Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Do Confidence Gates Survive Structured Communication Loss? A Frozen-Policy Audit of Residual Multi-UAV Coordination

Satish Mishra

satishmsr37@gmail.com

Netaji Subhas University of Technology

Manisha Khulbe

Netaji Subhas University of Technology

Research Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11729253/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Do Confidence Gates Survive Structured Communication Loss? A Frozen-Policy Audit of Residual Multi-UAV Coordination

Satish Kumar Mishra[1,*] · **Manisha Khulbe**[1]

[1]Department of Electronics and Communication Engineering, Netaji Subhas University of Technology, New Delhi 110078, India

*Corresponding author. E-mail: satishmsr37@gmail.com; ORCID: 0009-0008-4781-1894
Contributing author E-mail: manisha.khulbe@nsut.ac.in

Abstract Confidence signals are widely used to decide when a learned residual policy should refine a nominal controller, but in networked teams the signal is itself produced by the communication process. We audit this execution-time design choice in simulated multi-UAV maritime search and tracking, where a sigmoid gate scales a 450-parameter residual by post-fusion belief confidence. Using six independently trained gated/always-on policy pairs with frozen weights, evaluated on 23 held-out seeds, the gate's advantage over an always-on residual under a nominal disc channel is small (pooled margin $+0.009$, 95% CI $[-0.002, +0.020]$). Zero-shot transfer to a path-loss and co-channel-interference channel lowers the margin in all six pairs (mean shift -0.026, CI $[-0.048, -0.005]$) and makes it negative on average (-0.017), whereas the always-on residual shifts by only -0.006. Counterfactual interventions show that the mechanism is pair-dependent: replacing only the scalar trigger or opening the gate removes the deficit in two pairs but not in the others. In the case-study pair, fidelity-matched i.i.d. trigger corruption predicts a margin of -0.003 against the observed -0.036, whereas persistent per-UAV dropout and a shared cross-UAV outage regime each reproduce about 70% of the excess. Training on the structured channel recovers most of the loss in that pair but does not surpass the always-on residual. Gate evaluations should therefore include an always-on control and temporally persistent, spatially correlated link loss.

Keywords Multi-UAV search and tracking · Residual policy learning · Confidence gating · Structured communication loss · Evolution strategies · Robustness audit

1 Introduction

Maritime search and tracking requires more than static area coverage: drifting targets must be discovered, handed over, and re-acquired while unmanned aerial vehicles (UAVs) act under partial local information [1–3]. Combining a deterministic coverage-and-pursuit base controller with a compact learned residual policy is attractive in this setting because the nominal search structure is preserved while bounded velocity corrections refine tracking [4, 5].

A natural extension is to grant the residual *selective authority* by scaling it with a confidence signal, so that the learned correction is applied only when the agent's own state estimate is

trustworthy. In a networked team, however, local belief confidence is formed after fusing onboard detections with sparse, delayed, or missing neighbor messages. Execution-time confidence gating therefore couples the communication process directly into the closed-loop control path: a gate calibrated under an idealized communication model can change behavior when the temporal and spatial structure of packet loss changes, even though the residual policy itself is untouched.

We ask a deliberately narrow question: *when the communication process changes after training, does the benefit of a confidence-gated residual over an always-on residual persist?* We answer it with a frozen-policy audit in a multi-UAV maritime search-and-tracking simulator. Every comparison is made against an always-on residual trained with the same budget, using six independently trained policy pairs and training-disjoint evaluation seeds, because a single training run in this setting varies by about as much as the effects of interest (Section 5).

We define a gate-trigger sufficiency margin against a matched always-on residual and report it per policy pair and pooled, with run-level and seed-level inference on training-disjoint seeds. The gate's nominal advantage turns out to be small, and moving to a path-loss and interference channel lowers it in all six pairs while the always-on residual is far less affected. Frozen-policy interventions then locate the deficit. Trigger starvation explains it in two pairs; in others the evidence points to residual-action mismatch or co-adaptation between trigger and network. In a Run 1 case study, matched memoryless trigger corruption explains less than one sixth of the deficit, whereas per-UAV dropout persistence and a shared cross-UAV outage regime each reproduce about 70% of the excess.

Checkpoints, per-seed results, and evaluation scripts for all reported pairs are available on request (Data Availability).

2 Related Work

Residual reinforcement learning adds a learned correction to a hand-designed controller [4, 5], and uncertainty estimates have guided when and how residuals are trained [6]. Gating a learned component by confidence relates to mixtures of experts [7] and selective prediction [8], but neural confidence is often miscalibrated [9] and degrades under dataset shift [10]. Here the shift lies not in the data but in the communication process that forms the confidence.

The communication literature approaches the problem from the other side. Learned communication [11, 12] and message-dropout training [13] ask how a policy should use, or tolerate the loss of, messages; in our setting the messages never reach the network directly but alter a belief state whose scalar summary sets policy authority. Work on estimation and control over lossy networks [14, 15], burst-error channels [16, 17], and information freshness [18, 19] shows that the temporal correlation of loss, and not only its rate, governs performance. We test that lesson on a learned multi-UAV controller with fidelity-matched synthetic corruption.

3 System Model and Problem Formulation

3.1 Multi-UAV maritime search and tracking model

We model outer-loop planar velocity guidance above onboard attitude and velocity autopilots. The environment combines partial local sensing and event-triggered message passing [20] within a hybrid information structure: each UAV i observes its own kinematics together with local belief,

Table 1: Evaluated scenario suite. r_s and r_c are the sensing radius and primary-channel disc communication radius.

Scenario	UAVs	Victims	Horizon (steps)	r_s (m)	r_c (m)
Default	4	2	80	140	260
Medium	5	3	95	120	220
Hard	6	4	110	95	180

Table 2: Core task, policy, optimization, and diagnostic RF channel parameters shared across experiments.

Quantity	Setting
Arena size and step duration	$1000 \times 1000\,\mathrm{m}$; $\Delta t = 1.0\,\mathrm{s}$
Maximum UAV speed and track radius	$v_{\max} = 28\,\mathrm{m/step}$; $r_{\mathrm{track}} = 80\,\mathrm{m}$
Occupancy belief representation	20×20 grid; message fusion blend $\beta = 0.35$
Event-triggered transmit thresholds	conf. ≥ 0.14, $\lvert\Delta c\rvert \geq 0.025$, or stale ≥ 6 steps
Residual observation and network	25-D input; $25 \to 16 \to 2$ tanh MLP (450 params)
Residual authority and soft gate	scale $\alpha = 0.25$; slope $k = 20$, threshold $\tau = 0.08$
Evolutionary optimization (ES)	population 24, 6 elites, 100 iterations, $\sigma_0 = 0.12$
RF path-loss parameters	$\rho = 9.9 \times 10^{-5}$, $d_0 = 1.0\,\mathrm{m}$, $\eta = 2.2$, $P_{\mathrm{tx}} = 0.1\,\mathrm{W}$
RF noise and outage	$N_0 = 1.585 \times 10^{-20}\,\mathrm{W/Hz}$, $B = 1\,\mathrm{MHz}$, $\kappa_c = 10.0$
RF packet framing	$L_{\mathrm{msg}} = 12{,}800\,\mathrm{bits}$, protocol overhead $\tau_0 = 5\,\mathrm{ms}$
Evaluation seeds	30 per scenario (7 training, 23 training-disjoint)

coverage, and message-derived summaries, while its 25-dimensional residual input o_i also includes the peak coordinates and confidence of the team-wide mean belief grid, and base lane/target assignment uses shared state. The evaluated controller is therefore not strictly decentralized. Crucially, the scalar gate trigger c_i is computed strictly from UAV i's own post-fusion belief grid. At each 1 s step each UAV commands a horizontal velocity vector, and a victim is counted as tracked whenever at least one UAV lies within $r_{\mathrm{track}} = 80\,\mathrm{m}$; the tracking ratio is the fraction of victim–step pairs that are tracked.

Three scenario scales are evaluated (Table 1) with shared parameters in Table 2. Victims follow biased stochastic drift. The medium scenario is the primary audit regime because it is where the gate shows its largest nominal advantage; default and hard scenarios test scaling.

Each UAV maintains a 20×20 occupancy belief grid. Local detections update the grid directly, delivered neighbor messages fuse into the receiver's grid with weight $\beta = 0.35$, and all cells undergo drift prediction and negative-observation decay each step. Under the primary communication model, transmissions are event-triggered when local peak confidence exceeds 0.14, changes by at least 0.025, or reaches a 6-step staleness limit, and messages succeed deterministically within disc radius r_c (subject to optional independent drops).

3.2 Frontier–belief base controller and gated residual policy

In the deterministic base controller, detected victims receive direct pursuit priority; when no victim is locally assigned, a UAV steers toward a weighted combination of its lane center and the least-recently covered cell in its assigned lane, following frontier-based multi-robot exploration [21]. Priority ownership and handoff windows reduce redundant clustering on a single target.

Over this base controller, a $25 \to 16 \to 2$ tanh multilayer perceptron f_θ (450 parameters)

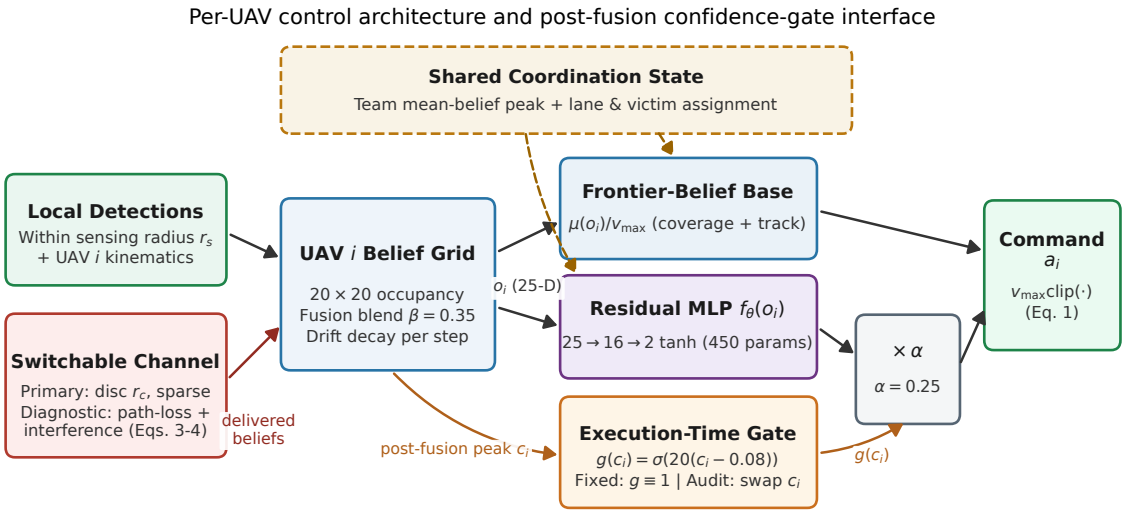


Figure 1: Per-UAV control architecture and information boundaries. Local detections and messages delivered across the switchable channel update UAV i's belief grid, whose post-fusion peak c_i drives the execution-time gate $g(c_i)$. Dashed orange paths indicate shared coordination features (team-wide mean-belief peak and assignment state).

outputs a bounded two-dimensional correction:

$$a_i = v_{\max} \operatorname{clip}\left(\frac{\mu(o_i)}{v_{\max}} + \alpha g(c_i) f_\theta(o_i), -1, 1 \right), \tag{1}$$

where $\mu(o_i)$ is the base velocity, $\alpha = 0.25$ bounds residual authority, and $c_i \in [0, 1]$ is the peak cell value of UAV i's belief grid after fusing delivered messages at the current step. The *always-on* (fixed) residual uses $g(c_i) \equiv 1$; the *soft-gated* policy uses

$$g(c_i) = \sigma(20(c_i - 0.08)), \tag{2}$$

with $\sigma(x) = (1 + e^{-x})^{-1}$ (a binary-gated baseline uses $\mathbb{I}[c_i \geq 0.08]$). Because $g(c_i)$ scales the issued correction online, any communication-induced distortion of c_i directly alters closed-loop actuation (Fig. 1). The gate parameters (k, τ) were fixed by hand and not tuned per channel.

3.3 Diagnostic RF channel and sufficiency margin

To evaluate sensitivity beyond the idealized disc channel, we use a switchable diagnostic channel with log-distance path loss, co-channel interference, and Rayleigh-fading outage [22, 23]:

$$G_{ij} = \rho(d_{ij} + d_0)^{-\eta}, \qquad \mathrm{SNR}_{ij} = \frac{P_{\mathrm{tx}} G_{ij}}{N_0 B + I_{ij}}, \tag{3}$$

$$R_{ij} = B \log_2(1 + \mathrm{SNR}_{ij}), \qquad P_{ij}^{\mathrm{succ}} = \exp(-\kappa_c / \mathrm{SNR}_{ij}), \tag{4}$$

where d_{ij} is inter-UAV distance, $I_{ij} = \sum_{m \neq i,j} P_{\mathrm{tx}} G_{mj}$ sums power received at UAV j from other simultaneously transmitting UAVs, and a delivered packet incurs transmission plus framing delay $\tau_{ij} = L_{\mathrm{msg}}/R_{ij} + \tau_0$ (deferred by one step if $\tau_{ij} > \Delta t$). Under concurrent transmissions the effective delivery range contracts into the 150–250 m regime (Fig. 2). This is an analytical engineering stress model, not a measured maritime radio. In the medium scenario the delivered fraction of sent messages falls from 0.38–0.54 under the disc channel to 0.16–0.17 under this

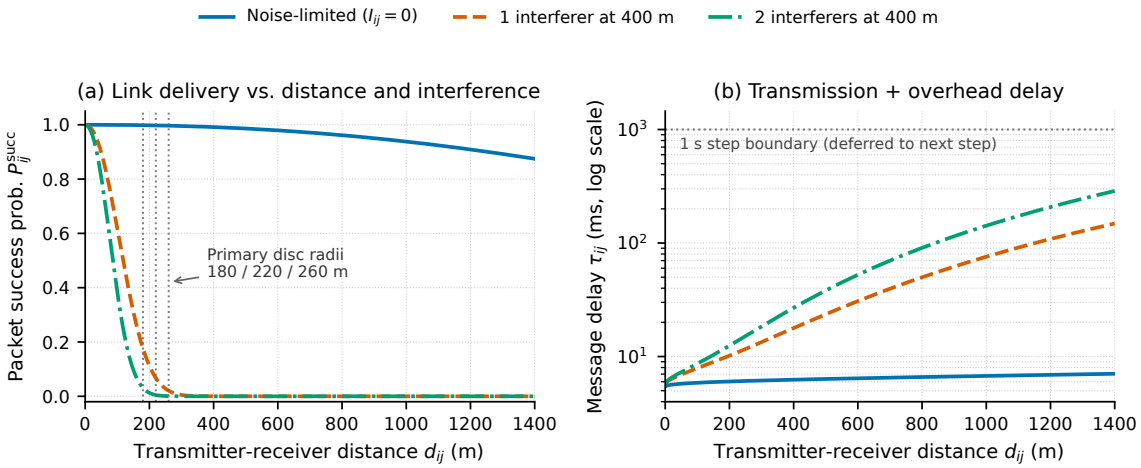


Figure 2: Diagnostic RF channel characteristics from Eqs. (3)–(4) (Table 2). (a) Packet success probability contracts into the 150–250 m regime when one or two co-channel interferers transmit at 400 m, spanning the primary disc radii. (b) Delivered-message delay remains two orders of magnitude below the 1 s simulation step.

channel (Run 1 policies), and no UAV exhausted its battery in any episode, so the energy model enabled in this condition is not binding.

For a policy pair, scenario, channel, and seed set, let $J(\cdot)$ be the mean tracking ratio. The *gate-trigger sufficiency margin* is

$$M(z) = J(\text{soft gate driven by trigger } z) - J(\text{always-on residual}), \tag{5}$$

so $M > 0$ means selective gating beats constant residual authority. The *channel shift* is $\Delta M = M_{\text{physical}} - M_{\text{primary}}$.

4 Experimental Protocol

4.1 Training, seeds, and inference

Fixed, binary-gated, and soft-gated residual networks are trained with an identical elite-selection evolution strategy [24] (population 24, 6 elites, 100 iterations, $\sigma_0 = 0.12$, seven training seeds; each candidate is evaluated on all three scenarios and the next mean is the unweighted average of the top six). Network topology, base controller, and optimization budget are held constant, isolating the gating rule. We train *six independent fixed/soft pairs* (Runs 1–6, different ES seeds) on the primary channel.

Evaluation uses 30 fixed seeds per scenario: the 7 training seeds plus 23 training-disjoint seeds. Because the training seeds are part of the 30, **all pooled and per-pair claims below use the 23 training-disjoint seeds only**; 30-seed values are reported where an analysis was run only at that size (Run 1 case study, Section 5.5) and are labelled. Differences are matched-seed paired, with 95% t-intervals. For pooled statements we give two intervals: *run-level* (treating the six pairs as the random sample, $n = 6$, the conservative choice for generalizing over training runs) and *seed-level* (averaging margins over pairs per seed, then $n = 23$). With only six pairs the run-level interval is wide; we state it rather than rely on the narrower one.

Table 3: Five-stage frozen-policy audit workflow.

Audit stage	Manipulated variable	Frozen controls	Diagnostic question
1. Nominal baseline	Policy architecture	Task, seeds, ES budget	Does the gate beat the always-on residual?
2. Channel transfer	Message delivery model	Trained weights, scenario, seeds	Does the margin survive structured RF loss?
3. Trigger substitution	Scalar input z to $g(\cdot)$	Residual weights, RF state and o_i	Is the corrupted confidence driving the loss?
4. Gate bypass / retune	Authority mapping $g(\cdot)$	Soft-gate weights, RF task, seeds	Is the residual still effective without suppression?
5. Synthetic corruption	Trigger persistence and cross-UAV synchrony	Primary task, weights, mean trigger fidelity	Do dark periods and synchrony govern the loss?

Role of Run 1. Run 1 is the pair on which the system was developed, and the counterfactual and mechanism analyses of Sections 5.3 and 5.5 were first carried out on it. It is also, together with Run 5, the pair with the largest channel shift, so those analyses should be read as a case study, not as a representative sample. Section 5.3 repeats the main interventions on all six pairs.

4.2 Audit workflow

Observing a tracking drop after switching channels cannot by itself separate invalid residual actions from mistimed gate suppression. Holding policy weights, task kinematics, and episode seeds fixed, the audit proceeds in five stages (Table 3).

5 Results

5.1 Nominal baseline

Table 4 gives the primary-channel baseline for the Run 1 pair (30 seeds). Training the 450-parameter MLP without the base controller (Pure MLP) yields poor tracking (0.148–0.393), confirming the need for the frontier–belief prior. The always-on residual improves on the hybrid base in all three scenarios (paired t-test: default $p < 0.001$, medium $p = 0.012$, hard $p < 0.001$); because both achieve 100% eventual victim discovery, the gain comes from faster first contact and tighter pursuit (hard-scenario mean first detection drops from 3.37 to 2.37 steps). The soft gate has the highest mean on default and medium, but its medium advantage over the always-on residual (+0.009 for this pair) is within the seed-to-seed spread, as the pooled analysis below shows.

5.2 Channel transfer across six independently trained pairs

Table 5 and Fig. 3 report the medium-scenario margin for all six pairs on the 23 training-disjoint seeds, with no retraining.

The nominal advantage of gating is small. The pooled margin is +0.009, its run-level interval includes zero, and only Run 5 is individually significant. The channel shift is the sturdier result: it is negative in every pair, and the pooled value of −0.026 has a run-level interval that excludes zero. On the physical channel the pooled margin is −0.017 (run-level interval just excluding zero), negative in four pairs and neutral in Runs 2 and 4.

Table 4: Primary sparse-communication tracking ratio, Run 1 pair (mean ± sample standard deviation across 30 seeds). Bold: highest mean per scenario.

Scenario	Hybrid base	Pure MLP	Always-on residual	Binary gate	Soft gate
Default	0.869±0.008	0.251±0.061	0.877±0.010	0.868±0.009	**0.881±0.008**
Medium	0.879±0.045	0.148±0.089	0.904±0.039	0.896±0.029	**0.913±0.019**
Hard	0.905±0.003	0.393±0.047	**0.912±0.003**	0.902±0.003	0.910±0.003

Table 5: Medium-scenario gate margin M and channel shift ΔM per policy pair, 23 training-disjoint seeds, no retraining. †: paired 95% t-interval excludes zero. The last row pools the six pairs; its interval treats the pairs as the random sample ($n = 6$).

Training pair	Primary M	Physical M	Shift ΔM
Run 1	$+0.011$	$-0.037^{\dagger}$	$-0.047^{\dagger}$
Run 2	$+0.009$	$+0.003$	-0.006
Run 3	-0.002	$-0.024^{\dagger}$	-0.022
Run 4	$+0.002$	$+0.001$	-0.001
Run 5	$+0.027^{\dagger}$	$-0.023^{\dagger}$	$-0.050^{\dagger}$
Run 6	$+0.010$	$-0.025^{\dagger}$	$-0.034^{\dagger}$
Pooled (run-level CI)	$+0.009$ $[-0.002, +0.020]$	-0.017 $[-0.034, -0.001]$	-0.026 $[-0.048, -0.005]$
Pooled (seed-level CI)	$[+0.000, +0.018]$	$[-0.029, -0.005]$	—
Pairs with negative value	1 of 6	4 of 6	6 of 6

The always-on residual is not immune but is far more stable: its primary-to-physical change is -0.006 on average (range -0.017 to $+0.014$ across pairs), compared with -0.033 on average for the soft gate. In Run 1 the always-on residual shifts by only -0.005 while the soft gate falls from 0.912 to 0.860 on the held-out seeds, to the level of the unlearned hybrid base. Across scenarios (Run 1 pair, 30 seeds, Table 6) the always-on residual is the strongest or tied policy on the physical channel in every regime.

Memoryless packet loss does not explain the shift. On the hard scenario under the primary channel, sweeping additional i.i.d. drop probability from 0% to 60% without retraining changes mean tracking by at most 0.001 for all three controllers, even though delivered messages per episode fall by more than 60% (220 → 85 for the always-on residual and 215 → 80 for the soft gate). Under the event-triggered, range-limited protocol, memoryless erasures barely matter. Under combined GPS noise (up to 5 m), range-dependent detection decay, and actuation lag on the hard scenario, the always-on residual leads under clean, mild, and moderate profiles and falls to 0.883 under the severe profile, where the hybrid base (0.895) and soft gate (0.894) are slightly higher. On the medium scenario the physical channel removes roughly 60–75% of delivered messages, a loss level comparable to the i.i.d. sweep (which was run on the hard scenario), yet it changes the gate margin where i.i.d. loss does not; Section 5.5 traces this to long, correlated dark periods.

5.3 Counterfactual interventions

A drop under the physical channel could arise because the soft-gated network outputs degraded actions under RF-corrupted observations o_i, or because the gate suppresses valid corrections when received confidence falls. We separate these with frozen weights and unchanged physical-channel dynamics and observations: (i) *trigger swap*: replace only the scalar c_i in Eq. (2) with

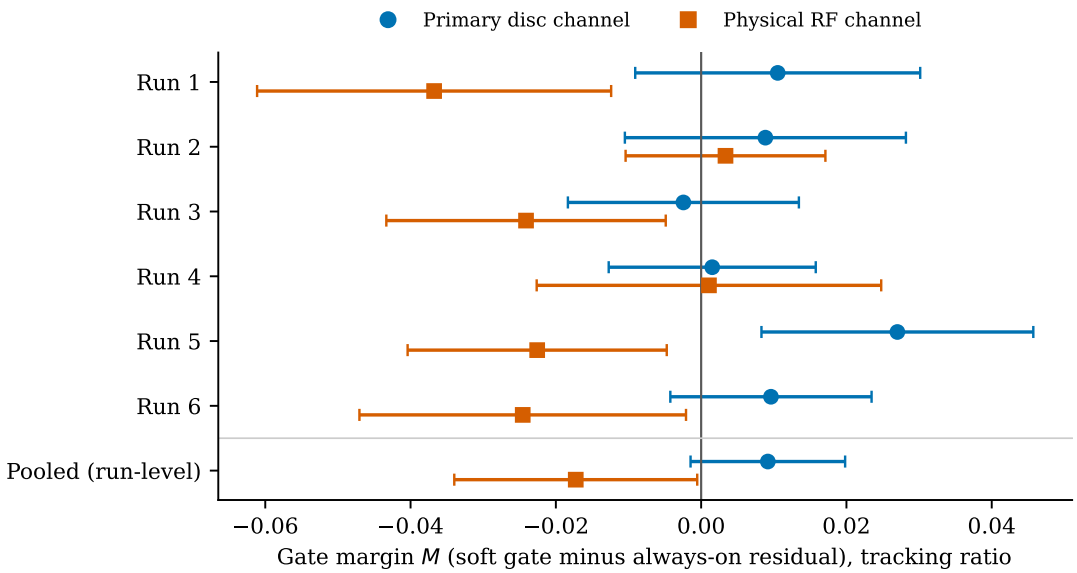


Figure 3: Gate margin M per policy pair under the primary disc channel and the physical RF channel (23 training-disjoint seeds, 95% paired t-intervals; pooled row uses run-level intervals). Positive values mean the gate beats the always-on residual.

Table 6: Run 1 pair: zero-shot tracking ratio across scenario and channel combinations (30 seeds, no retraining).

Channel	Scenario	Hybrid base	Always-on	Soft gate
Primary	Default	0.869	0.877	**0.881**
	Medium	0.879	0.904	**0.913**
	Hard	0.905	**0.912**	0.910
Physical	Default	0.867	**0.879**	0.877
	Medium	0.864	**0.899**	0.863
	Hard	0.905	**0.913**	0.909

its synchronized primary-channel counterfactual; (ii) *gate open*: $g \equiv 1$.

Run 1 case study (30 seeds). Table 7 and Fig. 4 show that swapping only the trigger restores tracking from 0.863 to 0.898 (recovering 97% of the deficit relative to the 0.899 physical always-on residual), and opening the gate reaches 0.914. Retuning the threshold τ after the fact recovers 0.909, but only by driving the gate toward near-constant activation, while a receiver-freshness fallback (open after four consecutive steps without reception, 0.914 on this pair) does not generalize across scenarios: it lowers physical-default tracking (0.877 → 0.862) and leaves physical-hard nearly unchanged (0.909 → 0.911). A primary-channel staleness sweep shows the margin stays positive up to 6 steps of artificial trigger delay and turns negative between 6 and 9, whereas zero-mean additive trigger noise has little effect.

Gate-calibration statistics (Table 8) show why suppression occurs. Under physical-channel delivery, mean post-fusion confidence falls from 0.209 to 0.106, and the fraction of UAV–steps with $c_i < 0.08$ *even though the local belief peak is already within sensing range of a victim* rises from 0.130 to 0.226. The counterfactual trigger reverses this under-confident suppression

Table 7: Run 1 pair, physical medium condition (30 seeds). Every row except the separately trained always-on reference runs the same soft-residual network.

Intervention on soft-gated policy	Tracking	Margin vs. always-on
Always-on residual reference (separately trained)	0.899	—
Unmodified physical-channel trigger c_i	0.863	−0.036
Substitute synchronized primary-channel scalar trigger	0.898	−0.001
Best post-hoc threshold retune	0.909	+0.010
Constant-open bypass ($g \equiv 1$)	**0.914**	+0.015
Receiver-freshness fallback (open after 4 dark steps)	**0.914**	+0.015

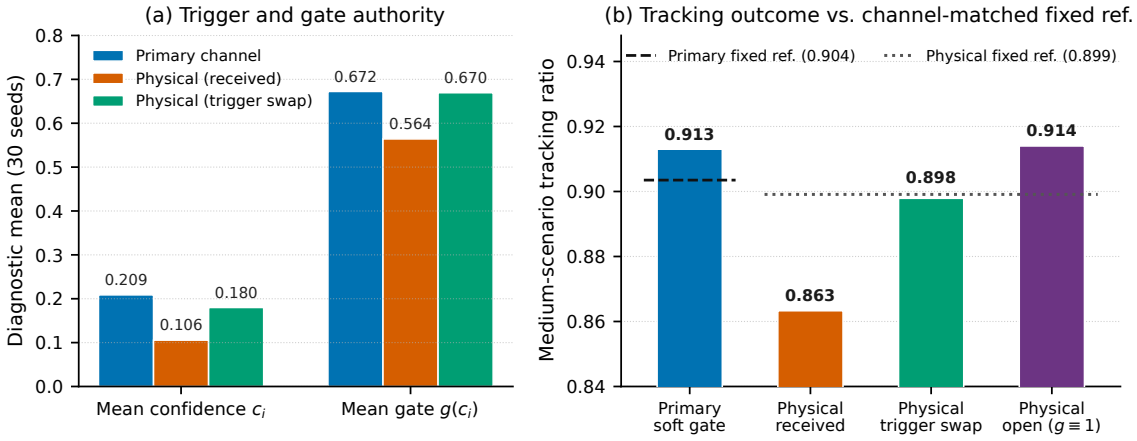


Figure 4: Run 1 pair, medium scenario (30 seeds). (a) Structured RF delivery halves mean post-fusion confidence c_i and depresses mean gate authority $g(c_i)$; substituting the primary-channel scalar trigger restores both while retaining physical-channel observations o_i. (b) Tracking ratios against channel-matched always-on references (0.904 primary, 0.899 physical).

(0.151).

All six pairs (23 training-disjoint seeds). Table 9 repeats the interventions across pairs. Trigger starvation accounts for the physical-channel deficit cleanly in Runs 1 and 6 (swap −0.004 and 0.000; open +0.016 and +0.020). It does not in the others: Run 5 degrades under the swap (−0.026) and recovers only partially when opened (−0.015); Run 4 degrades further when opened (−0.022); Run 3 improves only partially (−0.004 open, −0.012 swap); Run 2 has no deficit to explain. Pooled over the six pairs, opening the gate gives a margin of −0.001 (run-level interval $[-0.019, +0.016]$)—parity with the always-on residual, not a gain—and the trigger swap alone still leaves −0.011 ($[-0.021, -0.002]$; negative in all six pairs). The swap probably restores a trigger distribution the policy was optimized for, rather than information the policy lacks; whether the network and the gate co-adapt during ES training is not separated here and we do not claim it.

5.4 Robustness checks

Other belief-derived triggers. We tested belief concentration (1− normalized Shannon entropy) and top-9-cell belief mass as alternative triggers, keeping the trained weights fixed and calibrating sigmoid thresholds (30th percentile) and slopes (4/IQR) on the seven primary-channel

Table 8: Run 1 pair, medium scenario gate diagnostics (30 seeds, 14,250 UAV–steps). "Low-conf. aligned": steps with $c_i < 0.08$ while the local belief peak lies within r_s of a victim. "Mismatch" sums low-conf. aligned and high-conf. misaligned fractions.

Condition	Track.	Mean c_i	Mean $g(c_i)$	Low-conf. aligned	Mismatch
Primary channel	0.913	0.2088	0.6724	0.1300	0.1457
Physical, received	0.863	0.1056	0.5642	0.2260	0.2333
Physical, trigger swap	0.898	0.1800	0.6695	0.1508	0.1632

Table 9: Physical-medium counterfactual margins versus each pair's matched always-on residual, 23 training-disjoint seeds, no retraining. †: paired 95% t-interval excludes zero. Last row: pooled over pairs (run-level CI).

Training pair	Received trigger	Trigger swap	Constant open ($g \equiv 1$)
Run 1	$-0.037^{\dagger}$	-0.004	$+0.016$
Run 2	$+0.003$	-0.009	-0.003
Run 3	$-0.024^{\dagger}$	-0.012	-0.004
Run 4	$+0.001$	-0.017	$-0.022^{\dagger}$
Run 5	$-0.023^{\dagger}$	$-0.026^{\dagger}$	-0.015
Run 6	$-0.025^{\dagger}$	0.000	$+0.020^{\dagger}$
Pooled	-0.017 $[-0.034, -0.001]$	-0.011 $[-0.021, -0.002]$	-0.001 $[-0.019, +0.016]$

training seeds. A recalibrated peak-confidence control closely reproduces the Run 1 transfer (-0.037 vs. -0.036). In Runs 1–3 (30 seeds; Table 10) the physical-channel margin is lower than the primary-channel margin for all three trigger signals, with large drops in Run 1 (and Run 3 peak) and smaller ones for the entropy and mass triggers in Runs 2 and 3. The shift is therefore not specific to peak confidence.

Channel parameterizations. Over a 5×3 RF grid of path-loss exponent $\eta \in \{2.0, 2.2, 2.5, 2.8, 3.1\}$ and outage threshold $\kappa_c \in \{5, 10, 20\}$ (Runs 1–3, 30 seeds per cell), mean per-UAV consecutive no-delivery ("dark") run length spans 7.5 to 22.9 steps. The margin is negative in all 15 cells for Run 1 (mean -0.026) and Run 3 (mean -0.020) and near zero for Run 2 (mean $+0.004$, negative in 3 of 15 cells). Within Runs 1 and 3 longer dark runs, higher dark fractions, and stronger cross-UAV synchrony correlate moderately with worse margins ($r \approx -0.30$ to -0.46); pooling across runs weakens this ($|r| \leq 0.23$) because of between-checkpoint variance.

Retraining on the structured channel. We trained one physical-channel always-on residual (ES seed 301) and two physical-channel soft gates (ES seeds 302 and 312) with the same budget and seeds (Table 11). Retraining raises Run 1's medium-scenario soft-gate tracking from 0.863 to 0.887 (seed 302, margin -0.011, 95% CI $[-0.027, +0.005]$) and 0.900 (seed 312, margin $+0.001$, CI $[-0.010, +0.012]$) relative to the physical-trained always-on residual (0.899). On default and hard scenarios the physical-trained soft gate (seed 302) trails the always-on residual by 0.005 and 0.006. In-channel training mitigates the zero-shot loss without beating an always-on residual, so under this protocol the gate is worth keeping only if it is trained and validated on the deployment link statistics.

Table 10: Medium-scenario margin shift $M_{\text{primary}} \rightarrow M_{\text{physical}}$ across three post-fusion belief triggers (Runs 1–3, 30 seeds, no network retraining).

Belief trigger signal	Run 1	Run 2	Run 3
Peak confidence (recalibrated)	$+0.005 \rightarrow -0.037$	$+0.010 \rightarrow +0.000$	$-0.002 \rightarrow -0.021$
Concentration ($1 -$ entropy)	$+0.007 \rightarrow -0.031$	$-0.003 \rightarrow -0.008$	$-0.006 \rightarrow -0.014$
Top-9-cell belief mass	$+0.007 \rightarrow -0.030$	$+0.000 \rightarrow -0.001$	$-0.004 \rightarrow -0.016$

Table 11: Physical-channel medium-scenario evaluation (30 seeds): zero-shot primary-trained policies versus policies trained on the physical channel. Both retrained gates are compared with the physical-trained always-on residual (seed 301).

Training regime	Always-on	Soft gate	Margin M
Primary channel (zero-shot transfer, Run 1)	0.8991	0.8633	-0.0358
Physical channel (retrained, seed 302)	0.8985	0.8872	-0.0113
Physical channel (retrained, seed 312)	0.8985	0.8998	$+0.0013$

5.5 Why independent trigger corruption cannot explain the Run 1 deficit

This section uses the Run 1 pair (30 seeds) only; the synthetic-corruption models are calibrated to its physical-channel statistics, so the conclusions are about this pair.

Fidelity-matched independent corruption. To test whether the physical-channel deficit merely reflects lower average correlation with clean confidence, we corrupted only the primary-channel scalar trigger over a 6×4 grid of i.i.d. sample-and-hold staleness and additive Gaussian noise while executing on the primary channel. Across the 24 cells the margin follows a log-linear trend ($R^2 = 0.917$) in trigger fidelity φ, the squared Pearson correlation with clean primary-channel confidence. At the physical channel's empirical fidelity ($\varphi = 0.0646$), surrounding i.i.d. cells give margins of -0.006 to -0.0004 (midpoint -0.003), against -0.036 observed. Matching marginal fidelity under memoryless corruption is insufficient.

Persistent per-UAV dark periods. Physical delivery produces prolonged, heterogeneous per-UAV dark periods: the mean consecutive no-delivery length rises from 8.3 steps (90th percentile 17.6) under rate-matched Bernoulli loss to 13.3 steps (90th percentile 40.4), and across-UAV dispersion in dark-step fraction triples (0.081 vs. 0.027). Driving the gate trigger with a per-UAV two-state Markov dropout process calibrated to these statistics gives margins of -0.025 to -0.028; interpolated at $\varphi = 0.0646$, persistence accounts for roughly 70% of the excess deficit beyond i.i.d. corruption (Table 12).

Cross-UAV synchrony. Physical outages are synchronized across vehicles through shared geometry and co-channel interference: the probability that at least two UAVs receive packets given that at least one is connected is 0.551, versus 0.217 under circularly time-shifted independent sequences ($z = 25.4$, permutation $p < 0.002$ over 500 shuffles). Coupling per-UAV dropout to a shared team connectivity regime calibrated to this coincidence rate gives $M = -0.026$ at matched fidelity, versus -0.016 for an independent control with identical per-UAV marginals. In terms of the excess deficit, the independent-marginal control reproduces about 40% and the shared regime about 70%.

Table 12: Run 1 pair: causal mechanism tests for the medium-scenario gate collapse (30 seeds). Synthetic rows corrupt only the scalar gate trigger under primary-channel execution at matched trigger fidelity ($\varphi \approx 0.065$).

Trigger condition	Margin M
Primary channel, unmodified fresh trigger	$+0.009$
Fidelity-matched i.i.d. trigger corruption (surrounding grid cells)	-0.006 to -0.0004
Independent per-UAV connectivity control (matched marginals)	-0.016
Per-UAV Markov dropout matched to physical persistence	-0.026
Shared-regime cross-UAV connectivity synchrony (matched $z = 25.4$)	-0.026
Diagnostic physical RF channel, received trigger	-0.036

The persistence model and the shared-regime model are separate calibrations that each reach about 70%; we did not evaluate a joint model, so we do not claim that the two effects add, and about 30% of the excess deficit remains unexplained. What the evidence supports is that temporal persistence and cross-UAV correlation of link loss, rather than average delivery rate or memoryless erasure, are what a gate calibrated on independent drops misses.

6 Discussion and Limitations

What the evidence supports. Across six independently trained pairs, a belief-confidence gate gives at most a small nominal advantage over an always-on residual, and the change from a disc channel to a path-loss/interference channel lowers that advantage in every pair. The always-on residual, which does not read the communication-derived confidence, changes far less. Counterfactuals show that the cause differs between pairs: in two it is suppression by a starved trigger, in the others the swap or the open gate does not recover parity. This heterogeneity is itself a result: conclusions drawn from a single gated/always-on pair, including our own Run 1, can be misleading in either direction.

Design guidance. A confidence gate should be judged against an always-on residual of equal training budget, not only against the base controller, and on link models with temporally persistent and spatially correlated loss: i.i.d. erasure sweeps can leave tracking unchanged where a structured channel does not. Reporting the trigger's own distribution (mean confidence, and the fraction of low-confidence steps with the target in range) alongside task performance makes such failures visible. Fallbacks such as freshness-based opening need validation per team density, since this one did not transfer across scenarios.

Limitations. The simulator is a 2-D outer-loop model with 4–6 UAVs and no collision avoidance, and its information structure is hybrid: the team-wide mean-belief peak and assignment state are shared, so the controller is not fully decentralized. The RF channel is analytical (path loss, interference, Rayleigh outage) rather than an empirical trace with MAC scheduling or real maritime propagation. Six training pairs give wide run-level intervals, and only the shift ΔM is clearly separated from zero; the nominal gate advantage is not. The counterfactual and synthetic-corruption analyses were developed on Run 1 and are reported at 30 seeds including the training seeds, so their conclusions are specific to that pair. The gate itself is a single design, a sigmoid on peak confidence with hand-set k and τ paired with a 450-parameter residual; we

did not compare learned gates, calibrated or ensemble uncertainty, or other residual-training algorithms, and we did not train on a mixture of channels, the most natural remedy and a direct next step. Finally, all results are simulation-only and carry no claim about flight robustness.

7 Conclusion

In a multi-UAV search-and-tracking simulator, an execution-time confidence gate offers a small and statistically fragile advantage over an always-on residual under a nominal communication model, and that advantage is reduced in all six independently trained pairs under structured, path-loss and interference-driven loss, becoming negative on average. Frozen-policy interventions show that the mechanism varies across pairs, and fidelity-matched synthetic corruption in a case-study pair shows that persistence and cross-UAV correlation of outages, not memoryless packet loss, are what matter. Confidence gates driven by fused neighbor information should be evaluated against an always-on control, on training-disjoint seeds, and under temporally persistent and spatially correlated link loss.

Statements and Declarations

Funding No funds, grants, or other support were received.

Competing Interests The authors have no competing interests to declare.

Author Contributions Satish Kumar Mishra developed the simulation software, performed the experiments and analysis, and wrote the original draft. Manisha Khulbe supervised the study and reviewed and edited the manuscript. Both authors approved the final manuscript.

Data Availability The simulator source, saved policy checkpoints, per-seed result files, and evaluation scripts supporting this study are available from the corresponding author upon reasonable request.

References

[1] Cyril Robin and Simon Lacroix. Multi-robot target detection and tracking: Taxonomy and survey. *Autonomous Robots*, 40(4):729–760, 2016. doi: 10.1007/s10514-015-9491-7.

[2] Soon-Jo Chung, Aditya A. Paranjape, Philip Dames, Shaojie Shen, and Vijay Kumar. A survey on aerial swarm robotics. *IEEE Transactions on Robotics*, 34(4):837–855, 2018. doi: 10.1109/TRO.2018.2857475.

[3] Sonia Waharte and Niki Trigoni. Supporting search and rescue operations with UAVs. In *Proceedings of the International Conference on Emerging Security Technologies (EST)*, 2010.

[4] 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 *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6023–6029, 2019. doi: 10.1109/ICRA.2019.8794127.

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

[6] Lakshita Dodeja, Karl Schmeckpeper, Shivam Vats, Thomas Weng, Mingxi Jia, George Konidaris, and Stefanie Tellex. Accelerating residual reinforcement learning with uncertainty estimation. *arXiv preprint arXiv:2506.17564*, 2025.

[7] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.

[8] Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 4878–4887, 2017.

[9] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.

[10] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

[11] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, 2016.

[12] Sainbayar Sukhbaatar, Arthur Szlam, and Rob Fergus. Learning multiagent communication with backpropagation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, pages 2244–2252, 2016.

[13] Woojun Kim, Myungsik Cho, and Youngchul Sung. Message-dropout: An efficient training method for multi-agent deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[14] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I. Jordan, and Shankar S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, 2004. doi: 10.1109/TAC.2004.834121.

[15] Luca Schenato, Bruno Sinopoli, Massimo Franceschetti, Kameshwar Poolla, and Shankar S. Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, 2007. doi: 10.1109/JPROC.2006.887306.

[16] Edgar N. Gilbert. Capacity of a burst-noise channel. *Bell System Technical Journal*, 39 (5):1253–1265, 1960. doi: 10.1002/j.1538-7305.1960.tb03959.x.

[17] Edwin O. Elliott. Estimates of error rates for codes on burst-noise channels. *Bell System Technical Journal*, 42(5):1977–1997, 1963. doi: 10.1002/j.1538-7305.1963.tb00955.x.

[18] Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *Proceedings of IEEE INFOCOM*, pages 2731–2735, 2012. doi: 10.1109/INFCOM.2012.6195689.

[19] Roy D. Yates, Yin Sun, D. Richard Brown, Sanjit K. Kaul, Eytan Modiano, and Sennur Ulukus. Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, 39(5):1183–1210, 2021. doi: 10.1109/JSAC.2021.3065072.

[20] Frans A. Oliehoek and Christopher Amato. *A Concise Introduction to Decentralized POMDPs*. Springer, 2016. doi: 10.1007/978-3-319-28929-8.

[21] Brian Yamauchi. Frontier-based exploration using multiple robots. In *Proceedings of the International Conference on Autonomous Agents*, pages 47–53, 1998. doi: 10.1145/280765.280773.

[22] Theodore S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 2nd edition, 2002.

[23] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511841224.

[24] Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.