
| Pusher | ID | 116.6 (delay 2) | Fail | +75.2* | +98.4* | +41.4* |
| | ECA-m | 10.1 (velocity field) | Fail | +75.9* | +104.8* | +54.8* |
| | **ECA** | 1.2 (gain 0.5) | Fail | +75.9* | +104.7* | +54.7* |
| | ECA−ID | | | +0.7 [-0.1, +1.7] | +6.3 [+2.1, +12.9] | +13.4 [+9.1, +17.4] |

### 5.5 Held-out validation

ECA was evaluated once on the third cohort, trained after its complete setting had been recorded, against the criteria of Appendix D: preservation on the four preservation conditions, and recovery at $90^\circ$ and $120^\circ$ significantly better than the frozen policy and not significantly worse than ID. Two earlier versions of the gate were tested the same way on their own cohorts, and each failure drove one revision made on the development split alone. The second is why the library exists: on the second cohort the map-only gate lost $8.3$ return on the Pusher velocity field (chronology in Appendix D, records in Appendix E).

ECA met the recovery criterion on both plants and was ahead of ID, not behind it, at $90^\circ$, at $120^\circ$ and over the eight maps (Table 3). The condition that had failed for ECA-m is recovered: on the same episodes the Pusher velocity field gains $10.3$ return where ECA-m loses $10.1$, the coupling hypothesis being selected on almost every step. Preservation held on Reacher and failed on Pusher, on a different condition. Under gain loss ECA costs $1.2$ return, an interval whose lower bound falls just outside the recorded budget of 3; the loss comes from one policy of five and, within it, from one episode, whose trace Appendix E follows. ECA-m is identical there, so the cost belongs to the map gate, not to the added hypothesis. Comparators exceed ECA on individual cells, the largest being Bound ID's on four Pusher maps, bought with $37.8$ return of delay loss and failed preservation on both plants (Appendix E).

## 6 Conclusion

In this paper, we proposed **Efference-Copy Adaptation (ECA)**, which lets a frozen policy keep what an identified inverse recovers while giving up the interference that inverse causes when the fault is not a map. The rule that achieves this is evidential rather than conservative: compensation acts under the best-supported hypothesis, is withdrawn as soon as a zero-parameter alternative explains the data better, and is never larger than the command it corrects. On two MuJoCo manipulators this keeps the recovery of ungated identification, adds recovery where the ungated inverse amplifies its own estimation error, and removes its losses under a delay and on a healthy body; on a cohort trained after the method was frozen, the recorded recovery criterion is met on both plants.

The scope is narrow in ways our experiments make precise. ECA targets fixed linear input maps and their velocity-coupled extension, so it does not recover changes in inertia, nonlinear torques or actuator dynamics, and it can only reject a delay. Its gate is a consistency test and not a certificate: on one held-out episode it kept acting for forty steps after its own evidence had turned. And the evidence itself covers two simulated manipulators and five policies per split, with a third, contact-rich plant on which preservation fails outright, because there the identity no longer explains the model's residual better than a map does.

In the future, we plan to make withdrawal react faster than the informative window used here, to grow the hypothesis library without retuning the gate, and to take the method to hardware, where the healthy-body check that separates our three plants becomes a prerequisite rather than a diagnostic.

## REPRODUCIBILITY STATEMENT

Algorithm 1 specifies the complete method. Appendices A and B give proofs and implementation settings. The accompanying experiment records retain policy, model, and script hashes. Appendix B gives the identifiers under which the methods appear in the released code and records.

## AI USE STATEMENT

Language models assisted with manuscript restructuring, language editing, code drafting, figure layout, and consistency checks between the manuscript and implementation. The original results come from the recorded experiments. The additional conservative-ID study was run from a fixed candidate grid and its simulation records are retained separately.

## REFERENCES

Brian D. O. Anderson, Thomas S. Brinsmead, Daniel Liberzon, and A. Stephen Morse. Multiple model adaptive control with safe switching. *International Journal of Adaptive Control and Signal Processing*, 15(5):445–470, 2001.

Karl J Åström and Björn Wittenmark. *Adaptive Control*. Dover, 2nd edition, 2008.

Mogens Blanke, Michel Kinnaert, Jan Lunze, and Marcel Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer, 2 edition, 2006.

Josh Bongard, Victor Zykov, and Hod Lipson. Resilient machines through continuous self-modeling. *Science*, 314(5802):1118–1121, 2006.

Tao Chen, Adithyavairavan Murali, and Abhinav Gupta. Hardware conditioned policies for multi-robot transfer learning. In *Advances in Neural Information Processing Systems*, 2018.

Yikun Cheng, Pan Zhao, Fanxin Wang, Daniel J. Block, and Naira Hovakimyan. Improving the robustness of reinforcement learning policies with $\mathcal{L}_1$ adaptive control. *IEEE Robotics and Automation Letters*, 7(3):6574–6581, 2022. doi: 10.1109/LRA.2022.3169309. URL `https://arxiv.org/abs/2112.01953`.

Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. Robots that can adapt like animals. *Nature*, 521:503–507, 2015.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International Conference on Machine Learning (ICML)*, 2018.

João P. Hespanha, Daniel Liberzon, and A. Stephen Morse. Hysteresis-based switching algorithms for supervisory control of uncertain systems. *Automatica*, 39(2):263–272, 2003.

Naira Hovakimyan and Chengyu Cao. *$\mathcal{L}_1$ Adaptive Control Theory: Guaranteed Robustness with Fast Adaptation*. SIAM, 2010.

Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Vassilios Tsounis, Vladlen Koltun, and Marco Hutter. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26), 2019.

Nethmi Jayasinghe, Diana Gontero, Spencer T. Brown, Vinod K Sangwan, Mark C Hersam, and Amit Ranjan Trivedi. Cerebellar-inspired residual control for fault recovery: From inference-time adaptation to structural consolidation. *arXiv preprint arXiv:2602.07227*, 2026.

Tobias Johannink, Shikhar Bahl, Ashvin Nair, Jianlan Luo, Avinash Kumar, Matthias Loskyll, Juan Aparicio Ojea, Eugen Solowjow, and Sergey Levine. Residual reinforcement learning for robot control. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

Mitsuo Kawato, Kazunori Furukawa, and R Suzuki. A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, 57(3):169–185, 1987.

John W. Krakauer, Zachary M. Pine, Maria-Felice Ghilardi, and Claude Ghez. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, 20(23):8916–8924, 2000.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In *Robotics: Science and Systems*, 2021.

Yichao Liang, Kevin Ellis, and João Henriques. Rapid motor adaptation for robotic manipulator arms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, 2 edition, 1999.

Poramate Manoonpong and Florentin Wörgötter. Efference copies in neural control of dynamic biped walking. *Robotics and Autonomous Systems*, 57(11):1140–1153, 2009.

A. Stephen Morse. Supervisory control of families of linear set-point controllers—part i: Exact matching. *IEEE Transactions on Automatic Control*, 41(10):1413–1431, 1996.

Kumpati S. Narendra and Jeyendran Balakrishnan. Adaptive control using multiple models. *IEEE Transactions on Automatic Control*, 42(2):171–187, 1997.

Kouhei Ohnishi, Masaaki Shibata, and Toshiyuki Murakami. Motion control for advanced mechatronics. *IEEE/ASME Transactions on Mechatronics*, 1(1):56–67, 1996.

Abby O'Neill et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *IEEE International Conference on Robotics and Automation*, 2024.

Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

Reza Shadmehr and Ferdinando A Mussa-Ivaldi. Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, 14(5):3208–3224, 1994.

Tom Silver, Kelsey Allen, Josh Tenenbaum, and Leslie Kaelbling. Residual policy learning. *arXiv preprint arXiv:1812.06298*, 2018.

Jie Tan, Tingnan Zhang, Erwin Coumans, Atil Iscen, Yunfei Bai, Danijar Hafner, Steven Bohez, and Vincent Vanhoucke. Sim-to-real: Learning agile locomotion for quadruped robots. In *Robotics: Science and Systems*, 2018.

Gang Tao, Shuhao Chen, Xidong Tang, and Suresh M. Joshi. *Adaptive Control of Systems with Actuator Failures*. Springer, 2004.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

Mark Towers, Ariel Kwiatkowski, Jordan Terry, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.

Daniel M Wolpert and Mitsuo Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8):1317–1329, 1998.

Wenhao Yu, Jie Tan, C. Karen Liu, and Greg Turk. Preparing for the unknown: Learning a universal policy with online system identification. In *Robotics: Science and Systems*, 2017.

Youmin Zhang and Jin Jiang. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control*, 32(2):229–252, 2008.

Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *IEEE Symposium Series on Computational Intelligence*, 2020.

Kemin Zhou and John C. Doyle. *Essentials of Robust Control*. Prentice Hall, 1998.

**Algorithm 1** ECA: online adaptation of a frozen policy

**Require:** frozen policy $\pi$, nominal model $F_\phi$, action limit $a_{\max}$, hypotheses $H \in \{D, K\}$
both estimates start at the identity; the map hypothesis acts, the coupled one waits
**for** each control step **do**
  $a \leftarrow \pi(s)$ ▷ the policy's command, never modified
  **if** the acting hypothesis is open **then**
    $\bar{a} \leftarrow a - \mathrm{cap}(\hat{K}\dot{q})$ under the coupled hypothesis, $\bar{a} \leftarrow a$ under the map
    $u \leftarrow$ regularised inverse of the acting estimate applied to $\bar{a}$, of norm at most $\|\bar{a}\|_2$
  **else**
    $u \leftarrow a$ ▷ withdrawn: issue the policy's own command
  **end if**
  execute $u$ and form the command-space residual $\delta$ from the observed consequence (eq. 3)
  **if** the window is informative **then** update both estimates by recursive least squares (eq. 6)
  score both hypotheses, the identity and the delayed identities on the last $w$ transitions (eq. 4)
  open or close each gate by its ratio, with hysteresis, and let the better-supported hypothesis act
**end for**

# A PROOFS AND SCOPE OF THE LOCAL RESULTS

**Frozen feedback (Proposition 1).** Write the faulted loop as $x^+ = (A + BK)x + Bv$ with $v = \Delta Kx$. Schur stability of $A + BK$ makes the transfer $G$ stable. The small-gain theorem (Zhou & Doyle, 1998) gives stability when $\|\Delta\|_2\|G\|_\infty < 1$. For a rotation, $(R_\theta - I)^\top(R_\theta - I) = 4\sin^2(\theta/2)I$ on the rotated pair. Thus $\|\Delta\|_2 = 2|\sin(\theta/2)|$. Other channels contribute zero singular values. This proves the stated sufficient bound. It does not certify a time-varying linearisation or the full nonlinear trajectory. □

**Additive correction (Proposition 2).** Under the affine model and full column rank assumptions, $J^+J = I$. Substitution gives eq. (2), with fixed point $M^{-1}a - a$. Convergence from any initial correction requires $\rho(I - \lambda M) < 1$. For an eigenvalue $\nu_i = \alpha_i + i\beta_i$ of $M$,

$$|1 - \lambda\nu_i|^2 = 1 - 2\lambda\alpha_i + \lambda^2|\nu_i|^2 < 1 \quad \Longleftrightarrow \quad 0 < \lambda < 2\alpha_i/|\nu_i|^2.$$

A common positive gain exists exactly when every $\alpha_i > 0$. □

Leakage changes the spectrum and the target correction. For example, replacing the iteration matrix by $(1 - \kappa)I - \lambda M$ changes its stability region, so the half-plane condition cannot be applied unchanged. Clipping can prevent unbounded commands without restoring the desired fixed point. The two-timescale direct baseline and filtered control baselines are therefore empirical comparators, not exact instances of the proposition in all operating regimes. On the tested maps, their losses at large rotations and permutations agree with its qualitative prediction.

**Identification (Proposition 3).** The exact affine model gives $y - P(s, u) = J\Delta u$. An undamped left inverse yields $\delta = \Delta u$. For command matrix $U = [u_1, \ldots, u_T]$ and residual matrix $V = [\delta_1, \ldots, \delta_T]$, full row rank of $U$ implies $VU^\top(UU^\top)^{-1} = \Delta$. Recursive least squares solves the corresponding sequential regression (Ljung, 1999; Åström & Wittenmark, 2008). In the noiseless case, persistent excitation drives its initialisation error to zero. Exact compensation follows from $MM^{-1}a = a$. □

The result has no eigenvalue sign restriction on $M$. It still requires a full rank consequence Jacobian, informative commands, and adequate actuator authority. Finite-data error depends on the regressor Gram matrix and model error. With observation noise, fixed forgetting generally gives a tracking estimate with nonzero variance. Damping, surprise weighting, and innovation clipping can add bias. None of these practical effects is removed by invertibility alone.

## B IMPLEMENTATION AND EVALUATION SETTINGS

**Recursive estimate.** With regressor $\varphi$ ($\varphi = u$ for the map hypothesis, $\varphi = (u, \dot{q})$ for the coupling hypothesis), each transition updates

$$\begin{aligned} p &= C\varphi, & \ell &= p/(\beta + \varphi^\top p), \\ v &= \mathrm{cap}_{a_{\max}/2}\Big(g(\delta - \hat{E}_\varphi \varphi)\Big), & \hat{E}_\varphi^+ &= \hat{E}_\varphi + v\ell^\top, \\ C^+ &= \frac{C - pp^\top/(\beta + \varphi^\top p)}{\beta}. \end{aligned} \tag{6}$$

The surprise weight is $g = \max(0, 1 - 1/\max(\bar{m}, 10^{-6}))$, with $\bar{m} = \|(y - F_\phi(s,a) - \mu)/\sigma\|_2/\sqrt{m}$. The implementation symmetrises $C^+$ after each update.

**Nominal models.** Both models use healthy transitions from an earlier policy cohort, separate from every policy evaluated here. Both are small multilayer perceptrons of two hidden layers, of width $256$ on Reacher and $512$ on the higher-dimensional Pusher, fitted on $25{,}000$ and $120{,}000$ transitions collected with exploration noise and a fifth to a third of uniform commands, so that the model sees command directions a deterministic policy never visits. Each predicts the change in actuated joint velocity with a validation $R^2$ of about $0.97$ on policies it was not fitted on. On individual development policies the Pusher model explains distinctly less variance, between $0.58$ and $0.72$; that is the accuracy the compensator actually works against, and the reason a gate is needed rather than an unconditional inverse. The same frozen model file serves every comparator on a body, and no policy used for held-out evaluation contributed to fitting or selection. Algorithm 1 and eqs. (3)–(5) give the online parameters and operation order.

**Gate settings.** The gate has two thresholds and two counts, and each was set from a measured quantity rather than tuned for return. The opening and closing ratios, $0.25$ and $0.5$, sit in the gap between what a genuine map produces and what a body without one produces: on the development split the ratio was $0.05$–$0.2$ under rotations of at least $90^\circ$ and $0.8$–$1.1$ on healthy and delayed bodies. A threshold placed inside the second range does what one would expect: an earlier pair at $0.8/0.95$ left the delayed body compensated on $96\%$ of its steps, and was discarded. The informativeness floor is the same argument in the units of the residual, counting a window only when the identity's whitened error exceeds ten times its healthy level, which calibration fixes at $k$. Where in that range the line falls matters little, since $3k$ and $30k$ move no aggregate by more than one return unit (Table 7). Five buffered samples is the smallest count at which the test discriminated on development; three samples, or a window of ten instead of twenty, were not better. The lagged alternatives run one step beyond the longest delay tested, so that a delay is recognised rather than fitted. These settings were made on the development split and frozen before the second cohort was trained.

**Coupling hypothesis and selection.** The coupling hypothesis reads the joint-velocity coordinates from the observation (two on Reacher, seven on Pusher) and inherits the map hypothesis' innovation cap, surprise weight, forgetting factor, hold and thresholds, so the library adds three decisions rather than a second set of hyperparameters. The first is when it may act: its own gate starts closed and decides from ten samples rather than five, because it has twice as many parameters to fit and a velocity term will otherwise explain a short window on which nothing beats the identity. The second is how far it may reach: $\hat{K}\dot{q}$ is capped at $2a_{\max}$ before the inverse is taken, and the unit bound is applied to the corrected command $\bar{a}$, so the cap limits what the coupling term may claim while the bound still limits what is issued. The third is selection: on informative buffers of at least ten samples a ratio below $0.5$ hands the command to this hypothesis and a ratio above $1$ hands it back, the hysteresis preventing alternation at the boundary. Table 6 reports what each decision is worth by moving it: an initially open gate, five or twenty samples, a covariance prior of $I$ rather than $10I$, selection against the map alone, caps of $1a_{\max}$ and $3a_{\max}$, and thresholds $0.25/0.5$ and $0.5/2$. None of them changes the input-map results, so the choice was made on the preservation side, taking the setting with the smallest preservation losses among those within 2 return of the best field gain. A bound of $b = 2$ raises that gain further and was rejected, because one bound for both members is what makes the unit bound a property of the method rather than of a hypothesis. The setting and criteria were frozen before the third cohort was trained. Each control step evaluates one

Jacobian, two RLS updates, and seven batched model passes over the twenty buffered transitions. On one CPU thread, the full loop (policy, compensation, simulator, and identification) takes $8.9$ ms on Reacher and $13.9$ ms on Pusher. The frozen loop takes $0.8$ and $0.9$ ms, and ungated identification takes $2.4$ and $5.0$ ms. ECA therefore fits inside the $20$ ms and $50$ ms control periods with a factor of two to four in hand; these measurements include the simulator step that hardware would not charge to the compensator.

**DOB and $\mathcal{L}_1$ implementations.** Both use the same whitened consequence residual and damped Jacobian as the map identifier, with the same surprise weight. DOB filters the inferred command disturbance with gain $\alpha$ and subtracts it from the policy action. The $\mathcal{L}_1$ comparator maintains accumulated prediction error $\tilde{e}^+ = 0.5\tilde{e} + (J_w\hat{\sigma} - r)g$, updates $\hat{\sigma}^+ = \hat{\sigma} - (J_w^\top J_w + \rho_J I)^{-1} J_w^\top \tilde{e}^+$, and filters the correction with gain $\beta$. Corrections are clipped to $\pm 2a_{\max}$ before final action clipping. These are the implementations evaluated here, not a claim about every DOB or $\mathcal{L}_1$ design.

Filter gains $\alpha, \beta \in \{0.1, 0.3, 0.5, 0.8\}$ were compared on the development split. The selection rule maximised mean input-map gain summed over both bodies. It selected $\alpha = \beta = 0.1$. Higher gains sometimes improve moderate rotations but increase interference: DOB with $\alpha = 0.8$ gains $27.2$ return on Pusher at $90^\circ$ and loses $35.0$ under delay.

**Exact reference controls.** A healthy twin simulator is restored to the evaluated simulator's full MuJoCo integration state before each probe. It supplies the nominal consequence and a central-difference Jacobian with step $10^{-4}$. Probe transitions do not advance the evaluated episode. These privileged controls diagnose model error and are not deployable ECA components.

**Additive controls.** The direct additive comparator maintains fast and slow linear corrections, $d = (W_{\mathrm{f}} + W_{\mathrm{s}})a$. Its retention and learning rates are $(0.85, 0.5)$ and $(0.999, 0.15)$, respectively, with a surprise dead zone and correction clipping. The bounded version limits accumulation in the slow map. These controls use prediction error for the intended command, while ECA estimates a map from consequences of the issued command.

**Identifiers in the released code.** The released implementation and the stored records name the methods differently from this manuscript, because the names were fixed before the final algorithm existed. For reproduction: ECA is `lib4_c2_n10`, a `LibraryCompensator` whose map member is unchanged from ECA-m; ECA-m is `validated_v3`; the closed-start gate is `validated_v2`; ID is `id_inverse`, Bound ID is `amp_1`, and the direct additive control is `eca`, an identifier that predates the gate. The ablations of Table 6 appear under identifiers of the form `v3_`$\star$ (components of ECA-m) and `lib`$\star$, `aff_c2` (coupling hypothesis and selection). Renaming changed no recorded return and no implementation.

Table 4: Reference-model controls for ungated identification on development policies. Entries are paired first-100-step return differences; parentheses give final map error $\|\hat{D} - M\|_F$ (relative to $I$ for delay).

| Plant | Condition | Learned | Exact $J$ | Exact ref. + $J$ |
|---|---|---|---|---|
| Reacher | healthy | +0.0 (0.03) | +0.0 (0.03) | -0.0 (0.00) |
| | rot 120° | +20.7* (1.03) | +20.8* (1.01) | +20.8* (1.00) |
| | rot 180° | +38.0* (1.17) | +27.6 (1.32) | +16.2 (1.38) |
| | permutation | +7.7* (0.88) | +7.5* (0.89) | +7.7* (0.89) |
| | orthogonal mix | +19.1* (0.83) | +19.1* (0.83) | +19.1* (0.81) |
| | delay 2 | -44.6* (1.07) | -43.6* (1.14) | -42.0* (1.12) |
| Pusher | healthy | -1.8* (1.67) | -0.2 (1.25) | +0.0 (0.00) |
| | rot 120° | +100.3* (1.98) | +98.6* (1.74) | +102.5* (1.31) |
| | rot 180° | +60.6* (2.43) | +66.0* (2.15) | +69.0* (1.33) |
| | permutation | -24.0* (2.29) | +3.6 (2.09) | +21.5* (1.85) |
| | orthogonal mix | -2.8 (2.74) | +20.2* (2.08) | +23.1* (2.05) |
| | delay 2 | -113.2* (2.34) | -98.4* (2.68) | -96.7* (2.31) |

Table 5: Fraction of the first 100 steps on which the acting hypothesis is open, for ECA-m and ECA, on the development policies and on the third cohort. The closed-start gate, by construction, stays shut on the healthy body.

| | Development | | | | Held-out | | | |
|---|---|---|---|---|---|---|---|---|
| Condition | Reacher ECA-m | Reacher ECA | Pusher ECA-m | Pusher ECA | Reacher ECA-m | Reacher ECA | Pusher ECA-m | Pusher ECA |
| healthy | .83 | 1.00 | .95 | .96 | .86 | .93 | 1.00 | 1.00 |
| gain 0.5 | .95 | .96 | .94 | .94 | .88 | .97 | .96 | .96 |
| velocity field | .05 | .15 | .28 | .95 | .06 | .37 | .26 | .96 |
| delay 2 | .05 | .05 | .06 | .06 | .05 | .05 | .06 | .06 |
| rot 60° | .92 | .92 | .76 | .79 | .97 | .97 | 1.00 | 1.00 |
| rot 90° | .99 | .99 | .90 | .91 | 1.00 | 1.00 | .94 | .94 |
| rot 120° | .99 | .99 | .90 | .91 | 1.00 | 1.00 | 1.00 | 1.00 |
| rot 150° | .99 | .99 | .96 | .96 | 1.00 | 1.00 | .89 | .89 |
| rot 180° | .99 | .99 | .80 | .80 | .99 | .99 | .64 | .64 |
| permutation | .89 | .89 | .86 | .86 | .97 | .98 | .76 | .76 |
| orthogonal mix | .97 | .97 | .94 | .94 | 1.00 | 1.00 | .80 | .80 |

## C DIAGNOSTICS, TRACES AND ABLATIONS

**Reference accuracy and gate behaviour.** Table 4 isolates the effect of replacing the learned Jacobian and consequence reference. Large improvements for Pusher permutations and mixes support a model-error explanation. Exact references do not remove all large-angle losses, so reference accuracy alone is insufficient. These interventions do not uniquely distinguish weak excitation from clipping, trajectory changes, or estimator safeguards.

Deciding on every window, or updating the identifier on uninformative data, changes the map aggregates little but each adds about half a return unit of field loss to the map-only gate (Table 6). An exact nominal simulator (Appendix C) raises the ungated core's Pusher permutation and mix to $+21.5$ and $+23.1$, less than ECA recovers with the learned model, so the harm from the learned reference is mediated by amplification; even the exact model fails under delay.

Table 5 shows the gates' behaviour. ECA-m is open on the healthy body ($0.83$ and $0.95$), where with $\hat{E} \approx 0$ the open gate issues the policy's own command and nothing is lost; it is closed on the delay ($0.05$ and $0.06$) and the Reacher field ($0.05$), and open on $0.76$–$0.99$ of the map steps. Its Pusher field fraction of $0.28$ is the mis-fit of Section 5.3: the gate opens during the first twenty to thirty steps, while a contracting map explains the velocity-coupled torque, and closes once the fitted map stops explaining the data. ECA agrees with ECA-m on the maps and the delay and differs on the Pusher field, where the coupling hypothesis is selected and open on $0.95$ of the steps. On the healthy Reacher its fraction is $1.00$ rather than $0.83$: in a few episodes the coupling hypothesis, with a near-zero coupling term ($\|\hat{K}\dot{q}\| \leq 0.02$), passes its test on a window just above the informative threshold and is selected; the command it issues differs from the policy's by at most $0.04$ return per episode, which the unit bound guarantees to be small. The held-out cohort reproduces the pattern: ECA-m is open on $0.26$ of the Pusher field steps and $0.06$ of the Reacher ones, ECA on $0.96$ and

Table 6: Component ablations on development policies. Gain is the mean over the eight input maps, Field the velocity field, Worst the largest preservation loss with its condition, and Rule the recorded criterion. The second group removes one component of the map-only gate ECA-m; the third changes one setting of the coupling hypothesis or of the selection that turns ECA-m into ECA.

| Configuration | Reacher gain [CI] | Field | Worst | Rule | Pusher gain [CI] | Field | Worst | Rule |
|---|---|---|---|---|---|---|---|---|
| *Identification core and the earlier gate* | | | | | | | | |
| ID (ungated) | +15.1 [+10.0, +21.2] | -0.6 | 44.6 delay | Fail | +36.5 [+26.9, +46.2] | +12.5* | 113.2 delay | Fail |
| closed-start gate: waits, $b = 2$, probes | +13.2 [+7.4, +20.5] | +0.0 | 0.3 gain | Pass | +16.9 [+13.8, +20.4] | -3.6 | 3.6 field | Pass |
| *Map-only gate (ECA-m) and its components* | | | | | | | | |
| ECA-m: map hypothesis only | +19.7 [+13.8, +26.6] | +0.0 | 0.0 delay | Pass | +54.8 [+46.9, +62.7] | -4.3* | 4.3 field | Fail |
| − open start | +18.7 [+12.5, +26.0] | +0.0 | 0.0 healthy | Pass | +43.2 [+38.2, +48.4] | -3.2 | 3.2 field | Pass |
| − lagged hypotheses | +19.7 [+13.7, +26.5] | -0.1 | 0.1 field | Pass | +55.9 [+48.6, +63.3] | -5.2* | 5.2 field | Fail |
| − identifier hold | +19.7 [+13.9, +26.6] | +0.0 | 0.0 delay | Pass | +54.8 [+46.7, +62.8] | -4.8* | 4.8 field | Fail |
| − window decision rule | +19.7 [+13.8, +26.5] | +0.0 | 0.0 delay | Pass | +54.0 [+46.3, +61.5] | -4.7* | 4.7 field | Fail |
| bound $b = 2$ instead of $b = 1$ | +17.8 [+12.0, +24.8] | +0.0 | 0.8 delay | Pass | +36.8 [+25.2, +47.2] | +6.5* | 4.0 delay | Fail |
| *Hypothesis library and selection* | | | | | | | | |
| coupling hypothesis alone | -0.6 [-2.4, +1.3] | +2.6* | 1.5 delay | Fail | +22.2 [+15.4, +29.3] | +19.1* | 3.6 delay | Fail |
| coupling open at start | +19.6 [+13.5, +26.6] | -0.8* | 1.1 delay | Fail | +54.8 [+46.9, +62.2] | +17.9* | 2.2 delay | Fail |
| coupling decides at 5 samples | +19.6 [+13.8, +26.6] | -0.9* | 0.9 field | Fail | +54.8 [+46.9, +62.2] | +17.9* | 1.1 delay | Pass |
| coupling decides at 20 samples | +19.7 [+13.8, +26.6] | -0.1 | 0.1 field | Pass | +54.8 [+46.9, +62.6] | +11.4* | 1.1 delay | Pass |
| coupling prior $C_K = I$ | +19.7 [+13.7, +26.6] | +1.1* | 0.0 delay | Pass | +54.8 [+47.0, +62.7] | +16.3* | 1.1 delay | Pass |
| selection against the map alone | +19.6 [+13.5, +26.6] | -0.8* | 1.1 delay | Fail | +54.8 [+46.9, +62.2] | +17.9* | 2.2 delay | Fail |
| **ECA** | +19.7 [+13.7, +26.6] | +0.2 | 0.0 delay | Pass | +54.8 [+47.1, +62.6] | +15.9* | 1.1 delay | Pass |

0.37, both are closed on the delay (0.05 and 0.06), and their map fractions agree to 0.01. A low ratio establishes consistency only for the sampled commands. It cannot establish correctness of the full matrix or guarantee performance preservation.

**Failure traces.**

**Pusher at $150°$ (the closed-start gate).** In the recorded trace, the closed-start gate opens at step 10 with ratio 0.05 and closes near step 25. At closing, the map error is approximately 2.2, compared with 2.8 for the identity. Error later grows to 5–13. Four of the five episodes look like this, the gate open on about a fifth of the steps and the final map error between 5.5 and 7.8. In the fifth it stays open on 83% of them, ends with an error of 1.7, and performs like the ungated identifier: when the gate does not shut, the estimate converges. These traces show the self-locking failure: while the gate is shut, the frozen policy under the rotation drives the arm into states the nominal model describes poorly, the identifier regresses on that loop, and the estimate can no longer validate. ECA-m and ECA start open, so the identifier regresses on the compensated loop from the first step; the open fraction at $150°$ is 0.96 and the gain $+101.5$ and $+101.4$.

**Closed start.** The earlier gate started with the identity and opened only once the map hypothesis won the test. Under a $150°$ rotation the frozen policy drives the arm into states the nominal model describes poorly, the identifier fits that loop, and the estimate never validates again; the trace above is one instance. Starting open removes this self-locking failure, and it is safe in the sense that before the first decision the issued command is exactly ID's, so the gate can only withdraw compensation, never delay it.

**Reacher permutation.** Without a bound, the estimated map's second row is $(0.48, 0.51)$ instead of $(1, 0)$ despite a near-zero validation ratio. Commands in the observed direction cannot distinguish these rows. The inverse drives joint two into its limit at step 84; the estimated minimum singular value then falls from 0.39 to 0.02. Slower identification while open does not improve return ($-6.1$ to $-8.7$). The closed-start gate's factor-two bound prevents the joint-limit event and reduces the loss to 1.6; the unit bound recovers the condition ($+8.7$). The same mechanism, amplification along directions the loop has not excited, explains the ID losses on the Pusher permutation and mix.

**The Pusher velocity field and the map-only gate.** The field adds to each actuator a torque proportional to the velocity of a neighbouring joint, $\tau_i = \kappa\, \dot{q}_{i-1}$ with alternating signs, an additive, state-dependent disturbance. In the first twenty to thirty steps of an episode the map identifier fits it with a contracting map, whose ratio $q$ against the identity and the lagged identities is 0.02–0.4, so

Table 7: Further gate hypotheses for the map-only gate on development policies. Gain is the mean paired improvement over eight input maps; Field is the paired difference on the velocity field; Worst is the largest mean preservation loss; Rule is the recorded preservation criterion. Each row adds one change to ECA-m.

| Configuration | Reacher gain [95% CI] | Field | Worst | Rule | Pusher gain [95% CI] | Field | Worst | Rule |
|---|---|---|---|---|---|---|---|---|
| ECA-m | +19.7 [+13.8, +26.5] | -0.0 | 0.0 | Pass | +54.8 [+46.9, +62.7] | -4.3* | 4.3 | Fail |
| informative threshold $3k$ | +19.7 [+13.8, +26.6] | -0.0 | 0.0 | Pass | +54.2 [+46.5, +61.6] | -4.7* | 4.7 | Fail |
| informative threshold $30k$ | +19.7 [+13.8, +26.5] | -0.0 | 0.0 | Pass | +54.9 [+46.8, +62.6] | -4.8* | 4.8 | Fail |
| decisions from 3 samples | +19.6 [+13.6, +26.5] | -0.2 | 0.2 | Pass | +54.3 [+46.8, +61.9] | -4.4* | 4.4 | Fail |
| window 10 | +19.6 [+13.8, +26.5] | -0.1 | 0.1 | Pass | +51.3 [+44.0, +58.7] | -3.3 | 3.3 | Pass |
| probing while open ($0.05\,a_{\max}$) | +19.6 [+13.5, +26.4] | -0.7 | 0.7 | Pass | +52.6 [+45.6, +59.7] | -1.2 | 1.2 | Pass |
| + constant-offset hypothesis | +19.5 [+13.5, +26.2] | -0.4 | 0.4 | Pass | +47.5 [+39.8, +55.3] | -2.6 | 2.6 | Pass |
| + uniform-gain hypotheses $\{.5, .75, 1.5, 2\}$ | +19.7 [+13.8, +26.6] | -0.4 | 0.4 | Pass | +50.8 [+42.9, +58.7] | -1.5 | 1.5 | Pass |
| + uniform-gain hypotheses $\{.25, \ldots, 4\}$ | +19.5 [+13.6, +26.4] | -0.4 | 0.4 | Pass | +48.2 [+39.5, +56.9] | -2.5 | 2.5 | Pass |
| + amplification veto $\rho > 3$ | +17.2 [+11.6, +24.6] | -0.4 | 0.4 | Pass | +50.4 [+43.7, +57.1] | -4.2 | 4.2 | Pass |
| + amplification veto $\rho > 2$ | +13.1 [+8.1, +19.3] | -0.4 | 0.4 | Pass | +39.4 [+30.2, +48.9] | -1.6 | 1.6 | Pass |
| + amplification veto $\rho > 1.5$ | +8.5 [+4.3, +14.1] | -0.4 | 0.4 | Pass | +22.8 [+10.9, +34.8] | -0.4 | 0.5 | Pass |

the map gate is passed on data that do discriminate. The inverse of a contracting map amplifies; the unit bound withholds the amplification and applies only the redirection, which is wrong for this fault. With the map hypothesis alone the bound is the whole difference: Bound ID loses $29.8$ at $b = 1$ and $8.3$ at $b = 1.5$, and gains $5.0$ at $b = 2$; the gated ECA-m with $b = 2$ gains $6.5$. But $b = 2$ also returns the Pusher permutation to a loss ($-6.6$) and cuts the mix from $+65.1$ to $+10.6$ (Table 6). Table 7 lists the further gate hypotheses we tried on ECA-m in order to reject the field without paying on the input-map class. A constant-offset hypothesis and uniform-gain hypotheses reduce the field loss to $1.5$–$2.6$ but cut the Pusher gain by $4$–$7$ and, for the gain hypotheses, reject the gain-loss map itself (open fraction $0.06$). An amplification-demand veto, which closes the gate when the estimate's inverse would need to amplify the buffered policy commands by more than $\rho$, removes the field loss only at thresholds that also veto the partially identified rotations: the estimates of in-class maps are themselves contracting along unexcited directions, so amplification demand does not separate the two. Probing while open passes the preservation rule but costs $0.9$ on the healthy body and $5.5$ on the strength-one field. Rejecting the field was therefore impossible with the map hypothesis alone at no cost on the input-map class; the field is not a map, and a gate that tests only maps can at best refuse to act. The coupling hypothesis instead represents the fault: on the same windows its ratio against the alternatives is below $0.1$, it is selected within the first ten to fifteen steps, and the inverse of its near-identity map with the coupling term removed recovers $15.9$ return on the development split and $41$–$45$ at strength one. Replicated on a second set of development episodes, ECA and ECA-m differ by $+14.4$ and $+41.4$ on the two field strengths and by at most $1.3$ on every other condition (Table 8).

**Component ablations of the closed-start gate.** The factor-two bound and the probing were selected on ablations of that gate, which we summarise rather than tabulate, since the gate is superseded twice over. Removing both bound and probing gave $+10.4$ and $+15.7$ mean gain on the two plants; the reported setting gave $+13.2$ and $+16.9$; a delayed validation window (lag $10$) gave $+12.2$ and $+5.8$; dropping the probing and raising the closing threshold to $1$ gave $+13.8$ and $+25.8$ but failed preservation on Pusher, as did clipping the singular values of the estimate to $[.25, 2]$ at two of three threshold pairs. The pattern that survived into ECA-m is the one visible in Table 6: on this plant every change that recovers more input-map return spends it on the preservation set. The rows state their base explicitly, to avoid treating a sequence of development changes as several proposed algorithms. Aggregate gains include all eight tested input maps, including gain loss.

## D Recorded evaluation criteria

**Splits and seeds.** All policies are SAC seeds of the same trainer. The development split is policy seeds $200$–$204$ with episode seeds $70000+$, and a replication set of ten episodes per policy uses $73000+$. The three held-out cohorts are policy seeds $205$–$209$ (the first, for the closed-start gate, episodes $71000+$), $210$–$214$ (the second, for the map-only gate, episodes $72000+$) and $215$–$219$ (the third, for ECA, episodes $74000+$). The fresh-episode control grid reuses the first cohort with episodes $82000+$; reuse identifies on $74000+$ and evaluates on $75000+$; the mid-episode onset study

Table 8: Replication on a second set of development episodes, ten per policy: first-100-step return of the frozen policy and paired differences for ID, ECA-m, and ECA, with the paired difference ECA−ECA-m and its policy-bootstrap interval.

| Plant | Condition | Frozen | ID | ECA-m | ECA | ECA−ECA-m [CI] |
|---|---|---|---|---|---|---|
| Reacher | healthy | -5.3 | +0.0 | +0.0 | +0.0 | -0.0 [-0.0, +0.0] |
| | gain 0.5 | -5.6 | +0.2* | +0.0 | +0.0 | -0.0 [-0.0, +0.0] |
| | velocity field | -208.5 | +1.2* | +0.3* | +1.3* | +1.0 [+0.5, +1.4] |
| | delay 2 | -6.4 | -51.6* | -0.1 | -0.1 | +0.0 [+0.0, +0.0] |
| | rot 60° | -10.2 | +4.0* | +4.3* | +4.3* | +0.0 [-0.0, +0.0] |
| | rot 90° | -30.2 | +21.8* | +23.2* | +23.2* | +0.0 [-0.0, +0.0] |
| | rot 120° | -38.6 | +21.1* | +26.6* | +26.6* | -0.0 [-0.1, -0.0] |
| | rot 150° | -39.9 | +9.7* | +19.4* | +19.3* | -0.0 [-0.1, -0.0] |
| | rot 180° | -38.5 | -6.0 | +15.2* | +15.2* | +0.0 [+0.0, +0.1] |
| | permutation | -28.0 | +18.3* | +20.2* | +20.2* | -0.0 [-0.4, +0.2] |
| | orthogonal mix | -26.3 | +19.1* | +19.8* | +19.8* | -0.0 [-0.0, +0.0] |
| | delay 1 | -5.6 | -1.2* | +0.0 | +0.0 | +0.0 [+0.0, +0.0] |
| | delay 3 | -15.3 | -60.5* | +0.2 | +0.1 | -0.0 [-0.1, +0.0] |
| | field 1 | -215.8 | +0.7* | -0.6 | -0.4 | +0.2 [-0.4, +0.8] |
| | damping 2 | -5.5 | +0.0 | +0.0 | +0.0 | +0.0 [+0.0, +0.0] |
| | mix 11 | -27.9 | +18.2* | +20.5* | +20.3* | -0.2 [-0.6, +0.1] |
| Pusher | healthy | -31.6 | -2.9* | -0.2* | -0.3* | -0.0 [-0.1, +0.0] |
| | gain 0.5 | -33.6 | -0.4 | +0.1 | +0.1 | -0.0 [-0.0, +0.0] |
| | velocity field | -51.8 | +6.4* | -4.4* | +14.4* | +18.8 [+12.7, +24.4] |
| | delay 2 | -36.1 | -110.9* | -1.5* | -1.5* | +0.0 [+0.0, +0.0] |
| | rot 60° | -43.4 | +9.7* | +10.3* | +10.1* | -0.2 [-0.4, +0.0] |
| | rot 90° | -89.0 | +54.0* | +54.9* | +54.7* | -0.1 [-0.5, +0.1] |
| | rot 120° | -135.0 | +92.6* | +91.2* | +91.4* | +0.2 [-0.1, +0.7] |
| | rot 150° | -144.7 | +87.9* | +95.3* | +95.2* | -0.0 [-0.1, +0.0] |
| | rot 180° | -120.9 | +51.6* | +60.3* | +59.8* | -0.5 [-1.6, +0.2] |
| | permutation | -100.0 | -14.2 | +39.9* | +38.6* | -1.3 [-3.9, +0.1] |
| | orthogonal mix | -105.7 | -15.1* | +57.5* | +57.5* | +0.0 [+0.0, +0.0] |
| | delay 1 | -32.0 | -49.1* | -0.3* | -0.4* | -0.0 [-0.1, +0.0] |
| | delay 3 | -49.7 | -96.5* | -2.3* | -2.3* | +0.0 [+0.0, +0.0] |
| | field 1 | -133.6 | -18.8* | -2.8* | +41.4* | +44.3 [+27.8, +57.1] |
| | damping 2 | -37.7 | -3.8* | -0.1 | -0.2 | -0.1 [-0.2, -0.0] |
| | mix 11 | -100.8 | -21.2* | +52.9* | +52.8* | -0.1 [-0.3, -0.0] |

uses 76000+ and the healthy-drift measurement 78000+. Every number in the paper can be traced to one of these splits through the run records.

The original protocol records the following criteria before the first held-out evaluation:

1. *Preservation:* for healthy, gain 0.5, field, and delay conditions, the paired interval contains zero or its lower bound exceeds $-1$ on Reacher and $-3$ on Pusher.
2. *Recovery:* at 90° and 120°, the method is significantly better than the frozen policy and not significantly worse than ID.
3. *Reuse:* fixed transfer significantly exceeds reset and reaches at least 70% of the oracle gain at 90° and 120°.

The first rule is the original "no harm" criterion. We retain it for comparability but describe it as preservation: accepting an interval containing zero does not prove a small effect. Gain loss is a preservation condition within the input-map class.

Recorded amendments before that held-out evaluation include selection of the component configuration, replacement of the Pusher model using only the earlier cohort, and addition of DOB and $\mathcal{L}_1$ with development-selected gains. Development results were repeatedly inspected and are not presented as held out. The earlier protocol in the accompanying records explicitly requires significant reuse gains. Later summaries used the weaker phrase "beats reset". We retain the stricter original criterion here and report its incomplete satisfaction.

Table 9: Held-out first-100-step comparison of the closed-start gate on its own cohort, the first. Notation as in Table 1, with ECA denoting the closed-start gate; bold excludes True inverse. This diagnostic uses the known fault matrix $M$. Dashes mark conditions without a static input-map inverse.

| Plant | Condition | Frozen | Add. | DOB | $\mathcal{L}_1$ | ID | ECA | True inverse |
|---|---|---|---|---|---|---|---|---|
| Reacher | healthy | -2.7 | -0.7* | -0.0 | -0.0 | +0.0 | +0.0 | +0.0 |
| | gain 0.5 | -2.7 | +0.0 | -0.0 | -0.0 | **+0.1*** | -0.2* | +0.1* |
| | velocity field | -197.9 | **+3.4*** | +3.0* | +3.1* | -1.3 | +0.1 | – |
| | delay 2 | -3.2 | -56.3* | -2.7* | -6.2* | -44.7* | +0.0 | – |
| | rot 60° | -5.6 | +1.4 | -4.9* | -13.1* | **+2.5*** | +1.3* | +3.0* |
| | rot 90° | -32.6 | +19.4* | -18.9* | -87.9* | **+27.8*** | +25.3* | +29.9* |
| | rot 120° | -49.2 | -65.3* | -35.5* | -68.9* | **+38.4*** | +34.7* | +46.5* |
| | rot 150° | -51.6 | -61.8* | -41.8* | -68.9* | **+22.2*** | +17.6* | +48.9* |
| | rot 180° | -73.9 | -50.1* | -40.7* | -112.6* | +31.8* | **+34.4*** | +71.2* |
| | permutation | -40.2 | -70.0* | -32.4* | -115.2* | **+34.8*** | +32.0* | +37.5* |
| | orthogonal mix | -29.4 | +24.1* | -14.7* | -48.9* | **+25.5*** | +23.8* | +26.8* |
| Pusher | healthy | -27.0 | -0.1* | **+0.1** | -13.6* | -0.3 | +0.0 | +0.0 |
| | gain 0.5 | -28.6 | +0.6* | +0.6* | -5.2* | **+0.9*** | +0.2 | +1.6* |
| | velocity field | -43.7 | **+12.3*** | +2.7 | -15.1* | +4.1 | -0.5 | – |
| | delay 2 | -31.7 | -34.8* | -6.0* | -24.3* | -115.6* | -0.3 | – |
| | rot 60° | -36.2 | +6.6* | +5.0* | -8.8* | **+8.5*** | +5.2* | +9.3* |
| | rot 90° | -93.6 | +50.7* | +8.0 | -45.6* | **+64.6*** | +56.9* | +66.6* |
| | rot 120° | -135.5 | -33.9* | -10.9* | -57.2* | **+103.0*** | +59.6* | +108.6* |
| | rot 150° | -145.4 | -76.4* | -19.0* | -64.2* | **+92.4*** | +24.1* | +118.5* |
| | rot 180° | -117.6 | -100.3* | -24.9* | -104.5* | **+53.6*** | +4.1 | +90.6* |
| | permutation | -99.0 | -97.3* | -17.1* | -144.2* | -18.2 | -2.6 | +72.0* |
| | orthogonal mix | -108.9 | -46.8* | -19.1* | -69.3* | -3.3 | +3.4 | +81.9* |

**Revisions and the three cohorts.** The closed-start gate failed the second half of the recovery criterion on the first cohort. The gate was then revised on the development split only (Section 5.3, Table 7), and the revised map-only setting, ECA-m, was recorded, together with the preservation and recovery criteria, before the second cohort was trained. That cohort was evaluated once, with the comparators of Table 1; ECA-m met the recovery criterion on both plants and failed the preservation criterion on the Pusher velocity field (Section 5.5, Table 11). The coupling hypothesis and the selection rule were then designed and selected on the development split only (Table 6, Appendix B), with no run on the second cohort or on the held-out episodes of the first, and the complete setting was recorded, with the same two criteria, before the third cohort was trained. That cohort is evaluated once, with the comparators of Table 1; the results are those of Section 5.5. The second cohort was not re-run with ECA: its record is the held-out test of ECA-m and is reported as such. The reuse criterion was not re-evaluated on either later cohort: the identifier whose estimate is reused is unchanged from the closed-start gate, and no reuse runs were made on the fresh cohorts. The single runs of ECA-m and ECA on the fresh episodes of the first cohort (Section 5.2) were made after each setting was frozen; those policies had been inspected, so they are reported as additional paired evaluations. The manuscript does not create a new prospective registration beyond these recorded files.

## E Additional results

The following tables add the held-out evidence of the three cohorts, the per-condition fresh-episode comparison of Section 5.2, and longer-horizon checks without reproducing every development baseline at every horizon. The True inverse diagnostic tests recovery with the fault matrix supplied, removing estimation error. It does not test selective compensation under an unknown fault class, so it is excluded from the main comparison. Complete numerical outputs remain in the generated table files and original run records. All first-100-step significance marks use the policy bootstrap described in Section 5. Table 9 is the recorded held-out result of the closed-start gate, the first of the three cohorts and the one that sent the gate back to the development split.

Table 10: Fresh episodes of the first cohort: paired first-100-step differences for ID, Bound ID ($b = 1$), the closed-start gate, ECA-m, and ECA. Bold marks the largest positive mean on each input-map condition.

| Plant | Condition | Frozen | ID | Bound ID ($b = 1$) | Closed-start gate | ECA-m | **ECA** |
|---|---|---|---|---|---|---|---|
| Reacher | healthy | -4.2 | +0.0 | +0.0 | +0.0 | +0.0 | +0.0 |
| | gain 0.5 | -4.3 | +0.1* | +0.0 | -0.4* | +0.0 | +0.0 |
| | velocity field | -202.9 | -0.6 | -1.2 | +0.0 | +0.5 | +0.9 |
| | delay 2 | -5.0 | -44.9* | -0.1* | +0.0 | +0.0 | +0.0 |
| | rot 60° | -9.4 | +4.5* | +4.9* | +3.2* | **+4.9*** | +4.8* |
| | rot 90° | -30.5 | +23.2* | **+25.1*** | +19.8* | +25.0* | +25.0* |
| | rot 120° | -39.4 | +25.1* | **+30.6*** | +22.4* | +30.5* | +30.4* |
| | rot 150° | -58.7 | +27.6* | +38.2* | +35.6* | +38.3* | **+38.5*** |
| | rot 180° | -65.7 | +28.0* | **+45.6*** | +31.4* | +45.1* | +44.8* |
| | permutation | -26.9 | +19.8* | +20.5* | +18.4* | **+20.6*** | +20.4* |
| | orthogonal mix | -26.0 | +20.1* | **+21.0*** | +17.4* | +21.0* | +21.0* |
| Pusher | healthy | -27.8 | -0.5* | -1.0* | -0.1 | +0.0 | +0.0 |
| | gain 0.5 | -29.8 | +0.2 | -0.7 | +0.3* | -0.1 | -0.1 |
| | velocity field | -44.4 | +8.9* | -33.9* | +1.7 | -12.7* | +9.7* |
| | delay 2 | -30.7 | -107.0* | -36.5* | +0.0 | -1.8* | -1.8* |
| | rot 60° | -36.9 | +5.4* | **+8.6*** | +5.1* | +7.7* | +7.7* |
| | rot 90° | -92.3 | +62.7* | **+63.2*** | +50.0* | +62.8* | +62.8* |
| | rot 120° | -134.5 | +99.7* | **+102.1*** | +47.9* | +100.3* | +100.3* |
| | rot 150° | -140.6 | +84.4* | **+100.9*** | +11.1* | +90.8* | +90.8* |
| | rot 180° | -107.3 | +39.7* | **+65.5*** | -3.7 | +58.4* | +58.4* |
| | permutation | -99.3 | -33.0* | **+46.1*** | -10.0 | +34.6* | +32.7* |
| | orthogonal mix | -109.6 | -17.4 | **+61.9*** | +18.3 | +59.0* | +59.0* |

On these fresh episodes (Table 10) ECA matches Bound ID ($b = 1$) on Reacher, and on Pusher it is within 11 return of it at every rotation and within 14 on the permutation ($+32.7$ against $+46.1$), while Bound ID loses 33.9 on the field and 36.5 under delay. The gate's recovery cost is therefore at most a third of the recovery on any single condition and 4 return on the eight-map mean, while it removes 35 return of delay loss; the library turns the field from a loss of 12.7 (ECA-m) into a gain of 9.7, against 8.9 for ID.

Tables 11 and 12 are the record of the map-only gate on its cohort. ECA-m has the largest positive mean among the gated and additive comparators on all fourteen input-map cells; Bound ID, which never withdraws, is higher on six of them, by at most 0.3 return. The only cell on which a comparator is far better is the Pusher velocity field, where the direct additive update gains 8.9 and ID 6.3 while ECA-m loses 8.3. The closed-start gate preserves the field ($+1.2$) but recovers 22.9 on Pusher against ECA-m's 57.6. The oracle inverse of the known map gains 24.1 and 61.5 on the two plants; ECA-m is within 5 and 4 return of it.

Table 13 is the record of ECA on its cohort. Among the gated and additive comparators, ECA or ECA-m has the largest positive mean on thirteen of the fourteen input-map cells; at Pusher 180° ID gains 60.3 against ECA's 51.1, a difference whose interval contains zero. Bound ID, which applies the same inverse and bound but never withdraws, is higher on six Reacher cells by at most 0.3 return and on four Pusher cells by more: 150° ($+99.6$ against $+90.7$), 180° ($+72.4$ against $+51.1$), the permutation ($+61.2$ against $+46.8$), and the mix ($+75.6$ against $+60.7$), all four intervals excluding zero. These are windows on which ECA's gate or selection withdraws compensation; the same behaviour costs Bound ID 37.8 return under delay, 31.7 on the Pusher field, 27.1 on the strength-one field, and 2.9 on the Reacher field, and it fails the recorded preservation criterion on both plants. Outside the input-map class the comparators significantly better than ECA are the additive update, DOB and ID under gain loss on both plants ($+0.3$, $+0.1$, $+0.2$ against $+0.0$ on Reacher; about $+0.5$ against $-1.2$ on Pusher). The privileged oracle inverse of the known map gains 32.3 and 74.5 over the eight maps against ECA's 26.6 and 54.7, and exceeds ECA significantly on every map cell of this cohort; the gap is largest where the frozen policy loses most (Pusher permutation and mix, 40 and 39 return). ECA and ECA-m differ by at most 0.3 return on every condition except the Pusher

Table 11: The second cohort: the held-out evaluation of the map-only gate, run once before the coupling hypothesis existed. Notation as in Table 1. Bold marks the largest positive mean among the reported compensators on an input-map condition. ECA was not run on this cohort.

| Plant | Condition | Frozen | Add. | DOB | ID | Bound ID | ECA-m |
|---|---|---|---|---|---|---|---|
| Reacher | healthy | -4.1 | -0.1* | +0.0 | +0.0 | +0.0 | +0.0 |
| | gain 0.5 | -4.3 | +0.1* | +0.0 | +0.1* | +0.0 | +0.0 |
| | velocity field | -204.0 | +1.4* | +0.8* | +0.1 | -1.0* | +0.2 |
| | delay 2 | -5.0 | -35.5* | -1.2* | -44.3* | -0.1 | +0.0 |
| | rot 60° | -10.9 | +5.6* | -9.9* | +6.2* | **+6.5*** | +6.5* |
| | rot 90° | -29.6 | +5.6 | -19.7* | +22.7* | **+24.3*** | +24.3* |
| | rot 120° | -31.1 | -61.9* | -36.1* | +15.6* | **+22.3*** | +22.2* |
| | rot 150° | -48.8 | -50.2* | -38.7* | +15.6 | +29.5* | **+29.6*** |
| | rot 180° | -50.9 | -70.7* | -61.6* | +3.4 | +30.7* | **+31.1*** |
| | permutation | -22.4 | -53.7* | -26.0* | +15.6* | +16.4* | **+16.5*** |
| | orthogonal mix | -27.4 | +18.0* | -14.9* | +21.6* | **+22.6*** | +22.6* |
| | *input-map mean* | | -25.9* | -25.9* | +12.6* | +19.0* | +19.1* |
| | *worst loss* | | 35.5 | 1.2 | 44.3 | 1.0 | 0.0 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Pass |
| Pusher | healthy | -30.2 | +0.2 | -0.1* | -0.1 | +0.2 | +0.0 |
| | gain 0.5 | -31.4 | +0.7* | +0.5* | +1.1* | +0.1 | +0.4* |
| | velocity field | -42.2 | +8.9* | +3.0 | +6.3* | -34.2* | -8.3* |
| | delay 2 | -32.3 | -22.4* | -5.5* | -100.1* | -27.4* | -0.9* |
| | rot 60° | -35.7 | +3.9* | +2.1* | +5.5* | **+5.8*** | +5.7* |
| | rot 90° | -67.3 | +25.4* | -1.4 | +35.9* | **+36.9*** | +36.6* |
| | rot 120° | -133.4 | -22.5 | -8.2 | +98.5* | +101.2* | **+101.3*** |
| | rot 150° | -153.8 | -77.2* | -17.3* | +112.0* | +120.9* | **+120.9*** |
| | rot 180° | -115.4 | -93.8* | -23.1* | +66.5* | **+81.9*** | +81.8* |
| | permutation | -95.7 | -99.1* | -21.7* | -9.1* | +49.7* | **+49.7*** |
| | orthogonal mix | -100.6 | -62.1* | -16.9* | +1.8 | **+64.6*** | +64.4* |
| | *input-map mean* | | -40.6* | -10.8* | +39.0* | +57.6* | +57.6* |
| | *worst loss* | | 22.4 | 5.5 | 100.1 | 34.2 | 8.3 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Fail |

Table 12: The second cohort against the recorded criteria, for ID and ECA-m. Notation as in Table 3. the frozen returns at $90°$ and $120°$ are $-29.6$ and $-31.1$ on Reacher and $-67.3$ and $-133.4$ on Pusher.

| Plant | Method | Worst loss | Rule | $90°$ | $120°$ | Maps |
|---|---|---|---|---|---|---|
| Reacher | ID | 44.3 (delay 2) | Fail | +22.7* | +15.6* | +12.6* |
| | ECA-m | 0.0 (delay 2) | Pass | +24.3* | +22.2* | +19.1* |
| | ECA-m−ID | | | +1.5 [+1.3, +1.8] | +6.7 [+5.2, +8.2] | +6.5 [+5.1, +8.1] |
| Pusher | ID | 100.1 (delay 2) | Fail | +35.9* | +98.5* | +39.0* |
| | ECA-m | 8.3 (velocity field) | Fail | +36.6* | +101.3* | +57.6* |
| | ECA-m−ID | | | +0.7 [+0.0, +1.4] | +2.8 [+1.6, +3.9] | +18.6 [+15.0, +21.6] |

velocity field. Table 14 adds the two additive controls the main comparison omits, the bounded additive update and the $\mathcal{L}_1$ implementation.

**The gain-loss failure.** The gain-loss cost that denies ECA the preservation criterion on this cohort is concentrated rather than systematic: across the twenty-five episodes the median difference is $-0.2$ return, three episodes lose more than 2, and the mean excluding the single worst is $-0.6$. That worst episode, one episode of one policy at $-15.9$, shows the gate failing in the way that matters, and we report it in detail because it is the evidence against the method rather than for it. For the first forty steps the estimate explains the transitions well (ratio between $0.01$ and $0.63$) and nothing is lost. At step $46$ the ratio passes $1$: the identity now explains them better, and the gate's own test calls for withdrawal. The gate withdraws at step $86$. Of the $15.9$ return, $11.2$ is lost in those forty steps of acting on a hypothesis its test had already rejected, and $4.8$ afterwards, when the command

Table 13: The third cohort, the held-out evaluation of ECA, run once: first-100-step return of the frozen policy and paired differences for the comparators of Table 1, with the same summary rows. Bold marks the largest positive mean among the reported compensators on the input-map conditions.

| Plant | Condition | Frozen | Add. | DOB | ID | Bound ID | ECA-m | **ECA** |
|---|---|---|---|---|---|---|---|---|
| Reacher | healthy | -4.4 | -0.1* | +0.0 | +0.0 | +0.0 | +0.0 | +0.0 |
| | gain 0.5 | -4.7 | +0.3* | +0.1* | +0.2* | +0.0 | +0.0 | +0.0 |
| | velocity field | -205.0 | +0.6 | +1.3* | -1.0* | -2.9* | +0.2 | +1.4* |
| | delay 2 | -5.3 | -49.7* | -3.2* | -56.4* | -0.1* | +0.0 | +0.0 |
| | rot 60° | -10.2 | +4.6* | -9.3* | +5.1* | **+5.4*** | +5.4* | +5.4* |
| | rot 90° | -37.5 | +16.4* | -22.2* | +29.8* | **+31.7*** | +31.7* | +31.7* |
| | rot 120° | -55.9 | -55.1* | -34.6* | +37.8* | **+44.3*** | +44.3* | +44.3* |
| | rot 150° | -62.0 | -39.8* | -37.4* | +22.9* | +42.9* | **+42.9*** | +42.9* |
| | rot 180° | -59.8 | -48.7* | -55.3* | +9.0 | **+37.0*** | +36.9* | +36.9* |
| | permutation | -29.3 | -59.8* | -37.1* | +20.8* | **+23.1*** | +23.1* | +22.8* |
| | orthogonal mix | -34.2 | +24.7* | -17.6* | +28.0* | **+28.9*** | +28.8* | +28.8* |
| | *input-map mean* | | -19.7* | -26.7* | +19.2* | +26.7* | +26.6* | +26.6* |
| | *worst loss* | | 49.7 | 3.2 | 56.4 | 2.9 | 0.0 | 0.0 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Pass | Pass |
| Pusher | healthy | -28.8 | -0.3 | -0.1 | -0.6* | -0.3* | +0.0 | +0.0 |
| | gain 0.5 | -30.5 | +0.5 | +0.4* | +0.5 | -1.2* | -1.2* | -1.2* |
| | velocity field | -43.9 | +11.7* | +6.1* | +10.1* | -31.7* | -10.1* | +10.3* |
| | delay 2 | -32.5 | -27.7* | -6.3* | -116.6* | -37.8* | -1.1* | -1.1* |
| | rot 60° | -40.0 | +6.3* | +2.8* | +8.9* | **+9.6*** | +9.5* | +9.3* |
| | rot 90° | -106.8 | +44.4* | +11.8 | +75.2* | **+76.0*** | +75.9* | +75.9* |
| | rot 120° | -139.4 | -68.5* | -16.5* | +98.4* | +104.8* | **+104.8*** | +104.7* |
| | rot 150° | -143.2 | -78.8* | -20.4* | +78.7* | **+99.6*** | +90.6* | +90.7* |
| | rot 180° | -121.8 | -93.8* | -21.4* | +60.3* | **+72.4*** | +51.1* | +51.1* |
| | permutation | -115.7 | -93.6* | -19.4* | +3.0 | **+61.2*** | +46.8* | +46.8* |
| | orthogonal mix | -128.8 | -50.2* | -16.9* | +6.1 | **+75.6*** | +60.8* | +60.7* |
| | *input-map mean* | | -41.7* | -10.0* | +41.4* | +62.3* | +54.8* | +54.7* |
| | *worst loss* | | 27.7 | 6.3 | 116.6 | 37.8 | 10.1 | 1.2 |
| | *recorded rule* | | Fail | Fail | Fail | Fail | Fail | Fail |

is the policy's own again but the arm is no longer where the frozen policy would have taken it. The delay is the cost of the rule that revises the gate only on informative windows and with hysteresis (Section 4.2); under gain loss, where the bound forgoes the amplifying half of the inverse and the issued commands stay close to the policy's own, informative windows are rare. At step $46$ only $0.2$ return had been lost, so the delay accounts for essentially the whole episode, and that episode accounts for the criterion: without it the cohort's mean difference under gain loss is $-0.6$. A rule that withdrew as soon as the ratio crossed would plainly have been better here; we did not add one, because the setting was recorded before the cohort was trained and this is the first evidence that the withdrawal is too slow.

Table 15 repeats the development conditions over complete episodes. Over the eight input maps ECA is ahead of the ungated core by $+7.5$ $[+2.4, +13.9]$ on Reacher and $+24.8$ $[+17.3, +31.7]$ on Pusher, so the reported window understates the recovery as well as the interference: ECA's gain grows from $+52.2$ to $+179.2$ at Reacher $180°$ and from $+65.1$ to $+165.4$ on the Pusher mix, while ID's delay loss grows from $44.6$ to $99.7$ and from $113.2$ to $253.0$, where ECA's stays at $0.0$ and $0.1$. The comparison the design trades against is with Bound ID, the same inverse and bound with no gate, hold or second hypothesis: over complete episodes it recovers $145.0$ on the Pusher maps against ECA's $131.6$, and loses $49.2$ return under delay and $47.8$ on the velocity field, where ECA loses $0.1$ and gains $31.4$. The gate costs about thirteen return of recovery on that plant and removes about fifty of loss; on Reacher it costs nothing measurable ($+65.1$ against $+64.9$) and removes five. Three cells favour the ungated core over the whole episode, and all three are cells where the bound forgoes amplification: the Reacher velocity field ($+15.2$ against $-0.2$, the condition that exceeds the actuators' authority), Pusher $120°$ ($+236.4$ against $+225.5$) and the Reacher permutation ($+30.2$ against $+29.8$). Over the first $100$ steps the same three agree to within $0.6$ return; the divergence is

Table 14: Additional additive controls on development policies: paired first-100-step return differences. The shared frozen reference is in Table 1.

| Plant | Condition | Add.-b | $\mathcal{L}_1$ |
|---|---|---|---|
| Reacher | healthy | -0.3* | +0.0 |
| | gain 0.5 | +0.1 | +0.1* |
| | velocity field | +1.0 | +1.0 |
| | delay 2 | -33.5* | -2.3* |
| | rot 60° | +2.4* | -9.1* |
| | rot 90° | +13.8* | -87.3* |
| | rot 120° | -60.1* | -85.5* |
| | rot 150° | -72.2* | -70.4* |
| | rot 180° | -63.6* | -109.8* |
| | permutation | -28.1* | -132.5* |
| | orthogonal mix | +17.7* | -47.6* |
| Pusher | healthy | -0.3 | -10.4* |
| | gain 0.5 | +0.5* | -3.4* |
| | velocity field | +16.5* | -2.6 |
| | delay 2 | -39.3* | -23.4* |
| | rot 60° | +5.9* | -8.8* |
| | rot 90° | +42.8* | -38.3* |
| | rot 120° | -18.1* | -43.4* |
| | rot 150° | -63.3* | -69.3* |
| | rot 180° | -76.9* | -100.0* |
| | permutation | -96.0* | -131.4* |
| | orthogonal mix | -40.7* | -78.9* |

what the rest of the episode does with an amplified command that happens to be pointing the right way.

**Reuse across policies.** The estimate describes the body, so another policy for the same interface can reuse it. Policy A identifies a map in one episode; policy B is then evaluated with a fresh compensator state (reset), A's estimate with continued adaptation (transfer), or A's map held fixed (fixed), with a map estimated under a different fault (mismatched) as a control, over all 20 ordered pairs of five policies with five episodes per pair and the identifying policies as bootstrap clusters (Table 17). These runs used the closed-start configuration; the transferred identifier is unchanged in ECA, so reuse is a property of the estimate, not of the gate. On Reacher a fixed transferred map is close to the exact inverse and improves over reset; on Pusher reuse helps at $120°$ but not on the permutation, and continued adaptation does not consistently beat fixed reuse. The recorded criterion, which asks for a significant gain over reset and at least $70\%$ of the oracle gain at $90°$ and $120°$, is therefore met on one plant only. Which part of ECA does the reuse belong to? Table 16 applies policy A's fixed estimate through the ungated core, through ECA, and against versions that identify from scratch. On Reacher every mechanism gains from the transfer — the ungated core by $1.0$ to $11.1$ return and ECA by $0.7$ to $3.7$, every interval excluding zero — and the ungated core gains most, because it is the one that suffers most while identifying under its own compensation. On Pusher the transferred map helps the ungated core exactly where identification from scratch fails, the permutation ($+48.0$) and the mix ($+63.8$, interval excluding zero), and costs return at the rotations, where identification from scratch already succeeds. Reuse is therefore a property of the estimate rather than of the gate, and the records make that literal: with the estimate held fixed and the gate open, ECA issues exactly the bounded inverse of that estimate, and the two arms agree to $0.000$ return on all eight conditions and all twenty ordered pairs. What the gate changes is the arm that identifies from scratch, where it protects the estimate while it is still poor: on the Pusher permutation and mix, identifying from scratch costs the ungated core $113.6$ and $127.5$ return against $45.5$ and $49.5$ with the gate. On a second cohort of policies the asymmetry repeats: fixed reuse beats reset on every Reacher fault (up to $+8.4$) and on Pusher at $120°$ ($+31.5$), with the permutation and the mix positive but not significant. Over full episodes the Reacher gains persist on both splits, while on Pusher the development permutation turns negative ($-27.1$) and $120°$ and the mix stay positive. A map estimated under another fault often hurts Pusher, while on Reacher two composed rotations

Table 15: Complete episodes (250 steps on Reacher, 200 on Pusher) instead of the reported window, development policies, notation as in Table 1. The last summary row is the paired difference between ECA and the ungated core over the eight maps.

| Plant | Condition | Frozen | ID | Bound ID | ECA-m | **ECA** |
|---|---|---|---|---|---|---|
| Reacher | gain 0.5 | -5.0 | +0.1* | +0.0 | +0.0 | +0.0 |
| | rot 60° | -9.9 | +4.1* | +4.1* | +4.1* | +4.1* |
| | rot 90° | -64.7 | +56.5* | +57.5* | +57.4* | +57.4* |
| | rot 120° | -97.0 | +83.6* | +85.8* | +85.5* | +85.5* |
| | rot 150° | -153.5 | +83.9* | +116.8* | +117.3* | +116.3* |
| | rot 180° | -216.3 | +154.2* | +177.3* | +179.2* | +179.2* |
| | permutation | -40.8 | +30.2* | +29.1* | +29.7* | +29.8* |
| | orthogonal mix | -55.2 | +48.3* | +48.7* | +48.6* | +48.6* |
| | healthy | -4.8 | +0.0 | +0.0 | +0.0 | +0.0 |
| | velocity field | -504.5 | +15.2* | -5.1 | -0.9 | -0.2 |
| | delay 2 | -5.3 | -99.7* | -0.2* | +0.0 | +0.0 |
| | *input-map mean* | | +57.6* | +64.9* | +65.2* | +65.1* |
| | *worst loss* | | 99.7 | 5.1 | 0.9 | 0.2 |
| | *ECA−ID over the eight maps* | | +7.5 [+2.4, +13.9] | | | |
| Pusher | gain 0.5 | -46.2 | +0.8 | -4.4 | +0.4 | +0.3 |
| | rot 60° | -61.6 | +13.9* | +14.7* | +15.6* | +14.9* |
| | rot 90° | -183.5 | +134.8* | +137.8* | +136.8* | +137.2* |
| | rot 120° | -294.1 | +236.4* | +241.8* | +220.7* | +225.5* |
| | rot 150° | -312.2 | +230.7* | +256.2* | +238.6* | +241.4* |
| | rot 180° | -277.9 | +181.8* | +209.0* | +161.0* | +160.8* |
| | permutation | -194.4 | +12.6 | +125.2* | +109.3* | +107.0* |
| | orthogonal mix | -244.9 | +43.4* | +179.8* | +165.4* | +165.4* |
| | healthy | -43.3 | -3.1* | +0.1 | +0.0 | -0.1 |
| | velocity field | -83.4 | +25.5* | -47.8* | -0.9 | +31.4* |
| | delay 2 | -55.9 | -253.0* | -49.2* | -0.1 | -0.1 |
| | *input-map mean* | | +106.8* | +145.0* | +131.0* | +131.6* |
| | *worst loss* | | 253.0 | 49.2 | 0.9 | 0.1 |
| | *ECA−ID over the eight maps* | | +24.8 [+17.3, +31.7] | | | |

Table 16: Reuse of the estimate by three mechanisms, absolute first-100-step return over all 20 ordered pairs of development policies. The last two columns are the gain of the transferred estimate over identifying from scratch, bootstrapped over identifying policies.

| Plant | Fault | Untouched | ID reset | ID fixed | ECA reset | ECA fixed | Oracle | ID fixed−reset | ECA fixed−reset |
|---|---|---|---|---|---|---|---|---|---|
| Reacher | rot 90° | -31.0 | -8.6 | -6.1 | -7.2 | -5.6 | -5.1 | +2.4 [+2.0, +2.7] | +1.6 [+1.3, +1.8] |
| | rot 120° | -37.6 | -19.0 | -7.8 | -10.6 | -7.0 | -5.1 | +11.1 [+10.2, +12.0] | +3.7 [+3.1, +4.1] |
| | permutation | -24.5 | -11.0 | -5.6 | -7.8 | -5.4 | -5.1 | +5.5 [+4.7, +5.9] | +2.4 [+2.2, +2.6] |
| | orthogonal mix | -28.4 | -6.9 | -5.9 | -6.2 | -5.6 | -5.1 | +1.0 [+0.9, +1.1] | +0.7 [+0.6, +0.7] |
| Pusher | rot 90° | -77.0 | -31.5 | -42.3 | -31.1 | -41.5 | -28.9 | -10.8 [-35.4, +2.2] | -10.5 [-33.8, +1.8] |
| | rot 120° | -141.3 | -41.8 | -56.3 | -39.5 | -50.4 | -28.9 | -14.5 [-60.3, +11.7] | -10.9 [-50.9, +10.2] |
| | permutation | -94.3 | -113.6 | -65.6 | -45.5 | -52.0 | -28.9 | +48.0 [-16.7, +87.8] | -6.5 [-45.6, +17.3] |
| | orthogonal mix | -115.4 | -127.5 | -63.7 | -49.5 | -55.7 | -28.9 | +63.8 [+5.5, +98.0] | -6.2 [-52.1, +21.8] |

can remain within the policy's tolerance; reuse requires a compatible body and fault and sufficient identification coverage.

# F Boundary of the result

## F.1 Within-episode fault onset

A fault that appears while the policy is running is a different test from one that is present at reset: the estimate has already converged to the identity, and the gate is open on a body that the map explains perfectly. Table 18 shows that the recursive estimate follows the change without a change detector, and that the ordering of the methods is the one the rest of the paper reports, with two

Table 17: Reuse on development policies: absolute first-100-step return. The last column is a paired difference with a bootstrap over identifying policies. Rows marked mismatched apply a map identified under a different fault.

| Plant | Fault | Untouched | Reset | Transfer | Fixed | Oracle | Fixed−reset |
|---|---|---|---|---|---|---|---|
| Reacher | rot 90° | -33.9 | -15.8 | -8.1 | -8.1 | -6.3 | +7.7* |
| | rot 120° | -38.2 | -22.7 | -9.2 | -9.4 | -6.3 | +13.3* |
| | permutation | -29.2 | -13.6 | -7.4 | -7.4 | -6.3 | +6.1* |
| Pusher | rot 90° | -91.9 | -49.7 | -45.7 | -46.9 | -31.3 | +2.7 |
| | rot 120° | -134.9 | -81.8 | -76.0 | -57.2 | -31.3 | +24.6 |
| | permutation | -98.7 | -104.4 | -90.1 | -112.5 | -31.3 | -8.2 |

Table 18: Input maps switched on at step 30 instead of at the first step, development policies, full horizon. Returns are taken over the steps after the switch, Notation as in Table 1. The identifiers run from the first step, so the compensators spend thirty steps estimating a body that is still healthy.

| Plant | Condition | Frozen | ID | Bound ID | ECA-m | **ECA** |
|---|---|---|---|---|---|---|
| Reacher | rot 120° | -121.4 | +107.4* | +112.6* | +110.9* | +110.2* |
| | permutation | -49.2 | +40.3* | +41.9* | +41.3* | +40.8* |
| | orthogonal mix | -36.7 | +33.8* | +33.8* | +31.4* | +32.4* |
| | velocity field | -403.1 | +8.6 | +14.4* | -0.3 | +3.9 |
| Pusher | rot 120° | -157.0 | +109.1* | +126.9* | +106.9* | +105.5* |
| | permutation | -159.2 | +57.2* | +118.1* | +116.8* | +116.7* |
| | orthogonal mix | -175.1 | +37.6* | +127.2* | +128.5* | +126.4* |
| | velocity field | -44.4 | +13.5* | -9.1 | +0.3 | +13.6* |

Table 19: HalfCheetah-v5 ($k = 6$) under the protocol of the two manipulators, with its own nominal model ($R^2 = 0.99$). Ten episodes per policy are used here, because command differences of order $10^{-7}$ diverge within a few steps on this plant. Notation as in Table 1. The last column is a variant discussed below, not the method evaluated elsewhere in this paper.

| Plant | Condition | Frozen | ID | Bound ID | ECA-m | **ECA** | ECA, hold from step 0 |
|---|---|---|---|---|---|---|---|
| HalfCheetah | rot 60° | 180.1 | +66.0* | +67.2* | +53.9* | +53.9* | +24.5* |
| | rot 90° | 119.2 | +74.7* | +48.7* | +47.8* | +47.8* | +54.9* |
| | rot 120° | 93.0 | +0.0 | +3.4 | +12.2 | +12.2 | +17.8* |
| | rot 150° | 69.0 | -0.8 | -2.6 | -3.3 | -2.7 | +4.5* |
| | rot 180° | 61.3 | +19.6* | +21.0* | +19.8* | +21.1* | +7.1 |
| | permutation | 0.0 | +67.0* | +76.0* | +51.2* | +51.2* | +32.0* |
| | orthogonal mix | -10.4 | +81.1* | +93.5* | +89.6* | +89.6* | +78.1* |
| | gain 0.5 | 230.5 | -22.7* | -33.0* | -36.0* | -36.0* | -2.5* |
| | healthy | 345.5 | -16.1* | -15.9* | -20.3* | -20.3* | -0.7 |
| | velocity field | -37.4 | +0.9 | -1.3 | -0.8 | -0.8 | -1.3 |
| | delay 2 | 23.9 | -21.0* | -1.7 | -3.0 | -3.0 | -2.2 |
| | *input-map mean* | | +35.6* | +34.3* | +29.4* | +29.6* | +27.1* |
| | *worst loss* | | 22.7 | 33.0 | 36.0 | 36.0 | 2.5 |

differences. The ungated core loses most of its advantage on the Pusher permutation and mix ($+57.2$ and $+37.6$ against ECA's $+116.7$ and $+126.4$), because the map it must now identify is further from the estimate it starts with. The velocity field again separates the compensators: ECA gains $13.6$ where the bounded inverse without a gate loses $9.1$ and the map-only gate gains $0.3$. On the Reacher field no method recovers more than $15$ of $403$ return, as at the first step.

### F.2 A THIRD PLANT AND THE LIMITS OF TRANSFER

The two plants of the paper are fixed-base arms whose nominal model predicts the actuated joint velocities well. We repeated the whole protocol on a locomotion plant with contacts, HalfCheetah-v5, to see which half of the result transfers. Recovery does: over the eight input maps ECA gains

29.6 [+25.5, +33.8], including 51.2 on the permutation and 89.6 on the orthogonal mix, conditions on which the frozen policy scores 0.0 and −10.4 against 345.5 on the healthy body. Preservation does not. ECA loses 20.3 return on the healthy body and 36.0 under gain loss. On this plant the ungated core is better on both of the axes the paper is about: recovery (+35.6 against +29.6, paired +6.0 [+2.7, +9.8]) and the healthy body (−16.1 against −20.3, paired +4.1 [+0.7, +8.2]). Its worst preservation loss is smaller as well (22.7 against 36.0), though that gap is not resolved at five policies. The gate, which is the contribution, is the part that fails here. We report it as it stands.

The cause is measurable and specific, and we recorded it on all three plants, five policies and twenty healthy episodes each. On a healthy body the identity hypothesis' whitened error is 1.1, 2.3 and 1.0 times its calibrated level on Reacher, Pusher and HalfCheetah, in every case far below the informativeness floor of ten. No healthy window is informative on any of the three plants, so the gate neither opens nor closes on evidence and keeps the open state it starts in. That is harmless as long as the estimate the identifier leaves does not change the issued command. On the arms it does not: the command differs from the policy's by 0.0004 and 0.0073 on average, and the paired cost of running the compensator on a healthy body is −0.00 [−0.01, −0.00] and −0.06 [−0.19, +0.00] return. On HalfCheetah the same rule leaves an estimate whose bounded inverse moves the command by 0.125, and the paired cost is −22.7 [−36.6, −14.5]. The difference is not the size of the estimate — Pusher's $\|\hat{D} - I\|$ after a healthy episode averages 0.55 against HalfCheetah's 0.32 — but whether it points anywhere the command can follow. The model here is accurate in the usual sense ($R^2 = 0.99$) and still not accurate enough for this purpose. What ECA needs is not a small residual but one whose structure the identity explains better than a linear map does, and on a contact-rich body over a twenty-sample window that is not the case.

One line removes the failure on this plant: apply the identifier's own hold to the first window as well, which is the rule the rest of the method already follows. The worst preservation loss then falls from 36.0 to 2.5 and the healthy body to −0.7, against 2.5 return of the eight-map recovery, a difference this cohort cannot resolve ([−8.1, +5.4]). It is not a free improvement, and we do not adopt it. On the two manipulators the same change costs up to 93 return (Pusher 150°; −54.9 at 120°, −58.2 on the mix), because there the first window is exactly where the map is identified. The related ablation that waits for the gate before acting costs 11.6 return on Pusher, and the earlier closed-start gate 37.9 (Table 6). One further fact belongs here, because a reader will otherwise find it in the table: on this plant ECA and ECA-m return identical values on 548 of the 550 episodes, the two exceptions being single chaotic episodes at 150° and 180°. The coupling hypothesis is never selected, so the third plant tests the core and the gate, not the library. The setting is therefore plant-dependent at one identified point, and the honest summary is that ECA's preservation property is conditional on a nominal model whose error the identity explains better than a linear map does. That condition is cheap to check before any fault occurs, with healthy data and no fault knowledge: run the compensator on a healthy body and measure how far the issued command departs from the policy's. The three plants separate by two orders of magnitude on that quantity, and it tracks the paired healthy cost exactly. A plant on which the compensator quietly rewrites a seventh of the healthy command is one on which the gate cannot be trusted to withdraw, and the measurement costs twenty healthy episodes. That check, rather than the fix, is what this third plant taught us.

### F.3 Retraining and reward-driven alternatives

Training on the fault class is the alternative to compensating for it, and Table 20 runs it under conditions favourable to the baseline. Every episode of training draws a fresh input map: the identity with probability 0.2, and otherwise, with equal probability, a rotation of one actuator pair by $U(-\pi, \pi)$, a random permutation, a random orthogonal matrix, or a gain drawn from $U(0.3, 1)$ on half the channels. The evaluated faults are therefore inside the training support. It works, and it costs. On the eight input maps the randomised policy recovers 12 return on Reacher and 25 on Pusher relative to the healthy-trained policy, 59% and 46% of what ECA recovers on the healthy-trained policy, and it gives up 4.8 and 11.8 return on the healthy body, 6.3 and 12.7 under delay, and 4.5 and 11.4 under gain loss — the randomisation tax appears on exactly the conditions the preservation criterion protects. ECA on top of the randomised policy recovers part of the remaining gap at a cost of at most 0.55 return on the four preservation conditions of either plant: the eight-map gain is +3.3 on Reacher and +20.7 on Pusher, both intervals excluding zero. Identification of the body and robust-

Table 20: Domain-randomised policies against policies trained on the healthy body, absolute first-100-step return. The randomised policies share the architecture, seeds and training budget. Every episode of their training draws a fresh input map from the distribution given below, so the evaluated faults lie inside their training support. The last row of each plant is the paired gain of adding ECA to the randomised policy.

| Plant | Condition | Frozen | + ECA | Randomised | + ECA |
|---|---|---|---|---|---|
| Reacher | healthy | -3.4 | -3.4 | -8.2 | -8.2 |
| | gain 0.5 | -3.6 | -3.6 | -8.1 | -8.4 |
| | velocity field | -199.0 | -198.8 | -201.3 | -200.5 |
| | delay 2 | -3.9 | -3.9 | -10.2 | -10.2 |
| | rot 60° | -6.9 | -3.9 | -10.3 | -8.4 |
| | rot 90° | -26.9 | -4.7 | -17.6 | -10.1 |
| | rot 120° | -31.8 | -7.5 | -22.9 | -15.1 |
| | rot 150° | -45.1 | -17.4 | -21.1 | -19.5 |
| | rot 180° | -68.6 | -16.4 | -20.0 | -18.8 |
| | permutation | -14.9 | -6.2 | -12.5 | -12.3 |
| | orthogonal mix | -23.8 | -4.4 | -16.2 | -10.1 |
| | *input-map mean* | -27.7 | -8.0 | -16.1 | -12.8 |
| | *ECA on the randomised policy* | | | | +3.3 [+2.3, +4.2] |
| Pusher | healthy | -31.2 | -31.3 | -43.1 | -43.1 |
| | gain 0.5 | -33.4 | -33.1 | -44.8 | -43.7 |
| | velocity field | -52.7 | -36.9 | -56.3 | -45.9 |
| | delay 2 | -37.5 | -38.6 | -50.2 | -50.8 |
| | rot 60° | -41.3 | -33.4 | -54.0 | -43.7 |
| | rot 90° | -94.6 | -33.7 | -66.9 | -48.1 |
| | rot 120° | -141.5 | -41.7 | -79.5 | -54.5 |
| | rot 150° | -146.2 | -44.9 | -85.1 | -60.7 |
| | rot 180° | -123.9 | -63.7 | -87.6 | -54.7 |
| | permutation | -99.1 | -56.4 | -85.0 | -62.8 |
| | orthogonal mix | -113.0 | -47.8 | -87.1 | -56.5 |
| | *input-map mean* | -99.1 | -44.3 | -73.7 | -53.1 |
| | *ECA on the randomised policy* | | | | +20.7 [+15.4, +27.4] |

ness of the policy are therefore complements rather than substitutes; a policy that has been trained to tolerate the fault class still benefits from being told which member of it is present.

The reward baseline perturbs the command matrix antithetically, $W \pm \sigma E$ with $E$ Gaussian, and moves $W$ along $(R^+ - R^-)E$, normalised by the spread of recent returns and clipped to a trust region; an unnormalised step diverges, because returns are tens of units while the entries of $W$ are of order one. The scale and step size were selected per plant on a grid of fourteen $(\sigma, \alpha)$ pairs ($\sigma$ from $0.05$ to $0.5$, $\alpha$ from $0.03$ to $1$, forty episodes each) by mean return over the eight input maps: $\sigma = 0.3$, $\alpha = 1$ on Reacher and $\sigma = 0.2$, $\alpha = 0.3$ on Pusher. The selection matters: on Reacher the chosen setting ends $3.5$ return above the best of the smaller steps, so the baseline is reported at its own best setting rather than at a shared one. The reported runs use $120$ episodes, sixty antithetic pairs, and the final matrix is evaluated on five fresh episodes. Table 21 gives every condition. The baseline is strongest exactly where the search space is small: on Reacher, where $W$ has four entries, it ends within $0.5$ return of the privileged oracle inverse on all eight maps ($-3.70$ against $-3.44$ on the eight-map mean, against ECA's $-8.02$). On Pusher, with forty-nine entries, it closes $56$ of the $63$ return that separate the frozen policy from the oracle at $90°$ but falls short of ECA on the permutation ($-65.3$ against $-56.4$) and at $150°$ ($-65.6$ against $-44.9$). Its failure mode is the one a return-driven rule has to have: it optimises the conditions it is run on and has no representation for leaving a body alone. Under the two-step delay the matrix it converges to costs $32.9$ return on Reacher and $2.9$ on Pusher relative to the frozen policy, where ECA withdraws and costs $0.0$ and $1.1$; on the Reacher velocity field, which no compensator can help, it spends $120$ episodes moving $6$ return.

Table 21: Reward-driven online adaptation against ECA, development policies, absolute first-100-step return. The baseline keeps a command matrix across episodes and uses no model; ECA is reset every episode and uses no reward. "Episodes to ECA" is the median number of episodes before the baseline first matches ECA, with the number of policies reaching it in parentheses. It is shown only where ECA improves on the frozen policy.

| Plant | Condition | Frozen | ECA (one ep.) | Reward ep. 1 | Reward ep. 120 | Episodes to ECA |
|---|---|---|---|---|---|---|
| Reacher | gain 0.5 | -3.6 | -3.6 | -5.9 | -3.6 | |
| | rot 60° | -6.9 | -3.9 | -17.8 | -3.6 | 5 (5/5) |
| | rot 90° | -26.9 | -4.7 | -28.6 | -3.7 | 7 (5/5) |
| | rot 120° | -31.8 | -7.5 | -35.6 | -3.7 | 23 (5/5) |
| | rot 150° | -45.1 | -17.4 | -28.1 | -3.7 | 5 (5/5) |
| | rot 180° | -68.6 | -16.4 | -25.4 | -3.7 | 7 (5/5) |
| | permutation | -14.9 | -6.2 | -20.0 | -3.6 | 8 (5/5) |
| | orthogonal mix | -23.8 | -4.4 | -25.6 | -3.9 | 7 (5/5) |
| | healthy | -3.4 | -3.4 | -4.8 | -3.6 | |
| | velocity field | -199.0 | -198.8 | -204.6 | -192.9 | |
| | delay 2 | -3.9 | -3.9 | -5.4 | -36.8 | |
| Pusher | gain 0.5 | -33.4 | -33.1 | -28.9 | -31.9 | |
| | rot 60° | -41.3 | -33.4 | -46.2 | -34.6 | 3 (5/5) |
| | rot 90° | -94.6 | -33.7 | -123.9 | -38.5 | 22 (5/5) |
| | rot 120° | -141.5 | -41.7 | -137.6 | -49.8 | 82 (4/5) |
| | rot 150° | -146.2 | -44.9 | -132.1 | -65.6 | – (2/5) |
| | rot 180° | -123.9 | -63.7 | -104.7 | -48.5 | 44 (5/5) |
| | permutation | -99.1 | -56.4 | -98.2 | -65.3 | 82 (5/5) |
| | orthogonal mix | -113.0 | -47.8 | -101.9 | -56.8 | 68 (4/5) |
| | healthy | -31.2 | -31.3 | -27.4 | -31.4 | |
| | velocity field | -52.7 | -36.9 | -55.0 | -43.2 | |
| | delay 2 | -37.5 | -38.6 | -34.8 | -40.4 | |

Table 22: Conservative ID controls on fresh episodes. Gain is the mean over the eight input maps. Loss is the largest preservation loss, clipped below at zero. Extra is the same over the additional delay, field and damping settings. Parentheses give $\alpha$, the inverse ridge, or the norm bound $b$.

| Setting | Reacher gain [95% CI] | Loss | Extra | Pusher gain [95% CI] | Loss | Extra |
|---|---|---|---|---|---|---|
| Frozen | +0.0 [+0.0, +0.0] | 0.0 | 0.0 | +0.0 [+0.0, +0.0] | 0.0 | 0.0 |
| ID | +18.6 [+12.4, +24.7] | 44.9 | 65.8 | +30.2 [+21.3, +39.2] | 107.0 | 97.0 |
| Partial ID (0.01) | +0.7 [+0.3, +1.0] | 0.3 | 0.2 | +0.7 [+0.6, +0.8] | 0.2 | 0.1 |
| Partial ID (0.025) | +1.4 [+0.7, +2.6] | 0.3 | 0.6 | +1.9 [+1.5, +2.4] | 0.0 | 0.9 |
| Partial ID (0.05) | +2.8 [+1.3, +4.3] | 0.1 | 1.4 | +4.1 [+3.2, +5.5] | 0.4 | 0.3 |
| Partial ID (0.1) | +4.6 [+3.2, +6.0] | 0.2 | 3.5 | +7.9 [+6.1, +10.0] | 1.9 | 3.3 |
| Partial ID (0.25) | +13.6 [+9.6, +17.5] | 1.4 | 10.3 | +17.4 [+14.1, +20.6] | 5.6 | 5.5 |
| Partial ID (0.5) | +20.9 [+15.3, +26.8] | 11.8 | 24.9 | +32.4 [+28.4, +37.4] | 29.4 | 21.2 |
| Ridge ID (0.1) | +20.6 [+14.8, +26.3] | 17.8 | 33.7 | +48.0 [+37.4, +59.5] | 33.0 | 35.5 |
| Ridge ID (1) | +17.2 [+10.7, +23.8] | 0.9 | 6.7 | +39.1 [+30.5, +47.7] | 25.5 | 13.4 |
| Ridge ID (10) | +5.9 [+2.8, +9.4] | 0.4 | 0.6 | +17.3 [+14.0, +20.8] | 6.3 | 10.8 |
| Ridge ID (100) | +0.6 [+0.4, +1.0] | 0.0 | 0.0 | +3.4 [+2.2, +5.0] | 0.5 | 1.7 |
| Bound ID (1) | +23.2 [+17.3, +29.1] | 1.2 | 0.0 | +55.9 [+46.5, +65.4] | 36.5 | 24.9 |
| Bound ID (1.5) | +22.3 [+16.1, +28.5] | 7.9 | 29.6 | +43.8 [+34.4, +53.8] | 71.5 | 65.1 |
| Bound ID (2) | +20.7 [+14.4, +27.4] | 36.7 | 52.9 | +35.6 [+27.0, +44.1] | 100.2 | 84.3 |
| closed-start gate forced open | +19.9 [+14.2, +26.2] | 33.0 | 46.8 | +39.1 [+32.7, +46.2] | 99.1 | 87.8 |
| Closed-start gate | +18.5 [+12.1, +25.4] | 0.4 | 1.1 | +14.9 [+9.9, +21.8] | 0.1 | 14.6 |
| ECA-m | +23.2 [+17.3, +29.1] | 0.0 | 0.2 | +51.7 [+41.0, +62.3] | 12.7 | 7.9 |
| **ECA** | +23.1 [+17.3, +29.0] | 0.0 | 0.3 | +51.5 [+40.7, +62.1] | 1.8 | 0.3 |

## G CONSERVATIVE IDENTIFICATION CONTROLS

**Design and provenance.** This additional experiment addresses the comparison between ECA and the simpler identification core. We fix the policy, nominal model, RLS update, residual calibration,

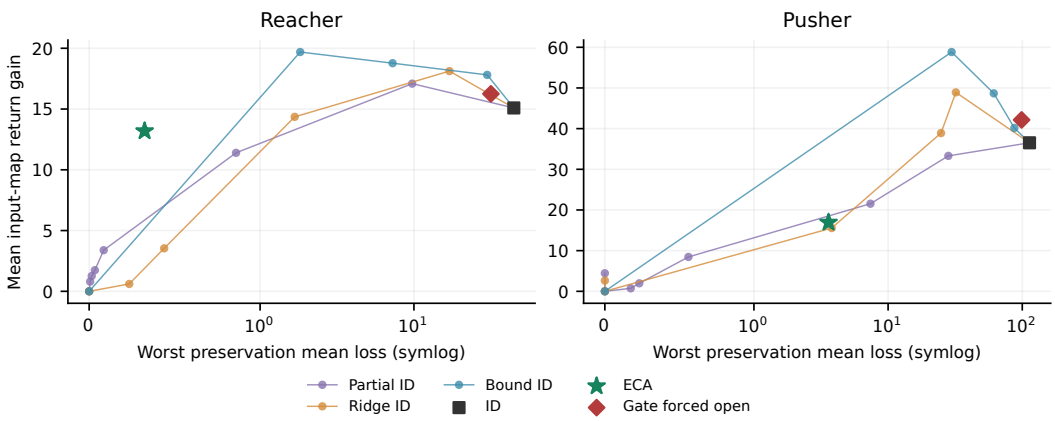



Figure 3: Development recovery and interference for the full control grid. Each point is one fixed setting. Lines connect settings within a baseline family; the horizontal axis uses a symmetric logarithmic scale with a linear region near zero.

action limits, and episode seeds across methods. The new runner reproduced all 66 checked original results exactly: frozen, ID, and ECA on both plants, eleven conditions, and the first episode of policy 200. New records store the policy, model, and script hashes.

Partial ID uses $u = a + \alpha(u_{\mathrm{ID}} - a)$ with $\alpha \in \{.01, .025, .05, .1, .25, .5\}$. Ridge ID replaces the inverse regularisation 0.01 with $\{.1, 1, 10, 100\}$ and leaves the estimator unchanged. Bound ID caps $\|u_{\mathrm{ID}}\|_2$ at $b\|a\|_2$ for $b \in \{1, 1.5, 2\}$. It adds no probing. The forced-open control uses the same inverse, norm bound, and probing amplitude as the closed-start gate, with its gate set to open before every action. This changes the timing of probing as a consequence of removing selective application. It is not a comparison with an identical noise schedule.

Development uses the development split and the fresh-episode evaluation the first cohort, with the seeds of Appendix D. All conditions have five episodes per policy and 100 steps. The original policy split was already inspected before this extension. No parameter of the closed-start gate was updated using these results, and ECA-m and ECA were each run once on them after their settings had been frozen on the development split.

**Selection and complete reporting.** The initial protocol selects a setting from each baseline family on development: maximise mean gain over eight input maps subject to a worst preservation mean loss of at most 1 on Reacher or 3 on Pusher. Frozen and ID are available endpoints. Before inspecting the new episode results, we expanded evaluation to all candidate settings. This avoids excluding a competitive setting just above a selection threshold. In particular, development Pusher ridge 10 has loss 3.8, close to ECA's 3.6, but neither satisfies the 3 screening budget. Table 22 therefore reports the full grid rather than only settings selected by that budget. Figure 3 plots the same grid on the recovery–interference plane, where the frontier is visible as a curve rather than a table. The budgets are screening choices, not safety guarantees. Protocols and amendments are retained with the run records.

**Metrics and uncertainty.** For each policy, we average paired gains across the eight input-map conditions, including gain loss. Intervals use 10,000 bootstrap resamples over the five policy means here and 4000 in the tables of Section 5; passing the preservation criterion is not a formal equivalence or a safety guarantee. The loss axis is the maximum of the four condition-wise mean losses relative to frozen, clipped below at zero. It measures observed interference, not tail risk or an upper confidence bound. The equal condition weights define a benchmark summary and do not estimate deployment fault frequencies. Complete condition-wise intervals and episode results are retained in the machine-readable supplement.

**Additional disturbance settings.** The five further tests are delays of one and three steps, a velocity field of strength 1, damping multiplied by 2, and a second orthogonal map. They do not participate

Table 23: Additional disturbances on the development split and on the third cohort: paired first-100-step gains for ID, Bound ID ($b = 1$), the closed-start gate, ECA-m, and ECA.

| Split | Plant | Method | Delay 1 | Delay 3 | Field 1 | Damping 2 | Mix 11 |
|---|---|---|---|---|---|---|---|
| Development | Reacher | ID | -0.4* | -60.9* | -0.1 | +0.0 | +8.6* |
| | Reacher | Bound ID ($b = 1$) | +0.0 | -1.1* | +0.3 | +0.0 | +9.3* |
| | Reacher | Closed-start gate | +0.0 | -1.3* | +0.3 | +0.0 | +0.7 |
| | Reacher | ECA-m | +0.0 | -0.3* | +0.5 | +0.0 | +9.3* |
| | Reacher | **ECA** | +0.0 | -0.3* | +0.8 | +0.0 | +9.5* |
| | Pusher | ID | -47.2* | -95.1* | -15.7 | -1.6* | -18.0* |
| | Pusher | Bound ID ($b = 1$) | -3.1* | -24.3* | -14.4* | -2.0* | +50.8* |
| | Pusher | Closed-start gate | +0.0 | -8.6* | -1.0 | -0.6* | +13.1* |
| | Pusher | ECA-m | -0.1* | -1.1 | +2.7 | +0.0 | +48.7* |
| | Pusher | **ECA** | -0.1 | -1.1 | +45.2* | +0.0 | +48.7* |
| Held-out | Reacher | ID | -2.2* | -72.4* | +0.4 | +0.0 | +20.3* |
| | Reacher | Bound ID ($b = 1$) | +0.0 | +0.2 | +0.6 | +0.0 | +22.5* |
| | Reacher | Closed-start gate | +0.0 | -1.9* | -0.1 | +0.0 | +17.5* |
| | Reacher | ECA-m | +0.0 | -0.4* | +0.4 | +0.0 | +22.5* |
| | Reacher | **ECA** | +0.0 | -0.4* | +0.2 | +0.0 | +22.3* |
| | Pusher | ID | -38.9* | -106.6* | -25.7* | -2.5* | -22.1* |
| | Pusher | Bound ID ($b = 1$) | -6.4* | -33.2* | -27.1* | -1.3* | +67.7* |
| | Pusher | Closed-start gate | +0.0 | -15.6* | -4.6 | -2.2* | +6.0 |
| | Pusher | ECA-m | -0.3* | -1.5 | -2.2 | -0.7 | +67.7* |
| | Pusher | **ECA** | -0.3* | -1.5 | +65.4* | -0.4 | +67.6* |

in baseline selection. They probe disturbance specificity and one additional map. Table 23 gives the same conditions on the development split, where the delayed-identity hypotheses were chosen, and on the third cohort; the fresh-episode results are in Table 24. The strength-one field and the doubled damping are both in the velocity-coupled class: on the development split ECA gains $45.2$ on the Pusher strength-one field where ECA-m gains $2.7$ and ID loses $15.7$, and neither gate loses on the damping. The held-out cohort repeats this: ECA gains $65.4$ on the Pusher strength-one field where ECA-m loses $2.2$, ID $25.7$, and Bound ID $27.1$; on Reacher, where the field exceeds the actuators' authority, it gains $0.2$. Doubled damping costs ECA $0.4$ on Pusher and nothing on Reacher, both intervals containing zero, and the two delays cost it at most $1.5$ where ID loses $38.9$ and $106.6$.

Table 24: Additional disturbances on fresh episodes of the first cohort, for the full control grid. Notation as in Table 1. The five conditions did not participate in any selection.

| Plant | Setting | Delay 1 | Delay 3 | Field 1 | Damping 2 | Mix 11 |
|---|---|---|---|---|---|---|
| Reacher | Frozen | +0.0 | +0.0 | +0.0 | +0.0 | +0.0 |
| | ID | -0.9* | -65.8* | +0.6 | -0.0 | +17.1* |
| | Partial ID (0.01) | -0.0* | -0.2* | -0.2 | -0.0 | +0.8* |
| | Partial ID (0.025) | -0.0* | -0.6* | +1.4* | -0.0 | +3.5* |
| | Partial ID (0.05) | -0.0* | -1.4* | +0.6 | -0.0 | +5.7* |
| | Partial ID (0.1) | -0.0* | -3.5* | +1.0* | -0.0 | +7.3* |
| | Partial ID (0.25) | -0.1* | -10.3* | +1.5* | -0.0 | +12.3* |
| | Partial ID (0.5) | -0.2* | -24.9* | +1.1 | -0.0 | +18.3* |
| | Ridge ID (0.1) | -0.5* | -33.7* | +1.3* | -0.0 | +15.5* |
| | Ridge ID (1) | -0.1* | -6.7* | +2.0* | -0.0 | +8.0* |
| | Ridge ID (10) | -0.0* | -0.6* | +1.6* | -0.0 | +4.3* |
| | Ridge ID (100) | -0.0* | -0.0* | +0.6 | -0.0 | +2.8* |
| | Bound ID (1) | -0.0* | +0.1 | +0.7 | +0.0* | +18.1* |
| | Bound ID (1.5) | -0.3* | -29.6* | +0.9 | -0.0 | +18.6* |
| | Bound ID (2) | -0.7* | -52.9* | +0.8 | -0.0 | +18.4* |
| | closed-start gate forced open | -1.5* | -46.8* | +0.5 | -0.7* | +17.9* |
| | Closed-start gate | +0.0 | -1.1 | +0.0 | +0.0 | +15.9* |
| | ECA-m | +0.0 | -0.2 | +1.0 | +0.0 | +18.0* |
| Pusher | Frozen | +0.0 | +0.0 | +0.0 | +0.0 | +0.0 |
| | ID | -29.9* | -97.0* | -31.8* | -3.1* | -15.6 |
| | Partial ID (0.01) | +0.0 | -0.1 | +1.2 | -0.1 | +0.9* |
| | Partial ID (0.025) | +0.0 | +0.6 | -0.9 | -0.1 | +1.7* |
| | Partial ID (0.05) | +0.0 | +1.1* | -0.1 | -0.3* | +5.9* |
| | Partial ID (0.1) | +0.0 | +1.2 | -3.3 | -0.3* | +10.7* |
| | Partial ID (0.25) | -0.5* | -3.3* | -5.5 | -0.4* | +20.2* |
| | Partial ID (0.5) | -3.9* | -21.2* | -14.1* | -0.8* | +23.8* |
| | Ridge ID (0.1) | -6.8* | -35.5* | -22.1* | -1.1* | +24.5* |
| | Ridge ID (1) | -0.6 | -0.4 | -13.4* | -0.4 | +36.9* |
| | Ridge ID (10) | +0.0 | +2.4* | -10.8* | -0.3 | +18.6* |
| | Ridge ID (100) | -0.5 | +0.2 | -1.7* | +0.0 | +3.8* |
| | Bound ID (1) | -7.2* | -24.8* | -24.9* | -1.0* | +50.4* |
| | Bound ID (1.5) | -14.0* | -65.1* | -28.3* | -1.6* | +6.0 |
| | Bound ID (2) | -20.0* | -84.3* | -30.3* | -2.0* | -14.1 |
| | closed-start gate forced open | -25.1* | -87.8* | -33.3* | -4.0* | -28.3* |
| | Closed-start gate | +0.0 | -14.6* | -4.9 | -1.8* | +1.6 |
| | ECA-m | -0.3 | +0.2 | -7.9* | -0.4 | +50.0* |