SIR-MARL: POST-TRAINING RECOVERY FROM GROUP-SPARSE OBSERVATION FAULTS IN MULTI-AGENT REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cooperative multi-agent policies trained with nominal observations can encounter sensor failures and perception errors after deployment. We study post-training observation recovery for frozen recurrent policies using clean trajectories, with unknown fault support and cardinality. Shared task dynamics produce related observations across agents, and recent history constrains their admissible clean values. Through a synchronized team-level recovery layer, SIR-MARL exploits this history-conditioned low-dimensional structure together with corruption that is sparse across agent groups. It predicts clean observations from temporal and cross-agent evidence, uses group residuals to guide correction, and refines recovery through leave-one-out contexts. Generic group masking and response-based supervision train the two-stage recovery process on nominal trajectories without matched deployment-fault data. Under structural, observability, and residual-separation conditions, our analysis bounds recovery error and its effect on the frozen actor's responses. Experiments on three StarCraft Multi-Agent Challenge (SMAC) tasks show improved faulted performance while largely preserving nominal performance. Structural and mechanistic diagnostics relate these gains to history-conditioned observation structure, sparse-error correction, and the quality of refined context.

1 INTRODUCTION

Cooperative multi-agent reinforcement learning (MARL) policies are typically trained under nominal observations, whereas sensor failures, perception errors, or faulty observation channels may arise after deployment. Such deployment faults can be difficult to characterize in advance, so representative matched fault data may be unavailable during source-policy training. Corruption of a few agents' local observations can alter current actions and propagate through recurrent hidden states to affect later decisions. This motivates protecting an already-trained source policy against deployment-time observation faults without retraining it for a prespecified fault model.

Recovering the underlying clean team observation from such faulty inputs is underdetermined without additional structure. Structured recovery addresses this ambiguity by combining restricted nominal structure with sparse corruption (Donoho, 2006; Candès et al., 2006). Cooperative agents observe different views of a shared evolving task: agent-specific observation processes map common task-relevant latent factors into correlated team observations. These factors can have far fewer degrees of freedom than the joint observation dimension, while recent history further restricts dynamically admissible states. This motivates a history-conditioned family of low-DoF clean-observation sets. Localized failures affect only a few agent streams, potentially densely within each stream, yielding agent-group-sparse corruption as the complementary structure for recovery.

SIR-MARL (Structured Inference and Recovery for MARL) uses a trusted recovery layer over synchronized team observations. Temporal and cross-agent predictions yield group residuals for initial correction; leave-one-out contexts from original and corrected observations refine each target. Nominal-trajectory training combines generic group masking for reconstruction with response-based supervision to preserve frozen-actor decisions.

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.

Our contributions are:

- SIR-MARL, a two-stage post-training recovery for frozen recurrent policies, using nominal-only training and history-conditioned structure under unknown group-sparse faults.
- Structural conditions for observation recoverability in MARL based on shared task dynamics and history, with error bounds for two-stage recovery and finite-horizon actor responses.
- SMAC evidence of improved faulted performance and largely preserved nominal performance, supported by structural diagnostics and tests across fault conditions.

2 Related Work

Robustness in MARL. Robust and adversarial MARL trains policies against observation uncertainty, action perturbations, and adversarial teammates using robustness objectives, regularization, or perturbation exposure (He et al., 2023; Bukharin et al., 2023; Li et al., 2025; Yuan et al., 2023; Lee et al., 2025). Communication-robust methods protect explicit inter-agent messages through filtering, ensembles, or selective transmission (Zhang et al., 2020; Sun et al., 2023; Mitchell et al., 2020). In contrast, SIR-MARL performs post-training recovery of local observations for a frozen recurrent source policy, using nominal trajectories without matched deployment-fault data.

Sparse and structured recovery. Structured recovery derives identifiability from restrictions on nominal signal classes and corruption geometry, with results for sparse and model-based signals, block/group sparsity, and low-rank signals with sparse errors (Donoho, 2006; Baraniuk et al., 2010; Eldar et al., 2010; Candès et al., 2011). Dynamic sparse recovery and robust subspace tracking further exploit temporal information to recover evolving supports and low-dimensional components or subspaces (Vaswani & Lu, 2010; Zhan et al., 2016; Narayanamurthy & Vaswani, 2018). However, shared task dynamics, agent-specific observation processes, and recent history H induce a state-dependent, potentially nonlinear, history-indexed low-DoF family $\{\mathcal{M}_t(H)\}_H$ of admissible team observations in cooperative MARL. Under agent-group-sparse faults, recovery outputs drive a frozen recurrent actor: similar reconstruction errors can affect actions differently and propagate through hidden states, motivating response-based supervision.

3 Problem Formulation

Frozen recurrent source policy. Consider a cooperative Dec-POMDP with n agents, latent state s_t, transition law P, shared reward r, and discount $\gamma \in [0, 1)$ (Oliehoek & Amato, 2016). Given agent i's D-dimensional clean observation $x_{t,i}$, its actor with frozen parameters ϕ updates hidden state $h^{\pi}_{t,i}$ and outputs action distribution $\pi_{t,i}$ over available actions $\mathcal{A}_{t,i}$, from which action $a_{t,i}$ is sampled:

$$(\pi_{t,i}, h^{\pi}_{t,i}) = \Pi_{\phi}(x_{t,i}, h^{\pi}_{t-1,i}, \mathcal{A}_{t,i}), \qquad a_{t,i} \sim \pi_{t,i}. \tag{1}$$

Group-sparse deployment faults. Let $X_t = [x_{t,1}, \ldots, x_{t,n}]^{\top}$ and $E_t = [e_{t,1}, \ldots, e_{t,n}]^{\top}$ in $\mathbb{R}^{n \times D}$ collect clean observations and agent-wise errors. The received observation is

$$Y_t = X_t + E_t, \qquad \mathcal{F}_t = \{i : \|e_{t,i}\|_2 > 0\}, \qquad \|E_t\|_{2,0} = |\mathcal{F}_t| \le s. \tag{2}$$

Here $\mathcal{F}_t$ indexes faulty agents, and $\|E_t\|_{2,0}$ counts their rows, each allowing dense corruption. Group-sparse faults here affect a minority of groups ($s < n/2$), hence a strict clean majority. Actual support, cardinality, and magnitude are unknown to recovery.

Post-training recovery problem. A trusted recovery operator with parameters θ estimates X_t from Y_t and the previous $T-1$ recovered frames $\widehat{H}_t = (\widehat{X}_{t-T+1}, \ldots, \widehat{X}_{t-1})$:

$$\widehat{X}_t = \mathcal{R}_{\theta}(\widehat{H}_t, Y_t). \tag{3}$$

Training uses nominal source-policy trajectories with internally generated group masks. Deployment assumes synchronized team measurements, reliable delivery, and a functioning recovery layer (Appendix C.1); recovered rows feed the frozen actors. Recovery aims to estimate X_t while preserving intact groups and reducing the return gap $\Delta J(\theta; \phi) := J_{\mathrm{clean}}(\phi) - J_{\mathrm{rec}}(\theta; \phi)$, where J_{clean} and J_{rec} are the frozen policy's expected discounted team returns on clean and recovered inputs, respectively.

4 METHOD

SIR-MARL uses team observation structure arising from shared environment dynamics to recover faulty local observations. Stage 1 combines recent recovered history and current cross-agent evidence to estimate clean observations, infer fault support from group residuals, and cancel part of the group-sparse error. Stage 2 uses original and corrected team conditions to refine the same clean target and update fault support. The recovered observations enter the frozen source actors (Figure 1).

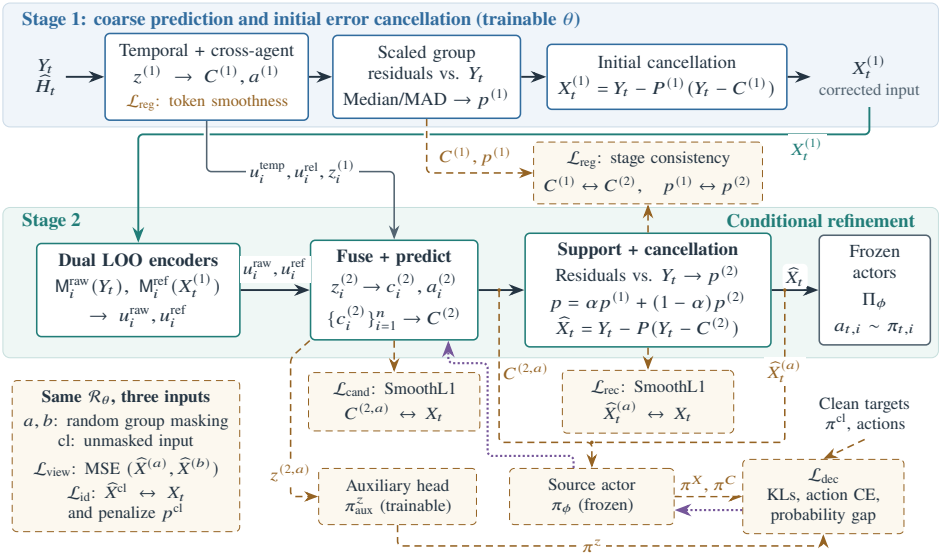


Figure 1: Solid arrows show recovery, amber dashed arrows training evaluations, and purple dotted arrows a representative input-gradient path through the frozen source actor. Here $P^{(1)} = \text{diag}(p^{(1)})$ and $P = \text{diag}(p)$; normalization and repeated inputs are omitted.

4.1 HISTORY-CONDITIONED STRUCTURE AND RECOVERY

Agents observe a shared evolving state through the joint clean-observation mapping $O(s) = [O_1(s), \ldots, O_n(s)]^\top$, where $x_{t,i} = O_i(s_t)$ is agent i's clean observation. Given clean history $H_t^\star = (X_{t-T+1}, \ldots, X_{t-1})$, let $\mathcal{S}_t(H_t^\star)$ contain current states reachable through trajectories compatible with this history and the Dec-POMDP dynamics. The admissible clean observations form

$$\mathcal{M}_t(H_t^\star) = \{O(s) : s \in \mathcal{S}_t(H_t^\star)\}. \tag{4}$$

By construction, $X_t \in \mathcal{M}_t(H_t^\star)$. Shared state constrains joint observation variation, and history further restricts reachable states. Under the structural conditions in Section 5.1, this history-indexed family has low degrees of freedom.

To recover target agent i, the other agents' current clean observations provide further constraints on the history-conditioned family $\mathcal{M}_t(H_t^\star)$. Denote these observations by $X_{t,-i}$ and the possible target values under both constraints by $\mathcal{C}_{t,i}(H_t^\star, X_{t,-i}) \subseteq \mathbb{R}^D$. When this conditional set is a singleton, history and the other agents' observations uniquely determine the target, with ideal conditional completion map $G_{t,i}^\star(H_t^\star, X_{t,-i}) = x_{t,i}$. The nominal family constrains clean observations, while the group-sparse fault model restricts corruption across agent groups.

The ideal map $G_{t,i}^\star$ requires clean conditions $(H_t^\star, X_{t,-i})$, while deployment provides recovered history $\widehat{H}_t$ and current team input Y_t. Target prediction thus uses conditions that themselves require recovery, while correcting those conditions relies on predictions of their clean values. SIR-MARL connects these two tasks through group-sparse error cancellation: Stage 1 forms coarse clean predictions, infers fault support from group residuals, and corrects the team input. Stage 2 combines the original and corrected conditions to refine its estimate of the same clean target.

4.2 Initial Prediction and Support Inference

Recovery uses fixed feature-wise normalization from clean trajectories, denoted Norm, with bars marking normalized quantities. It preserves agent-row support and yields $\bar{Y}_t = \bar{X}_t + \bar{E}_t$. For each agent i, a trainable temporal encoder processes recovered history and the current observation to produce $u_{t,i}^{\mathrm{temp}}$. A trainable cross-agent encoder produces $u_{t,i}^{\mathrm{rel}}$ from current team observations, excluding i from its own keys and values. These features provide recent temporal constraints and simultaneous evidence from other agents. A trainable fusion mapping f_1 forms $z_{t,i}^{(1)} = f_1([u_{t,i}^{\mathrm{temp}}; u_{t,i}^{\mathrm{rel}}])$, where semicolons denote concatenation. Trainable prediction heads D_1 and S_1 produce a D-dimensional clean candidate and a positive scalar residual scale:

$$\bar{c}_{t,i}^{(1)} = D_1(z_{t,i}^{(1)}), \qquad a_{t,i}^{(1)} = \mathrm{softplus}(S_1(z_{t,i}^{(1)})) + \delta_a. \tag{5}$$

Here $\delta_a > 0$ is a fixed offset ensuring positivity. Together, the encoders, fusion mapping, and candidate head estimate the clean target specified by $G_{t,i}^{\star}$ from the available history and team input. Appendix B gives the normalization and full encoder and prediction equations.

Fault locations and cardinality are unknown at deployment. Support inference compares each candidate with its received observation using a scaled group root-mean-square (RMS) residual and a robust team reference:

$$d_{t,i}^{(1)} = \frac{\sqrt{D^{-1}\|\bar{y}_{t,i} - \bar{c}_{t,i}^{(1)}\|_2^2 + \epsilon_r}}{a_{t,i}^{(1)}}, \qquad p_{t,i}^{(1)} = \mathrm{sigmoid}\left(\frac{d_{t,i}^{(1)} - \tau_t^{(1)}}{T_s}\right). \tag{6}$$

The learned scale supplies state-dependent residual normalization. Here $\tau_t^{(1)} = m_t^{(1)} + \beta(q_t^{(1)} + \epsilon_\tau)$, where $m_t^{(1)}$ and $q_t^{(1)}$ are the median and median absolute deviation (MAD) of the team scores $d_{t,j}^{(1)}$. The positive constants $\epsilon_r, \epsilon_\tau, \beta, T_s$ provide numerical stability, a threshold offset, MAD scaling, and a temperature, respectively. In the $s < n/2$ regime, a strict clean majority lets the median and MAD limit the influence of extreme faulty-group scores on the team reference. Larger soft support $p_{t,i}^{(1)}$ indicates stronger evidence of a faulty observation and increases correction strength.

4.3 Initial Group-Sparse Error Cancellation

Stack candidates $\bar{c}_{t,i}^{(1)}$ by row into $\bar{C}_t^{(1)}$ and set $P_t^{(1)} = \mathrm{diag}(p_{t,1}^{(1)}, \ldots, p_{t,n}^{(1)})$. The estimated error and intermediate recovery are

$$\widehat{\bar{E}}_t^{(1)} = P_t^{(1)}(\bar{Y}_t - \bar{C}_t^{(1)}), \qquad \bar{X}_t^{(1)} = \bar{Y}_t - \widehat{\bar{E}}_t^{(1)}. \tag{7}$$

Low-support groups largely retain their observations, while high-support groups receive larger candidate corrections. Substituting $\bar{Y}_t = \bar{X}_t + \bar{E}_t$ into Eq. (7) gives the exact error decomposition

$$\bar{X}_t^{(1)} - \bar{X}_t = (I_n - P_t^{(1)})\bar{E}_t + P_t^{(1)}(\bar{C}_t^{(1)} - \bar{X}_t) \tag{8}$$

where I_n is the identity matrix. The first term is the deployment corruption left after cancellation, while the second is the error introduced by correcting observations with imperfect clean predictions. Within each group, stronger correction suppresses the former while giving more weight to the latter. The resulting corrected other-agent rows $\bar{X}_{t,-i}^{(1)}$ serve as conditions for Stage 2.

4.4 Conditional Refinement and Final Recovery

Following the initial recovery, Stage 2 combines original and corrected team observations to refine its estimate of the clean target specified by $G_{t,i}^{\star}$. When encoding these inputs, a context encoder may directly reuse the target's original observation or preliminary reconstruction. In particular, when the latter is close to the clean target, reuse can yield a small reconstruction error with little use of other-agent information. Stage 2 therefore forms two leave-one-out (LOO) contexts, $Z_{t,i}^{\mathrm{raw}} = \mathsf{M}_i^{\mathrm{raw}}(\bar{Y}_t)$ and $Z_{t,i}^{\mathrm{ref}} = \mathsf{M}_i^{\mathrm{ref}}(\bar{X}_t^{(1)})$, to draw on other-agent evidence from both inputs. The LOO rule fixes the masked position to target agent i. Each context encoder replaces that target's observation embedding with its own trainable mask vector. Two encoders with independent parameters aggregate the other

agents' representations in these masked contexts, producing $u_{t,i}^{\mathrm{raw}}$ and $u_{t,i}^{\mathrm{ref}}$, respectively. A trainable fusion mapping f_2 combines these context features with $u_{t,i}^{\mathrm{temp}}$, $u_{t,i}^{\mathrm{rel}}$, and $z_{t,i}^{(1)}$ to form $z_{t,i}^{(2)}$, retaining the target's temporal and Stage 1 information. Separate trainable heads D_2 and S_2 map $z_{t,i}^{(2)}$ to the refined candidate $\bar{c}_{t,i}^{(2)}$ and positive residual scale $a_{t,i}^{(2)}$ using Eq. (5) with superscript (2).

Applying Eq. (6) with superscript (2) again scores residuals against original $\bar{y}_{t,i}$. With fixed $\alpha \in [0,1]$, row-stacked candidates $\bar{C}_t^{(2)}$, and $P_t = \mathrm{diag}(p_{t,1}, \ldots, p_{t,n})$, support fusion and final recovery are

$$p_{t,i} = \alpha p_{t,i}^{(1)} + (1-\alpha) p_{t,i}^{(2)}, \qquad \widehat{\bar{X}}_t = \bar{Y}_t - P_t(\bar{Y}_t - \bar{C}_t^{(2)}). \tag{9}$$

The inverse normalization gives $\widehat{X}_t = \mathrm{Norm}^{-1}(\widehat{\bar{X}}_t)$.

4.5 Decision-Aware Training on Clean Trajectories

The trainable recovery parameters θ comprise the weights of the encoders, fusion mappings, and candidate and scale heads, together with the mask vectors. We run the frozen source policy under nominal observations to collect samples $(H_t^\star, X_t, \mathbf{a}_t, h_t^-, \mathcal{A}_t, \pi_t^{\mathrm{cl}})$. Here $(H_t^\star, X_t)$ is a clean observation window, $\mathbf{a}_t$ records the joint action, and h_t^-, $\mathcal{A}_t$, and π_t^{cl} collect the agents' incoming recurrent states, available-action sets, and source-policy action distributions, respectively.

With fault support and cardinality unknown at deployment, we construct two independently masked views a, b of each clean window by replacing selected time–agent groups with clean feature means, and retain the unmasked version cl. These three input versions of each sample are used to train the same recovery network with shared parameters θ. Every input is processed by both stages, including their internal LOO operations.

For view a, in original observation coordinates, $\mathcal{L}_{\mathrm{cand}}$ and $\mathcal{L}_{\mathrm{rec}}$ are the SmoothL1 errors of the Stage 2 candidate $C_t^{(2,a)}$ and final recovery $\widehat{X}_t^{(a)}$ against X_t, respectively, supervising the clean prediction and resulting correction. Between views a and b, the consistency loss $\mathcal{L}_{\mathrm{view}} = \mathrm{MSE}(\widehat{X}_t^{(a)}, \widehat{X}_t^{(b)})$ encourages recovery of the same clean observation from different masking patterns. For the unmasked view cl, $\mathcal{L}_{\mathrm{id}} = \mathrm{SmoothL1}(\widehat{X}_t^{\mathrm{cl}}, X_t) + \eta_0 n^{-1} \sum_i p_{t,i}^{\mathrm{cl}}$, where $p_{t,i}^{\mathrm{cl}}$ is its fused support score and η_0 a penalty weight. This identity regularizer penalizes changes to clean inputs and positive support scores on those inputs. The observation objective is $\mathcal{L}_{\mathrm{sig}} = \lambda_{\mathrm{cand}}\mathcal{L}_{\mathrm{cand}} + \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{view}}\mathcal{L}_{\mathrm{view}} + \lambda_{\mathrm{id}}\mathcal{L}_{\mathrm{id}}$.

Even small reconstruction errors can change the frozen policy's selected action, motivating decision-aware supervision. We evaluate recovered observations and candidates through the original frozen actor, whose action-distribution mapping is denoted π_ϕ. Gradients through these inputs provide decision supervision while actor parameters remain fixed. For view a, let $\widehat{x}_{t,i}^{(a)}$ and $c_{t,i}^{(2,a)}$ be the original-coordinate recovery and candidate. A trainable auxiliary action mapping π_{aux}^z uses the Stage 2 representation $z_{t,i}^{(2,a)}$ for this view. All action distributions are restricted to the recorded available-action set $\mathcal{A}_{t,i}$. Suppressing this argument throughout and the recorded incoming recurrent state $h_{t,i}^-$ in π_ϕ, the response distributions are

$$\pi_{t,i}^X = \pi_\phi(\cdot \mid \widehat{x}_{t,i}^{(a)}), \quad \pi_{t,i}^C = \pi_\phi(\cdot \mid c_{t,i}^{(2,a)}), \quad \pi_{t,i}^z = \pi_{\mathrm{aux}}^z(\cdot \mid z_{t,i}^{(2,a)}). \tag{10}$$

With $\ell_{t,i}^{\mathrm{resp}}(\pi) = D_{\mathrm{KL}}(\pi_{t,i}^{\mathrm{cl}} \| \pi)$, the response loss is

$$\mathcal{L}_{\mathrm{resp}} = \frac{1}{n} \sum_{i=1}^{n} \left[\lambda_X \ell_{t,i}^{\mathrm{resp}}(\pi_{t,i}^X) + \lambda_C \ell_{t,i}^{\mathrm{resp}}(\pi_{t,i}^C) + \lambda_z \ell_{t,i}^{\mathrm{resp}}(\pi_{t,i}^z) \right]. \tag{11}$$

This loss aligns the action responses of the final recovery, refined candidate, and learned representation with the source policy's clean responses. The decision objective is $\mathcal{L}_{\mathrm{dec}} = \mathcal{L}_{\mathrm{resp}} + \lambda_{\mathrm{act}}\mathcal{L}_{\mathrm{act}} + \lambda_{\mathrm{margin}}\mathcal{L}_{\mathrm{margin}}$, computed on view a. Here $\mathcal{L}_{\mathrm{act}}$ is cross-entropy between $\pi_{t,i}^X$ and recorded actions, encouraging recovery to preserve the source policy's action choices. The SmoothL1 term $\mathcal{L}_{\mathrm{margin}}$ matches the gap between the two largest probabilities in $\pi_{t,i}^X$ to that in $\pi_{t,i}^{\mathrm{cl}}$, preserving the separation of leading action probabilities.

The regularizer $\mathcal{L}_{\text{reg}}$ sums three weighted MSE penalties over views $v \in \{a, b, \text{cl}\}$. The stage-consistency terms $\text{MSE}(\bar{C}_t^{(1,v)}, \bar{C}_t^{(2,v)})$ and $\text{MSE}(p_t^{(1,v)}, p_t^{(2,v)})$ align the normalized candidates and support vectors, respectively, coordinating estimates of the same clean observation and fault support. The temporal term penalizes differences between adjacent observation embeddings, encouraging continuity in the representations used by the history encoder. The full objective is

$$\mathcal{L} = \mathcal{L}_{\text{sig}} + \lambda_{\text{dec}}\mathcal{L}_{\text{dec}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}, \tag{12}$$

With loss weights λ, this objective jointly trains both recovery stages and the auxiliary action mapping end to end, keeping source-actor parameters frozen. Appendix B.4 specifies the losses and coefficients. At deployment, each recovered $\widehat{X}_t$ feeds the original frozen actors and updates the recovered history.

5 Theoretical Analysis

5.1 History Constraints and Conditional Completion

Section 4.1 constructs $\mathcal{M}_t(H)$ from observations compatible with the environment dynamics and clean history. The following proposition bounds its degrees of freedom and gives a condition under which the other agents' observations uniquely and stably determine the clean target. We use the group norm $\|V\|_{\text{G}} = \|V\|_F/\sqrt{D}$.

Proposition 5.1 (History-conditioned structure and completion)**.** *Suppose clean observations depend on $k_s \ll nD$ task-relevant continuous factors and finitely many observation regimes. Under the trajectory-regularity conditions in Appendix A.4,*

$$\dim_{\text{H}} \mathcal{M}_t(H) \le \max_{j \in \mathcal{J}_t(H)} \min\{k_s, q_j - r_j(H)\}. \tag{13}$$

If, additionally, $\|X - X'\|_{\text{G}} \le \kappa_1 \|X_{-i} - X'_{-i}\|_{\text{G}}$ for all $X, X' \in \mathcal{M}_t(H)$ and targets i, with finite κ_1, then $\mathcal{C}_{t,i}(H, X_{-i}) = \{x_{t,i}\}$ and the ideal completion $G^\star_{t,i}$ is κ_1-Lipschitz in the retained clean observations at fixed H.

Here $\dim_{\text{H}}$ is Hausdorff dimension, and $\mathcal{J}_t(H)$ indexes compatible trajectory branches. In branch j, q_j counts the continuous degrees of freedom of admissible trajectories before conditioning on history. The history map takes these trajectory variables to past observations through the environment's transition and observation processes; its Jacobian rank $r_j(H)$ counts the independent constraints imposed by fixing history (Appendix A.4). The additional group-restricted observability condition guarantees the unique, stable ideal completion $G^\star_{t,i}$ introduced in Section 4.1 (Appendices A.5–A.6).

5.2 Support Inference and Cancellation Error Bounds

Sections 4.2–4.3 use scaled group RMS residuals to set correction weights. Theorem 5.2 quantifies when this correction reduces observation error. For the recovery bounds below, we fix a time step and suppress time indices and normalization bars. Let $\mathcal{F}$ be the fault support, $\mathcal{F}^c$ the clean groups, and $\Delta = \|Y - X\|_{\text{G}}$ the input error.

Theorem 5.2 (Cancellation error bound)**.** *Let $X^+ = Y - P(Y - C)$ for a candidate C and $P = \text{diag}(p)$, with $0 \le p_i \le 1$. If $p_i \le \eta_{\text{c}}$ for $i \in \mathcal{F}^c$ and $1 - p_i \le \eta_{\text{f}}$ for $i \in \mathcal{F}$, with $\eta_{\text{c}}, \eta_{\text{f}} \in [0, 1]$, then*

$$\|X^+ - X\|_{\text{G}} \le \eta_{\text{f}}\Delta + \sqrt{\|(C - X)_{\mathcal{F}}\|_{\text{G}}^2 + \eta_{\text{c}}^2 \|(C - X)_{\mathcal{F}^c}\|_{\text{G}}^2}. \tag{14}$$

For Stage 1, $(C, P) = (C^{(1)}, P^{(1)})$ and $X^+ = X^{(1)}$. The square-root term bounds candidate error injected by correction, with clean-group errors downweighted by η_{c}. An upper bound below Δ certifies strictly improved observations (Appendix A.10).

The support bounds η_{f} and η_{c} quantify missed correction on faulty groups and correction on clean groups, respectively. For the sigmoid gate in Section 4.2, positive residual margins $\gamma_{\text{c}}, \gamma_{\text{f}}$ satisfying $d_i \le \tau - \gamma_{\text{c}}$ on clean groups and $d_i \ge \tau + \gamma_{\text{f}}$ on faulty groups yield $\eta_{\text{c}} = (1 + e^{\gamma_{\text{c}}/T_s})^{-1}$ and $\eta_{\text{f}} = (1 + e^{\gamma_{\text{f}}/T_s})^{-1}$. For the median–MAD gate, Appendices A.8–A.9 give lower bounds for $\gamma_{\text{c}}, \gamma_{\text{f}}$ under a strict clean majority, $|\mathcal{F}| < n/2$, together with conditions on prediction errors, residual scales, and fault amplitudes.

5.3 Stage 2 Prediction Error

Section 4.4 forms the Stage 2 candidate $c_i^{(2)}$ for clean target x_i using both Y_{-i} and the initially corrected observations $X_{-i}^{(1)}$. Let $\Delta_{-i} = \|(Y - X)_{-i}\|_{\mathrm{G}}$ and $\Delta_i^{\mathrm{ref}} = \|(X^{(1)} - X)_{-i}\|_{\mathrm{G}}$ denote their errors relative to X_{-i}. Applying Theorem 5.2 to the Stage 1 correction on groups $j \neq i$ gives sufficient conditions for $\Delta_i^{\mathrm{ref}} < \Delta_{-i}$. The next theorem bounds the Stage 2 candidate error entering Theorem 5.2, relating it to errors in the original and corrected observations, recovered history, and inherited features.

Theorem 5.3 (Stage 2 context sensitivity)**.** *Let $\|v\|_{\mathrm{g}} = \|v\|_2/\sqrt{D}$. Under the local stability conditions of the learned Stage 2 predictor in Appendix A.7,*

$$\|c_i^{(2)} - x_i\|_{\mathrm{g}} \leq e_i^{\mathrm{clean}} + L_{2H,i}\delta_H + L_{\mathrm{raw},i}\Delta_{-i} + L_{\mathrm{ref},i}\Delta_i^{\mathrm{ref}} + L_{\mathrm{aux},i}\delta_{\mathrm{aux},i}. \tag{15}$$

Here e_i^{clean} is the predictor's error with clean history, both LOO contexts formed from X, and temporal, relational, and Stage 1 features recomputed from these clean inputs. The history error δ_H is the root-sum-square of past recovery errors in $\widehat{H}$; $\delta_{\mathrm{aux},i}$ is the Euclidean deviation of $[u_i^{\mathrm{temp}}; u_i^{\mathrm{rel}}; z_i^{(1)}]$ from its clean-input counterpart. Appendix A.7 specifies the local sensitivity bounds L for the respective inputs. It also bounds e_i^{clean} by admissible-target uncertainty in $\mathcal{C}_{t,i}$ plus network approximation error, and $\delta_{\mathrm{aux},i}$ by history error and faults in the target's and other agents' observations. For $L_{\mathrm{ref},i} > 0$, reducing Δ_i^{ref} lowers the term $L_{\mathrm{ref},i}\Delta_i^{\mathrm{ref}}$ in the bound (Appendix A.11).

5.4 Final Recovery and Frozen-Actor Responses

The final correction in Section 4.4 uses the Stage 2 candidate $C^{(2)}$ and fused support P. Substituting these into Theorem 5.2 gives the following recovery bound, whose candidate-error term is bounded by Theorem 5.3.

Corollary 5.4 (Final recovery)**.** *If each stage $r \in \{1, 2\}$ satisfies the support bounds $\eta_{r,\mathrm{c}}, \eta_{r,\mathrm{f}}$, define $\eta_{\mathrm{u}} = \alpha\eta_{1,\mathrm{u}} + (1 - \alpha)\eta_{2,\mathrm{u}}$ for $\mathrm{u} \in \{\mathrm{c}, \mathrm{f}\}$. Then*

$$\|\widehat{X} - X\|_{\mathrm{G}} \leq \eta_{\mathrm{f}}\Delta + B_2. \tag{16}$$

Here B_2 is the square-root term in Eq. (14), evaluated with $C^{(2)}$ and fused η_{c}. When $B_2 < (1-\eta_{\mathrm{f}})\Delta$, final recovery strictly reduces observation error (Appendix A.12). Recovered observations then enter the recurrent source actor in Section 4.5. Restoring time indices, the next result bounds both their immediate effect on actions and the accumulated effect through hidden states.

Corollary 5.5 (Finite-horizon actor response)**.** *Under the local actor and recurrent-update stability conditions in Appendix A.14, with equal initial hidden states and a common action and availability sequence,*

$$\mathrm{TV}(\pi_{t,i}^{\mathrm{cl}}, \pi_{t,i}^{\mathrm{rec}}) \leq L_{\pi x,i}\epsilon_{t,i} + L_{\pi h,i}L_{F,i}\sum_{k=0}^{t-1}\kappa_{h,i}^{t-1-k}\epsilon_{k,i}. \tag{17}$$

Here TV is total variation distance, $\pi_{t,i}^{\mathrm{rec}}$ is the frozen actor's response to recovered inputs, and $\epsilon_{t,i} = \|\widehat{x}_{t,i} - x_{t,i}\|_2$ is original-coordinate recovery error. The constants $L_{\pi x,i}, L_{\pi h,i}$ bound action sensitivity to observations and hidden states; $L_{F,i}, \kappa_{h,i}$ bound recurrent-update sensitivity to the same inputs. Past observation recovery errors receive decaying weights when $\kappa_{h,i} < 1$. Appendix A.13 bounds recovered-history propagation. The local actor KL analysis in Appendix A.15 relates error direction to action sensitivity, supporting the decision-aware objective in Section 4.5. Section 6.2 and Appendix C provide finite-sample diagnostics for several assumptions underlying the preceding theoretical results in the SMAC environments `8m`, `3s5z`, and `2s3z`.

6 Experiments

We evaluate closed-loop robustness on SMAC `8m`, `3s5z`, and `2s3z` under per-step Gaussian faults affecting $s \in \{1, 2, 3, 4\}$ uniformly sampled agent groups, with per-group noise scale independent of s (Appendix C.1.2). Recovery operates with unknown fault support and cardinality. The primary metric is team win rate; results aggregate five policy seeds, averaging recovery seeds within each

Table 1: Closed-loop win rate (%) under per-step group-sparse Gaussian faults, with per-group noise scale independent of s. Entries are mean±std over five policy seeds; SIR-MARL first averages five recovery seeds within each policy, whereas PCA-SIR is a deterministic clean-only fit.

Method	Clean	$s1$	$s2$	$s3$	$s4$
8m: 8 homogeneous agents; 12.5–50% *corrupted*					
Corrupted MAPPO	84.38 ± 11.43	42.50 ± 4.34	12.19 ± 6.09	2.81 ± 4.74	0.94 ± 1.40
EIR	59.06 ± 15.44	21.56 ± 12.71	5.31 ± 5.25	0.63 ± 1.40	0.00 ± 0.00
MIR3	82.19 ± 24.20	43.44 ± 26.32	10.63 ± 9.40	3.75 ± 4.50	0.31 ± 0.70
ERNIE	71.56±9.47	61.88±10.63	53.75±18.74	30.31±38.22	28.13±32.53
ROMANCE	92.50±4.87	65.63±8.34	22.81±8.60	3.75±4.36	0.00±0.00
WALL	91.25±4.50	47.19±13.59	13.44±11.41	3.44±5.99	0.31±0.70
PCA-SIR	83.13±11.45	31.88±7.54	0.00±0.00	0.00±0.00	0.00±0.00
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
3s5z: 8 heterogeneous agents; 12.5–50% *corrupted*					
Corrupted MAPPO	82.50 ± 13.32	49.69 ± 16.44	24.06 ± 13.83	9.38 ± 10.07	3.13 ± 6.99
EIR	70.94 ± 19.19	25.00 ± 15.15	8.75 ± 6.50	2.19 ± 2.37	0.31 ± 0.70
MIR3	96.25 ± 3.92	66.56 ± 20.69	54.38 ± 13.68	24.69 ± 16.18	3.44 ± 3.01
ERNIE	66.88±9.97	50.63±17.90	47.50±16.63	38.75±17.76	25.94±33.72
ROMANCE	88.13±6.31	65.63±23.49	36.56±20.78	12.81±15.29	2.81±6.29
WALL	83.44±17.77	64.38±24.59	30.94±19.47	9.38±11.54	2.81±5.46
PCA-SIR	79.38±11.13	51.88±21.24	15.00±15.01	1.25±2.80	0.00±0.00
SIR-MARL	82.19 ± 12.33	73.44 ± 14.53	68.25 ± 18.88	56.06 ± 23.63	36.13 ± 25.12
2s3z: 5 heterogeneous agents; 20–80% *corrupted*					
Corrupted MAPPO	82.50 ± 13.68	49.38 ± 15.13	14.69 ± 5.13	1.56 ± 1.10	1.56 ± 2.71
EIR	92.50 ± 4.87	48.13 ± 9.14	6.88 ± 7.04	0.94 ± 2.10	0.00 ± 0.00
MIR3	83.44 ± 6.01	53.44 ± 10.09	9.38 ± 4.56	0.31 ± 0.70	0.00 ± 0.00
ERNIE	60.31±11.35	57.19±14.43	23.75±29.39	3.44±3.56	0.00±0.00
ROMANCE	98.13±1.31	69.06±6.94	14.06±12.45	0.63±1.40	0.00±0.00
WALL	95.94±5.01	68.44±25.20	21.56±17.79	2.50±3.60	0.00±0.00
PCA-SIR	85.94±10.07	25.94±9.28	1.56±1.91	0.00±0.00	0.00±0.00
SIR-MARL	83.75 ± 12.01	79.75 ± 16.48	69.50 ± 24.08	13.19 ± 3.56	0.38 ± 0.34

source policy (Table 1). Policy and recovery checkpoint-selection criteria, setup, and diagnostic metrics are specified in Appendix C.2, with source-paired comparisons in Appendix G.4.

We compare corrupted MAPPO with recurrent MAPPO implementations of EIR, ERNIE, MIR3, ROMANCE, and WALL, evaluating transfer of their training mechanisms to the benchmark faults. SIR-MARL freezes the source policy and trains recovery on nominal traces with generic masking. We construct PCA-SIR as a linear structured-recovery baseline using the same frozen source policies and nominal data: it projects onto a clean-fitted subspace, detects support from group residuals, and preserves observations outside the estimated support. Baseline configurations and comparison protocols are detailed in Appendix C.2; additional MPE results appear in Appendix D.

6.1 Robustness to Observation Faults

Table 1 shows substantial gains over corrupted MAPPO except at 80% corruption on 2s3z, with near-source clean performance. On the eight-agent maps, gains extend to $s4$, the boundary of the median–MAD sufficient condition $s < n/2$ (Section 5.2). On 2s3z, gains are stronger under a clean majority than at $s3$/$s4$, which exceed this guarantee with fewer clean groups to anchor residual-based support inference. On Google Research Football (GRF), SIR-MARL improves mean win rate over the frozen source by 16.5 percentage points across $s1$–$s5$ (Appendix D.2).

6.2 Observation Structure and Recovery Diagnostics

With sample counts matched in Figure 2A, the smaller k_{95} for history-selected observations supports history-conditioned concentration of clean-observation variation. In B, residual separation under a clean majority supports selective correction that preserves clean inputs. Panel C shows that Stage 1 improves the conditions supplied to Stage 2, which further lowers mean error on all three maps

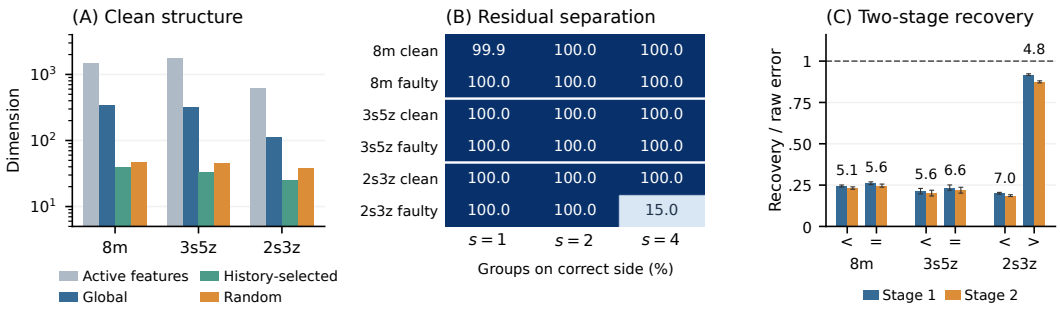


Figure 2: (A) k_{95}: PCA components explaining 95% variance of clean team observations using all samples (Global), samples with similar histories (History-selected), or equally many random samples (Random). Active features exclude constant or near-zero coordinates. (B) Clean/faulty groups below/above the Stage 1 residual threshold (%). (C) Recovery error divided by input corruption error over other agents, in original coordinates (mean±SD; 1: unchanged error). Paired samples include other-agent faults. Labels: Stage 2 reductions relative to Stage 1 (%); ticks: s versus $n/2$. Details: Appendix C.

Table 2: Module ablations on `8m` (win rate, %). Entries are mean±std over five source-policy aggregates, with recovery seeds averaged within each source policy. Each ablated variant is trained separately.

Setting	Clean	$s1$	$s2$	$s3$	$s4$
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
w/o temporal features	83.81 ± 10.97	70.31 ± 21.86	56.94 ± 25.66	43.94 ± 22.70	28.81 ± 19.75
w/o relational features	83.06 ± 12.42	73.94 ± 12.76	60.94 ± 19.43	44.75 ± 21.33	30.25 ± 17.49
w/o LOO contexts	83.88 ± 11.14	71.44 ± 11.99	56.88 ± 13.64	38.38 ± 12.07	22.56 ± 8.75
w/o second stage	83.31 ± 10.78	73.56 ± 12.51	57.19 ± 15.22	40.88 ± 16.01	22.94 ± 11.69
w/o decision supervision	84.50 ± 10.58	67.62 ± 11.18	49.38 ± 15.13	27.62 ± 12.98	16.50 ± 12.06

under a clean majority and at the eight-agent boundary. This progression supports refinement after initial recovery. On `2s3z` at $s4$, fault coverage exceeds the $s < n/2$ range of the median–MAD sufficient guarantee, with weaker residual separation and limited error reduction.

6.3 Ablation Study

Table 2 evaluates separately trained variants on `8m`. Temporal and relational variants suppress the corresponding features; the temporal variant also removes temporal regularization. Other variants remove the dedicated LOO contexts, the second recovery stage, or decision supervision (Appendix C.2). The Stage 2 and LOO comparisons support refinement that uses other-agent evidence after initial correction, consistent with reducing direct reuse of a target's faulty observation. Relational features further support faulted performance by supplying simultaneous cross-agent information. Decision supervision is most beneficial at higher fault counts, consistent with prioritizing action-relevant recovery errors when reconstruction is harder. Temporal features provide historical information for recovery alongside the current cross-agent inputs. Additional ablations and sensitivity results appear in Appendix E.

7 Conclusion

SIR-MARL protects frozen MARL policies through two-stage structured recovery and decision-aware training. The analysis connects history-conditioned observation structure and residual separation to recovery and finite-horizon actor-response bounds. Experiments support post-training robustness under the evaluated deployment faults.

AI USE STATEMENT

Generative AI tools assisted with drafting and revising portions of the manuscript, refining experimental designs, implementing and editing code, analyzing and visualizing results, and checking and revising mathematical arguments and proofs. The authors take responsibility for the final content.

REPRODUCIBILITY STATEMENT

An anonymized code package containing SIR-MARL training, inference, and diagnostic scripts is provided in the supplementary material. The appendices specify the implementation, training, evaluation, and statistical procedures underlying the reported results. Appendix B defines the recovery architecture, normalization, inference computations, and training objectives with their coefficients. Appendix C specifies data splits, optimization settings, checkpoint selection, baseline implementations, random-seed conventions, evaluation budgets, and aggregation across policy and recovery seeds. It also defines the sampling, interventions, norms, and aggregation used in the finite-sample theoretical diagnostics. Appendix E specifies ablation operations, information-assisted controls, sensitivity settings, and computational profiling. Appendices D, F, and G document additional environments, fault protocols, and comparison settings; Appendix G.4 defines the paired and source-normalized statistics. Appendix A provides explicit assumptions, theorem dependencies, and proofs.

REFERENCES

Richard G. Baraniuk, Volkan Cevher, Marco F. Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010. doi: 10.1109/TIT.2010.2040894.

Alexander Bukharin, Yan Li, Yue Yu, Qingru Zhang, Zhehui Chen, Simiao Zuo, Chao Zhang, Songan Zhang, and Tuo Zhao. Robust multi-agent reinforcement learning via adversarial regularization: Theoretical foundation and stable algorithms. In *Advances in Neural Information Processing Systems*, volume 36, pp. 68121–68133, 2023.

Emmanuel J. Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. doi: 10.1002/cpa.20124.

Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011. doi: 10.1145/1970392.1970395.

David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. doi: 10.1109/TIT.2006.871582.

Yonina C. Eldar, Patrick Kuppinger, and Helmut Bölcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Transactions on Signal Processing*, 58(6):3042–3054, 2010. doi: 10.1109/TSP.2010.2044837.

Sihong He, Songyang Han, Sanbao Su, Shuo Han, Shaofeng Zou, and Fei Miao. Robust multi-agent reinforcement learning with state uncertainty. *Transactions on Machine Learning Research*, 2023.

Sunwoo Lee, Jaebak Hwang, Yonghyeon Jo, and Seungyul Han. Wolfpack adversarial attack for robust multi-agent reinforcement learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 33025–33056. PMLR, 2025.

Simin Li, Jun Guo, Jingqiao Xiu, Ruixiao Xu, Xin Yu, Jiakai Wang, Aishan Liu, Yaodong Yang, and Xianglong Liu. Byzantine robust cooperative multi-agent reinforcement learning as a bayesian game. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/86c408308bec72ecce4d86edf243634c-Abstract-Conference.html.

Simin Li, Ruixiao Xu, Jingqiao Xiu, Yuwei Zheng, Pu Feng, Yuqing Ma, Bo An, Yaodong Yang, and Xianglong Liu. Robust multi-agent reinforcement learning by mutual information regularization.

IEEE Transactions on Neural Networks and Learning Systems, 36(10):18118–18132, 2025. doi: 10.1109/TNNLS.2025.3577259.

Rupert Mitchell, Jan Blumenkamp, and Amanda Prorok. Gaussian process based message filtering for robust multi-agent cooperation in the presence of adversarial communication. *arXiv preprint arXiv:2012.00508*, 2020.

Praneeth Narayanamurthy and Namrata Vaswani. Nearly optimal robust subspace tracking. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 3701–3709. PMLR, 2018. URL https://proceedings.mlr.press/v80/narayanamurthy18a.html.

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

Yanchao Sun, Ruijie Zheng, Parisa Hassanzadeh, Yongyuan Liang, Soheil Feizi, Sumitra Ganesh, and Furong Huang. Certifiably robust policy learning against adversarial communication in multi-agent systems. In *International Conference on Learning Representations*, 2023.

Namrata Vaswani and Wei Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Transactions on Signal Processing*, 58(9):4595–4607, 2010. doi: 10.1109/TSP.2010.2051150.

Lei Yuan, Ziqian Zhang, Ke Xue, Hao Yin, Feng Chen, Cong Guan, Lihe Li, Chao Qian, and Yang Yu. Robust multi-agent coordination via evolutionary generation of auxiliary adversarial attackers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11753–11762, 2023. doi: 10.1609/aaai.v37i10.26388.

Jinchun Zhan, Brian Lois, Han Guo, and Namrata Vaswani. Online (and offline) robust PCA: Novel algorithms and performance guarantees. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pp. 1488–1496. PMLR, 2016. URL https://proceedings.mlr.press/v51/zhan16.html.

Sai Qian Zhang, Qi Zhang, and Jieyu Lin. Succinct and robust multi-agent communication with temporal message control. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17271–17282, 2020.

A Theoretical Analysis and Proofs

This appendix develops a deterministic recovery analysis for SIR-MARL that follows the deployed computation graph. The analysis treats the two stages as distinct learned operators. Stage 1 is an initial predictor that can access the current target observation through its temporal branch. Stage 2 uses two parameter-distinct leave-one-out (LOO) contexts, formed with different learned mask tokens and different encoders, while retaining auxiliary target-visible features inherited from the temporal, relational, and Stage 1 pathways. The scale heads are treated only as positive residual scales. All support and recovery statements are conditional on explicit environment, fault-channel, and learned-operator properties.

A.1 Dependency structure and scope

The two environment-level structural routes are distinct:

1. latent dynamics, history fibers, and finite regular observation regimes induce the history-conditioned low-DoF family;
2. cross-agent observational redundancy supplies group-restricted observability or, when exact identifiability fails, a bounded conditional ambiguity.

Low intrinsic dimension and LOO identifiability are separate properties. Their roles meet in the analysis of the learned conditional estimator. The complete proof chain is

$$\begin{gathered}\text{piecewise-regular latent/reachable-state structure}\\ \Downarrow\\ \boldsymbol{X}_t \in \mathcal{M}_t(H_t^*), \quad \dim_{\mathrm{H}} \mathcal{M}_t(H_t^*) \le k_h(H_t^*) \le k_\star \ll nD\\ \text{group-restricted observability or conditional ambiguity}\\ \Downarrow\\ \text{well-posed clean LOO target}\\ \text{Stage 1 clean accuracy and self/cross/history finite-secant gains}\\ \Downarrow\\ \text{standardized-residual envelopes and a median–MAD support certificate}\\ \Downarrow\\ \text{initial sparse-error cancellation}\\ \Downarrow\\ \text{Stage 2 raw/refined LOO candidate bound and fixed support fusion}\\ \Downarrow\\ \text{final recovery and strict-improvement bounds}\\ \Downarrow\\ \text{recovered-history and recurrent-actor stability corollaries.}\end{gathered}$$

The family $\{\mathcal{M}_t(H)\}_H$ is induced by the Dec-POMDP dynamics, the history-conditioned reachable set, and finite observation regimes. It varies with history, and the learned candidates are analyzed as approximate conditional estimators. Exact or near-exact support recovery requires a nonzero residual margin. The weak-fault theorem retains undetected small corruptions according to their signal energy.

A.2 Dec-POMDP, normalization, histories, and group geometry

Dec-POMDP. We consider a cooperative Dec-POMDP

$$\mathfrak{G} = (\mathcal{N}, \mathcal{S}, \{\mathcal{A}_i\}_{i=1}^n, P, r, \{\Omega_i\}_{i=1}^n, \mathsf{O}, \rho_0, \gamma),$$

where $\mathcal{N} = \{1, \dots, n\}$ is the agent set, $s_t \in \mathcal{S}$ is the latent environment state, $\boldsymbol{a}_t = (a_{t,1}, \dots, a_{t,n}) \in \prod_i \mathcal{A}_i$ is the joint action, $P(\mathrm{d}s' \mid s, \boldsymbol{a})$ is the transition kernel, $r(s, \boldsymbol{a})$ is the shared reward, $\Omega_i \subseteq \mathbb{R}^D$ is the clean local-observation space, $\mathsf{O}(\mathrm{d}\boldsymbol{X} \mid s)$ is the joint observation kernel, ρ_0 is the initial-state distribution, and $\gamma \in [0, 1)$ is the discount factor. The clean joint observation is

$$\boldsymbol{X}_t = [\boldsymbol{x}_{t,1}, \dots, \boldsymbol{x}_{t,n}]^\top \in \mathbb{R}^{n \times D}.$$

The recovery target is the clean team observation $\boldsymbol{X}_t$.

Feature-wise normalization. For each task, clean training traces provide $\boldsymbol{\mu}_{\mathrm{cl}} \in \mathbb{R}^D$ and $\boldsymbol{\sigma}_{\mathrm{cl}} \in \mathbb{R}^D_+$, with $\sigma_{\mathrm{cl},j} \geq 10^{-4}$. Define

$$\mathrm{Norm}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{cl}}) \oslash \boldsymbol{\sigma}_{\mathrm{cl}}.$$

If the physical-coordinate deployment observation is $\boldsymbol{Y}_t^{\mathrm{phys}} = \boldsymbol{X}_t^{\mathrm{phys}} + \boldsymbol{E}_t^{\mathrm{phys}}$, then normalization gives

$$\boldsymbol{Y}_t = \boldsymbol{X}_t + \boldsymbol{E}_t, \qquad \boldsymbol{E}_t = \boldsymbol{E}_t^{\mathrm{phys}} \oslash \boldsymbol{\sigma}_{\mathrm{cl}}.$$

Every coordinate scale is strictly positive, so normalization preserves the corrupted agent-group support. All subsequent variables are in normalized coordinates unless marked "phys." The transformation standardizes coordinates using a diagonal scale matrix.

Group norms and projections. For $\boldsymbol{v} \in \mathbb{R}^D$, define the group RMS norm

$$\|\boldsymbol{v}\|_{\mathrm{g}} := D^{-1/2} \|\boldsymbol{v}\|_2 .$$

For $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]^\top \in \mathbb{R}^{n \times D}$, define

$$\|\boldsymbol{V}\|_{\mathrm{G}} := \left(\sum_{i=1}^{n} \|\boldsymbol{v}_i\|_{\mathrm{g}}^2 \right)^{1/2} = D^{-1/2} \|\boldsymbol{V}\|_F .$$

For $J \subseteq \mathcal{N}$, $\Pi_J \boldsymbol{V}$ retains the rows indexed by J and sets all other rows to zero. Write $\Pi_{-i} = \Pi_{\mathcal{N} \setminus \{i\}}$. The group-counting functional is

$$\|\boldsymbol{V}\|_{2,0} := |\{i : \|\boldsymbol{v}_i\|_2 > 0\}| .$$

For $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]^\top$, define

$$F := \{i : \|\boldsymbol{e}_i\|_2 > 0\}, \qquad \boldsymbol{E}_i := \Pi_{\{i\}} \boldsymbol{E}, \qquad \boldsymbol{E}_{-i} := \Pi_{-i} \boldsymbol{E},$$
$$\delta_i := \|\boldsymbol{e}_i\|_{\mathrm{g}} , \qquad \Delta := \|\boldsymbol{E}\|_{\mathrm{G}} , \qquad \Delta_{-i} := \|\boldsymbol{E}_{-i}\|_{\mathrm{G}} .$$

We use $[z]_+ = \max\{z, 0\}$ and $\mathrm{sigm}(z) = (1 + e^{-z})^{-1}$.

Clean and recovered histories. The history length is $T = 12$. Away from episode initialization,

$$\boldsymbol{H}_t^* = (\boldsymbol{X}_{t-T+1}, \ldots, \boldsymbol{X}_{t-1}), \qquad \widehat{\boldsymbol{H}}_t = (\widehat{\boldsymbol{X}}_{t-T+1}, \ldots, \widehat{\boldsymbol{X}}_{t-1}).$$

For histories $H = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(T-1)})$ and $H' = (\boldsymbol{X}'^{(1)}, \ldots, \boldsymbol{X}'^{(T-1)})$, define

$$\|H - H'\|_{\mathcal{H}} := \left(\sum_{k=1}^{T-1} \left\| \boldsymbol{X}^{(k)} - \boldsymbol{X}'^{(k)} \right\|_{\mathrm{G}}^2 \right)^{1/2} .$$

At a fixed time, write $\delta_H := \left\| \widehat{\boldsymbol{H}} - \boldsymbol{H}^* \right\|_{\mathcal{H}}$. Episode-start padding is treated in Section A.13.

Observation coordinates and branch-specific embedding masks. For branch $r \in \{\mathrm{raw}, \mathrm{ref}\}$, let $d_{\mathrm{emb},r}$ be its token dimension, let $\phi_{r,j} : \mathbb{R}^D \to \mathbb{R}^{d_{\mathrm{emb},r}}$ denote the observation embedding at slot j, and let $\boldsymbol{m}_r \in \mathbb{R}^{d_{\mathrm{emb},r}}$ be its learned mask vector. The mask vector is shared across target slots and time within a branch; the raw and refined branches have distinct learned vectors and encoder parameters. Define the actual observation-to-masked-token map

$$\mathsf{M}_i^r : \mathbb{R}^{n \times D} \longrightarrow \mathcal{Z}_r := \mathbb{R}^{n \times d_{\mathrm{emb},r}}, \qquad \big(\mathsf{M}_i^r(\boldsymbol{V})\big)_j = \begin{cases} \boldsymbol{m}_r, & j = i, \\ \phi_{r,j}(\boldsymbol{v}_j), & j \neq i. \end{cases}$$

Thus the target embedding is replaced after embedding. Within one branch the two copies of $\boldsymbol{m}_r$ cancel, giving the token-space identity

$$\|\mathsf{M}_i^r(\boldsymbol{V}) - \mathsf{M}_i^r(\boldsymbol{W})\|_F^2 = \sum_{j \neq i} \|\phi_{r,j}(\boldsymbol{v}_j) - \phi_{r,j}(\boldsymbol{w}_j)\|_2^2 ,$$

where $\boldsymbol{w}_j$ is row j of $\boldsymbol{W}$. If $k_{r,i}^{\mathrm{emb}} \geq 0$ bounds the Euclidean finite-secant gains of the visible-slot embeddings on the compared inputs, set $K_{r,i}^{\mathrm{emb}} = \sqrt{D}\, k_{r,i}^{\mathrm{emb}}$. Then

$$\|\mathsf{M}_i^r(\boldsymbol{V}) - \mathsf{M}_i^r(\boldsymbol{W})\|_F \leq K_{r,i}^{\mathrm{emb}} \|\Pi_{-i}(\boldsymbol{V} - \boldsymbol{W})\|_{\mathrm{G}} , \qquad r \in \{\mathrm{raw}, \mathrm{ref}\}. \tag{18}$$

The factor $\sqrt{D}$ follows from $\|\Pi_{-i}(\boldsymbol{V} - \boldsymbol{W})\|_F = \sqrt{D}\, \|\Pi_{-i}(\boldsymbol{V} - \boldsymbol{W})\|_{\mathrm{G}}$. These are local gains on the compared inputs. Below, raw and refined context errors are measured in normalized observation coordinates before embedding, and their gains include the embedding and subsequent encoder sensitivities. Token-space differences use the Frobenius norm on $\mathcal{Z}_r$.

A.3 EXACT DEPLOYED RECOVERY EQUATIONS

At a fixed time, Stage 1 is represented by

$$\boldsymbol{C}^{(1)} = \mathcal{G}_1(\widehat{\boldsymbol{H}}, \boldsymbol{Y}).$$

This notation includes the temporal branch, which can access the current target row $\boldsymbol{y}_i$, and the self-masked relational branch. Stage 1 therefore includes a target-visible path. Let

$$a_0 := 0.05, \qquad a_i^{(1)} = \mathrm{softplus}(\ell_i^{(1)}) + a_0.$$

With

$$\zeta = 10^{-4}, \qquad c_\tau = 2.05, \qquad b_\tau = 0.03, \qquad T_\tau = 0.09,$$

Stage 1 computes

$$r_i^{(1)} = \boldsymbol{y}_i - \boldsymbol{c}_i^{(1)}, \tag{19}$$

$$g_i^{(1)} = \sqrt{\left\| r_i^{(1)} \right\|_{\mathrm{g}}^2 + \zeta^2}, \tag{20}$$

$$d_i^{(1)} = g_i^{(1)} / a_i^{(1)}, \tag{21}$$

$$m^{(1)} = \mathrm{median}_i \, d_i^{(1)}, \tag{22}$$

$$q^{(1)} = \mathrm{median}_i \left| d_i^{(1)} - m^{(1)} \right|, \tag{23}$$

$$\tau^{(1)} = m^{(1)} + c_\tau (q^{(1)} + b_\tau), \tag{24}$$

$$p_i^{(1)} = \mathrm{sigm}\left((d_i^{(1)} - \tau^{(1)}) / T_\tau \right). \tag{25}$$

Let $\boldsymbol{P}^{(1)} = \mathrm{diag}(p_1^{(1)}, \ldots, p_n^{(1)})$. The first corrected observation is

$$\boldsymbol{X}^{(1)} = \boldsymbol{Y} - \boldsymbol{P}^{(1)} (\boldsymbol{Y} - \boldsymbol{C}^{(1)}), \tag{26}$$

which gives the exact identity

$$\boldsymbol{X}^{(1)} - \boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{P}^{(1)}) \boldsymbol{E} + \boldsymbol{P}^{(1)} (\boldsymbol{C}^{(1)} - \boldsymbol{X}). \tag{27}$$

For target group i, Stage 2 receives the masked token tensors

$$\boldsymbol{Z}_i^{\mathrm{raw}} = \mathsf{M}_i^{\mathrm{raw}}(\boldsymbol{Y}), \qquad \boldsymbol{Z}_i^{\mathrm{ref}} = \mathsf{M}_i^{\mathrm{ref}}(\boldsymbol{X}^{(1)}),$$

in $\mathcal{Z}_{\mathrm{raw}}$ and $\mathcal{Z}_{\mathrm{ref}}$, respectively, through parameter-distinct encoders. The remaining temporal, relational, and Stage 1-derived inputs are grouped as

$$\boldsymbol{V}_i(\widehat{\boldsymbol{H}}, \boldsymbol{Y}) = [\boldsymbol{u}_i^{\mathrm{temp}}; \boldsymbol{u}_i^{\mathrm{rel}}; \boldsymbol{z}_i^{(1)}].$$

The deployed Stage 2 candidate is

$$\boldsymbol{c}_i^{(2)} = \mathcal{G}_{2,i}\left(\widehat{\boldsymbol{H}}, \boldsymbol{Z}_i^{\mathrm{raw}}, \boldsymbol{Z}_i^{\mathrm{ref}}, \boldsymbol{V}_i(\widehat{\boldsymbol{H}}, \boldsymbol{Y}) \right), \qquad \boldsymbol{C}^{(2)} = [\boldsymbol{c}_1^{(2)}, \ldots, \boldsymbol{c}_n^{(2)}]^\top. \tag{28}$$

The auxiliary feature is explicit because it can contain target-visible information even though both dedicated context encoders are LOO. Stage 2 uses

$$a_i^{(2)} = \mathrm{softplus}(\ell_i^{(2)}) + a_0$$

and computes $r_i^{(2)}, g_i^{(2)}, d_i^{(2)}, m^{(2)}, q^{(2)}, \tau^{(2)}, p_i^{(2)}$ by the same formulas and constants as Stage 1, with superscript (2).

For the realized finite forward pass, define

$$a_{r,\max} := \max_{i \in \mathcal{N}} a_i^{(r)}, \qquad r \in \{1, 2\}. \tag{29}$$

Every finite network output gives $a_{r,\max} < \infty$, while the architecture gives $a_i^{(r)} > a_0$. The single-step analysis uses the scale values realized at the evaluated input.

The final support score is

$$p_i = \alpha p_i^{(1)} + (1-\alpha) p_i^{(2)}, \qquad \alpha = \tfrac{1}{4}. \tag{30}$$

Both stagewise scores lie in $(0,1)$, so the implementation clamp to $[0,1]$ is algebraically inactive in exact arithmetic and serves only numerical safety. With $\boldsymbol{P} = \mathrm{diag}(p_1, \dots, p_n)$,

$$\widehat{\boldsymbol{X}} = \boldsymbol{Y} - \boldsymbol{P}(\boldsymbol{Y} - \boldsymbol{C}^{(2)}), \tag{31}$$

and therefore

$$\widehat{\boldsymbol{X}} - \boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{E} + \boldsymbol{P}(\boldsymbol{C}^{(2)} - \boldsymbol{X}). \tag{32}$$

Equations (27) and (32) are unconditional algebraic interfaces.

A.4 Primitive assumptions

The single-step analysis fixes a time t, a clean pair $(H, \boldsymbol{X}) = (\boldsymbol{H}_t^*, \boldsymbol{X}_t)$, a deployed history $\widehat{H}$, and a corruption $\boldsymbol{E}$. Unless a result explicitly states uniformity over an input family, network errors and finite-secant gains below are local constants for this tuple and for the line segments joining the displayed inputs.

Assumption A.1 (Low-dimensional task-relevant factor and finite observation regimes)**.** *The latent state admits a decomposition $s_t = (\xi_t, \nu_t)$. There are an integer k_s, a prescribed dimension budget $k_\star$, and a small fixed ratio $\varepsilon_{\mathrm{dof}} \in (0,1)$ such that*

$$k_s \le k_\star \le \varepsilon_{\mathrm{dof}} nD. \tag{33}$$

The task-relevant factor satisfies $\xi_t \in \Xi_{b_t} \subseteq \mathbb{R}^{k_s}$, where the observation regime b_t belongs to a finite set $\mathcal{B}$, and

$$\boldsymbol{X}_t = \mathcal{O}_{b_t}(\xi_t). \tag{34}$$

For each $b \in \mathcal{B}$, the analyzed clean reachable subset Ξ_b is compact and $\mathcal{O}_b : \Xi_b \to \mathbb{R}^{n\times D}$ is L_b-Lipschitz. When a manifold statement is invoked, the relevant restriction of $\mathcal{O}_b$ is additionally a C^1 embedding.

The finite regime variable covers piecewise-regular observation construction such as visibility patterns, alive/dead patterns, entity appearance, collision modes, or other finite observation modes. The model permits discontinuities across regime changes. Random clean quantities that affect $\boldsymbol{X}_t$ may be included in ξ_t; the current recovery target is invariant to the nuisance component ν_t for fixed (ξ_t, b_t).

Assumption A.2 (Finite hybrid trajectory charts and history-fiber regularity)**.** *For every analyzed time and clean history, compatible trajectories are covered by a finite collection of hybrid charts $j \in \mathcal{J}_t$. Chart j has an open set $U_j \subseteq \mathbb{R}^{q_j}$, a compact set $K_j \subset U_j$, a current observation regime $b(j) \in \mathcal{B}$, and C^1 maps defined on U_j,*

$$\mathscr{H}_j : U_j \to (\mathbb{R}^{n\times D})^{T-1}, \qquad \Gamma_j : U_j \to \Xi_{b(j)}.$$

The clean history and current task-relevant factor generated by a parameter $\theta \in K_j$ are $\mathscr{H}_j(\theta)$ and $\Gamma_j(\theta)$. For every nonempty compact fiber

$$\Theta_j(H) := K_j \cap \mathscr{H}_j^{-1}(H), \tag{35}$$

there is an open neighborhood $W_j(H) \subset U_j$ of $\Theta_j(H)$ on which $D\mathscr{H}_j$ has constant rank $r_j(H)$. Continuous actions or innovations remaining free after conditioning on history are included among the q_j chart coordinates.

A hybrid chart may encode a finite sequence of discrete transition and observation regimes over the history window. The open-domain extension in Assumption A.2 is the domain required by the constant-rank theorem; the compact set K_j only restricts the analyzed reachable trajectories.

Connection to environment dynamics. The trajectory coordinates in branch j parameterize trajectories admitted by the initial-state distribution, transition kernel, action sequence, and observation process in Appendix A.2. They include the free initial-state coordinates, continuous actions, and transition or observation innovations needed to generate those trajectories. For a parameter value θ, the history map stacks the resulting clean observations: $\mathscr{H}_j(\theta) = (\boldsymbol{X}_{t-T+1}(\theta), \ldots, \boldsymbol{X}_{t-1}(\theta))$. Thus $r_j(H) = \operatorname{rank} D\mathscr{H}_j(\theta)$ on the constant-rank neighborhood is the rank of the stacked observation Jacobian with respect to the free trajectory coordinates. In a differentiable realization of the transition and observation kernels, this Jacobian follows by composing their derivatives along the trajectory. Assumption A.2 specifies the branchwise regularity required for this construction; Proposition A.1 uses that rank to quantify the freedom remaining after history is fixed.

Assumption A.3 (Group-restricted observability on the full nominal union)**.** *At a fixed clean history H, the nominal family $\mathcal{M}_t(H)$ defined in Equation* (45) *satisfies* $\mathrm{RO}(q, \kappa_q; H)$ *if, for every $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{M}_t(H)$ and every $J \subseteq \mathcal{N}$ with $|J| \le q$,*

$$\|\boldsymbol{X} - \boldsymbol{X}'\|_{\mathrm{G}} \le \kappa_q \, \|\Pi_{J^c}(\boldsymbol{X} - \boldsymbol{X}')\|_{\mathrm{G}} \,. \tag{36}$$

The quantifier covers pairs in the same regime and pairs in different regimes of the finite union. A statement uniform over histories additionally requires a common finite upper bound on $\kappa_q(H)$.

The main Stage 2 theorem uses only $q = 1$ when exact LOO identifiability is invoked. Without $\mathrm{RO}(1)$, the theorem retains an explicit conditional-ambiguity term.

Assumption A.4 (Agent-group-sparse fault channel)**.** *At deployment,*

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}, \qquad \|\boldsymbol{E}\|_{2,0} = |F| \le s. \tag{37}$$

The support F, realized cardinality, and fault amplitudes are unknown to the recovery operator. The $s < n/2$ regime in Section 3 supplies the strict clean majority used by the median–MAD sufficient guarantee.

Assumption A.5 (Pointwise Stage 1 clean accuracy and history sensitivity)**.** *For each group i, there are $\varepsilon_{1,i} \ge 0$ and $L_{1H,i} \ge 0$ such that*

$$\|\mathcal{G}_{1,i}(H, \boldsymbol{X}) - \boldsymbol{x}_i\|_{\mathrm{g}} \le \varepsilon_{1,i}, \tag{38}$$

$$\left\|\mathcal{G}_{1,i}(\widehat{H}, \boldsymbol{X} + \boldsymbol{E}) - \mathcal{G}_{1,i}(H, \boldsymbol{X} + \boldsymbol{E})\right\|_{\mathrm{g}} \le L_{1H,i}\delta_H. \tag{39}$$

The quantity $\varepsilon_{1,i}$ is a clean initial-prediction error. Training supervises $\boldsymbol{C}^{(1)}$ indirectly through downstream and consistency losses, so this property is measured on held-out clean windows.

Assumption A.6 (Pointwise Stage 1 target-self and cross-group finite-secant gains)**.** *For each group i, there are $0 \le \eta_{1,i} < 1$ and $L_{1\times,i} \ge 0$ such that*

$$\|\mathcal{G}_{1,i}(H, \boldsymbol{X} + \boldsymbol{E}) - \mathcal{G}_{1,i}(H, \boldsymbol{X} + \boldsymbol{E}_{-i})\|_{\mathrm{g}} \le \eta_{1,i}\delta_i, \tag{40}$$

$$\|\mathcal{G}_{1,i}(H, \boldsymbol{X} + \boldsymbol{E}_{-i}) - \mathcal{G}_{1,i}(H, \boldsymbol{X})\|_{\mathrm{g}} \le L_{1\times,i}\Delta_{-i}. \tag{41}$$

The first inequality controls the actual target-visible path after all other group faults are held fixed. The strict bound $\eta_{1,i} < 1$ is a sufficient non-copy condition for detectable strong faults. The second inequality permits corruption propagation from other agents. Both inequalities specify predictor properties independently of the support estimator.

The Stage 2 assumptions are stated after the environment-level nominal family and conditional ambiguity are constructed.

A.5 History-conditioned low-degree-of-freedom structure

For a clean history H, let

$$\mathcal{J}_t(H) := \{j \in \mathcal{J}_t : \Theta_j(H) \neq \varnothing\}.$$

Define the compatible task-relevant latent set and its branch components by

$$\Xi_{t,j}(H) := \Gamma_j(\Theta_j(H)), \tag{42}$$

$$\Xi_t(H) := \bigcup_{j \in \mathcal{J}_t(H)} \Xi_{t,j}(H). \tag{43}$$

The corresponding branchwise and full nominal observation families are

$$\mathcal{M}_{t,j}(H) := \mathcal{O}_{b(j)}(\Xi_{t,j}(H)), \tag{44}$$

$$\mathcal{M}_t(H) := \bigcup_{j \in \mathcal{J}_t(H)} \mathcal{M}_{t,j}(H) \subseteq \mathbb{R}^{n \times D}. \tag{45}$$

By construction, every clean trajectory covered by the assumptions obeys $\boldsymbol{X}_t \in \mathcal{M}_t(\boldsymbol{H}_t^*)$.

Proposition A.1 (MDP-induced conditional low-degree-of-freedom family)**.** *Under Assumptions A.1 and A.2, define*

$$k_h(H) := \max_{j \in \mathcal{J}_t(H)} \min\{k_s, q_j - r_j(H)\}. \tag{46}$$

Then

$$\dim_{\mathrm{H}} \Xi_t(H) \le k_h(H), \qquad \dim_{\mathrm{H}} \mathcal{M}_t(H) \le k_h(H) \le k_s \le k_\star \le \varepsilon_{\mathrm{dof}} nD. \tag{47}$$

Thus the nominal family is low-dimensional relative to the ambient joint-observation dimension whenever the prescribed ratio $\varepsilon_{\mathrm{dof}}$ is small. If a compatible component is a $k_j(H)$-dimensional C^1 manifold and the composition $\mathcal{O}_{b(j)} \circ \Gamma_j$ is a C^1 embedding on that component, its image is a $k_j(H)$-dimensional embedded manifold. Branchwise Lipschitz regularity yields a compact low-DoF set; a smooth-manifold characterization requires additional regularity.

Proof. Fix $j \in \mathcal{J}_t(H)$. By Assumption A.2, $\mathscr{H}_j$ is C^1 on the open set U_j, and its derivative has constant rank $r_j(H)$ on an open neighborhood of the compact fiber $\Theta_j(H)$. The constant-rank theorem therefore represents $\mathscr{H}_j^{-1}(H)$ locally as a C^1 submanifold of dimension $q_j - r_j(H)$. Hausdorff dimension is monotone under intersection with K_j, and compactness yields a finite cover by such local representations. Hence

$$\dim_{\mathrm{H}} \Theta_j(H) \le q_j - r_j(H). \tag{48}$$

Because Γ_j is C^1 on an open neighborhood of the compact fiber, its derivative is bounded on a sufficiently small compact neighborhood contained in U_j. A mean-value argument in finitely many coordinate neighborhoods therefore makes Γ_j Lipschitz on each member of a finite cover of $\Theta_j(H)$. Hausdorff dimension is nonincreasing under Lipschitz maps, so

$$\dim_{\mathrm{H}} \Xi_{t,j}(H) \le \min\{q_j - r_j(H), k_s\}.$$

The observation map $\mathcal{O}_{b(j)}$ is Lipschitz on $\Xi_{b(j)}$, and hence

$$\dim_{\mathrm{H}} \mathcal{M}_{t,j}(H) \le \dim_{\mathrm{H}} \Xi_{t,j}(H).$$

Both $\Xi_t(H)$ and $\mathcal{M}_t(H)$ are finite unions. Hausdorff dimension of a finite union is the maximum of the component dimensions, which gives the first two inequalities in (47); the remaining inequalities follow from (33). Compactness follows from compactness of every fiber, continuity of Γ_j and $\mathcal{O}_{b(j)}$, and finiteness of the union. The manifold statement follows from the embedded-submanifold theorem on any component where the stated embedding condition holds. □

Remark A.2 (When history conditioning strictly reduces dimension)**.** *Let $\mathcal{M}_{\mathrm{glob}}$ be the finite union generated by all globally reachable task-relevant factors and regimes, and write $k_{\mathrm{glob}} = \dim_{\mathrm{H}} \mathcal{M}_{\mathrm{glob}}$. If*

$$k_h(H) < k_{\mathrm{glob}},$$

then Proposition A.1 gives $\dim_{\mathrm{H}} \mathcal{M}_t(H) < \dim_{\mathrm{H}} \mathcal{M}_{\mathrm{glob}}$. A sufficient chart-level condition is that the history map removes enough independent continuous degrees of freedom so that $q_j - r_j(H) < k_{\mathrm{glob}}$ on every compatible chart. Uninformative history or unconstrained innovations can yield equality with the globally reachable dimension.

Remark A.3 (Empirical role of effective-rank diagnostics)**.** *The observed global $k95$ values $344/1480$ on 8m, $113/620$ on 2s3z, and $323/1784$ on 3s5z, together with lower history-conditioned local $k95$, support the plausibility of a small $k_\star$ and a further reduction to $k_h(H)$. These are empirical variance diagnostics; the dimension statement relies on the specified reachable-state and branchwise regularity assumptions.*

A.6 GROUP-RESTRICTED IDENTIFIABILITY AND CONDITIONAL AMBIGUITY

For target group i, define the environment-consistent target fiber

$$\mathcal{F}_i(H, \boldsymbol{X}_{-i}) := \{\boldsymbol{x}_i' : \boldsymbol{X}' \in \mathcal{M}_t(H),\ \Pi_{-i}\boldsymbol{X}' = \Pi_{-i}\boldsymbol{X}\}\,. \tag{49}$$

The fiber is nonempty for the realized $\boldsymbol{X}$ and compact because $\mathcal{M}_t(H)$ is compact. Define its intrinsic conditional ambiguity by its group-norm diameter,

$$\omega_i(H, \boldsymbol{X}_{-i}) := \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{F}_i(H,\boldsymbol{X}_{-i})} \|\boldsymbol{u} - \boldsymbol{v}\|_{\mathrm{g}}\,. \tag{50}$$

This quantity is determined by the environment-level nominal family. It is zero exactly when the target fiber is a singleton. The learned network is characterized by its ideal-input error relative to this ambiguity.

Proposition A.4 (Stable LOO completion under restricted observability)**.** *If* $\mathrm{RO}(1, \kappa_1; H)$ *holds on the full union* $\mathcal{M}_t(H)$*, then for every target* i*, the projection*

$$\Pi_{-i} : \mathcal{M}_t(H) \to \Pi_{-i}\mathcal{M}_t(H)$$

is injective. Consequently, $\mathcal{F}_i(H, \boldsymbol{X}_{-i}) = \{\boldsymbol{x}_i\}$*,* $\omega_i(H, \boldsymbol{X}_{-i}) = 0$*, and there is a unique completion map* G_i^* *with* $G_i^*(H, \boldsymbol{X}_{-i}) = \boldsymbol{x}_i$*. For any* $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{M}_t(H)$*,*

$$\left\|G_i^*(H, \boldsymbol{X}_{-i}) - G_i^*(H, \boldsymbol{X}'_{-i})\right\|_{\mathrm{g}} \le \kappa_1 \left\|\Pi_{-i}(\boldsymbol{X} - \boldsymbol{X}')\right\|_{\mathrm{G}}\,. \tag{51}$$

Proof. If $\Pi_{-i}\boldsymbol{X} = \Pi_{-i}\boldsymbol{X}'$, then (36) with $J = \{i\}$ gives $\|\boldsymbol{X} - \boldsymbol{X}'\|_{\mathrm{G}} = 0$, proving injectivity and the singleton-fiber claim. For stability,

$$\begin{aligned}\left\|G_i^*(H, \boldsymbol{X}_{-i}) - G_i^*(H, \boldsymbol{X}'_{-i})\right\|_{\mathrm{g}} &= \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_{\mathrm{g}} \\ &\le \|\boldsymbol{X} - \boldsymbol{X}'\|_{\mathrm{G}} \\ &\le \kappa_1 \left\|\Pi_{-i}(\boldsymbol{X} - \boldsymbol{X}')\right\|_{\mathrm{G}}\,.\end{aligned}$$

□

Remark A.5 (Low dimension and restricted observability are independent)**.** *A one-dimensional nominal set may fail LOO identifiability. For example, with* $n = 2, D = 1$*, the set* $\mathcal{M} = \{(u, 0) : u \in [0, 1]\}$ *has Hausdorff dimension one, but deleting group 1 maps every point to zero. Assumption A.3 therefore supplies an additional identifiability condition beyond Proposition A.1. Conversely, restricted observability is imposed on the entire finite union, so it covers both within-regime and cross-regime pairs.*

Definition A.6 (Minimum group distance)**.** *For* $\mathcal{M} \subseteq \mathbb{R}^{n\times D}$*, define*

$$d_{\mathrm{grp}}(\mathcal{M}) := \inf_{\substack{\boldsymbol{X},\boldsymbol{X}'\in\mathcal{M}\\ \boldsymbol{X}\neq\boldsymbol{X}'}} \|\boldsymbol{X} - \boldsymbol{X}'\|_{2,0}\,.$$

Proposition A.7 (Relation to group-sparse unique decoding)**.** *For a fixed history* H*:*

- *(i) If* $d_{\mathrm{grp}}(\mathcal{M}_t(H)) > 1$*, deleting any one group is injective on* $\mathcal{M}_t(H)$*.*
- *(ii) If* $d_{\mathrm{grp}}(\mathcal{M}_t(H)) > s$*, then any known erased set* F *with* $|F| \le s$ *leaves enough groups to identify* $\boldsymbol{X} \in \mathcal{M}_t(H)$ *uniquely.*
- *(iii) If* $d_{\mathrm{grp}}(\mathcal{M}_t(H)) > 2s$*, then* $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}$*, with* $\boldsymbol{X} \in \mathcal{M}_t(H)$ *and* $\|\boldsymbol{E}\|_{2,0} \le s$*, has a unique decomposition.*

These are qualitative conditions. A finite κ_q *in* $\mathrm{RO}(q, \kappa_q; H)$ *is the corresponding quantitative secant condition.*

Proof. For (i), agreement outside one group would produce a nonzero nominal secant supported on at most one group, contradicting the distance condition. The same argument with at most s erased groups proves (ii). For (iii), if $\boldsymbol{X} + \boldsymbol{E} = \boldsymbol{X}' + \boldsymbol{E}'$, then $\boldsymbol{X} - \boldsymbol{X}' = \boldsymbol{E}' - \boldsymbol{E}$, whose group support has size at most $2s$. The distance condition forces $\boldsymbol{X} = \boldsymbol{X}'$, and then $\boldsymbol{E} = \boldsymbol{E}'$. □

Remark A.8 (Why the main theorem uses only $\mathrm{RO}(1)$)**.** *Stage 2 predicts one target from a target-masked context. Thus* $\mathrm{RO}(1)$ *is the weakest restricted-observability condition matching the ideal LOO problem. The stronger* $\mathrm{RO}(s)$ *or* $\mathrm{RO}(2s)$ *conditions are needed only for joint completion from a known* s*-group erasure pattern or arbitrary blind* s*-group unique decoding. A branchwise tangent lower bound can serve as a local sufficient condition within a smooth component, while cross-branch pairs require the full-union secant condition.*

A.7 Stage 2 Approximation and Stability Assumptions

Define the two distinct ideal clean masked-token contexts

$$\boldsymbol{R}^*_{i,\mathrm{raw}} := \mathsf{M}^{\mathrm{raw}}_i(\boldsymbol{X}), \qquad \boldsymbol{R}^*_{i,\mathrm{ref}} := \mathsf{M}^{\mathrm{ref}}_i(\boldsymbol{X}), \qquad \boldsymbol{V}^*_i := \boldsymbol{V}_i(H, \boldsymbol{X}). \tag{52}$$

Assumption A.7 (Pointwise Stage 2 ideal-context accuracy)**.** *For each target* i*, there is a learned excess error* $\varepsilon^{\mathrm{net}}_{2,i} \geq 0$ *such that*

$$\left\|\mathcal{G}_{2,i}(H, \boldsymbol{R}^*_{i,\mathrm{raw}}, \boldsymbol{R}^*_{i,\mathrm{ref}}, \boldsymbol{V}^*_i) - \boldsymbol{x}_i\right\|_{\mathrm{g}} \leq \omega_i(H, \boldsymbol{X}_{-i}) + \varepsilon^{\mathrm{net}}_{2,i}. \tag{53}$$

The term ω_i is environment-intrinsic conditional ambiguity; $\varepsilon^{\mathrm{net}}_{2,i}$ is learned excess error on the ideal clean inputs. Under $\mathrm{RO}(1)$, $\omega_i = 0$. Assumption A.7 permits any estimator satisfying the stated ideal-input error bound on a possibly non-singleton fiber.

Assumption A.8 (Pointwise Stage 2 context, auxiliary, and history stability)**.** *For* $\boldsymbol{Z}^{\mathrm{raw}}_i = \mathsf{M}^{\mathrm{raw}}_i(\boldsymbol{Y})$ *and* $\boldsymbol{Z}^{\mathrm{ref}}_i = \mathsf{M}^{\mathrm{ref}}_i(\boldsymbol{X}^{(1)})$*, there are nonnegative constants* $L_{2H,i}$*,* $L_{\mathrm{raw},i}$*,* $L_{\mathrm{ref},i}$*, and* $L_{\mathrm{aux},i}$ *such that the composed branches satisfy the following bound in pre-embedding observation coordinates:*

$$\begin{aligned}
&\left\|\mathcal{G}_{2,i}(\widehat{H}, \boldsymbol{Z}^{\mathrm{raw}}_i, \boldsymbol{Z}^{\mathrm{ref}}_i, \boldsymbol{V}_i(\widehat{H}, \boldsymbol{X} + \boldsymbol{E})) - \mathcal{G}_{2,i}(H, \boldsymbol{R}^*_{i,\mathrm{raw}}, \boldsymbol{R}^*_{i,\mathrm{ref}}, \boldsymbol{V}^*_i)\right\|_{\mathrm{g}} \\
&\quad \leq L_{2H,i}\delta_H + L_{\mathrm{raw},i}\left\|\Pi_{-i}(\boldsymbol{Y} - \boldsymbol{X})\right\|_{\mathrm{G}} + L_{\mathrm{ref},i}\left\|\Pi_{-i}(\boldsymbol{X}^{(1)} - \boldsymbol{X})\right\|_{\mathrm{G}} \\
&\qquad + L_{\mathrm{aux},i}\left\|\boldsymbol{V}_i(\widehat{H}, \boldsymbol{X} + \boldsymbol{E}) - \boldsymbol{V}^*_i\right\|_2.
\end{aligned} \tag{54}$$

The auxiliary path satisfies

$$\left\|\boldsymbol{V}_i(\widehat{H}, \boldsymbol{X} + \boldsymbol{E}) - \boldsymbol{V}_i(H, \boldsymbol{X})\right\|_2 \leq K_{H,i}\delta_H + K_{\mathrm{self},i}\delta_i + K_{\times,i}\Delta_{-i}. \tag{55}$$

The constants $L_{\mathrm{raw},i}$ and $L_{\mathrm{ref},i}$ are gains of the full composed branches with respect to normalized observation inputs. To obtain them from token-space bounds, let $\widetilde{L}_{r,i} \geq 0$ bound the downstream sensitivity to branch r's token tensor in Frobenius norm. Combining that bound with (18) permits $L_{r,i} = \widetilde{L}_{r,i}K^{\mathrm{emb}}_{r,i}$, where $L_{r,i}$ denotes $L_{\mathrm{raw},i}$ or $L_{\mathrm{ref},i}$. Hence embedding sensitivity is included once in (54). The target-self auxiliary term is necessary because the complete Stage 2 computation retains target-visible features: the temporal feature and $\boldsymbol{z}^{(1)}_i$ can depend on $\boldsymbol{y}_i$. The raw and refined gains are kept separate because only the refined context is changed by Stage 1 cancellation.

Dependence on observations, faults, and the learned network. All deviations in Assumption A.8 compare the deployed computation with clean observations from the same environment trajectory. In particular, δ_H sums past recovery errors, δ_i measures the target's fault, and Δ_{-i} measures faults in the other agents. Before embedding, the raw-context deviation is exactly $\|\Pi_{-i}(\boldsymbol{Y} - \boldsymbol{X})\|_{\mathrm{G}} = \Delta_{-i}$, and the refined-context deviation is $\left\|\Pi_{-i}(\boldsymbol{X}^{(1)} - \boldsymbol{X})\right\|_{\mathrm{G}}$, bounded by initial cancellation. Equation (18) controls the corresponding token deviations through the embedding gains already included in $L_{\mathrm{raw},i}$ and $L_{\mathrm{ref},i}$. Equation (55) expresses the change in the inherited temporal, relational, and Stage 1 features using these same history and fault errors. The constants L and K are sensitivities of the specified learned mappings on these inputs. Hence the Stage 2 bound combines environment-level conditional ambiguity, approximation by the trained predictor, and propagation of actual observation errors through the deployed network.

A.8 Stage 1 residual separation

Lemma A.9 (Block-Jacobian conditions sufficient for the Stage 1 finite-secant gains)**.** *Assume* $\mathcal{G}_{1,i}(H, \cdot)$ *is continuously differentiable along the two segments*

$$\boldsymbol{X} + \boldsymbol{E}_{-i} + \theta\boldsymbol{E}_i, \qquad \boldsymbol{X} + \theta\boldsymbol{E}_{-i}, \qquad \theta \in [0, 1].$$

Let $J_{ij}(\boldsymbol{Z})$ denote the Euclidean block Jacobian from input group j to output group i at $\boldsymbol{Z}$. If

$$\sup_{\theta\in[0,1]} \|J_{ii}(\boldsymbol{X}+\boldsymbol{E}_{-i}+\theta\boldsymbol{E}_i)\|_{2\to2} \le \eta_{1,i} < 1, \tag{56}$$

$$\sup_{\theta\in[0,1]} \left(\sum_{j\neq i} \|J_{ij}(\boldsymbol{X}+\theta\boldsymbol{E}_{-i})\|_{2\to2}^2\right)^{1/2} \le L_{1\times,i}, \tag{57}$$

then (40)–(41) *hold.*

Proof. For the target-self segment, the fundamental theorem of calculus gives

$$\mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}) - \mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}_{-i}) = \int_0^1 J_{ii}(\boldsymbol{X}+\boldsymbol{E}_{-i}+\theta\boldsymbol{E}_i)\boldsymbol{e}_i \,\mathrm{d}\theta.$$

The group RMS norm scales both the input and output Euclidean norms by the same factor $D^{-1/2}$, so the block operator norm is unchanged. Equation (56) yields (40). For the cross segment,

$$\mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}_{-i}) - \mathcal{G}_{1,i}(H,\boldsymbol{X}) = \int_0^1 \sum_{j\neq i} J_{ij}(\boldsymbol{X}+\theta\boldsymbol{E}_{-i})\boldsymbol{e}_j \,\mathrm{d}\theta.$$

Cauchy–Schwarz over the agent-group index and (57) give

$$\|\mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}_{-i}) - \mathcal{G}_{1,i}(H,\boldsymbol{X})\|_{\mathrm{g}} \le L_{1\times,i}\left(\sum_{j\neq i}\delta_j^2\right)^{1/2} = L_{1\times,i}\Delta_{-i}.$$

□

The finite-secant assumptions are weaker than uniform Jacobian bounds and can be estimated directly by target-only and cross-group interventions.

Lemma A.10 (Stage 1 candidate error)**.** *Under Assumptions A.5 and A.6,*

$$\left\|\boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \le b_{1,i} := \varepsilon_{1,i} + L_{1H,i}\delta_H + \eta_{1,i}\delta_i + L_{1\times,i}\Delta_{-i}. \tag{58}$$

For a clean group $i \notin F$,

$$\left\|\boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \le b_{1,i}^{\mathrm{c}} := \varepsilon_{1,i} + L_{1H,i}\delta_H + L_{1\times,i}\Delta_{-i}. \tag{59}$$

Proof. Insert the exact-history output at the corrupted input, the output with only cross-group faults, and the clean output:

$$\begin{aligned}\left\|\boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \le{}& \left\|\mathcal{G}_{1,i}(\widehat{H},\boldsymbol{X}+\boldsymbol{E}) - \mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E})\right\|_{\mathrm{g}} \\ &+ \|\mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}) - \mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}_{-i})\|_{\mathrm{g}} \\ &+ \|\mathcal{G}_{1,i}(H,\boldsymbol{X}+\boldsymbol{E}_{-i}) - \mathcal{G}_{1,i}(H,\boldsymbol{X})\|_{\mathrm{g}} \\ &+ \|\mathcal{G}_{1,i}(H,\boldsymbol{X}) - \boldsymbol{x}_i\|_{\mathrm{g}}.\end{aligned}$$

Apply (39), (40), (41), and (38). If $i \notin F$, then $\delta_i = 0$. □

Proposition A.11 (Derived Stage 1 standardized-residual envelopes)**.** *Under Assumptions A.5 and A.6, every clean group satisfies*

$$d_i^{(1)} \le \frac{\sqrt{(b_{1,i}^{\mathrm{c}})^2+\zeta^2}}{a_0}. \tag{60}$$

Every faulty group satisfies

$$d_i^{(1)} \ge \frac{\sqrt{(\underline{r}_{1,i}^{\mathrm{f}})^2+\zeta^2}}{a_{1,\max}}, \tag{61}$$

where

$$\underline{r}^{\mathrm{f}}_{1,i} := [(1-\eta_{1,i})\delta_i - \varepsilon_{1,i} - L_{1H,i}\delta_H - L_{1\times,i}\Delta_{-i}]_+ . \tag{62}$$

Consequently, define the realized envelope values

$$\ell_1 := \zeta / a_{1,\max}, \tag{63}$$

$$u_1 := \max_{i\notin F} \frac{\sqrt{(b^{\mathrm{c}}_{1,i})^2 + \zeta^2}}{a_0}, \tag{64}$$

$$v_1 := \min_{i\in F} \frac{\sqrt{(\underline{r}^{\mathrm{f}}_{1,i})^2 + \zeta^2}}{a_{1,\max}}. \tag{65}$$

All clean discrepancies lie in $[\ell_1, u_1]$, and all faulty discrepancies are at least v_1. When $F = \varnothing$, only ℓ_1 and u_1 are used.

Proof. For $i \notin F$, $\boldsymbol{y}_i = \boldsymbol{x}_i$, and Lemma A.10 gives

$$\left\| \boldsymbol{y}_i - \boldsymbol{c}_i^{(1)} \right\|_{\mathrm{g}} \le b^{\mathrm{c}}_{1,i}.$$

Because $a_i^{(1)} > a_0$, division gives (60). For $i \in F$,

$$\begin{aligned}
\left\| \boldsymbol{y}_i - \boldsymbol{c}_i^{(1)} \right\|_{\mathrm{g}} &= \left\| \boldsymbol{e}_i + (\boldsymbol{x}_i - \boldsymbol{c}_i^{(1)}) \right\|_{\mathrm{g}} \\
&\ge \delta_i - \left\| \boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i \right\|_{\mathrm{g}} \\
&\ge (1-\eta_{1,i})\delta_i - \varepsilon_{1,i} - L_{1H,i}\delta_H - L_{1\times,i}\Delta_{-i}.
\end{aligned}$$

Taking the positive part and using $a_i^{(1)} \le a_{1,\max}$ proves (61). Finally, every $g_i^{(1)} \ge \zeta$ and every $a_i^{(1)} \le a_{1,\max}$, so every discrepancy is at least ℓ_1. □

Remark A.12 (Why target self-copy attenuation is required)**.** *If $\mathcal{G}_{1,i}(H, \boldsymbol{Y}) = \boldsymbol{y}_i$ holds for every history and input, the target-self gain is one and the unstandardized residual $r_i^{(1)} = \boldsymbol{y}_i - \boldsymbol{c}_i^{(1)}$ is zero for every target corruption. Thus $g_i^{(1)} = \zeta$, whereas $d_i^{(1)} = \zeta / a_i^{(1)}$ can vary through the input-dependent scale. Exact copying removes the growth of the unstandardized residual with target-fault amplitude; the sufficient strict self-gain condition excludes this case.*

A.9 Soft-Support Bounds and Median–MAD Conditions

All cancellation and final-recovery results below use only deterministic false-positive and false-negative bounds. The next lemma is the primary support interface. The median–MAD proposition that follows is one robust, interpretable sufficient route to certify its margins; its interval width supplies a worst-case sufficient condition for those margins.

Lemma A.13 (Soft-support bounds from realized discrepancy margins)**.** *Let $p_i = \mathrm{sigm}((d_i - \tau)/T_\tau)$. If a clean group satisfies $d_i \le \tau - \gamma_{\mathrm{c}}$ for $\gamma_{\mathrm{c}} > 0$, then*

$$p_i \le \mathrm{sigm}(-\gamma_{\mathrm{c}}/T_\tau). \tag{66}$$

If a faulty group satisfies $d_i \ge \tau + \gamma_{\mathrm{f}}$ for $\gamma_{\mathrm{f}} > 0$, then

$$1 - p_i \le \mathrm{sigm}(-\gamma_{\mathrm{f}}/T_\tau). \tag{67}$$

Proof. The sigmoid is monotone increasing. The clean result follows by direct substitution, and the faulty result follows from $1 - \mathrm{sigm}(z) = \mathrm{sigm}(-z)$. □

Proposition A.14 (Deterministic median–MAD threshold control)**.** *Let $d_1, \dots, d_n \ge 0$. Suppose a set $C \subseteq \mathcal{N}$ with $|C| > n/2$ satisfies $d_i \in [\ell, u]$ for every $i \in C$. Define*

$$m = \mathrm{median}_i\, d_i, \qquad q = \mathrm{median}_i\, |d_i - m|, \qquad \tau = m + c_\tau(q + b_\tau).$$

Then

$$\ell \le m \le u, \tag{68}$$
$$0 \le q \le u - \ell, \tag{69}$$
$$\ell + c_\tau b_\tau \le \tau \le u + c_\tau(u - \ell + b_\tau). \tag{70}$$

Consequently, every $i \in C$ satisfies

$$\tau - d_i \ge \gamma_{\mathrm{c}}(\ell, u) := \ell + c_\tau b_\tau - u, \tag{71}$$

and every index with $d_i \ge v$ satisfies

$$d_i - \tau \ge \gamma_{\mathrm{f}}(\ell, u, v) := v - u - c_\tau(u - \ell + b_\tau). \tag{72}$$

Proof. A strict majority of values lies in $[\ell, u]$, so a median lies in $[\ell, u]$. Since both d_i and m lie in $[\ell, u]$ for every $i \in C$, more than half of the absolute deviations are at most $u - \ell$; hence their median is at most $u - \ell$. Substitution gives the threshold bounds, and combining those bounds with $d_i \le u$ or $d_i \ge v$ gives the two margins. □

Remark A.15 (Conservative interval certificate)**.** *The interval argument certifies a positive margin for every clean group when*

$$u - \ell < c_\tau b_\tau = 2.05 \times 0.03 = 0.0615. \tag{73}$$

This is a worst-case sufficient condition. Positive realized values of $\tau - d_i$ may hold even when the interval width exceeds 0.0615, and benchmark validation should report the realized margins directly.

Theorem A.16 (Stage 1 soft-support guarantee from the median–MAD certificate)**.** *Under all hypotheses of Proposition A.11, assume $|F| \le s < n/2$. Define*

$$\gamma_{1,\mathrm{c}} := \ell_1 + c_\tau b_\tau - u_1, \tag{74}$$
$$\gamma_{1,\mathrm{f}} := v_1 - u_1 - c_\tau(u_1 - \ell_1 + b_\tau) \quad \text{when } F \ne \varnothing. \tag{75}$$

If the applicable margins are positive, then

$$p_i^{(1)} \le \eta_{1,\mathrm{c}} := \mathrm{sigm}(-\gamma_{1,\mathrm{c}}/T_\tau), \qquad i \notin F, \tag{76}$$
$$1 - p_i^{(1)} \le \eta_{1,\mathrm{f}} := \mathrm{sigm}(-\gamma_{1,\mathrm{f}}/T_\tau), \qquad i \in F. \tag{77}$$

Proof. The clean set has strict majority. Proposition A.11 supplies $[\ell_1, u_1]$ for clean discrepancies and the lower bound v_1 for faulty discrepancies. Apply Proposition A.14, then Lemma A.13. □

Corollary A.17 (Pointwise sufficient Stage 1 detectable amplitude)**.** *Under all hypotheses of Proposition A.11, let*

$$\tau_{1,\max} := u_1 + c_\tau(u_1 - \ell_1 + b_\tau),$$

and fix $\bar\gamma_{1,\mathrm{f}} > 0$. For the realized forward pass, a sufficient condition for $d_i^{(1)} \ge \tau_{1,\max} + \bar\gamma_{1,\mathrm{f}}$ is

$$\delta_i \ge \frac{\varepsilon_{1,i} + L_{1H,i}\delta_H + L_{1\times,i}\Delta_{-i} + \sqrt{\left[a_{1,\max}^2(\tau_{1,\max} + \bar\gamma_{1,\mathrm{f}})^2 - \zeta^2\right]_+}}{1 - \eta_{1,i}}. \tag{78}$$

Proof. The condition makes $\underline{r}_{1,i}^{\mathrm{f}}$ at least the displayed square root. Substitution into (61) gives the desired discrepancy margin. □

The condition is pointwise and uses the realized $a_{1,\max}$. A uniform amplitude corollary over an input family $\mathcal{D}$ may additionally assume $\sup_{(H,\boldsymbol{Y})\in\mathcal{D}} a_{1,\max} \le \bar a_1$ and uniform bounds on the remaining terms. These uniform bounds are additional conditions on the chosen input family. Minimum amplitude is needed only for a uniform support statement in the presence of a nonzero approximation floor; the weak-fault recovery theorem retains smaller faults by their energy.

A.10 INITIAL SPARSE-ERROR CANCELLATION

Lemma A.18 (Deterministic Stage 1 cancellation bound)**.** *Assume the Stage 1 support interface*

$$p_i^{(1)} \le \eta_{1,\mathrm{c}} \quad (i \notin F), \qquad 1 - p_i^{(1)} \le \eta_{1,\mathrm{f}} \quad (i \in F), \tag{79}$$

and candidate bounds $\left\| \boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i \right\|_{\mathrm{g}} \le b_{1,i}$*. Define*

$$B_1 := \left[\sum_{i \in F} b_{1,i}^2 + \eta_{1,\mathrm{c}}^2 \sum_{i \notin F} b_{1,i}^2 \right]^{1/2}. \tag{80}$$

Then

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \eta_{1,\mathrm{f}} \Delta + B_1. \tag{81}$$

For target i, define

$$B_{1,-i} := \left[\sum_{\substack{j \in F \\ j \ne i}} b_{1,j}^2 + \eta_{1,\mathrm{c}}^2 \sum_{\substack{j \notin F \\ j \ne i}} b_{1,j}^2 \right]^{1/2}. \tag{82}$$

The two LOO context errors, measured before embedding in normalized observation coordinates, are

$$\Delta_i^{\mathrm{raw}} := \|\Pi_{-i}(\boldsymbol{Y} - \boldsymbol{X})\|_{\mathrm{G}} = \|\Pi_{-i}\boldsymbol{E}\|_{\mathrm{G}} = \Delta_{-i}, \tag{83}$$

$$\begin{aligned} \Delta_i^{\mathrm{ref}} &:= \left\| \Pi_{-i}(\boldsymbol{X}^{(1)} - \boldsymbol{X}) \right\|_{\mathrm{G}} \\ &\le \eta_{1,\mathrm{f}} \Delta_{-i} + B_{1,-i}. \end{aligned} \tag{84}$$

Proof. The exact identity (27) and the triangle inequality give

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \left\| (\boldsymbol{I} - \boldsymbol{P}^{(1)})\boldsymbol{E} \right\|_{\mathrm{G}} + \left\| \boldsymbol{P}^{(1)}(\boldsymbol{C}^{(1)} - \boldsymbol{X}) \right\|_{\mathrm{G}}. \tag{85}$$

The first term has nonzero rows only on F, so

$$\left\| (\boldsymbol{I} - \boldsymbol{P}^{(1)})\boldsymbol{E} \right\|_{\mathrm{G}}^2 = \sum_{i \in F} (1 - p_i^{(1)})^2 \delta_i^2 \le \eta_{1,\mathrm{f}}^2 \Delta^2.$$

For the candidate-injection term,

$$\begin{aligned} \left\| \boldsymbol{P}^{(1)}(\boldsymbol{C}^{(1)} - \boldsymbol{X}) \right\|_{\mathrm{G}}^2 &= \sum_i (p_i^{(1)})^2 \left\| \boldsymbol{c}_i^{(1)} - \boldsymbol{x}_i \right\|_{\mathrm{g}}^2 \\ &\le \sum_{i \in F} b_{1,i}^2 + \eta_{1,\mathrm{c}}^2 \sum_{i \notin F} b_{1,i}^2 = B_1^2. \end{aligned}$$

This proves (81). Since $\boldsymbol{Y} - \boldsymbol{X} = \boldsymbol{E}$, projecting away target row i gives (83). Apply the same projection to (27) and repeat the row-wise calculation to obtain

$$\left\| \Pi_{-i}(\boldsymbol{X}^{(1)} - \boldsymbol{X}) \right\|_{\mathrm{G}} \le \eta_{1,\mathrm{f}} \Delta_{-i} + B_{1,-i},$$

which is (84). The associated token deviations satisfy

$$\left\| \boldsymbol{Z}_i^{\mathrm{raw}} - \boldsymbol{R}_{i,\mathrm{raw}}^* \right\|_F \le K_{\mathrm{raw},i}^{\mathrm{emb}} \Delta_i^{\mathrm{raw}}, \qquad \left\| \boldsymbol{Z}_i^{\mathrm{ref}} - \boldsymbol{R}_{i,\mathrm{ref}}^* \right\|_F \le K_{\mathrm{ref},i}^{\mathrm{emb}} \Delta_i^{\mathrm{ref}},$$

by (18), with each branch compared to its own ideal masked-token input. □

Corollary A.19 (Partitioned Stage 1 cancellation for weak faults)**.** *Let $F = F_{\mathrm{d}} \dot\cup F_{\mathrm{w}}$. Suppose $1 - p_i^{(1)} \le \eta_{1,\mathrm{f}}$ is certified only for $i \in F_{\mathrm{d}}$, while the trivial bound $1 - p_i^{(1)} \le 1$ is used on F_{w}. Retain the clean false-positive and candidate bounds of Lemma A.18. Define*

$$\Delta_{\mathrm{d}} = \|\Pi_{F_{\mathrm{d}}}\boldsymbol{E}\|_{\mathrm{G}}, \qquad \Delta_{\mathrm{w}} = \|\Pi_{F_{\mathrm{w}}}\boldsymbol{E}\|_{\mathrm{G}},$$

and define $\Delta_{\mathrm{d},-i}$ and $\Delta_{\mathrm{w},-i}$ by additionally deleting target i. Then

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \sqrt{\eta_{1,\mathrm{f}}^2 \Delta_{\mathrm{d}}^2 + \Delta_{\mathrm{w}}^2} + B_1, \tag{86}$$

$$\Delta_i^{\mathrm{ref}} \le \sqrt{\eta_{1,\mathrm{f}}^2 \Delta_{\mathrm{d},-i}^2 + \Delta_{\mathrm{w},-i}^2} + B_{1,-i}. \tag{87}$$

Proof. The missed-error rows on F_{d} and F_{w} are disjoint, and therefore their squared group norm is at most $\eta_{1,\mathrm{f}}^2\Delta_{\mathrm{d}}^2 + \Delta_{\mathrm{w}}^2$. The candidate-injection term is unchanged because every faulty row already used $p_i^{(1)} \le 1$. Deleting target row i gives the second inequality. □

Corollary A.20 (When one cancellation is a contraction)**.** *If* $\Delta > 0$ *and*

$$B_1 < (1 - \eta_{1,\mathrm{f}})\Delta, \tag{88}$$

then

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} < \|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathrm{G}}.$$

Equivalently, with

$$\rho_1 := \eta_{1,\mathrm{f}} + B_1/\Delta,$$

condition (88) *gives* $\rho_1 < 1$ *and*

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \rho_1 \Delta.$$

For target i, the refined LOO context has strictly smaller pre-embedding observation error than the raw LOO context whenever

$$B_{1,-i} < (1 - \eta_{1,\mathrm{f}})\Delta_{-i}. \tag{89}$$

Proof. The global statements follow directly from (81). Under (89), (84) and (83) give $\Delta_i^{\mathrm{ref}} < \Delta_i^{\mathrm{raw}}$. □

Corollary A.21 (A uniform primitive sufficient condition for Stage 1 contraction)**.** *Let* $m = |F| \ge 1$*. Suppose a specified fault family admits common constants satisfying*

$$\varepsilon_{1,i} + L_{1H,i}\delta_H \le \varepsilon_{1H}, \qquad \eta_{1,i} \le \overline{\eta}_1, \qquad L_{1\times,i} \le \overline{L}_{1\times}.$$

Then

$$B_1 \le \left[\overline{\eta}_1 + \overline{L}_{1\times}\left(\sqrt{m-1} + \eta_{1,\mathrm{c}}\sqrt{n-m}\right)\right]\Delta + \left(\sqrt{m} + \eta_{1,\mathrm{c}}\sqrt{n-m}\right)\varepsilon_{1H}. \tag{90}$$

Consequently,

$$\left\| \boldsymbol{X}^{(1)} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \rho_{1,\mathrm{lin}}\Delta + B_{1,0}, \tag{91}$$

where

$$\rho_{1,\mathrm{lin}} := \eta_{1,\mathrm{f}} + \overline{\eta}_1 + \overline{L}_{1\times}\left(\sqrt{m-1} + \eta_{1,\mathrm{c}}\sqrt{n-m}\right), \tag{92}$$

$$B_{1,0} := \left(\sqrt{m} + \eta_{1,\mathrm{c}}\sqrt{n-m}\right)\varepsilon_{1H}. \tag{93}$$

If $\rho_{1,\mathrm{lin}} < 1$ *and* $\Delta > B_{1,0}/(1 - \rho_{1,\mathrm{lin}})$*, Stage 1 is a strict contraction throughout that family.*

Proof. For $i \in F$,

$$b_{1,i} \le \varepsilon_{1H} + \overline{\eta}_1\delta_i + \overline{L}_{1\times}\Delta_{-i},$$

whereas for $i \notin F$, $b_{1,i} \le \varepsilon_{1H} + \overline{L}_{1\times}\Delta$. Apply Minkowski's inequality to the weighted vector defining B_1. The target-self term has norm $\overline{\eta}_1\Delta$. Moreover,

$$\sum_{i\in F}\Delta_{-i}^2 = \sum_{i\in F}(\Delta^2 - \delta_i^2) = (m-1)\Delta^2.$$

The cross terms on clean rows have weighted norm $\eta_{1,\mathrm{c}}\overline{L}_{1\times}\sqrt{n-m}\Delta$, and the constant terms contribute at most $(\sqrt{m} + \eta_{1,\mathrm{c}}\sqrt{n-m})\varepsilon_{1H}$. This proves (90); combine it with (81). □

Remark A.22 (Single-fault boundary case)**.** *If target i is the only faulty group, then* $\Delta_{-i} = 0$*, so the raw LOO context for estimating target i is already clean. The refined context may still contain Stage 1 candidate injection through* $B_{1,-i}$*. The relative context quality therefore depends on the remaining fault energy and candidate injection. Stage 2 retains both inputs, and Assumption A.8 quantifies their separate influence.*

A.11 STAGE 2 CONDITIONAL REFINEMENT

Theorem A.23 (Stage 2 raw/refined LOO candidate bound)**.** *Assume Assumptions A.7 and A.8, and all hypotheses of Lemma A.18: the Stage 1 support interface* (79) *and, for every group j, a candidate bound $\left\|\boldsymbol{c}_j^{(1)} - \boldsymbol{x}_j\right\|_{\mathrm{g}} \le b_{1,j}$ with $b_{1,j} \ge 0$. Let $B_{1,-i}$ be defined by* (82) *using these same candidate bounds and the same clean-support constant $\eta_{1,\mathrm{c}}$. When the explicit choice* (58) *is used for $b_{1,j}$, also assume Assumptions A.5 and A.6, which justify it by Lemma A.10. Then*

$$\left\|\boldsymbol{c}_i^{(2)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \le \beta_{2,i}, \tag{94}$$

where

$$\begin{aligned}\beta_{2,i} := \omega_i + \varepsilon_{2,i}^{\mathrm{net}} &+ (L_{2H,i} + L_{\mathrm{aux},i}K_{H,i})\delta_H \\ &+ L_{\mathrm{raw},i}\Delta_{-i} + L_{\mathrm{ref},i}\left(\eta_{1,\mathrm{f}}\Delta_{-i} + B_{1,-i}\right) \\ &+ L_{\mathrm{aux},i}\left(K_{\mathrm{self},i}\delta_i + K_{\times,i}\Delta_{-i}\right),\end{aligned} \tag{95}$$

and $\omega_i = \omega_i(H, \boldsymbol{X}_{-i})$. Under $\mathrm{RO}(1, \kappa_1; H)$, Proposition A.4 gives $\omega_i = 0$.

Proof. Insert only the ideal clean-context network output:

$$\begin{aligned}\left\|\boldsymbol{c}_i^{(2)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \le &\left\|\boldsymbol{c}_i^{(2)} - \mathcal{G}_{2,i}(H, \boldsymbol{R}_{i,\mathrm{raw}}^*, \boldsymbol{R}_{i,\mathrm{ref}}^*, \boldsymbol{V}_i^*)\right\|_{\mathrm{g}} \\ &+ \left\|\mathcal{G}_{2,i}(H, \boldsymbol{R}_{i,\mathrm{raw}}^*, \boldsymbol{R}_{i,\mathrm{ref}}^*, \boldsymbol{V}_i^*) - \boldsymbol{x}_i\right\|_{\mathrm{g}}.\end{aligned}$$

The second term is at most $\omega_i + \varepsilon_{2,i}^{\mathrm{net}}$ by (53). For the first term, apply the observation-coordinate stability bound (54). Its raw contribution uses

$$\|\Pi_{-i}(\boldsymbol{Y} - \boldsymbol{X})\|_{\mathrm{G}} = \|\Pi_{-i}\boldsymbol{E}\|_{\mathrm{G}} = \Delta_{-i}.$$

The assumed Stage 1 support and candidate bounds establish every hypothesis of Lemma A.18, so its refined-context bound gives

$$\left\|\Pi_{-i}(\boldsymbol{X}^{(1)} - \boldsymbol{X})\right\|_{\mathrm{G}} \le \eta_{1,\mathrm{f}}\Delta_{-i} + B_{1,-i}.$$

The embedding gains are already included in $L_{\mathrm{raw},i}$ and $L_{\mathrm{ref},i}$ as specified in Assumption A.8. Apply (55) and collect terms to obtain (95). □

Corollary A.24 (What the intermediate cancellation improves)**.** *Under all hypotheses of Theorem A.23, if* (89) *holds for target i, then, whenever $L_{\mathrm{ref},i} > 0$,*

$$L_{\mathrm{ref},i}\left(\eta_{1,\mathrm{f}}\Delta_{-i} + B_{1,-i}\right) < L_{\mathrm{ref},i}\Delta_{-i}.$$

Thus the refined observation-error contribution to the candidate bound is strictly smaller than $L_{\mathrm{ref},i}\Delta_{-i}$. The coefficient includes the refined embedding and encoder sensitivities. The complete Stage 2 candidate also depends on the raw branch and target-visible auxiliary path through their respective error terms.

Proof. Multiply (89) by $L_{\mathrm{ref},i} > 0$. □

For Stage 2 support analysis, define the target-visible self gain

$$\chi_i := L_{\mathrm{aux},i}K_{\mathrm{self},i} \tag{96}$$

and the remaining contribution

$$\begin{aligned}A_{2,i}^{\neg\mathrm{self}} := \omega_i + \varepsilon_{2,i}^{\mathrm{net}} &+ (L_{2H,i} + L_{\mathrm{aux},i}K_{H,i})\delta_H \\ &+ L_{\mathrm{raw},i}\Delta_{-i} + L_{\mathrm{ref},i}\left(\eta_{1,\mathrm{f}}\Delta_{-i} + B_{1,-i}\right) + L_{\mathrm{aux},i}K_{\times,i}\Delta_{-i}.\end{aligned} \tag{97}$$

Then

$$\beta_{2,i} = A_{2,i}^{\neg\mathrm{self}} + \chi_i\delta_i. \tag{98}$$

Remark A.25 (Fixed-depth interpretation)**.** *The proved benefit concerns the refined-context term: one candidate is used to cancel part of the sparse error, and a parameter-distinct conditional estimator then receives the corrected LOO context. Two stages are the smallest fixed-depth realization of "estimate–cancel–re-estimate." The two stages have distinct parameters; additional refinement depth is an empirical design choice.*

Proposition A.26 (Derived Stage 2 standardized-residual envelopes)**.** *Under all hypotheses of Theorem A.23, every clean group satisfies*

$$d_i^{(2)} \le \frac{\sqrt{\beta_{2,i}^2 + \zeta^2}}{a_0}. \tag{99}$$

Every faulty group satisfies

$$d_i^{(2)} \ge \frac{\sqrt{(\underline{r}_{2,i}^{\mathrm{f}})^2 + \zeta^2}}{a_{2,\max}}, \tag{100}$$

where

$$\underline{r}_{2,i}^{\mathrm{f}} := \left[(1-\chi_i)\delta_i - A_{2,i}^{\neg\mathrm{self}}\right]_+. \tag{101}$$

Define

$$\ell_2 := \zeta / a_{2,\max}, \tag{102}$$

$$u_2 := \max_{i \notin F} \frac{\sqrt{\beta_{2,i}^2 + \zeta^2}}{a_0}, \tag{103}$$

$$v_2 := \min_{i \in F} \frac{\sqrt{(\underline{r}_{2,i}^{\mathrm{f}})^2 + \zeta^2}}{a_{2,\max}}. \tag{104}$$

Then all clean Stage 2 discrepancies lie in $[\ell_2, u_2]$, and all faulty Stage 2 discrepancies are at least v_2.

Proof. For $i \notin F$, $\boldsymbol{y}_i = \boldsymbol{x}_i$, so (94) and $a_i^{(2)} > a_0$ give (99). For $i \in F$,

$$\begin{aligned}\left\|\boldsymbol{y}_i - \boldsymbol{c}_i^{(2)}\right\|_{\mathrm{g}} &\ge \delta_i - \left\|\boldsymbol{c}_i^{(2)} - \boldsymbol{x}_i\right\|_{\mathrm{g}} \\ &\ge \delta_i - A_{2,i}^{\neg\mathrm{self}} - \chi_i\delta_i.\end{aligned}$$

Take the positive part and use $a_i^{(2)} \le a_{2,\max}$. The common lower endpoint follows from $g_i^{(2)} \ge \zeta$. □

Theorem A.27 (Stage 2 soft-support guarantee from the median–MAD certificate)**.** *Assume all hypotheses of Proposition A.26, including the full conditions of Theorem A.23, and $|F| \le s < n/2$. Define*

$$\gamma_{2,\mathrm{c}} := \ell_2 + c_\tau b_\tau - u_2, \tag{105}$$

$$\gamma_{2,\mathrm{f}} := v_2 - u_2 - c_\tau(u_2 - \ell_2 + b_\tau) \quad \textit{when } F \neq \varnothing. \tag{106}$$

If the applicable margins are positive, then

$$p_i^{(2)} \le \eta_{2,\mathrm{c}} := \mathrm{sigm}(-\gamma_{2,\mathrm{c}}/T_\tau), \qquad i \notin F, \tag{107}$$

$$1 - p_i^{(2)} \le \eta_{2,\mathrm{f}} := \mathrm{sigm}(-\gamma_{2,\mathrm{f}}/T_\tau), \qquad i \in F. \tag{108}$$

Proof. Apply Proposition A.14 to the envelopes in Proposition A.26, followed by Lemma A.13. □

Corollary A.28 (Stage 2 target-visible gain and detectable amplitude)**.** *Under all hypotheses of Proposition A.26, let*

$$\tau_{2,\max} := u_2 + c_\tau(u_2 - \ell_2 + b_\tau),$$

and fix $\bar{\gamma}_{2,\mathrm{f}} > 0$. For the realized forward pass, the condition $\chi_i < 1$ together with

$$\delta_i \ge \frac{A_{2,i}^{\neg\mathrm{self}} + \sqrt{\left[a_{2,\max}^2(\tau_{2,\max} + \bar{\gamma}_{2,\mathrm{f}})^2 - \zeta^2\right]_+}}{1 - \chi_i} \tag{109}$$

is sufficient for $d_i^{(2)} \geq \tau_{2,\max} + \bar{\gamma}_{2,\mathrm{f}}$.

For a uniform corollary over a specified fault family, suppose additionally

$$A_{2,i}^{\neg\mathrm{self}} \leq \overline{A}_{2,i} + \chi_{\times,i}\delta_i, \qquad \tau_{2,\max} \leq \overline{\tau}_2, \qquad a_{2,\max} \leq \overline{a}_2, \qquad \chi_i + \chi_{\times,i} < 1.$$

Then it is sufficient that

$$\delta_i \geq \frac{\overline{A}_{2,i} + \sqrt{\left[\overline{a}_2^2(\overline{\tau}_2 + \bar{\gamma}_{2,\mathrm{f}})^2 - \zeta^2\right]_+}}{1 - \chi_i - \chi_{\times,i}}. \tag{110}$$

Proof. Equation (100) yields the desired margin once

$$(1 - \chi_i)\delta_i - A_{2,i}^{\neg\mathrm{self}} \geq \sqrt{\left[a_{2,\max}^2(\tau_{2,\max} + \bar{\gamma}_{2,\mathrm{f}})^2 - \zeta^2\right]_+}.$$

Rearrangement proves (109). Substitute the three uniform bounds and collect the δ_i terms to obtain (110). □

Because $A_{2,i}^{\neg\mathrm{self}}$ can depend on δ_i through cross-group and Stage 1 pathways, a uniform amplitude theorem additionally requires control of these pathways. When the total target-corruption gain is at least one, establishing residual growth with amplitude requires a sharper directional residual condition.

A.12 Support fusion, final recovery, and strict improvement

Lemma A.29 (Fixed support-fusion bounds)**.** *Suppose the two stages satisfy the deterministic support interfaces*

$$p_i^{(r)} \leq \eta_{r,\mathrm{c}} \quad (i \notin F), \qquad 1 - p_i^{(r)} \leq \eta_{r,\mathrm{f}} \quad (i \in F), \qquad r \in \{1, 2\}.$$

For $p_i = \alpha p_i^{(1)} + (1 - \alpha)p_i^{(2)}$*, define*

$$\eta_{\mathrm{c}} := \alpha\eta_{1,\mathrm{c}} + (1 - \alpha)\eta_{2,\mathrm{c}}, \tag{111}$$
$$\eta_{\mathrm{f}} := \alpha\eta_{1,\mathrm{f}} + (1 - \alpha)\eta_{2,\mathrm{f}}. \tag{112}$$

Then

$$p_i \leq \eta_{\mathrm{c}} \quad (i \notin F), \qquad 1 - p_i \leq \eta_{\mathrm{f}} \quad (i \in F).$$

The implementation uses the fixed fusion weight $\alpha = 1/4$*.*

Proof. Both statements follow by expanding the convex combination. The implementation clamp is inactive because the convex combination already lies in $[0, 1]$. □

Theorem A.30 (Final normalized observation-recovery bound)**.** *Suppose the Stage 2 candidate bounds* (94) *hold for every target and the stagewise support interfaces in Lemma A.29 are available. In particular, using Theorem A.23 to obtain the explicit bounds* (95) *requires all of that theorem's hypotheses, including the Stage 1 candidate bounds. Define*

$$B_2 := \left[\sum_{i \in F} \beta_{2,i}^2 + \eta_{\mathrm{c}}^2 \sum_{i \notin F} \beta_{2,i}^2\right]^{1/2}. \tag{113}$$

Then

$$\boxed{\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{G}} \leq \eta_{\mathrm{f}}\Delta + B_2.} \tag{114}$$

The median–MAD Theorems A.16 and A.27 provide one sufficient certification of the required support interfaces; the recovery statement applies to any estimator satisfying those interfaces.

Proof. Apply the triangle inequality to (32):

$$\left\|\widehat{\boldsymbol{X}}-\boldsymbol{X}\right\|_{\mathrm{G}} \le \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{E}\|_{\mathrm{G}} + \left\|\boldsymbol{P}(\boldsymbol{C}^{(2)}-\boldsymbol{X})\right\|_{\mathrm{G}}. \tag{115}$$

For the missed-error term,

$$\|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{E}\|_{\mathrm{G}}^2 = \sum_{i\in F}(1-p_i)^2\delta_i^2 \le \eta_{\mathrm{f}}^2\Delta^2.$$

For candidate injection,

$$\begin{aligned}\left\|\boldsymbol{P}(\boldsymbol{C}^{(2)}-\boldsymbol{X})\right\|_{\mathrm{G}}^2 &= \sum_i p_i^2\left\|\boldsymbol{c}_i^{(2)}-\boldsymbol{x}_i\right\|_{\mathrm{g}}^2 \\ &\le \sum_{i\in F}\beta_{2,i}^2 + \eta_{\mathrm{c}}^2\sum_{i\notin F}\beta_{2,i}^2 = B_2^2.\end{aligned}$$

Substitute both bounds into (115). □

Corollary A.31 (Interpretable decomposition of candidate injection)**.** *Under Theorem A.30, with $\beta_{2,i}$ given by* (95) *under all hypotheses of Theorem A.23, define the exact faulty-candidate and clean false-support components*

$$B_{2,F} := \left(\sum_{i\in F}\beta_{2,i}^2\right)^{1/2}, \tag{116}$$

$$B_{2,C}^{\mathrm{fp}} := \eta_{\mathrm{c}}\left(\sum_{i\notin F}\beta_{2,i}^2\right)^{1/2}. \tag{117}$$

Then

$$B_2 = \sqrt{B_{2,F}^2 + (B_{2,C}^{\mathrm{fp}})^2} \le B_{2,F} + B_{2,C}^{\mathrm{fp}}. \tag{118}$$

Here $B_{2,F}$ bounds replacement error on truly faulty groups, while $B_{2,C}^{\mathrm{fp}}$ is candidate error injected into clean groups through false-positive support.

For a componentwise expansion, let

$$w_i := \begin{cases}1, & i\in F,\\ \eta_{\mathrm{c}}, & i\notin F.\end{cases}$$

Define

$$A_{\mathrm{amb}} := \|(w_i\omega_i)_{i=1}^n\|_2, \tag{119}$$

$$A_{\mathrm{model}} := \left\|(w_i\varepsilon_{2,i}^{\mathrm{net}})_{i=1}^n\right\|_2, \tag{120}$$

$$A_{\mathrm{hist}} := \left\|(w_i(L_{2H,i}+L_{\mathrm{aux},i}K_{H,i})\delta_H)_{i=1}^n\right\|_2, \tag{121}$$

$$A_{\mathrm{raw}} := \left\|(w_iL_{\mathrm{raw},i}\Delta_{-i})_{i=1}^n\right\|_2, \tag{122}$$

$$A_{\mathrm{ref}} := \left\|(w_iL_{\mathrm{ref},i}(\eta_{1,\mathrm{f}}\Delta_{-i}+B_{1,-i}))_{i=1}^n\right\|_2, \tag{123}$$

$$A_{\mathrm{aux}} := \left\|(w_iL_{\mathrm{aux},i}(K_{\mathrm{self},i}\delta_i+K_{\times,i}\Delta_{-i}))_{i=1}^n\right\|_2. \tag{124}$$

Then

$$B_2 \le A_{\mathrm{amb}}+A_{\mathrm{model}}+A_{\mathrm{hist}}+A_{\mathrm{raw}}+A_{\mathrm{ref}}+A_{\mathrm{aux}}, \tag{125}$$

and therefore

$$\begin{aligned}\left\|\widehat{\boldsymbol{X}}-\boldsymbol{X}\right\|_{\mathrm{G}} \le\ & \underbrace{\eta_{\mathrm{f}}\Delta}_{\text{missed sparse error}} + \underbrace{A_{\mathrm{amb}}}_{\text{conditional ambiguity}} + \underbrace{A_{\mathrm{model}}}_{\text{learned approximation}} + \underbrace{A_{\mathrm{hist}}}_{\text{history mismatch}} \\ &+ \underbrace{A_{\mathrm{raw}}}_{\text{raw LOO context}} + \underbrace{A_{\mathrm{ref}}}_{\text{refined LOO context}} + \underbrace{A_{\mathrm{aux}}}_{\text{target-visible auxiliary path}}.\end{aligned} \tag{126}$$

Under RO(1)*, $A_{\mathrm{amb}}=0$.*

Proof. The exact split (118) follows from (113). Also $B_2 = \|(w_i\beta_{2,i})_{i=1}^n\|_2$. Substitute (95) and apply Minkowski's inequality to its six component vectors, proving (125); combine with Theorem A.30. □

Corollary A.32 (Final strict improvement and contraction)**.** *Under all hypotheses of Theorem A.30, let* $\Delta = \|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathrm{G}} = \|\boldsymbol{E}\|_{\mathrm{G}} > 0$*. If*

$$B_2 < (1 - \eta_{\mathrm{f}})\Delta, \tag{127}$$

then

$$\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{G}} < \|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathrm{G}}\,. \tag{128}$$

Equivalently, define

$$\rho_{\mathrm{final}} := \eta_{\mathrm{f}} + B_2/\Delta. \tag{129}$$

Condition (127) *gives* $\rho_{\mathrm{final}} < 1$ *and*

$$\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{G}} \le \rho_{\mathrm{final}}\Delta. \tag{130}$$

A more conservative but fully decomposed sufficient condition is

$$A_{\mathrm{amb}} + A_{\mathrm{model}} + A_{\mathrm{hist}} + A_{\mathrm{raw}} + A_{\mathrm{ref}} + A_{\mathrm{aux}} < (1 - \eta_{\mathrm{f}})\Delta. \tag{131}$$

Proof. Substitute (127) into (114). Equation (129) is an algebraic rearrangement. The decomposed condition implies the exact condition through (125). □

Remark A.33 (Stage 1 history error inside the refined contribution)**.** *When* $b_{1,j}$ *is chosen as* (58) *under Assumptions A.5 and A.6, the term* $B_{1,-i}$ *contains Stage 1 history error through* $b_{1,j}$*. For a completely separated history coefficient, write*

$$b_{1,j} \le \widetilde{b}_{1,j} + L_{1H,j}\delta_H, \qquad \widetilde{b}_{1,j} = \varepsilon_{1,j} + \eta_{1,j}\delta_j + L_{1\times,j}\Delta_{-j},$$

and apply Minkowski's inequality to (82)*. This adds*

$$L_{\mathrm{ref},i}\left[\sum_{\substack{j\in F\\ j\neq i}} L_{1H,j}^2 + \eta_{1,\mathrm{c}}^2 \sum_{\substack{j\notin F\\ j\neq i}} L_{1H,j}^2\right]^{1/2} \delta_H$$

to the explicit history component for target i*.*

Theorem A.34 (Signal recovery with detected and weak fault groups)**.** *Let* $F = F_{\mathrm{d}} \dot\cup F_{\mathrm{w}}$*. Assume Assumptions A.7 and A.8, and all hypotheses of Corollary A.19. Explicitly, for every clean group* $j \notin F$*, require* $p_j^{(1)} \le \eta_{1,\mathrm{c}}$*; for every group* j*, require* $\left\|\boldsymbol{c}_j^{(1)} - \boldsymbol{x}_j\right\|_{\mathrm{g}} \le b_{1,j}$ *with* $b_{1,j} \ge 0$*; and on* F_{d}*, require* $1 - p_j^{(1)} \le \eta_{1,\mathrm{f}}$*. Define* $B_{1,-i}$ *by* (82) *from these same quantities, and use* $\Delta_{\mathrm{d},-i}$ *and* $\Delta_{\mathrm{w},-i}$ *from Corollary A.19. If* $b_{1,j}$ *is chosen as* (58)*, also assume Assumptions A.5 and A.6 as required by Lemma A.10.*

For the final fused scores, assume $p_j \le \eta_{\mathrm{c}}$ *on* F^c *and* $1 - p_j \le \eta_{\mathrm{f}}$ *on* F_{d}*. Only* $0 \le p_j^{(1)}, p_j \le 1$ *is required on* F_{w}*. Define*

$$\begin{aligned}\widetilde{\beta}_{2,i} := \omega_i + \varepsilon_{2,i}^{\mathrm{net}} &+ (L_{2H,i} + L_{\mathrm{aux},i}K_{H,i})\delta_H + L_{\mathrm{raw},i}\Delta_{-i}\\ &+ L_{\mathrm{ref},i}\left(\sqrt{\eta_{1,\mathrm{f}}^2\Delta_{\mathrm{d},-i}^2 + \Delta_{\mathrm{w},-i}^2} + B_{1,-i}\right)\\ &+ L_{\mathrm{aux},i}\left(K_{\mathrm{self},i}\delta_i + K_{\times,i}\Delta_{-i}\right),\end{aligned} \tag{132}$$

and

$$\widetilde{B}_2 := \left[\sum_{i\in F}\widetilde{\beta}_{2,i}^2 + \eta_{\mathrm{c}}^2\sum_{i\notin F}\widetilde{\beta}_{2,i}^2\right]^{1/2}. \tag{133}$$

Then

$$\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{G}} \le \sqrt{\eta_{\mathrm{f}}^2\left\|\Pi_{F_{\mathrm{d}}}\boldsymbol{E}\right\|_{\mathrm{G}}^2 + \left\|\Pi_{F_{\mathrm{w}}}\boldsymbol{E}\right\|_{\mathrm{G}}^2} + \widetilde{B}_2. \tag{134}$$

The uniform support guarantee uses a minimum-amplitude condition, while the signal-recovery guarantee accounts for undetected weak faults through their energy.

Proof. The stated Stage 1 clean-support, candidate-error, and detected-fault bounds establish every hypothesis of Corollary A.19. It gives the pre-embedding refined-context bound (87). Applying Assumptions A.7 and A.8 as in the proof of Theorem A.23, with that partitioned bound in place of (84), yields

$$\left\| \boldsymbol{c}_i^{(2)} - \boldsymbol{x}_i \right\|_{\mathrm{g}} \le \widetilde{\beta}_{2,i}.$$

This step uses the partitioned support hypotheses, with no false-negative requirement on F_{w}. In the final missed-error term, the rows in F_{d} and F_{w} are disjoint, so

$$\|(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{E}\|_{\mathrm{G}}^2 \le \eta_{\mathrm{f}}^2 \, \|\Pi_{F_{\mathrm{d}}} \boldsymbol{E}\|_{\mathrm{G}}^2 + \|\Pi_{F_{\mathrm{w}}} \boldsymbol{E}\|_{\mathrm{G}}^2 .$$

The candidate-injection term is at most $\widetilde{B}_2$: faulty rows use $p_i \le 1$, and clean rows use $p_i \le \eta_{\mathrm{c}}$. Apply (32). □

Remark A.35 (Weak-fault compatibility with strict improvement)**.** *Because F_{d} and F_{w} are disjoint,*

$$\Delta^2 = \|\Pi_{F_{\mathrm{d}}} \boldsymbol{E}\|_{\mathrm{G}}^2 + \|\Pi_{F_{\mathrm{w}}} \boldsymbol{E}\|_{\mathrm{G}}^2 .$$

Under all hypotheses of Theorem A.34, strict signal improvement follows whenever its right-hand side is smaller than Δ. For F_{w}, the condition bounds residual energy and candidate injection while allowing missed support.

Corollary A.36 (Clean-input error and exact fixed points)**.** *If $\boldsymbol{E} = 0$ and every group satisfies $p_i \le \eta_{\mathrm{c}}$, then*

$$\left\| \widehat{\boldsymbol{X}} - \boldsymbol{X} \right\|_{\mathrm{G}} \le \eta_{\mathrm{c}} \left\| \boldsymbol{C}^{(2)} - \boldsymbol{X} \right\|_{\mathrm{G}} . \tag{135}$$

In particular, $\widehat{\boldsymbol{X}} = \boldsymbol{X}$ if either $\boldsymbol{P} = 0$ or $\boldsymbol{C}^{(2)} = \boldsymbol{X}$. If both the clean support and candidate are exact, $\boldsymbol{X}$ is an exact fixed point.

Proof. When $\boldsymbol{E} = 0$, (32) becomes $\widehat{\boldsymbol{X}} - \boldsymbol{X} = \boldsymbol{P}(\boldsymbol{C}^{(2)} - \boldsymbol{X})$. Square the group norm and use $p_i \le \eta_{\mathrm{c}}$. □

Corollary A.37 (Return to physical feature coordinates)**.** *Let $\sigma_{\max} = \max_j \sigma_{\mathrm{cl},j}$. Then*

$$\left\| \widehat{\boldsymbol{X}}^{\mathrm{phys}} - \boldsymbol{X}^{\mathrm{phys}} \right\|_F \le \sigma_{\max} \sqrt{D} \, \left\| \widehat{\boldsymbol{X}} - \boldsymbol{X} \right\|_{\mathrm{G}} . \tag{136}$$

Normalization preserves group support but changes feature-wise error weights.

Proof. For normalized entrywise error $\epsilon_{ij} = \widehat{X}_{ij} - X_{ij}$,

$$\left\| \widehat{\boldsymbol{X}}^{\mathrm{phys}} - \boldsymbol{X}^{\mathrm{phys}} \right\|_F^2 = \sum_{i,j} \sigma_{\mathrm{cl},j}^2 \epsilon_{ij}^2 \le \sigma_{\max}^2 \left\| \widehat{\boldsymbol{X}} - \boldsymbol{X} \right\|_F^2 = \sigma_{\max}^2 D \left\| \widehat{\boldsymbol{X}} - \boldsymbol{X} \right\|_{\mathrm{G}}^2 .$$

□

A.13 Recovered-history error propagation

Let $\mathcal{R}(H, \boldsymbol{Y})$ denote the complete deployed two-stage recovery operator, including support estimation and fusion. The next result is conditional on a local Lipschitz property of this full operator with respect to its history argument; the preceding single-step results provide the base error when the clean history is supplied.

Assumption A.9 (Full recovery-operator history stability)**.** *On the analyzed sequential-deployment domain,*

$$\left\| \mathcal{R}(\widehat{H}, \boldsymbol{Y}) - \mathcal{R}(H, \boldsymbol{Y}) \right\|_{\mathrm{G}} \le L_{\mathcal{R},H} \left\| \widehat{H} - H \right\|_{\mathcal{H}} . \tag{137}$$

For an episode indexed from $t = 0$, let

$$e_t := \left\| \widehat{\boldsymbol{X}}_t - \boldsymbol{X}_t \right\|_{\mathrm{G}} .$$

Let $r_t = \min\{t, T-1\}$ be the number of available past recovered observations and $m_t = T-1-r_t$ the number of missing history slots. The implementation fills each missing slot with the current raw $\boldsymbol{Y}_t$. Define the matching clean reference padding by filling the same slots with $\boldsymbol{X}_t$. Then

$$\delta_{H,t} \le \left(\sum_{k=1}^{r_t} e_{t-k}^2 + m_t \Delta_t^2 \right)^{1/2}, \qquad \Delta_t = \|\boldsymbol{E}_t\|_{\mathrm{G}}. \tag{138}$$

Corollary A.38 (Finite-window history propagation)**.** *Suppose the exact-reference-history single-step bound at time t is b_t, meaning*

$$\left\| \mathcal{R}(H_{t,\mathrm{pad}}^*, \boldsymbol{Y}_t) - \boldsymbol{X}_t \right\|_{\mathrm{G}} \le b_t,$$

where $H_{t,\mathrm{pad}}^*$ uses the matching clean padding described above. Under Assumption A.9,

$$e_t \le b_t + L_{\mathcal{R},H} \left(\sum_{k=1}^{r_t} e_{t-k}^2 + m_t \Delta_t^2 \right)^{1/2}. \tag{139}$$

After the history window is full, $m_t = 0$, and with

$$M_{t-1} := \max_{0 \le u < t} e_u, \qquad \lambda_H := L_{\mathcal{R},H} \sqrt{T-1},$$

we have

$$e_t \le b_t + \lambda_H M_{t-1}. \tag{140}$$

If $b_t \le B$ and $\lambda_H < 1$, then

$$\sup_{t \ge 0} e_t \le \max \left\{ \max_{0 \le t < T-1} e_t, \frac{B}{1-\lambda_H} \right\}. \tag{141}$$

If $\lambda_H \ge 1$, (140) *gives a horizon-dependent bound; a uniform long-horizon guarantee requires the additional contraction condition.*

Proof. Insert the recovery output computed with the matching clean reference history:

$$\begin{aligned} e_t &\le \left\| \mathcal{R}(\widehat{H}_t, \boldsymbol{Y}_t) - \mathcal{R}(H_{t,\mathrm{pad}}^*, \boldsymbol{Y}_t) \right\|_{\mathrm{G}} + \left\| \mathcal{R}(H_{t,\mathrm{pad}}^*, \boldsymbol{Y}_t) - \boldsymbol{X}_t \right\|_{\mathrm{G}} \\ &\le L_{\mathcal{R},H} \delta_{H,t} + b_t. \end{aligned}$$

Equation (138) gives (139). For a full window,

$$\left(\sum_{k=1}^{T-1} e_{t-k}^2 \right)^{1/2} \le \sqrt{T-1}\, M_{t-1},$$

which proves (140). Set

$$M^* = \max \left\{ \max_{0 \le t < T-1} e_t, \frac{B}{1-\lambda_H} \right\}.$$

If all previous errors are at most M^*, then

$$e_t \le B + \lambda_H M^* \le M^*.$$

Induction proves (141). □

Remark A.39 (Startup padding domain)**.** *Before the window is full, the matching clean padded history is a synthetic reference formed by repeating the current clean observation. The exact-reference quantity b_t in Corollary A.38 must therefore be established on this padded input domain separately. Proposition A.1 applies directly to genuine clean histories; the padding rule can be included as an additional deterministic branch if a startup low-degree-of-freedom statement is required.*

Remark A.40 (Explicit finite-horizon growth)**.** *If $b_t \le B$ and the coarse recursion $M_t \le B + \lambda_H M_{t-1}$ is used, then for $\lambda_H > 1$,*

$$M_t \le \lambda_H^t M_0 + B \frac{\lambda_H^t - 1}{\lambda_H - 1},$$

and for $\lambda_H = 1$, $M_t \le M_0 + tB$. These bounds explain why dynamically changing faults and recovered-history feedback can be harder than isolated single-step corruption through the explicit feedback of recovered history.

A.14 FROZEN RECURRENT ACTOR STABILITY

The policy notation $\pi_{\phi,i}(\cdot \mid \boldsymbol{x}, \boldsymbol{h}, \text{avail})$ includes the recurrent feature update, with $\boldsymbol{h}$ the incoming state. In $F_{\phi,i}(\boldsymbol{h}, \boldsymbol{x}, a)$, the common action argument is retained for the general statement; the deployed MAPPO update is invariant to this argument. The following result compares the frozen recurrent actor under clean and recovered observations along a common action and availability sequence. It is a response-stability statement; a closed-loop return comparison requires additional transition and occupancy stability assumptions.

Assumption A.10 (Local frozen-actor response and hidden-state stability)**.** *For agent i, the frozen policy and recurrent update satisfy*

$$\mathrm{TV}(\pi_{\phi,i}(\cdot \mid \boldsymbol{x}, \boldsymbol{h}, \text{avail}), \pi_{\phi,i}(\cdot \mid \boldsymbol{x}', \boldsymbol{h}', \text{avail})) \le L_{\pi x,i} \left\|\boldsymbol{x} - \boldsymbol{x}'\right\|_2 + L_{\pi h,i} \left\|\boldsymbol{h} - \boldsymbol{h}'\right\|_2, \quad (142)$$

$$\left\|F_{\phi,i}(\boldsymbol{h}, \boldsymbol{x}, a) - F_{\phi,i}(\boldsymbol{h}', \boldsymbol{x}', a)\right\|_2 \le \kappa_{h,i} \left\|\boldsymbol{h} - \boldsymbol{h}'\right\|_2 + L_{F,i} \left\|\boldsymbol{x} - \boldsymbol{x}'\right\|_2. \quad (143)$$

Corollary A.41 (Frozen recurrent actor response bound)**.** *Let*

$$\epsilon_{t,i} := \left\|\widehat{\boldsymbol{x}}_{t,i}^{\text{phys}} - \boldsymbol{x}_{t,i}^{\text{phys}}\right\|_2, \qquad h_{t,i}^{\Delta} := \left\|\widehat{\boldsymbol{h}}_{t,i} - \boldsymbol{h}_{t,i}\right\|_2.$$

Along a common action and availability sequence,

$$h_{t,i}^{\Delta} \le \kappa_{h,i}^{t} h_{0,i}^{\Delta} + L_{F,i} \sum_{k=0}^{t-1} \kappa_{h,i}^{t-1-k} \epsilon_{k,i}. \quad (144)$$

Consequently,

$$\mathrm{TV}\left(\pi_{\phi,i}^{\text{clean},t}, \pi_{\phi,i}^{\text{rec},t}\right) \le L_{\pi x,i}\epsilon_{t,i} + L_{\pi h,i}\kappa_{h,i}^{t} h_{0,i}^{\Delta} + L_{\pi h,i} L_{F,i} \sum_{k=0}^{t-1} \kappa_{h,i}^{t-1-k} \epsilon_{k,i}. \quad (145)$$

If $h_{0,i}^{\Delta} = 0$, $\epsilon_{t,i} \le \bar{\epsilon}_i$, and $\kappa_{h,i} < 1$, then

$$\mathrm{TV}\left(\pi_{\phi,i}^{\text{clean},t}, \pi_{\phi,i}^{\text{rec},t}\right) \le \left(L_{\pi x,i} + \frac{L_{\pi h,i} L_{F,i}}{1 - \kappa_{h,i}}\right) \bar{\epsilon}_i. \quad (146)$$

For $\kappa_{h,i} \ge 1$, the applicable result is the finite-horizon bound (145)*; the uniform bound* (146) *requires $\kappa_{h,i} < 1$.*

Proof. Equation (143) gives the one-step recursion

$$h_{t+1,i}^{\Delta} \le \kappa_{h,i} h_{t,i}^{\Delta} + L_{F,i}\epsilon_{t,i}.$$

Iterating it proves (144). Substitute this expression into (142) to obtain (145). If $\kappa_{h,i} < 1$, bound the geometric sum by $(1 - \kappa_{h,i})^{-1}$ to obtain (146). □

Remark A.42 (Conditions for extending actor-response bounds to returns)**.** *Once clean and recovered actors choose different actions, their recurrent states, environment states, available-action masks, and future observation distributions can diverge. A cumulative-return theorem would additionally require bounded rewards and quantitative stability of the transition kernel and occupancy distributions under action-distribution perturbations. Monotonic return improvement would further require a connection between these perturbation bounds and policy performance.*

A.15 LOCAL INTERPRETATION OF DECISION-AWARE TRAINING

The observation-recovery theorems depend on the deployed operator properties. Decision-aware training uses the frozen actor to weight observation errors by their effect on action responses. The following result gives the local actor metric and an approximation bound for the restricted-access variant in Appendix C.6.

Proposition A.43 (Actor Fisher metric and bounded surrogate bias)**.** *Fix a nonempty available-action set $\mathcal{A}_{\mathrm{av}}$. Let $\pi_\phi(\cdot \mid \boldsymbol{x}, \boldsymbol{h}, \mathcal{A}_{\mathrm{av}})$ be a fixed differentiable categorical actor whose probabilities are strictly positive on $\mathcal{A}_{\mathrm{av}}$, and suppress the fixed availability argument below. Let $\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{h})$ be its available-action logits,*

$$J_\ell(\boldsymbol{x}, \boldsymbol{h}) = \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{x}}, \qquad \mathcal{F}_\pi(\boldsymbol{x}, \boldsymbol{h}) = \mathrm{diag}(\pi_\phi) - \pi_\phi \pi_\phi^\top.$$

For a sufficiently small $\delta\boldsymbol{x}$,

$$\mathrm{KL}(\pi_\phi(\cdot \mid \boldsymbol{x}, \boldsymbol{h}) \,\|\, \pi_\phi(\cdot \mid \boldsymbol{x} + \delta\boldsymbol{x}, \boldsymbol{h})) = \frac{1}{2} \delta\boldsymbol{x}^\top J_\ell^\top \mathcal{F}_\pi J_\ell \delta\boldsymbol{x} + o(\|\delta\boldsymbol{x}\|_2^2). \tag{147}$$

For the restricted-access variant, let $q(\cdot \mid \boldsymbol{x}, \boldsymbol{h})$ abbreviate its frozen differentiable surrogate $\widetilde{\pi}_\psi$ (Appendix C.6), with the common available-action set suppressed. On a neighborhood $\mathcal{U}$ of $\boldsymbol{x}$, assume

$$\min_a \pi_{\phi,a}(\boldsymbol{x}', \boldsymbol{h}) \geq \varrho, \qquad \min_a q_a(\boldsymbol{x}', \boldsymbol{h}) \geq \varrho, \qquad \|q(\cdot \mid \boldsymbol{x}', \boldsymbol{h}) - \pi_\phi(\cdot \mid \boldsymbol{x}', \boldsymbol{h})\|_1 \leq \varepsilon_q$$

for every $\boldsymbol{x}' \in \mathcal{U}$, where $\varrho > 0$. Then, whenever $\boldsymbol{x} + \delta\boldsymbol{x} \in \mathcal{U}$,

$$|\mathrm{KL}(\pi_\phi(\cdot \mid \boldsymbol{x}, \boldsymbol{h}) \,\|\, q(\cdot \mid \boldsymbol{x} + \delta\boldsymbol{x}, \boldsymbol{h})) - \mathrm{KL}(\pi_\phi(\cdot \mid \boldsymbol{x}, \boldsymbol{h}) \,\|\, \pi_\phi(\cdot \mid \boldsymbol{x} + \delta\boldsymbol{x}, \boldsymbol{h}))| \leq \frac{\varepsilon_q}{\varrho}. \tag{148}$$

Proof. For the softmax actor, KL divergence between equal nearby logit parameters has zero first derivative and Hessian $\mathcal{F}_\pi$. Taylor expansion in $\delta\boldsymbol{l} = J_\ell \delta\boldsymbol{x} + o(\|\delta\boldsymbol{x}\|_2)$ proves (147). For the surrogate result, let $p = \pi_\phi(\cdot \mid \boldsymbol{x}, \boldsymbol{h})$, $\pi' = \pi_\phi(\cdot \mid \boldsymbol{x} + \delta\boldsymbol{x}, \boldsymbol{h})$, and $q' = q(\cdot \mid \boldsymbol{x} + \delta\boldsymbol{x}, \boldsymbol{h})$. The mean-value theorem for $\log z$ on $[\varrho, 1]$ gives

$$|\log \pi'_a - \log q'_a| \leq \frac{|\pi'_a - q'_a|}{\varrho}.$$

Therefore

$$\left| \sum_{a \in \mathcal{A}_{\mathrm{av}}} p_a \log \frac{\pi'_a}{q'_a} \right| \leq \frac{1}{\varrho} \sum_{a \in \mathcal{A}_{\mathrm{av}}} p_a |\pi'_a - q'_a| \leq \frac{\|\pi' - q'\|_1}{\varrho} \leq \frac{\varepsilon_q}{\varrho}.$$

□

Equation (147) describes the local geometry of direct actor-response supervision: observation directions are weighted by their effect on the action distribution. In the restricted-access variant, Eq. (148) bounds the corresponding loss discrepancy by surrogate fidelity and the probability floor. A mismatch at the clean input can also introduce constant and linear terms in the local expansion. Candidate-response KL, recorded-action cross-entropy, action-margin matching, and the latent action-head loss supply the additional signals specified in Appendix B.4.

A.16 Counterexamples illustrating the theorem conditions

The following elementary constructions test the necessity and placement of the main conditions.

- **A. Low dimension without LOO observability.** The set $\{(u, 0) : u \in [0, 1]\} \subset \mathbb{R}^2$ is one-dimensional, but deleting its first group destroys all information. Conditional recovery therefore additionally requires target identifiability from the retained groups.
- **B. Stage 1 unit self-copy gain.** If $\mathcal{G}_{1,i}(H, \boldsymbol{Y}) = \boldsymbol{y}_i$ for every history and input, the target-self gain is one and $\boldsymbol{y}_i - \boldsymbol{c}_i^{(1)} = 0$ at every target-fault amplitude. With the RMS floor, $g_i^{(1)} = \zeta$ and $d_i^{(1)} = \zeta / a_i^{(1)}$; the input-dependent scale can change the standardized score. The example establishes that exact copying prevents the unstandardized residual from growing with target-fault amplitude. This is the equality case excluded by the strict self-gain condition.
- **C. Vanishing fault amplitude.** With a nonzero initial-prediction or history error floor, clean and faulty discrepancy distributions can overlap as $\delta_i \to 0$. Uniform exact support recovery therefore needs a separation condition, while Theorem A.34 still controls signal error by the weak-fault energy.

D. Loss of clean majority. For $n = 3$, two faulty groups can place their discrepancies at an arbitrary common value and determine the median and MAD. The median certificate therefore stops at $|F| < n/2$; the implementation may still run outside that regime.

E. Improved refined context without full Stage 2 improvement. Let the refined input be exact but let the Stage 2 output depend only on the raw branch with a large gain. The refined contribution improves while the raw-branch error persists in the full candidate. Corollary A.24 therefore claims improvement only for the refined term.

F. Noncontractive recurrent state. For $h_{t+1} = \lambda h_t + x_t$ with $|\lambda| \geq 1$, a persistent observation perturbation can accumulate with the horizon. Corollary A.41 accordingly retains a finite-horizon result when $\kappa_h \geq 1$.

B METHOD AND TRAINING DETAILS

This appendix specifies the conditional target, normalization, and learned mappings used in Section 4, together with the grouped training losses. Appendix B.4 gives the effective training coefficients and source-actor selection for each trace. All observation coordinates in this appendix are original coordinates unless marked with a bar.

B.1 CONDITIONAL NOMINAL TARGETS

For the clean history $H_t^\star$ and nominal family $\mathcal{M}_t(H_t^\star)$ in Eq. (4), define

$$\mathcal{C}_{t,i}(H_t^\star, X_{t,-i}) = \{z \in \mathbb{R}^D : \exists X' \in \mathcal{M}_t(H_t^\star),\ X'_{-i} = X_{t,-i},\ x'_i = z\}. \tag{149}$$

Here $X_{t,-i}$ removes target row i, and x'_i is the corresponding row of X'. History first restricts the team to the nominal family; the current clean observations of the other agents impose additional compatibility constraints. This set collects the resulting possible target observations. It corresponds to the target conditional set analyzed in Section 5.1 and Appendix A.6. When it is a singleton, the ideal conditional completion map is

$$G^\star_{t,i}(H_t^\star, X_{t,-i}) = x_{t,i}.$$

The ideal map $G^\star_{t,i}$ requires clean conditions $(H_t^\star, X_{t,-i})$, while deployment provides recovered history $\widehat{H}_t$ and current team input Y_t. Target prediction thus uses conditions that themselves require recovery, while correcting those conditions relies on predictions of their clean values. SIR-MARL connects these two tasks through group-sparse error cancellation: Stage 1 forms coarse clean predictions, infers fault support from group residuals, and corrects the team input. Stage 2 combines the original and corrected conditions to refine its estimate of the same clean target. The sufficient structural and observability conditions are given in Section 5.1 and Appendix A.

B.2 NORMALIZATION AND PREDICTION MAPPINGS

Clean source-policy trajectories provide feature means $\mu_{\mathrm{cl}} \in \mathbb{R}^D$ and standard deviations $\sigma_{\mathrm{cl}} \in \mathbb{R}^D_{\geq 0}$. With a positive floor ϵ_{norm}, let $s_{\mathrm{cl}} = \max(\sigma_{\mathrm{cl}}, \epsilon_{\mathrm{norm}}\mathbf{1})$ coordinatewise. The forward and inverse transformations are

$$\bar{x} = \mathrm{Norm}(x) = (x - \mu_{\mathrm{cl}}) \oslash s_{\mathrm{cl}}, \qquad \mathrm{Norm}^{-1}(\bar{x}) = \mu_{\mathrm{cl}} + s_{\mathrm{cl}} \odot \bar{x}, \tag{150}$$

where $\oslash$ and $\odot$ are elementwise division and multiplication. The same transformation applies rowwise to every agent, with $\bar{e} = e \oslash s_{\mathrm{cl}}$ for errors. Thus $\bar{Y}_t = \bar{X}_t + \bar{E}_t$, preserving agent-row support. Normalized recovered history is $\widehat{\bar{H}}_t$, and $\widehat{\bar{H}}_{t,i}$ denotes its sequence for agent i. Missing startup history slots are filled with the current team input.

A temporal encoder T_θ processes each agent's recovered history and current observation. A self-masked cross-agent encoder R_θ uses current team observations, excluding agent i from its own keys and values:

$$u^{\mathrm{temp}}_{t,i} = T_\theta(\widehat{\bar{H}}_{t,i}, \bar{y}_{t,i}), \qquad u^{\mathrm{rel}}_{t,i} = R_\theta(\bar{Y}_t; i).$$

With trainable fusion mapping f_1, candidate head D_1, and scalar scale head S_1, Stage 1 computes

$$\begin{aligned} z^{(1)}_{t,i} &= f_1([u^{\mathrm{temp}}_{t,i}; u^{\mathrm{rel}}_{t,i}]), \\ \bar{c}^{(1)}_{t,i} &= D_1(z^{(1)}_{t,i}), \qquad a^{(1)}_{t,i} = \mathrm{softplus}(S_1(z^{(1)}_{t,i})) + \delta_a. \end{aligned} \tag{151}$$

Semicolons denote concatenation, $z^{(1)}_{t,i}$ is the fused feature, and $\delta_a > 0$ ensures a positive residual-normalization scale. Writing $r^{(1)}_{t,i} = \bar{y}_{t,i} - \bar{c}^{(1)}_{t,i}$ and $g^{(1)}_{t,i} = \sqrt{D^{-1}\|r^{(1)}_{t,i}\|_2^2 + \epsilon_r}$, Eq. (6) uses $d^{(1)}_{t,i} = g^{(1)}_{t,i}/a^{(1)}_{t,i}$. Explicitly, its robust statistics are

$$\begin{aligned} m^{(1)}_t &= \mathrm{median}_j\, d^{(1)}_{t,j}, \qquad q^{(1)}_t = \mathrm{median}_j\, |d^{(1)}_{t,j} - m^{(1)}_t|, \\ \tau^{(1)}_t &= m^{(1)}_t + \beta(q^{(1)}_t + \epsilon_\tau), \qquad p^{(1)}_{t,i} = \mathrm{sigmoid}((d^{(1)}_{t,i} - \tau^{(1)}_t)/T_s). \end{aligned} \tag{152}$$

Both stages compute medians across the agent dimension. For sorted values $v_{(1)} \le \cdots \le v_{(n)}$, the implementation uses $\mathrm{median}(v) = v_{(\lceil n/2 \rceil)}$, the lower of the two middle values when n is even. The MAD uses this same convention after taking absolute deviations from the selected median.

Stage 2 uses the target-masked inputs $Z^{\mathrm{raw}}_{t,i}$ and $Z^{\mathrm{ref}}_{t,i}$ defined in Section 4.4. The LOO position is fixed to target i. Each independently parameterized context encoder replaces the target observation embedding with its own trainable mask vector, shared across agents and time steps:

$$u^{\mathrm{raw}}_{t,i} = \mathrm{Enc}_{\mathrm{raw}}(Z^{\mathrm{raw}}_{t,i}), \qquad u^{\mathrm{ref}}_{t,i} = \mathrm{Enc}_{\mathrm{ref}}(Z^{\mathrm{ref}}_{t,i}).$$

Trainable fusion f_2 and prediction heads D_2, S_2 compute

$$\begin{aligned} z^{(2)}_{t,i} &= f_2([u^{\mathrm{temp}}_{t,i}; u^{\mathrm{rel}}_{t,i}; u^{\mathrm{raw}}_{t,i}; u^{\mathrm{ref}}_{t,i}; z^{(1)}_{t,i}]), \\ \bar{c}^{(2)}_{t,i} &= D_2(z^{(2)}_{t,i}), \qquad a^{(2)}_{t,i} = \mathrm{softplus}(S_2(z^{(2)}_{t,i})) + \delta_a. \end{aligned} \tag{153}$$

The same residual and support rule is applied with superscript (2) and original input $\bar{y}_{t,i}$. The dedicated LOO branches mask the current target observation, while temporal and Stage 1 features retain target-related information already present in the recovery computation. Section 4.4 specifies support fusion and final recovery from the original input.

Network architecture. The default recovery network uses a 12-step input window (11 history slots and the current observation), hidden width $h = 224$, and latent width $z = 80$. Table 3 specifies the layer widths. An MLP listed as $d_0 \to \cdots \to d_L$ contains L affine layers, with GELU and LayerNorm after each hidden affine layer and a linear output layer. The temporal encoder is a single-layer GRU; its final output is used for each agent. The raw and refined LOO encoders have separate parameters, each with two pre-LayerNorm Transformer layers, four attention heads, feed-forward width 896, GELU activation, and zero dropout. Each LOO branch reads the encoded token at the masked target position. Token mappings and prediction heads share weights across agents; the two stages have distinct fusion, decoder, and scale-head parameters. The model adds no positional or agent-ID embedding beyond information already present in the observations.

Table 3: Default SIR-MARL recovery-network dimensions. D is the per-agent observation dimension and A is the actor's action-space size. The auxiliary action head is used during training.

Component	Dimensions and operation
Raw / refined token MLPs	Each: $D \to 224 \to 224$
Temporal encoder	GRU: input and hidden widths 224, one layer
Relational query / key	Each: bias-free $224 \to 224$ linear map
Raw / refined LOO encoders	Each: 2 Transformer layers, 4 heads, width 224
Transformer feed-forward block	$224 \to 896 \to 224$
Stage 1 fusion f_1	$448 \to 224 \to 224 \to 80$
Stage 2 fusion f_2	$976 \to 224 \to 224 \to 80$
Candidate heads D_1, D_2	Each: $80 \to 224 \to 224 \to D$
Scale heads S_1, S_2	Each: $80 \to 224 \to 1$
Auxiliary action head	$80 \to 224 \to A$

B.3 Grouped training losses and response distributions

For each independently masked training view $v \in \{a, b\}$, let $\bar{C}^{(r,v)}_t$ and $p^{(r,v)}_t$ denote the row-stacked candidates and support vector at stage $r \in \{1, 2\}$. Let $p^{(v)}_t$ be the fused support vector and $\widehat{\bar{X}}^{(v)}_t$ the final normalized recovery. Define original-coordinate outputs $C^{(2,v)}_t = \mathrm{Norm}^{-1}(\bar{C}^{(2,v)}_t)$ and $\widehat{X}^{(v)}_t = \mathrm{Norm}^{-1}(\widehat{\bar{X}}^{(v)}_t)$. Superscript cl denotes an additional unmasked clean view. The observation objective $\mathcal{L}_{\mathrm{sig}}$ in Section 4.5 groups the following losses:

$$\mathcal{L}_{\mathrm{cand}} = \mathrm{SmoothL1}(C^{(2,a)}_t, X_t), \qquad \mathcal{L}_{\mathrm{rec}} = \mathrm{SmoothL1}(\widehat{X}^{(a)}_t, X_t),$$

$$\mathcal{L}_{\mathrm{view}} = \mathrm{MSE}(\widehat{X}^{(a)}_t, \widehat{X}^{(b)}_t),$$

$$\mathcal{L}_{\mathrm{id}} = \mathrm{SmoothL1}(\widehat{X}^{\mathrm{cl}}_t, X_t) + \eta_0 n^{-1} \sum_i p^{\mathrm{cl}}_{t,i}.$$

Here η_0 is a nonnegative weight. Cross-view consistency encourages recovery of the same clean observation from different masking patterns. Masking probabilities, the visible-agent fallback, loss reductions, and effective coefficients appear in Appendix B.4. The masks independently hide complete time–agent groups; all three input versions are processed by the same recovery network with shared parameters. Among the masked inputs, view a receives direct candidate and recovery supervision; the unmasked input receives the identity loss.

Let π_ϕ denote the action-distribution component of the frozen recurrent actor Π_ϕ, $h^-_{t,i}$ its incoming hidden state along the clean trace, and $\mathcal{A}_{t,i}$ the current available-action set. The clean and training response distributions are

$$\begin{aligned}
\pi^{\mathrm{cl}}_{t,i} &= \pi_\phi(\cdot \mid x_{t,i}, h^-_{t,i}, \mathcal{A}_{t,i}),\\
\pi^{X}_{t,i} &= \pi_\phi(\cdot \mid \widehat{x}^{(a)}_{t,i}, h^-_{t,i}, \mathcal{A}_{t,i}),\\
\pi^{C}_{t,i} &= \pi_\phi(\cdot \mid c^{(2,a)}_{t,i}, h^-_{t,i}, \mathcal{A}_{t,i}),\\
\pi^{z}_{t,i} &= \pi^z_{\mathrm{aux}}(\cdot \mid z^{(2,a)}_{t,i}, \mathcal{A}_{t,i}).
\end{aligned} \tag{154}$$

The recovery and candidate are evaluated by the original source actor with its parameters ϕ fixed. Gradients through these observation inputs train the recovery network; the auxiliary mapping π^z_{aux} is trained jointly. All action distributions are restricted to the available-action set $\mathcal{A}_{t,i}$. The decision objective is

$$\mathcal{L}_{\mathrm{dec}} = \mathcal{L}_{\mathrm{resp}} + \lambda_{\mathrm{act}}\mathcal{L}_{\mathrm{act}} + \lambda_{\mathrm{margin}}\mathcal{L}_{\mathrm{margin}},$$

where $\mathcal{L}_{\mathrm{resp}}$ is Eq. (11); Appendix B.4 defines recorded-action cross-entropy and probability-gap matching.

With nonnegative internal weights $\lambda_{C,\mathrm{reg}}, \lambda_{p,\mathrm{reg}}, \lambda_{\mathrm{tok}}$, the grouped regularizer is

$$\begin{aligned}
\mathcal{L}_{\mathrm{reg}} = \sum_{v\in\{a,b,\mathrm{cl}\}} \Big[&\lambda_{C,\mathrm{reg}}\,\mathrm{MSE}(\bar{C}^{(2,v)}_t, \bar{C}^{(1,v)}_t)\\
&+ \lambda_{p,\mathrm{reg}}\,\mathrm{MSE}(p^{(2,v)}_t, p^{(1,v)}_t) + \lambda_{\mathrm{tok}}\mathcal{L}^{(v)}_{\mathrm{tok}}\Big].
\end{aligned} \tag{155}$$

Candidate consistency uses normalized coordinates. The temporal-token term $\mathcal{L}^{(v)}_{\mathrm{tok}}$ is defined in Eq. (156); the outer weight λ_{reg} and all internal weights are given in Appendix B.4.

B.4 Exact training objective and coefficients

This section specifies the training objective for the recovery operator analyzed in Appendix A. For a clean window, two views $v \in \{a, b\}$ independently mask each time–agent group with probability $p_{\mathrm{mask}} = 0.28$ and replace the selected group by the clean feature mean. If every group in the current slice is selected, agent 0 remains visible. An additional view $v = \mathrm{cl}$ is left unmasked. Denote the stagewise candidates, support scores, final recoveries, and temporal-token regularizer by $\bar{\boldsymbol{C}}^{(1,v)}, \bar{\boldsymbol{C}}^{(2,v)}, \boldsymbol{p}^{(1,v)}, \boldsymbol{p}^{(2,v)}, \widehat{\boldsymbol{X}}^{(v)}$, and $\mathcal{L}^{(v)}_{\mathrm{tok}}$, respectively. Bars denote normalized observation coordinates. The signal-space candidate $\boldsymbol{C}^{(2,v)} = \mathrm{Norm}^{-1}(\bar{\boldsymbol{C}}^{(2,v)})$, the final recovery $\widehat{\boldsymbol{X}}^{(v)}$, and the clean target $\boldsymbol{X}$ are in original observation coordinates. The inverse acts rowwise. The fused support vector is $\boldsymbol{p}^{(v)} = \alpha\boldsymbol{p}^{(1,v)} + (1-\alpha)\boldsymbol{p}^{(2,v)}$, with the deployed fixed coefficient $\alpha = 0.25$. The expressions below describe one window. SmoothL1 (with unit transition) and MSE average over all entries of their arguments; action losses average over agents, and the full objective is averaged over training windows.

For the masked or clean input view v, let $\zeta^{(v)}_{\ell i} \in \mathbb{R}^{d_{\mathrm{tok}}}$ denote its learned input token for time ℓ and agent i, obtained by embedding the normalized local observation before temporal encoding. Here d_{tok} is the token feature dimension. For the training windows, $T \geq 2$, and the temporal-token term is

$$\mathcal{L}^{(v)}_{\mathrm{tok}} = \frac{1}{(T-1)nd_{\mathrm{tok}}} \sum_{\ell=t-T+2}^{t}\sum_{i=1}^{n} \left\| \zeta^{(v)}_{\ell i} - \zeta^{(v)}_{\ell-1,i} \right\|_2^2. \tag{156}$$

Thus it averages squared differences between adjacent input tokens over time, agents, and token coordinates.

The signal-space terms are

$$\mathcal{L}_{\text{cand}} = \text{SmoothL1}(\boldsymbol{C}^{(2,a)}, \boldsymbol{X}), \qquad \mathcal{L}_{\text{rec}} = \text{SmoothL1}(\widehat{\boldsymbol{X}}^{(a)}, \boldsymbol{X}),$$
$$\mathcal{L}_{\text{view}} = \text{MSE}(\widehat{\boldsymbol{X}}^{(a)}, \widehat{\boldsymbol{X}}^{(b)}),$$
$$\mathcal{L}_{\text{id},X} = \text{SmoothL1}(\widehat{\boldsymbol{X}}^{(\text{cl})}, \boldsymbol{X}), \qquad \mathcal{L}_{\text{id},p} = n^{-1} \sum_i p_i^{(\text{cl})}.$$

For a clean sample from source actor q, let $\boldsymbol{h}^-$ denote its recorded incoming recurrent state, $\pi^{\text{cl},q}$ its recorded clean action distribution, and $\pi_\phi^{(q)}$ the action-distribution mapping of that frozen source actor. The source identifier selects the actor associated with each trace, and $\mathcal{A}$ is its recorded available-action set. Define

$$\begin{aligned}
\mathcal{L}_{\text{KL},X} &= \text{KL}(\pi^{\text{cl},q} \| \pi_\phi^{(q)}(\widehat{\boldsymbol{x}}^{(a)}, \boldsymbol{h}^-, \mathcal{A})), \\
\mathcal{L}_{\text{KL},C} &= \text{KL}(\pi^{\text{cl},q} \| \pi_\phi^{(q)}(\boldsymbol{c}^{(2,a)}, \boldsymbol{h}^-, \mathcal{A})), \\
\mathcal{L}_{\text{KL},z} &= \text{KL}(\pi^{\text{cl},q} \| \pi_{\text{aux}}^{z}(\boldsymbol{z}^{(2,a)}, \mathcal{A})), \\
\mathcal{L}_{\text{act}} &= \text{CE}(\pi_\phi^{(q)}(\widehat{\boldsymbol{x}}^{(a)}, \boldsymbol{h}^-, \mathcal{A}), a_{\text{recorded}}), \\
\mathcal{L}_{\text{margin}} &= \text{SmoothL1}(m(\pi^{\text{cl},q}), m(\pi_\phi^{(q)}(\widehat{\boldsymbol{x}}^{(a)}, \boldsymbol{h}^-, \mathcal{A}))),
\end{aligned}$$

where π_{aux}^z is the trainable auxiliary action mapping. For the full masked action-probability vector, unavailable actions have probability zero and $m(\pi) = \pi_{(1)} - \pi_{(2)}$, with $\pi_{(1)} \geq \pi_{(2)}$ its two largest probabilities. If only one action is available, $\pi_{(2)} = 0$. This loss matches the size of the probability gap; the response KL and recorded-action cross-entropy constrain action identities. The effective scalar loss is

$$\begin{aligned}
\mathcal{L} = {} & 0.9\mathcal{L}_{\text{cand}} + \mathcal{L}_{\text{rec}} + 2.4\mathcal{L}_{\text{KL},X} + 1.2\mathcal{L}_{\text{KL},C} + 0.8\mathcal{L}_{\text{KL},z} \\
& + 0.15\mathcal{L}_{\text{act}} + 0.35\mathcal{L}_{\text{margin}} + 0.25\mathcal{L}_{\text{view}} \\
& + 1.2\mathcal{L}_{\text{id},X} + 0.96\mathcal{L}_{\text{id},p} \\
& + \sum_{v \in \{a,b,\text{cl}\}} \left[0.0048\,\text{MSE}(\bar{\boldsymbol{C}}^{(2,v)}, \bar{\boldsymbol{C}}^{(1,v)}) + 0.0024\,\text{MSE}(\boldsymbol{p}^{(2,v)}, \boldsymbol{p}^{(1,v)}) + 0.0036\mathcal{L}_{\text{tok}}^{(v)} \right].
\end{aligned} \tag{157}$$

In the grouped notation of Section 4.5, these coefficients correspond to

$$\lambda_{\text{cand}} = 0.9, \quad \lambda_{\text{view}} = 0.25, \quad \lambda_{\text{id}} = 1.2, \quad \eta_0 = 0.8,$$
$$\lambda_{\text{dec}} = 1, \quad (\lambda_X, \lambda_C, \lambda_z) = (2.4, 1.2, 0.8), \quad \lambda_{\text{act}} = 0.15, \quad \lambda_{\text{margin}} = 0.35,$$
$$\lambda_{\text{reg}} = 0.12, \quad \lambda_{C,\text{reg}} = 0.04, \quad \lambda_{p,\text{reg}} = 0.02, \quad \lambda_{\text{tok}} = 0.03.$$

The three KL terms above are the agent averages of the corresponding $\ell_{t,i}^{\text{resp}}$ terms in the main text. Inter-stage candidate regularization uses normalized coordinates, while candidate, recovery, cross-view observation, and clean-identity supervision use original coordinates.

Only recovery-network parameters and the auxiliary action mapping are updated by this objective. Source actors remain frozen, with gradients retained through their observation inputs. The auxiliary action mapping is used only during training; deployment feeds recovered observations to the original frozen source actors.

C Experimental Setup and Recovery Diagnostics

C.1 Deployment interface and fault protocol

C.1.1 Trusted team-level recovery

SIR-MARL assumes that a trusted layer can gather synchronized local observations before invoking the frozen actors. The deployed architecture combines centralized team-level recovery with local actor execution. After preprocessing, each actor retains its original parameters, input dimension,

recurrent update, and action rule. The post-training interface preserves the deployed actors and adds observation preprocessing.

The interface requires delivery of nD observation scalars per synchronized team step and return of nD recovered scalars when recovery is hosted separately. These are logical payload counts. Evaluation assumes reliable synchronization, communication, and operation of the recovery layer, and measures observation-fault recovery and local computation cost. Network latency and traffic remain deployment-specific quantities requiring system-level measurement.

C.1.2 GROUP-SPARSE GAUSSIAN FAULT PROTOCOL

At each team step t, let $X_t \in \mathbb{R}^{n\times D}$ be the clean team observation. The runtime samples a support S_t uniformly from all subsets of $\{1,\dots,n\}$ of size s, and draws independent vectors $z_{t,i} \sim \mathcal{N}(0, I_D)$ for $i \in S_t$. It computes one noise scale from the clean team observation and applies it to every selected group:

$$\sigma_t = 3\big(\mathrm{std}(\mathrm{vec}\, X_t) + 10^{-6}\big), \qquad Y_{t,i} = \begin{cases} x_{t,i} + \sigma_t z_{t,i}, & i \in S_t, \\ x_{t,i}, & i \notin S_t. \end{cases} \tag{158}$$

Here I_D is the D-dimensional identity matrix, and the standard deviation uses all nD coordinates with population normalization. Conditional on X_t, the scale is identical across selected groups and independent of s; the sweep changes the number of corrupted groups. The scale adapts to the clean observation at each step. Supports and Gaussian draws are resampled at every step using the evaluation's fault seed. The recovery layer receives Y_t and its recovered history, with S_t, s, and σ_t used only by the evaluation generator.

C.2 BENCHMARK SETUP AND COMPARISON PROTOCOLS

Tasks and faults. We use SMAC `8m` (eight homogeneous Marines), `3s5z` (eight heterogeneous agents), and `2s3z` (five heterogeneous agents). At each team step, exactly $s \in \{1,2,3,4\}$ groups are sampled uniformly without replacement. Independent Gaussian noise corrupts every coordinate of each selected group, using a per-group scale independent of s. Support and noise are resampled each step under a deterministic fault seed; Appendix C.1.2 gives the generator. Thus $s4$ corrupts 50% of an eight-agent team and 80% of `2s3z`; the recovery model operates with unknown fault support and cardinality.

Training and model selection. Policy-training comparisons use recurrent actor–critic implementations with method-specific robustness mechanisms, as detailed below. Each policy checkpoint is fixed across clean and faulted evaluation conditions. Best-checkpoint selection uses periodic nominal-evaluation win rate. The SMAC EIR and ERNIE protocols each allocate 10 million environment steps to each of five training seeds per map. Checkpoints are selected by deterministic nominal evaluation over 32 episodes, retaining the earliest checkpoint attaining the highest observed win rate. Final evaluation uses 64 episodes per checkpoint under Clean and $s1$–$s4$, with the fixed-scale generator in Appendix C.1.2 and paired fault seeds. The selected training step and checkpoint hashes are recorded with the results. Frozen-source MAPPO evaluations, including SIR-MARL and its recovery controls, use deterministic actions: at each step the actor selects the highest-probability available action. Available-action masks are retained, recurrent states are reset to zero at each episode start, and recovered observations drive the actor's recurrent updates during rollout. These evaluations use one environment instance and 64 completed episodes per checkpoint and condition. SIR-MARL freezes the source actor and critic before recovery training and constructs training and validation inputs from nominal traces with generic masking. SMAC recovery training uses AdamW for 180 epochs, batch size 256, learning rate 3.5×10^{-4}, weight decay 10^{-4}, and gradient clipping at norm 5. Checkpoint selection minimizes $V = E_{\mathrm{mask}} + 0.2E_{\mathrm{view}} + 0.5E_{\mathrm{clean}} + 0.8S_{\mathrm{clean}}$, where the first three terms are validation masked-recovery MSE, MSE between the two recovered views, and clean-recovery MSE, and S_{clean} is the mean support score on clean inputs. The earliest checkpoint improving the score by more than 10^{-7} is retained; all 180 epochs are completed.

Our SMAC SIR-MARL results evaluate five independently initialized recovery models per map, each trained on pooled nominal traces from source policies 1–4 and selected using policy 5. Each recovery model is evaluated with all five frozen source policies. Recovery seeds are first averaged

within each source-policy seed, after which we report mean and standard deviation over the five policy-level aggregates. Policy-training baselines are reported over five policy seeds. Each final benchmark evaluation contains 64 independently reset episodes.

Nominal recovery data and windows. Table 4 reports the nominal data used by the recovery models. Each SMAC source policy contributes 128 episodes; policies 1–4 provide 512 training episodes and policy 5 provides 128 validation episodes. A training example is a team-observation window ending at one retained record. For SMAC and MPE, windows advance by one environment step within each episode and contain the current observation and up to 11 preceding observations. Missing initial slots repeat the episode's first observation. Windows remain within their trace file, episode, and data split. SMAC uses all listed windows, shuffled once per epoch, with the final partial minibatch retained. This gives 59, 139, and 102 updates per epoch on `8m`, `3s5z`, and `2s3z`, respectively. Masked views are generated during minibatch processing. The cross-domain sampling and training budgets are specified in Appendix D.

Table 4: Nominal recovery data. Episode counts describe collected trajectories; window counts describe the training examples and validation examples constructed from them. SMAC and MPE pool source policies 1–4 for training and use policy 5 for validation. Each GRF row describes a separately fitted source-policy recovery model, with the record split specified in Appendix D.2.

Environment	Clean episodes	Training windows	Validation windows
8m	640	15,002	3,994
3s5z	640	35,452	7,400
2s3z	640	25,972	6,128
MPE5	1,000	20,000	5,000
GRF source 1	96	3,524	880
GRF source 2	96	7,229	1,807
GRF source 3	96	7,037	1,759

Policy-training baselines on SMAC. Corrupted MAPPO executes the nominal source actor directly on faulty observations. EIR uses the released EIR-MAPPO implementation; ERNIE, MIR3, ROMANCE, and WALL use adaptations of their training mechanisms to our recurrent MAPPO implementation. These comparisons evaluate transfer from method-specific training perturbations or regularization to the common observation-fault benchmark, with each method retaining its specified policy inputs and training objective.

EIR (Li et al., 2024) uses an environment-compatible copy of the released EIR-MAPPO code (commit `fcc96e8272c9`), retaining its recurrent belief network, belief-conditioned defender, and adversarial-training procedure. During evaluation, each agent predicts its belief from the available observation and recurrent belief state, then supplies that prediction to the defender actor. Actor and belief parameters are loaded together from the selected checkpoint and held fixed. Teacher forcing is disabled in all nominal and faulted evaluations, including model selection; a runtime assertion checks that the actor receives the predicted belief. Training retains the released teacher-forcing schedule, adversary updates, and belief optimization. Gaussian evaluation faults act on physical observation coordinates, and the noise scale is computed from those coordinates; appended belief/type slots are preserved during corruption. The comparison therefore retains EIR's released implementation and method-specific architecture, with the stated environment-compatibility, evaluation, and checkpoint-selection procedures.

ERNIE denotes an ERNIE-style adaptation of adversarial response regularization to recurrent MAPPO. The regularizer is recomputed on each PPO minibatch and added with coefficient 0.1 to the clipped actor objective and entropy term. Gradients update actor parameters directly through the regularizer; environment rewards, return computation, and the state-value critic use the standard MAPPO objectives.

Each minibatch initializes an observation perturbation with Gaussian noise of standard deviation 0.001, then applies three differentiable projected ascent steps within an observation-coordinate ℓ_∞ radius of 0.1. Each step uses RMS-normalized input gradients with step size $0.1/3$. The discrepancy is the Euclidean distance between clean and perturbed categorical action-probability vectors, averaged over active agents, with the same available-action mask applied to both responses. The outer

actor gradient retains the dependence of all three adversarial updates on actor parameters, using recurrent kernels that support higher-order derivatives. This transfers ERNIE's direct adversarial regularization and unrolled-response mechanism to MAPPO; the categorical policy output, minibatch reduction, perturbation parameterization, and state-value critic specify the MAPPO adaptation.

MIR3 estimates dependence between the shared state and joint action using a categorical CLUB upper-bound estimator and a MINE lower-bound diagnostic estimator. The CLUB penalty, weighted by 0.01, is subtracted from rollout rewards before MAPPO return computation. Both estimators learn from on-policy rollout samples; the joint action is represented by concatenated one-hot agent actions. This realizes the information-regularization mechanism through MAPPO return shaping.

ROMANCE uses a population of four action attackers and an archive of up to fifteen attackers, with sparse attack selection and population-diversity training. The per-episode attack budget is six. *WALL* uses a transformer-based attack-timing planner and an L2-based rule to select follow-up agents. Its training configuration uses attack duration three, period twenty, a budget of twelve, and two follow-up agents. Both mechanisms modify action execution during robust-policy training; benchmark evaluation executes the resulting policy under the observation-fault generator in Appendix C.1.2.

Recovery and augmentation comparisons. PCA-SIR fits a deterministic linear subspace to nominal team observations pooled across the five source policies. The fit targets 95% explained variance with rank capped at 192, using at most $250{,}000$ clean observations. Group residual thresholds are calibrated at the 99.5th percentile on clean data. At deployment, the projection supplies candidate observations and group residuals determine support; raw observations outside the estimated support are preserved. This compares linear structured recovery with SIR-MARL under the same frozen source policies and clean-data interface.

SIR-Direct trains a one-stage raw-context belief network using the same clean-only training interface and unknown-cardinality support rule; Appendix G.2 gives its architecture comparison. SparseAug-Mix provides an observation-augmentation policy-training control, described in Appendix G.3.

Metrics. The primary metric is team win rate. Observation error, support F1, actor KL, and action agreement diagnose the mechanism. Appendix C defines source-normalized gap recovery and the diagnostic aggregation. Paired source-level differences and normalized performance are in Appendix G.4.

Cross-domain validation on MPE and GRF is provided in Appendix D.

Ablation definitions. Table 2 examines the main components on $\texttt{8m}$. The temporal variant sets the temporal feature to zero and removes $\mathcal{L}_{\mathrm{tok}}$; the relational variant sets the relational feature to zero. The LOO variant sets both raw and refined LOO context features to zero, preserving the fusion-layer input dimensions and the temporal, relational, and Stage 1 pathways. The decision-supervision variant sets the recovery-response KL, candidate-response KL, latent-response KL, recorded-action cross-entropy, and margin-loss coefficients to zero; reconstruction, output-view consistency, clean identity, and the applicable stage and temporal regularizers remain active. SIR-Direct is a separately constructed one-stage raw-context belief network using the clean-only recovery interface; Appendix G.2 describes that architecture comparison.

Single-stage supervision. For the *w/o second stage* variant, each view $v \in \{a, b, \mathrm{cl}\}$ supplies

$$C_{\mathrm{out}}^{(v)} = C^{(1,v)}, \qquad p_{\mathrm{out}}^{(v)} = p^{(1,v)}, \qquad z_{\mathrm{out}}^{(v)} = z^{(1,v)}, \qquad \widehat{x}_i^{(v)} = (1 - p_i^{(1,v)}) y_i^{(v)} + p_i^{(1,v)} c_i^{(1,v)}.$$

Candidate reconstruction and candidate-response KL use $C^{(1,a)}$; recovery reconstruction, recovery-response KL, recorded-action cross-entropy, and margin matching use $\widehat{X}^{(a)}$. The auxiliary action head receives $z^{(1,a)}$ for latent-response KL. These terms retain their respective coefficients in Appendix B.4. Output-view consistency compares the two single-stage recoveries, and clean identity uses the single-stage clean recovery and support. Both inter-stage consistency terms are zero; temporal-token regularization remains active. Thus Stage 1 receives the final-output supervision when refinement is removed.

Training and evaluation settings for the module variants. The module variants use nominal traces from source policies 1–4 for training and policy 5 for validation, with feature normaliza-

tion fitted on the training observations. Recovery seeds are 1–5. The masking probability is 0.28, with independently generated masked views and an unmasked clean view. Training uses AdamW, learning rate 3.5×10^{-4}, weight decay 10^{-4}, batch size 256, gradient clipping at norm 5, and 180 epochs. Minibatch sampling and the two masking streams use generators seeded by the recovery seed; checkpoints are selected from nominal validation diagnostics. For source policy q, evaluation uses environment seed $50{,}000q$, fault seed $700{,}000 + q$, and 64 episodes per condition, with the deterministic actor and fault protocol specified above.

C.3 Benchmark metrics

Metrics. The primary metric is team win rate. For recovery relative to the same frozen source policy we additionally use normalized gap recovery,

$$\mathrm{GR} = \frac{R_{\mathrm{rec}} - R_{\mathrm{corr}}}{R_{\mathrm{clean}} - R_{\mathrm{corr}}}, \tag{159}$$

where R_{clean}, R_{corr}, and R_{rec} refer to nominal, corrupted, and recovered execution of the corresponding source checkpoint. Observation error, support F1, actor KL, and action agreement are used as mechanistic diagnostics alongside closed-loop performance.

C.4 Observation structure and recovery diagnostics

C.4.1 Clean-observation structure

Global PCA uses 8,000 clean team observations per SMAC map from five source policies, after removing constant or near-zero coordinates. The retained coordinate count is D_{active}, labeled active features in Figure 2A. We define k_{95} as the smallest number of principal components explaining at least 95% of the sample variance. Local PCA uses 48 query neighborhoods, each containing 64 observations selected by 11-step history; random controls contain the same number of observations. We report the mean local k_{95} across queries. Small neighborhoods impose a rank ceiling of 63, so the equal-size control isolates the contribution of history-based selection. Effective PCA rank summarizes empirical variance concentration; the theoretical dimension and observability conditions are stated in Appendix A.

Table 5: Clean-observation dimensionality used in Figure 2.

Map	D_{active}	Global k_{95}	History k_{95}	Random k_{95}
8m	1480	344	39.73	46.81
3s5z	1784	323	33.29	46.04
2s3z	620	113	24.98	38.54

C.4.2 Residual separation

Figure 2B reports realized Stage 1 residual separation for SIR-MARL. Using the scores and threshold in Eq. (6), a clean group is on the correct side when $d_{t,i}^{(1)} < \tau_t^{(1)}$, and a faulty group when $d_{t,i}^{(1)} > \tau_t^{(1)}$. Each SMAC map uses five source-policy trajectory seeds and three recovery checkpoints shared across the sources, with 512 input windows at each evaluated fault count $s \in \{1, 2, 4\}$. Fractions are computed within each policy/recovery cell using the relevant clean or faulty group counts, averaged across recovery seeds within policy, then averaged across policies. The plotted values are percentages of groups satisfying these strict inequalities.

C.4.3 Paired Stage 1 and Stage 2 recovery

Figure 2C uses a paired offline replay of saved nominal trajectories. For each map, five source-policy trajectory seeds are evaluated with three recovery checkpoints at $s \in \{1, 2, 4\}$, using 512 sampled windows per combination. The same three checkpoints per map are reused across the five trajectory seeds. At each replayed step, exactly s agent groups receive Gaussian faults with scale computed once from the clean team observation as specified in Appendix C.1.2. History is updated recursively

with preceding final recoveries; missing initial history is padded with the current faulty input. Both stages are recorded in the same forward pass, sharing the clean target, faulty input, and recovered history. Stage 1 uses the intermediate correction in Eq. (7); Stage 2 uses the final output in Eq. (9) with fused support.

Figure 2C measures errors in original observation coordinates, matching the coordinate space of the reconstruction supervision in Section 4.5. All observation coordinates are retained. With time indices suppressed, define the other-agent recovery ratio

$$R_i(V) = \frac{\|V_{-i} - X_{-i}\|_{\mathrm{G}}}{\|Y_{-i} - X_{-i}\|_{\mathrm{G}}}, \qquad V \in \{X^{(1)}, \widehat{X}\}, \qquad \|W\|_{\mathrm{G}} = \|W\|_F/\sqrt{D}. \tag{160}$$

The Stage 1 ratio quantifies the corrected other-agent conditions available to Stage 2; the Stage 2 ratio measures final recovery on those same rows. We retain targets with at least one other faulty group and use the same rows for both stages. Raw and Stage 1 error norms must exceed 10^{-8} for the paired statistics; all 460,800 eligible target records satisfy this threshold in both original and feature-normalized coordinates. Norms and ratios use no additive stabilizer. Ratios are averaged within each trajectory-seed/checkpoint/fault-count cell, then across the three recovery checkpoints within each trajectory seed. Figure 2C reports the mean and sample standard deviation across the five trajectory seeds. Regime summaries equally average the evaluated fault counts within each seed: the interior ($s < n/2$) contains $s1$ and $s2$; $s4$ is the boundary ($s = n/2$) for eight agents and beyond it ($s > n/2$) for five agents. Let μ_1 and μ_2 be the resulting mean ratios for the two stages. Labels above each bar pair give $100(1 - \mu_2/\mu_1)$, the percentage reduction in the mean ratio relative to Stage 1.

Table 6: Original-coordinate paired errors in Figure 2C (mean±SD). The final column reports the percentage of target records with strictly smaller Stage 2 error, averaged with the same hierarchy as the ratios.

Map	Regime	Stage 1/raw	Stage 2/raw	Stage 2 improves (%)
8m	$s < n/2$	0.2445 ± 0.0061	0.2322 ± 0.0072	79.0 ± 4.0
8m	$s = n/2$	0.2614 ± 0.0081	0.2469 ± 0.0093	83.0 ± 3.4
3s5z	$s < n/2$	0.2135 ± 0.0164	0.2016 ± 0.0175	78.2 ± 9.9
3s5z	$s = n/2$	0.2347 ± 0.0163	0.2191 ± 0.0178	87.4 ± 10.4
2s3z	$s < n/2$	0.2003 ± 0.0056	0.1863 ± 0.0054	81.7 ± 3.2
2s3z	$s > n/2$	0.9184 ± 0.0051	0.8746 ± 0.0063	37.3 ± 1.3

In original coordinates, Stage 2 lowers the mean error ratio under a clean majority on all three maps and improves most target records (Table 6). Both patterns persist at the eight-agent boundary. These results support refinement of the remaining error after the initial correction. At `2s3z` $s4$, the small average change and lower improvement frequency characterize refinement beyond the clean-majority regime.

For comparison with the theoretical error geometry, Table 7 reports feature-normalized ratios on exactly the same paired records. They are obtained from Eq. (160) by replacing X, Y, V with their checkpoint-normalized counterparts, retaining all coordinates. Coordinatewise normalization weights errors by the inverse clean feature scales, whereas the original-coordinate statistic measures errors in the coordinates used for reconstruction supervision. The sample selection and arithmetic aggregation are identical in the two tables.

Feature-normalized ratios quantify recovery in the error geometry of the theoretical analysis. On the eight-agent maps, both stages reduce input error substantially, with a further small mean reduction at Stage 2. On `2s3z` under a clean majority, Stage 2 improves 64.6% of target records; its mean ratio is 0.00563, compared with 0.000464 at Stage 1. At $s2$, approximately 1.90% of target records have a Stage 2 ratio above 0.1, illustrating the sensitivity of this coordinate-weighted mean to larger residual errors. The original-coordinate and normalized statistics report recovery in their respective error geometries.

C.4.4 Additional Stage 1 diagnostics

Table 8 reports additional Stage 1 diagnostics on fixed inputs. The team ratio $\rho_1^{\mathrm{emp}} = \|X^{(1)} - X\|_{\mathrm{G}}/\|Y - X\|_{\mathrm{G}}$ measures Stage 1 error relative to input error, using normalized coordinates as

Table 7: Feature-normalized counterpart of the paired errors in Figure 2C. Error ratios are mean±SD across five trajectory seeds after recovery averaging. The final column is the similarly aggregated percentage of target records with strictly smaller Stage 2 error.

Map	Regime	Stage 1/raw	Stage 2/raw	Stage 2 improves (%)
8m	$s < n/2$	$(10.13 \pm 0.58) \times 10^{-4}$	$(10.12 \pm 0.59) \times 10^{-4}$	51.9 ± 4.1
8m	$s = n/2$	$(9.77 \pm 0.52) \times 10^{-4}$	$(9.75 \pm 0.53) \times 10^{-4}$	54.2 ± 3.0
3s5z	$s < n/2$	$(5.93 \pm 0.70) \times 10^{-4}$	$(5.87 \pm 0.71) \times 10^{-4}$	72.1 ± 3.6
3s5z	$s = n/2$	$(5.78 \pm 0.64) \times 10^{-4}$	$(5.69 \pm 0.64) \times 10^{-4}$	84.0 ± 4.6
2s3z	$s < n/2$	$(4.64 \pm 0.29) \times 10^{-4}$	$(56.26 \pm 20.94) \times 10^{-4}$	64.6 ± 4.0
2s3z	$s > n/2$	0.9023 ± 0.0048	0.8596 ± 0.0062	37.0 ± 1.6

in Section 5. The LOO ratio $\Delta_i^{\mathrm{ref}}/\Delta_{-i}$ measures the corresponding other-agent condition error. Sample means and fractions weight cells by their observation counts; the LOO proxy is a sample-count-weighted average of diagnostic batch medians. Each evaluation cell is a policy/recovery-seed and fault-count combination. These diagnostics use the fault counts and regime partition above; Figure 2C uses arithmetic means of Eq. (160) from the paired replay.

Table 8: Regime-specific correction statistics. LOO rows require at least one other faulty group ($m \geq 1$); N_1 and N_{LOO} count diagnostic observations; policy seeds define the independent source units.

Environment	Regime	N_1	Mean ρ_1^{emp}	$\Pr(\rho_1^{\mathrm{emp}} < 1)$	N_{LOO}	LOO proxy
8m	interior	15360	0.00101	1.000	115200	0.000935
8m	boundary	7680	0.000974	1.000	61440	0.000909
3s5z	interior	15360	0.000592	1.000	115200	0.000537
3s5z	boundary	7680	0.000577	1.000	61440	0.000534
2s3z	interior	15360	0.000461	1.000	69120	0.000422
2s3z	beyond	7680	0.904	0.366	38400	0.963
MPE5	interior	15360	7.77e-05	1.000	69120	7.33e-05
MPE5	beyond	7680	0.812	0.466	38400	0.917

C.5 Diagnostics of theoretical conditions

Table 9 reports finite-sample diagnostics of prediction error, directional network sensitivity, and observation structure. Direct errors quantify candidate accuracy; local gain proxies describe responses to specified interventions; secants and residual margins examine observability and support separation. The definitions below link each measurement to its input space and theoretical role.

Sampling and aggregation. Recovery diagnostics use five source policies and three recovery seeds per task, with 512 observation windows sampled without replacement per policy–recovery–fault-count cell. Fault counts are $s \in \{1, 2, 4\}$ on SMAC and $s \in \{1, 2, 3\}$ on MPE5. Windows follow the nominal trajectories; Gaussian faults are applied to the current observations and recovered observations are fed back into the history, with episode-start padding as in Appendix A.13. Window selection uses seed $20260905 + 10000q + 101s$, where q is the source-policy seed. Samples are ordered by trajectory and time and processed in batches of 16 on SMAC and 32 on MPE5. For batch/group records b containing N_b applicable samples, error and gain summaries are $\sum_b N_b \,\mathrm{median}(v_b)/\sum_b N_b$; proportions are $\sum_b N_b \,\mathrm{mean}(\mathbf{1}_{A_b})/\sum_b N_b$. History gains use full windows with nonzero history deviation. The margin summaries pool stages 1 and 2 over the stated fault counts. The RO and actor diagnostics use the separate samples specified below.

Coordinates and Stage 1 interventions. All recovery inputs and candidate outputs below use checkpoint-normalized observation coordinates. Write $r(V) = \|V\|_F/\sqrt{|V|}$ for entrywise RMS, where $|V|$ is the number of entries, and V_{-i} for the tensor with target row i deleted. Let H be the clean history, $\widehat{H}$ its recovered counterpart, and $Y = X + E$. Direct clean error is $e_{1,i} =$

Table 9: Finite-sample theoretical diagnostics. Error and gain summaries average batch medians with sample-count weights; event entries are sample proportions. RO uses weighted within-source/target fifth percentiles. Realized-margin proportions pool faulty-group–stage samples; empirical-envelope proportions use window–stage samples.

Primitive	8m	3s5z	2s3z	MPE5
RO(1): 5th-percentile secant proxy	0.954	0.940	0.883	0.880
Clean prediction error e_1	1.944	1.435	1.300	0.582
$\Pr(\widehat{\eta}_1 < 1)$	1.000	1.000	1.000	1.000
Cross secant $g_{1,\times}$	0.042	0.042	0.067	0.067
History secant $g_{1,H}$	0.023	0.051	0.026	0.079
Positive realized fault margin	1.000	1.000	0.529	0.650
Positive empirical envelope $\widehat{\gamma}_{\mathrm{f}}$	1.000	1.000	0.993	1.000
$\Pr(\rho_1^{\mathrm{emp}} < 1)$	1.000	1.000	0.789	0.822
Ideal-context prediction error e_2^*	1.938	1.427	1.274	0.263
Raw-context feature gain g_{raw}	0.087	0.067	0.101	0.112
Refined-context feature gain g_{ref}	0.095	0.078	0.075	0.114
Stage 1 latent-path gain g_z	0.057	0.085	0.055	0.074
History RMS screen $\Pr(\sqrt{11}g_{\mathcal{R},H} < 1)$	0.997	1.000	0.852	0.798
$\Pr(\widehat{\kappa}_h < 1)$	0.209	0.250	0.016	0.000
Global $k_{95}/D_{\mathrm{active}}$	0.232	0.181	0.182	0.209

$r(\mathcal{G}_{1,i}(H,X) - x_i)$. The directional finite-secant proxies are

$$\widehat{\eta}_{1,i} = \frac{r(\mathcal{G}_{1,i}(H,Y) - \mathcal{G}_{1,i}(H, X + E_{-i}))}{r(e_i)},$$
$$g_{1,\times,i} = \frac{r(\mathcal{G}_{1,i}(H,X + E_{-i}) - \mathcal{G}_{1,i}(H,X))}{r(E_{-i}^{\mathrm{del}})},$$
$$g_{1,H,i} = \frac{r(\mathcal{G}_{1,i}(\widehat{H},Y) - \mathcal{G}_{1,i}(H,Y))}{r(\widehat{H} - H)}.$$

Here E_{-i}^{del} deletes row i, whereas $E_{-i} = \Pi_{-i}E$ retains the full input shape. The self intervention holds the other faults and history fixed and is evaluated on faulty targets; the cross intervention holds the target and history clean; the history intervention holds current Y fixed. Each forward pass recomputes all Stage 1 features from its specified input. Finite-difference denominators use an additive 10^{-12}; history samples require $r(\widehat{H} - H) > 10^{-10}$. Zero cross perturbations contribute zero. In the theoretical norms, the corresponding nonzero secants are $\widehat{\eta}_{1,i}$, $g_{1,\times,i}/\sqrt{n-1}$, and $g_{1,H,i}/\sqrt{n(T-1)}$, respectively, up to the denominator regularizer. The realized cancellation ratio is $\rho_1^{\mathrm{emp}} = r(X^{(1)} - X)/r(Y - X)$.

Ideal prediction and encoded-context gains. For target i, write $D_i(t, g, q_{\mathrm{raw}}, q_{\mathrm{ref}}, z)$ for the Stage 2 latent map and decoder, with temporal feature t, relational feature g, encoded raw/refined LOO contexts $q_{\mathrm{raw}}, q_{\mathrm{ref}}$, and Stage 1 latent z. Its ideal input a_i^* is computed from (H, X) with both LOO branches formed directly from X. The measured ideal-context error is

$$e_{2,i}^* = r(D_i(a_i^*) - x_i),$$

the prediction error on the left side of Assumption A.7, whose upper bound is $\omega_i + \varepsilon_{2,i}^{\mathrm{net}}$. For $r \in \{\mathrm{raw}, \mathrm{ref}, z\}$, let $a_i^{[r]}$ replace just that feature in a_i^*. The raw replacement is encoded from Y; the refined replacement is encoded from $X^{(1)}(H,Y)$; the latent replacement is $z_i^{(1)}(H,Y)$. All other features remain at their ideal values. Table 9 reports

$$g_{r,i} = \frac{r(D_i(a_i^{[r]}) - D_i(a_i^*))}{r(a_{i,r}^{[r]} - a_{i,r}^*) + 10^{-12}}.$$

These are encoded-feature-to-candidate secants; $g_{z,i}$ isolates the Stage 1 latent component of the auxiliary path. For an encoded context of width h_r, its downstream gain in Euclidean input norm

and group-RMS output norm is $g_{r,i}/\sqrt{h_r}$. The observation-coordinate gains in Assumption A.8 additionally compose this downstream mapping with the branch embedding, target mask, and context encoder. The table measures the downstream response at the displayed feature interventions.

Restricted clean secants. For each source policy q, 512 clean query windows and a disjoint pool of up to 1,024 reference windows are sampled without replacement using seed $20260905 + q$. Each query is paired with its 64 nearest references in Euclidean distance between flattened normalized histories, excluding references from the same episode. For $\Delta X = X - X'$ with $r(\Delta X) \geq 10^{-6}$, define

$$\widehat{\mu}_i(X, X') = \frac{r(\Delta X_{-i})}{r(\Delta X) + 10^{-12}}, \qquad \widehat{\mu}_{.05} = \frac{\sum_b N_b Q_{.05}(\widehat{\mu}_b)}{\sum_b N_b}.$$

Here b indexes source-policy/target-group pair sets, N_b counts valid pairs, and $Q_{.05}$ is the empirical fifth percentile with linear interpolation. For nonzero secants, the ratio in the group norms of Assumption A.3 is $\sqrt{(n-1)/n}\,\widehat{\mu}_i$, up to the denominator regularizer. History-neighbor secants characterize local visibility of clean variation; the global $k_{95}/D_{\mathrm{active}}$ statistic summarizes variance concentration as defined in Appendix C.

Residual margins. For each sampled window and stage $r \in \{1, 2\}$, the realized faulty-group margin is $d_i^{(r)} - \tau^{(r)}$. Define the empirical residual endpoints

$$\ell^{(r)} = \min_{i \notin F} d_i^{(r)}, \qquad u^{(r)} = \max_{i \notin F} d_i^{(r)}, \qquad v^{(r)} = \min_{i \in F} d_i^{(r)}.$$

The fault-side envelope diagnostic is

$$\widehat{\gamma}_{\mathrm{f}}^{(r)} = v^{(r)} - u^{(r)} - c_\tau\big(u^{(r)} - \ell^{(r)} + b_\tau\big).$$

The table pools faulty-group samples for the realized margin and window–stage samples for the envelope diagnostic, including all stated fault counts. The median–MAD sufficient condition applies on the clean-majority subset $s < n/2$; these empirical endpoints describe the observed residual range. The theoretical envelopes in Proposition A.11 are derived from the predictor error and sensitivity bounds.

History and recurrent-actor gains. The complete recovery intervention holds Y fixed and recomputes both stages, support estimates, and fusion under H and $\widehat{H}$. Its RMS secant is

$$g_{\mathcal{R},H} = \frac{r(\mathcal{R}(\widehat{H}, Y) - \mathcal{R}(H, Y))}{r(\widehat{H} - H) + 10^{-12}}.$$

Conversion to the history norm in Assumption A.9 gives $g_{\mathcal{R},H}/\sqrt{T-1}$, up to the denominator regularizer. The table reports the conservative RMS screen $\sqrt{T-1}\,g_{\mathcal{R},H} < 1$, with $T = 12$; the corresponding finite-secant coefficient in the theoretical maximum-error recursion is $g_{\mathcal{R},H}$. For each source actor q, 64 observation/hidden-state pairs are sampled without replacement from nominal traces using seed $20260905+q$. With the observation fixed and the recurrent reset mask set to one, $\widehat{\kappa}_h$ estimates the Euclidean spectral norm of the hidden-to-hidden Jacobian $D_h F(h, x)$ using six power iterations implemented by Jacobian-vector and vector-Jacobian products. Fractions below one and source-wise medians are aggregated with sample-count weights. The table separates direct errors, directional gain proxies, and structural or sufficient-condition diagnostics; together they support the plausibility of the analyzed mechanisms on the sampled inputs.

Realized and sufficient residual margins. Figure 2B measures the realized margins used by the soft-support bound. The interval certificate in Appendix A additionally controls the spread of clean residuals through $\gamma_{\mathrm{c}} > 0$. On the sampled SMAC inputs with $s < n/2$, the clean-side interval margin is nonpositive, while realized clean and faulty residuals remain separated by the learned threshold. These measurements characterize the realized support interface and the conservatism of its interval-envelope sufficient condition.

Recurrent stability. Local hidden-to-hidden Jacobian medians are approximately 1.479, 1.058, 1.269, and 2.363 on `2s3z`, `3s5z`, `8m`, and MPE5. 1.6%, 25.0%, 20.9%, and 0% of samples have gain below one. The evaluated Jacobians therefore predominantly occupy a noncontractive regime. The finite-horizon bound in Appendix A.14 allows noncontractive gains; an infinite-horizon conclusion requires its additional assumptions.

C.6 RESPONSE-SUPERVISION COMPARISON

Actor-access interface. Direct actor supervision uses the source actor's differentiable observation-to-action mapping. We also consider an interface in which the recovery trainer receives nominal observation windows, recorded actions, incoming recurrent states, available-action sets, and action distributions, while the actor implementation and its input gradients remain accessible only to the policy owner. Under this interface, a differentiable surrogate $\widetilde{\pi}_\psi$ is fitted to the recorded action distributions and then frozen. Replacing $\pi_\phi^{(q)}$ by the assigned $\widetilde{\pi}_\psi$ in the recovery-response KL, candidate-response KL, recorded-action cross-entropy, and margin losses supplies gradients for recovery training. The latent action head and observation losses retain their definitions. At deployment, the policy owner applies the original frozen actor to the recovered observations.

In this response-supervision comparison, the fitted-response arm uses a shared surrogate fitted to pooled source traces on `8m` and a policy-specific surrogate bank on `3s5z` and `2s3z`. Surrogates are fitted for 40 epochs. The shared fit uses source policies 1–4 and validates on policy 5; each bank member uses a within-policy training/validation split. Recovery training uses policies 1–4 and recovery validation uses policy 5 in all three arms.

Response-supervision configurations. We compare direct frozen-actor supervision, the fitted-response arm, and recovery trained with the decision-loss coefficients set to zero. The recovery configurations use the same nominal-data split and frozen source actors, and are evaluated under the same benchmark fault protocol. Five recovery seeds and five source actors give 25 evaluation cells per map, arm, and condition. Each cell contains 64 closed-loop episodes under Clean or $s1$–$s4$. Actor parameters remain fixed; direct supervision differentiates only with respect to the recovered observation and candidate inputs. Source-actor responses are checked against the recorded clean distributions before training, and parameter hashes are checked after training.

Supervision interface and optimization. The direct-actor and fitted-response configurations use the same two-stage recovery architecture. Each training sample supplies its source identifier, recorded clean action distribution, recorded action, incoming recurrent state, and available-action mask. The direct-actor interface evaluates the original frozen actor on the recovery and candidate; the fitted-response interface evaluates the frozen surrogate on those inputs. Recovery-response KL, candidate-response KL, recorded-action cross-entropy, and margin matching propagate gradients through the selected response mapping to the observation inputs. Their coefficients are 2.4, 1.2, 0.15, and 0.35, respectively. Latent-response KL uses the recovery model's auxiliary action head with coefficient 0.8 and the recorded clean distribution as its target. The no-decision configuration sets all five coefficients to zero.

All three configurations use AdamW with learning rate 3.5×10^{-4}, weight decay 10^{-4}, batch size 256, gradient clipping at norm 5, and 180 recovery-training epochs. Surrogate fitting has its separate 40-epoch budget described above. Source actors and fitted response models remain frozen during recovery training, and nominal reconstruction and clean-input diagnostics govern checkpoint selection. At deployment, every configuration feeds its recovered observations to the original frozen source actor.

Table 10: Closed-loop response-supervision comparison (win rate, %). Entries are mean±SD across five source policies after averaging five recovery seeds within each source.

Map	Supervision	Clean	$s1$	$s2$	$s3$	$s4$
8m	No decision losses	84.50 ± 10.58	67.62 ± 11.18	49.38 ± 15.13	27.62 ± 12.98	16.50 ± 12.06
8m	Surrogate	83.38 ± 10.57	71.00 ± 12.75	53.75 ± 13.42	39.62 ± 14.03	23.88 ± 9.62
8m	Direct actor	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
3s5z	No decision losses	83.06 ± 11.77	61.06 ± 13.69	38.44 ± 10.43	17.56 ± 6.42	5.69 ± 2.56
3s5z	Surrogate	81.06 ± 13.49	70.62 ± 14.42	58.44 ± 12.51	41.88 ± 11.84	22.19 ± 10.93
3s5z	Direct actor	82.19 ± 12.33	73.44 ± 14.53	68.25 ± 18.88	56.06 ± 23.63	36.13 ± 25.12
2s3z	No decision losses	83.62 ± 12.77	70.38 ± 18.48	51.12 ± 16.54	8.44 ± 6.21	0.50 ± 0.52
2s3z	Surrogate	84.69 ± 11.79	77.19 ± 15.89	65.94 ± 18.78	12.31 ± 5.98	0.44 ± 0.36
2s3z	Direct actor	83.75 ± 12.01	79.75 ± 16.48	69.50 ± 24.08	13.19 ± 3.56	0.38 ± 0.34

The direct-actor recovery configuration has higher mean $s1$–$s4$ win rate than the surrogate configuration by 9.28, 10.19, and 1.73 percentage points on `8m`, `3s5z`, and `2s3z`, respectively. The source-level change is positive for four of five sources on `8m` and `2s3z`, and all five on `3s5z`. Surrogate supervision also improves the severity-averaged win rate over training without decision supervision on every source policy in all three maps. These results support direct access to the frozen actor for decision-aware training and demonstrate the utility of fitted responses under the specified access restriction.

Table 11: Closed-loop recovery diagnostics averaged over $s1$–$s4$, recovery seeds, and source policies with equal weight at each level. MSE uses original observation coordinates. Actor KL compares the original actor on clean and recovered observations at the recorded evaluation hidden state; agreement compares their selected actions.

Map	Supervision	MSE	Actor KL	Agreement (%)	Support F1 (%)
8m	No decision losses	0.01425	0.4629	89.46	99.91
8m	Surrogate	0.01568	0.3758	90.48	99.71
8m	Direct actor	0.01474	0.2727	91.54	99.96
3s5z	No decision losses	0.02001	0.3220	90.41	99.37
3s5z	Surrogate	0.01879	0.2456	92.46	99.96
3s5z	Direct actor	0.01569	0.1511	93.62	99.77
2s3z	No decision losses	0.34717	0.6887	81.45	64.22
2s3z	Surrogate	0.32467	0.6496	82.27	69.32
2s3z	Direct actor	0.30951	0.6053	82.92	71.34

Direct actor supervision yields the lowest mean actor KL and highest action agreement on all three maps. On `8m`, training without decision supervision has slightly smaller observation MSE than direct actor supervision, while its action agreement and task performance are lower. The diagnostics thus distinguish reconstruction accuracy from preservation of action responses. Each arm's closed-loop measurements describe the states visited by that arm.

Fidelity on recovery-generated observations. We evaluate the fitted response mapping against the original actor on policy-5 traces used for recovery validation. Up to 4,096 clean windows per recovery seed supply clean targets and generated candidates and recoveries. The generic-mask probe uses masking probability 0.28. Fault probes corrupt only the current observation with s randomly selected groups and scale $3\max(\mathrm{std}(X_t), 10^{-6})$, keeping preceding observations clean. Table 12 reports final recoveries from the surrogate-trained arm. KL is $D_{\mathrm{KL}}(\pi_\phi \| \widetilde{\pi}_\psi)$ on the same input, incoming clean recurrent state, and available-action set, using the training implementation's probability floor. Agreement is between the two distributions' highest-probability actions. Values average the five recovery-seed probe means. Policy 5 is held out from recovery training; its bank surrogate is fitted using the within-policy split described above.

Table 12: Surrogate fidelity on clean inputs and recovery-generated inputs. Agreement is a percentage.

Map	Input	Actor-to-surrogate KL	Top-1 agreement (%)
8m	Clean input	1.3241	54.59
8m	Masked recovery	1.0746	59.25
8m	$s1$ recovery	1.2990	54.82
8m	$s4$ recovery	1.2583	54.69
3s5z	Clean input	0.0037	98.68
3s5z	Masked recovery	1.2607	51.45
3s5z	$s1$ recovery	0.2077	91.20
3s5z	$s4$ recovery	0.8056	69.68
2s3z	Clean input	0.0037	98.94
2s3z	Masked recovery	1.5313	44.34
2s3z	$s1$ recovery	0.3425	86.88
2s3z	$s4$ recovery	1.4757	48.68

The policy-specific surrogates closely match clean responses on `3s5z` and `2s3z`, while fidelity changes on recovery-generated inputs. This input dependence motivates evaluating response approximation on the observations used by the recovery objective in addition to nominal samples.

Recurrent trace replay. For 20 policy-5 episodes per map, we replay the fixed clean environment trajectories while propagating recovered observation history and the original actor's recurrent state. At each step, the clean trace supplies the reference observation, action availability, and clean recurrent state. Table 13 compares clean-response KL when the recovered observation is evaluated with this clean state and with the propagated state. The fault generator uses the probe scale above; results average time steps within each recovery-seed replay and then the five recovery seeds. This fixed-trajectory diagnostic isolates the effect of recurrent state propagation alongside the closed-loop task evaluations.

Table 13: Clean-response KL under clean and propagated recurrent states on fixed environment traces.

Map	Faults	Supervision	Clean state	Propagated state
8m	$s1$	Surrogate	0.3381	0.3926
8m	$s1$	Direct actor	0.3168	0.3694
8m	$s4$	Surrogate	1.4205	1.6244
8m	$s4$	Direct actor	1.3258	1.5070
3s5z	$s1$	Surrogate	0.1383	0.1858
3s5z	$s1$	Direct actor	0.1023	0.1319
3s5z	$s4$	Surrogate	0.5637	0.8879
3s5z	$s4$	Direct actor	0.4349	0.6585
2s3z	$s1$	Surrogate	0.0834	0.1292
2s3z	$s1$	Direct actor	0.0773	0.1194
2s3z	$s4$	Surrogate	0.8744	1.6861
2s3z	$s4$	Direct actor	0.8498	1.6427

At $s4$, propagated states increase clean-response KL on all three maps. Direct actor supervision yields lower KL than surrogate supervision with either state choice in these replays, connecting the response-training comparison to the finite-horizon recurrent analysis in Appendix A.14.

D Cross-Domain Validation

The MPE and GRF experiments evaluate the recovery method and training objective in Section 4 with frozen source policies in each environment.

Cross-domain checkpoint selection. Both environments select recovery checkpoints by minimizing

$$V_{\mathrm{MPE}} = V_{\mathrm{GRF}} = E_{\mathrm{mask}} + 0.2E_{\mathrm{view}} + 0.5E_{\mathrm{clean}} + 0.8S_{\mathrm{clean}} + 1.2K_{\mathrm{mask}} + 0.35M_{\mathrm{mask}} + 0.5K_{\mathrm{clean}}. \tag{161}$$

The observation and support terms follow Appendix C.2: E_{mask} and E_{clean} are final-recovery MSE against the clean target for masked view a and the unmasked view, respectively; E_{view} is the MSE between the final recovered observations from masked views a and b; S_{clean} is the mean fused support score on the unmasked view. Observation MSE uses original coordinates and averages over validation windows, agents, and observation dimensions.

For each validation record and agent i, let q_i be the recorded nominal action distribution and $\widehat{q}_i^{(v)}$ the original frozen actor's distribution on the final recovery from view $v \in \{a, \mathrm{cl}\}$. Responses use the recorded incoming recurrent state and available-action mask. The terms K_{mask} and K_{clean} average $D_{\mathrm{KL}}(q_i \| \widehat{q}_i^{(a)})$ and $D_{\mathrm{KL}}(q_i \| \widehat{q}_i^{(\mathrm{cl})})$, respectively. Writing $m(q) = q_{(1)} - q_{(2)}$ for the difference between the two largest action probabilities, M_{mask} averages $|m(\widehat{q}_i^{(a)}) - m(q_i)|$. Support and response terms average over validation windows and agents.

D.1 FIVE-AGENT MPE

We use five-agent Simple Spread with 25-step episodes and local reward ratio 0.5. Reported return sums rewards over all five agents and all time steps; comparison with three-agent Spread requires a matching reward and aggregation convention. The distance component is shared and repeated in the team sum. An occupied landmark is within 0.1 of an agent.

Recovery data and optimization. Each of the five source policies supplies 200 nominal episodes of 25 steps. Policies 1–4 supply 800 training episodes (20,000 windows), and policy 5 supplies 200 validation episodes (5,000 windows), using the window construction in Appendix C.2. Recovery seeds are 1–3. Each epoch draws 20,000 training windows with replacement using inverse source-policy-frequency weights. AdamW uses batch size 192, learning rate 3.5×10^{-4}, weight decay 10^{-4}, and gradient clipping at norm 5. The budget is at most 180 epochs, with a 28-epoch non-improvement patience on the nominal validation score; the three runs complete 100, 155, and 48 epochs. Checkpoint selection uses Eq. (161) and retains the earliest score improvement exceeding 10^{-7}.

Table 14: MPE recovery results. Recovery seeds are averaged within source policy before source-level mean and standard deviation; MSE and F1 are diagnostic cell means.

Method	s	Source seeds	Team return	MSE	Support F1
clean	0	5	-141.79 ± 2.03	0.000	–
corrupted	1	5	-144.35 ± 1.09	0.834	0.000
corrupted	2	5	-149.29 ± 3.78	3.497	0.000
corrupted	3	5	-154.44 ± 3.68	11.702	0.000
PCA-SIR	0	5	-141.73 ± 1.35	7.29e-04	–
PCA-SIR	1	5	-150.24 ± 3.18	0.105	0.763
PCA-SIR	2	5	-166.87 ± 4.24	0.690	0.609
PCA-SIR	3	5	-176.34 ± 4.88	2.397	0.750
SIR-Direct	0	5	-141.59 ± 2.31	0.002	–
SIR-Direct	1	5	-143.01 ± 3.51	0.034	0.993
SIR-Direct	2	5	-148.59 ± 3.08	0.087	1.000
SIR-Direct	3	5	-155.84 ± 4.70	6.374	0.452
SIR-MARL	0	5	-141.33 ± 1.95	9.14e-04	–
SIR-MARL	1	5	-142.85 ± 2.73	0.036	0.992
SIR-MARL	2	5	-148.21 ± 3.10	0.090	1.000
SIR-MARL	3	5	-154.63 ± 3.14	5.516	0.478

The clean source-policy return is approximately -141.79. SIR-MARL reduces observation MSE by approximately 96%–97% at $s1$/$s2$, with smaller corresponding task-return gains. At $s3$, beyond the majority regime, support and task recovery degrade. Signal fidelity and task utility are therefore distinct outcomes.

Source-policy quality. A separate 1,000-episode paired-reset evaluation gives mean return -143.33 for the five MAPPO seeds versus -177.13 for random actions. Two privileged-state Hungarian assignment heuristics obtain -63.30 and -63.04 and serve as reference policies. MAPPO's mean final occupied-landmark count is 0.206, versus 4.371 and 4.348 for the references. The source policies outperform random actions while achieving substantially lower landmark occupancy than the heuristic references. This source-quality evaluation uses an independent set of environment reset seeds.

Table 15: MPE continuation on three source seeds, each evaluated over 1,000 paired reset seeds.

Seed	Before	After +2M	Paired change	95% CI
1	-142.69	-143.24	-0.55	[-2.04, 0.89]
2	-142.26	-143.34	-1.08	[-2.76, 0.59]
3	-143.44	-141.95	+1.50	[-0.13, 3.09]

None of the three continuation runs met the prespecified improvement criterion (gain above 2 and a positive CI lower bound) over the additional 2M environment steps. Each continuation restores weights with a fresh Adam optimizer.

D.2 GOOGLE RESEARCH FOOTBALL

GRF uses `academy_corner`, ten controlled agents, and three frozen RMAPPO source policies trained for 50M steps each. Each evaluation cell contains 500 episodes with distinct environment seeds and paired environment/fault seed streams across methods. SIR-MARL and SIR-Direct each use three recovery seeds per source. Recovery-seed values are averaged within source before taking the mean and sample standard deviation across the three sources. Clean evaluates each method with its deployed inference procedure. Cardinalities $s1$–$s4$ are interior to $s < n/2$, and $s5$ is the 50% boundary.

Recovery data and optimization. Each source policy supplies 96 nominal episodes. A separate recovery model is fitted for each source and recovery seed using the data sizes in Table 4. For recovery seed $r \in \{1, 2, 3\}$, a NumPy generator seeded by r permutes the N observation records; the first $\lfloor 0.2N \rfloor$ form validation data and the remainder form training data. Within each split and episode, records are ordered by their original time indices, and each window contains the current record and up to 11 preceding retained records, repeating the first retained observation for initial padding. Each epoch visits every training window once in shuffled order. AdamW uses batch size 256, learning rate 3.5×10^{-4}, weight decay 10^{-4}, and gradient clipping at norm 5. Training uses a maximum of 180 epochs and a 28-epoch non-improvement patience; the nine runs complete 35–111 epochs. Nominal validation uses Eq. (161) and selects the earliest score improvement exceeding 10^{-7}.

Fault and comparison protocols. At each step, the GRF fault generator samples s distinct agents uniformly and corrupts their observations in the sampled order. Let $g_1, \ldots, g_s$ denote this order and $Y_t^{[0]} = X_t$ the clean team observation. Before the jth injection, it sets $\sigma_{t,j} = 3\big(\mathrm{std}(\mathrm{vec}\, Y_t^{[j-1]}) + 10^{-6}\big)$ and updates $Y_{t,g_j}^{[j]} = Y_{t,g_j}^{[j-1]} + \sigma_{t,j} z_{t,j}$, where $z_{t,j} \sim \mathcal{N}(0, I_D)$ independently; the other rows are preserved. The standard deviation uses population normalization. All GRF methods use this sequentially scaled Gaussian channel, with support and noise resampled at each step.

Table 16 compares recovery with task-specific recurrent MAPPO adaptations of ERNIE, ROMANCE, and WALL, labeled ERNIE-MAPPO, ROMANCE-MAPPO, and WALL-MAPPO. The GRF ERNIE adaptation applies differentiable adversarial response regularization directly to the PPO actor objective, following the minibatch regularization procedure in Appendix C.2. ROMANCE and WALL transfer their action-adversarial training mechanisms to this observation-fault evaluation. PCA-SIR and SIR-Direct use the frozen-source recovery interface described in Appendix C.2; SparseAug-Mix uses observation augmentation during policy training as described in Appendix G.3.

Table 16: GRF closed-loop win rate (%). Mean±std over three source-policy seeds, with recovery seeds averaged within source; 500 episodes per evaluation cell.

Method	Clean	$s1$	$s2$	$s3$	$s4$	$s5$
RMAPPO	47.80±2.46	36.00±1.91	24.67±4.46	17.07±5.03	10.73±3.35	7.73±1.80
ERNIE-MAPPO	48.27±0.83	25.87±3.31	13.53±1.01	8.27±1.68	4.73±0.70	2.73±0.50
ROMANCE-MAPPO	46.47±1.01	29.73±5.13	19.07±5.35	12.40±4.13	7.93±3.40	5.13±2.34
WALL-MAPPO	41.13±7.09	27.67±2.97	19.53±4.28	12.60±2.62	7.60±3.34	4.47±0.92
PCA-SIR	47.73±2.37	1.00±0.35	1.13±0.81	0.60±0.00	0.60±0.00	0.47±0.31
SIR-Direct	47.80±2.51	45.49±0.92	43.53±1.03	37.62±1.43	32.33±2.30	23.84±2.48
SIR-MARL	47.78±2.44	45.31±1.90	42.62±2.26	36.60±3.19	30.62±1.93	23.69±3.86
SparseAug-Mix	48.73±20.60	48.73±21.62	44.27±21.96	41.27±20.22	35.07±16.08	28.67±16.57

The neural recovery methods retain near-source clean performance and improve win rates over the unprotected source and PCA-SIR throughout the fault sweep. This supports post-training recovery on football observations as well as SMAC. SparseAug-Mix combines higher average faulted scores with larger variation across policy seeds.

Table 17: GRF performance and observation-recovery diagnostics. Win rates (%) use source-level aggregates. MSE and F1 average over recorded team steps within each evaluation cell, then over recovery and source seeds.

Method	s	Win rate (%)	MSE	Support F1
Source	0	47.80 ± 2.46	0.0000	–
SIR-MARL	0	47.78 ± 2.44	0.0014	–
Corrupted	1	36.00 ± 1.91	0.1545	–
SIR-MARL	1	45.31 ± 1.90	0.0038	0.952
Corrupted	2	24.67 ± 4.46	0.4324	–
SIR-MARL	2	42.62 ± 2.26	0.0068	0.980
Corrupted	3	17.07 ± 5.03	0.9258	–
SIR-MARL	3	36.60 ± 3.19	0.0112	0.991
Corrupted	4	10.73 ± 3.35	1.7953	–
SIR-MARL	4	30.62 ± 1.93	0.0164	0.996
Corrupted	5	7.73 ± 1.80	3.5216	–
SIR-MARL	5	23.69 ± 3.86	0.0274	0.997

Action-response preservation. On the SIR-MARL closed-loop trajectories, we evaluate the frozen actor on clean, corrupted, and recovered observations at the same visited state and incoming recurrent state. Table 18 reports KL divergence from the clean-observation action distribution and greedy-action agreement with that response. Recovery lowers action-distribution deviation and increases action agreement across the evaluated fault counts. These measurements connect observation correction to the source actor's decisions.

Table 18: Frozen-actor response diagnostics on GRF. KL and action agreement are averaged over agent decisions within each SIR-MARL evaluation cell, then over recovery and source seeds.

s	Corrupted KL	Recovered KL	Corrupted agreement (%)	Recovered agreement (%)
0	0.0000	0.0035	100.00	98.32
1	0.0345	0.0143	91.62	93.98
2	0.0701	0.0260	83.05	89.68
3	0.1073	0.0413	74.35	84.42
4	0.1434	0.0575	65.47	78.48
5	0.1797	0.0785	56.60	71.10

Per-source recovery variability. Table 19 reports each recovery replicate and its source-matched corrupted reference. The gain averages the three recovery replicates before subtracting the corresponding source score.

Table 19: Source/recovery GRF win rates (%) and mean paired gains (percentage points). Every cell uses 500 episodes.

Source	s	Corrupted	Recovery 1	Recovery 2	Recovery 3	Mean gain
1	1	35.80	47.20	47.60	47.60	+11.67
1	2	22.40	43.80	45.40	44.80	+22.27
1	3	12.40	38.40	38.40	36.40	+25.33
1	4	8.60	31.40	30.60	27.80	+21.33
1	5	6.00	25.20	22.20	22.00	+17.13
2	1	38.00	44.00	45.00	42.60	+5.87
2	2	29.80	40.60	42.80	37.20	+10.40
2	3	22.40	32.40	37.80	28.80	+10.60
2	4	14.60	29.20	35.00	23.20	+14.53
2	5	9.60	18.80	28.20	13.40	+10.53
3	1	34.20	46.40	44.00	43.40	+10.40
3	2	21.80	41.60	42.80	44.60	+21.20
3	3	16.40	43.60	37.20	36.40	+22.67
3	4	9.00	36.60	32.20	29.60	+23.80
3	5	7.60	29.60	28.80	25.00	+20.20

E Ablations, Sensitivity, and Computational Cost

E.1 Architecture ablations

The full architecture ablation matrix below reports five source-policy aggregates. Removing masked-team completion zeros both dedicated raw and corrected target-masked context features while preserving the other pathways. Removing the second stage deploys $C^{(1)}$ and $p^{(1)}$ directly. Temporal and relational ablations suppress their corresponding features. These are separately trained variants. The tables compare separately trained recovery configurations on the same frozen source policies under the benchmark evaluation protocol in Appendix C.2. Appendix C.2 also specifies feature zeroing, the single-stage output-to-loss mapping, and training and evaluation settings.

Table 20: Complete 8m architecture ablations; mean and standard deviation over five policy aggregates.

Setting	Clean	$s1$	$s2$	$s3$	$s4$
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
w/o decision supervision	84.50 ± 10.58	67.62 ± 11.18	49.38 ± 15.13	27.62 ± 12.98	16.50 ± 12.06
w/o relational features	83.06 ± 12.42	73.94 ± 12.76	60.94 ± 19.43	44.75 ± 21.33	30.25 ± 17.49
w/o masked-team completion	83.88 ± 11.14	71.44 ± 11.99	56.88 ± 13.64	38.38 ± 12.07	22.56 ± 8.75
w/o second stage	83.31 ± 10.78	73.56 ± 12.51	57.19 ± 15.22	40.88 ± 16.01	22.94 ± 11.69
w/o temporal features	83.81 ± 10.97	70.31 ± 21.86	56.94 ± 25.66	43.94 ± 22.70	28.81 ± 19.75

Decision losses and masking design. Table 21 removes each response KL, recorded-action cross-entropy, or action-margin loss in turn. The Stage 1 LOO variant replaces the target's current observation with its training mean during initial prediction, retaining its past history and the Stage 2 architecture. The zero-token variant fixes both Stage 2 mask tokens to zero. The variants use nominal-data training and the same benchmark evaluation protocol.

The three response KL terms contribute to robustness as fault coverage increases, supporting supervision of the recovered observation, candidate, and latent representation. Retaining the target input in Stage 1 gives higher average faulted win rates than masking it in both stages, consistent with preserving current local information for initial prediction. Cross-entropy, margin matching, and the mask-token parameterization have smaller effects in this comparison, with the full and corresponding variants attaining similar average performance.

Table 21: Decision-loss and masking-design ablations on 8m (win rate, %). Five recovery seeds are averaged within each of five source policies; entries report mean±SD over the source aggregates.

Setting	Clean	$s1$	$s2$	$s3$	$s4$
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
w/o recovery KL	84.69 ± 10.04	73.94 ± 11.58	60.19 ± 19.70	43.69 ± 18.08	29.38 ± 14.74
w/o candidate KL	84.00 ± 11.37	72.25 ± 13.38	60.13 ± 15.24	44.88 ± 17.25	29.88 ± 18.70
w/o latent KL	84.06 ± 11.50	73.19 ± 12.72	59.88 ± 16.80	45.13 ± 22.30	30.13 ± 17.37
w/o action CE	84.38 ± 9.53	73.81 ± 11.60	63.31 ± 15.29	49.31 ± 20.05	33.69 ± 18.47
w/o action margin	83.13 ± 10.56	72.69 ± 12.92	65.44 ± 16.33	50.75 ± 19.81	37.00 ± 20.08
Stage 1 also LOO	84.50 ± 10.64	70.38 ± 13.14	60.00 ± 17.77	47.81 ± 21.28	28.94 ± 15.64
Zero mask tokens	83.19 ± 12.01	72.69 ± 12.81	62.31 ± 18.75	52.31 ± 18.61	34.56 ± 15.25

E.2 Additional refinement and hyperparameters

The following sweeps vary refinement depth, masking probability, history length, decision-loss weight, or inference constants. Training sweeps use source policy 4 for nominal training, policy 5 for validation, and three recovery seeds evaluated with policies 4 and 5. Inference sweeps use recovery seed 1 with all five source policies. Each table uses the default trained and evaluated within its own subset. Recovery seeds are averaged within source before source averaging.

In the setting labels, α denotes support fusion, β the residual-threshold multiplier, and temp the support temperature. Suffixes 05, 10, 15, and 20 denote 0.5, 1, 1.5, and 2 times the default; $\alpha 0$ sets the fusion weight to zero. Training labels specify decision-weight multipliers, masking probabilities, or history lengths. The default masking probability is 0.28 and history length is 12.

Additional refinement improves the 8m subset at $s2$–$s4$, while the two-stage model has a slightly higher average faulted win rate on the 3s5z subset. These depth sweeps characterize task-dependent returns from another refinement step. The masking and history sweeps characterize the balance between reconstructing unavailable observations and retaining informative context.

Table 22: Refinement-depth sweep on 8m. Win rates (%) average recovery seeds within source first; N_p and N_r are the evaluated policy and recovery seed counts.

Setting	N_p	N_r	$s1$	$s2$	$s3$	$s4$
two_stage	2	3	77.34	58.85	38.54	25.78
k3_natural	2	3	75.00	64.32	47.92	30.21

Table 23: Refinement-depth sweep on 3s5z. Win rates (%) average recovery seeds within source first; N_p and N_r are the evaluated policy and recovery seed counts.

Setting	N_p	N_r	$s1$	$s2$	$s3$	$s4$
two_stage	2	3	72.66	61.98	47.92	29.43
k3_natural	2	3	74.48	58.85	44.01	30.21

E.3 Support controls and target error

Support localization and target reconstruction. On 8m, the standard unknown-cardinality estimator has support F1 of 99.92%, 99.99%, 99.99%, and 99.95% from $s1$ to $s4$. Final observation MSE increases from 0.0059 to 0.0235, and actor KL from 0.102 to 0.451. Accurate localization accompanies residual error in reconstructing decision-relevant information as fewer intact observations remain. The known-k and oracle controls provide fault cardinality and support, respectively.

Final-support interventions. Each support control is applied at evaluation time with recovery-network parameters fixed. Given an input window, the network first computes its ordinary Stage 1 correction, Stage 2 candidate $C^{(2)}$, and fused support p. The known-k control selects the k largest

Table 24: Inference sensitivity on 8m. Win rates (%) average recovery seeds within source first; N_p and N_r are the evaluated policy and recovery seed counts.

Setting	N_p	N_r	$s1$	$s2$	$s3$	$s4$
inf_alpha0	5	1	78.75	66.25	53.75	40.31
inf_alpha05	5	1	76.88	68.44	52.50	40.31
inf_alpha10	5	1	72.50	65.94	53.75	40.31
inf_alpha20	5	1	77.19	68.13	54.69	40.31
inf_beta05	5	1	77.81	65.94	48.75	40.31
inf_beta10	5	1	72.50	65.94	53.75	40.31
inf_beta15	5	1	77.81	66.88	52.81	39.06
inf_beta20	5	1	76.56	66.56	51.25	39.06
inf_temp05	5	1	75.00	66.25	52.50	39.06
inf_temp10	5	1	72.50	65.94	53.75	40.31
inf_temp20	5	1	74.69	66.25	51.56	40.31

Table 25: Inference sensitivity on 3s5z. Win rates (%) average recovery seeds within source first; N_p and N_r are the evaluated policy and recovery seed counts.

Setting	N_p	N_r	$s1$	$s2$	$s3$	$s4$
inf_alpha0	5	1	79.69	69.69	52.19	36.88
inf_alpha05	5	1	76.88	67.81	50.94	36.88
inf_alpha10	5	1	73.44	66.56	51.56	36.88
inf_alpha20	5	1	76.56	67.50	55.94	36.88
inf_beta05	5	1	76.88	70.31	62.50	35.00
inf_beta10	5	1	73.44	66.56	51.56	36.88
inf_beta15	5	1	75.00	70.94	56.56	38.44
inf_beta20	5	1	75.00	68.13	54.69	38.44
inf_temp05	5	1	75.63	68.13	54.69	38.44
inf_temp10	5	1	73.44	66.56	51.56	36.88
inf_temp20	5	1	73.75	67.19	58.44	35.63

Table 26: Training sensitivity on 8m. Win rates (%) average recovery seeds within source first; N_p and N_r are the evaluated policy and recovery seed counts.

Setting	N_p	N_r	$s1$	$s2$	$s3$	$s4$
sens_lamdec05	2	3	75.26	61.72	46.09	26.56
sens_lamdec20	2	3	73.18	59.64	39.58	21.35
sens_mask014	2	3	73.18	53.39	30.47	9.90
sens_mask028	2	3	77.34	58.85	38.54	25.78
sens_mask042	2	3	77.08	59.38	44.79	28.13
sens_t24	2	3	73.18	50.00	25.52	13.80
sens_t6	2	3	74.74	61.20	42.71	26.30

entries of this fused support; the oracle control uses the injected support set F. Their hard weights and final recovery are

$$p_i^{\text{known-}k} = \mathbf{1}\{i \in \text{TopK}(p, k)\}, \qquad p_i^{\text{oracle}} = \mathbf{1}\{i \in F\}, \qquad \widehat{x}_i = (1 - p_i^{\text{control}})y_i + p_i^{\text{control}} c_i^{(2)}.$$

Only the final mixing weights are replaced. For that input window, both stages retain their ordinary computations and the intervention reuses the resulting candidate. These controls require no recovery training or fine-tuning. During closed-loop evaluation, each control feeds its own recovered outputs into subsequent history windows.

Estimated support attains high support F1, and the recovery errors are close to those of the oracle-support reference. A separate oracle-context control supplies clean observations to the refined LOO branch:

$$M_i^{\text{ref}}(X^{(1)}) \longrightarrow M_i^{\text{ref}}(X).$$

The target slot remains masked. Given the same input window, the raw branch still uses Y, and the temporal features, relational features, and Stage 1 outputs retain their ordinary values. With network parameters fixed, Stage 2 recomputes its candidate and residual-based support from the substituted context, then applies the ordinary support-fusion and recovery rules. Its mean final MSE spans 0.0059–0.0244 over $s1$–$s4$, close to ordinary recovery. Together, these controls distinguish accurate support localization from the remaining conditional-prediction error.

Table 27: 8m support estimates and information-assisted recovery controls. Entries are diagnostic means across source and recovery seeds.

Support mode	s	Recovery seeds	Win rate (%)	F1	MSE
estimated	s1	5	74.38	0.999	0.006
estimated	s2	5	64.88	1.000	0.012
estimated	s3	5	50.75	1.000	0.018
estimated	s4	5	35.38	1.000	0.024
known_k	s1	5	74.88	1.000	0.006
known_k	s2	5	66.13	1.000	0.012
known_k	s3	5	49.50	1.000	0.018
known_k	s4	5	35.13	1.000	0.024
oracle	s1	5	74.88	1.000	0.006
oracle	s2	5	65.75	1.000	0.012
oracle	s3	5	49.50	1.000	0.018
oracle	s4	5	35.13	1.000	0.024

E.4 Deployment cost

For the `8m` input configuration ($n = 8$, $D = 204$, and $T = 12$), the full recovery module contains $3{,}764{,}904$ parameters, including the auxiliary action head. Profiling uses one FP32 zero-valued input of shape $1 \times 19{,}584$, evaluation mode, and disabled gradients. A CPU-side `torch.profiler` forward pass with `with_flops=True` counts $298{,}636{,}003$ FLOPs from supported operators. This is a partial operator count for the full module, including its auxiliary outputs. The frozen actor, environment stepping, and communication are outside the measured computation.

CUDA-event timing of the same module on an NVIDIA GeForce RTX 4090 gives a mean of 4.33 ms per team step over 30 forwards following 10 warm-up forwards, with CUDA synchronization before and after the timed block. Inputs and model parameters are already on the GPU. The implementation environment uses Python 3.8.20, PyTorch 2.4.1 with CUDA 11.8, and cuDNN 9.1.0. Parameter count and supported-operator FLOPs describe the profiled architecture; elapsed time also depends on device load and execution conditions. The reported model uses the full architecture.

F Dynamic and Unseen-Family Faults

These evaluations examine recovery under temporally persistent supports and different corruption families. DYN2 brackets three two-agent support segments with clean phases; DYN3 uses successive cardinalities 1, 4 (3 on `2s3z`), and 2; DYN4 inserts five short bursts with randomly selected

cardinality. Support is fixed within each segment. Fault-family tests use one support for the entire episode at $s2$ or $s4$.

F0 adds a Gaussian vector shared by the faulty groups at each step, with scale $3(\mathrm{std}(\mathrm{vec}\, X_t)+10^{-6})$ computed from the clean team observation. This evaluates persistent support with correlated corruption across agents. The dynamic tests use the same corruption rule within their active segments. F1 replaces faulty observations with their nominal feature means; F2 adds independent Gaussian noise scaled by nominal feature standard deviations; F3 holds the sampled Gaussian bias fixed throughout a segment; F4 uses each target's clean observation from four steps earlier; F5 replaces each target with a sample from its nominal observation pool. F2 and F3 use unit intensity in Figure 3.

All tests use recovery seed 1 and source seeds 1/3/4 (`8m`), 2/3/4 (`3s5z`), and 1/2/5 (`2s3z`), selected by low, median, and high nominal source performance. Corrupted MAPPO and SIR-MARL are evaluated for 64 episodes per condition with matched source checkpoints and fault manifests. Fault streams vary by episode; the environment seed initializes each evaluation process. Figure 3 averages $s2/s4$ within source and then the three sources. Table 28 separates the two cardinalities and gives absolute corrupted and recovered performance together with clean-source and clean-recovery references for the same source subset. Each cell is computed from integer wins divided by the actual completed episodes, then summarized as the mean and sample standard deviation across the three sources; rounding is applied only to the displayed values.

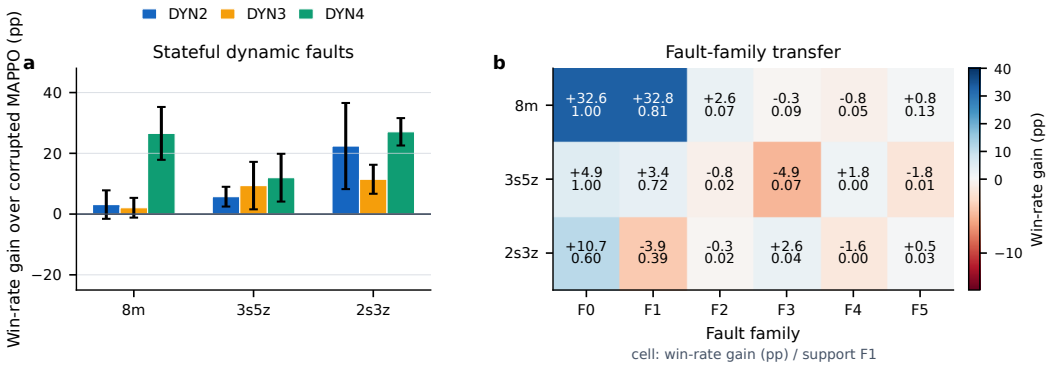


Figure 3: Recovery under temporal and distributional changes. (a) Mean±SD of source-paired win-rate gains for support switching (DYN2), cardinality switching (DYN3), and bursts (DYN4). (b) Mean win-rate gain (pp, upper line) and support F1 (lower line) for the persistent-support families F0–F5 defined above. Each setting uses three source policies and one recovery checkpoint; family cells average $s2/s4$.

Recovery under temporal regime changes. SIR-MARL improves mean win rate over corrupted MAPPO in all nine map/regime combinations. Bursts yield gains of 26.6, 12.0, and 27.1 percentage points on `8m`, `3s5z`, and `2s3z`. Support and cardinality switching also give positive mean gains. These comparisons support adapting correction to changing fault coverage with fixed recovery parameters.

Fault structure and recovery. Persistent-support F0 yields positive mean gains on all three maps. On `8m`, F0 and mean replacement (F1) produce substantial gains at both cardinalities, although recovered win rates remain below the clean-source reference. The smaller gains for other families occur alongside substantial clean-to-corrupted performance losses: F2/F3 retain more source performance than F0/F1 at $s2$, whereas stale observations (F4) and nominal-pool replacement (F5) can severely degrade the policy with little recovery. On `8m`, both corrupted and recovered win rates are zero for F4/F5 at $s4$, compared with a 90.62% clean-source reference.

Recovery also varies with fault cardinality. On `2s3z`, F0 recovery reaches 27.08% at $s2$ but 0.52% at $s4$, which corrupts four of five agents and exceeds the clean-majority regime. For F1 on `3s5z`, recovery improves the $s2$ mean but slightly lowers the $s4$ mean. The absolute results and support F1 jointly characterize task- and family-dependent transfer.

Table 28: Absolute win rates (%) for the selected stress-test source policies. Entries are mean±sample SD across three sources, with one recovery checkpoint and 64 episodes per source/method/condition. Clean rows use separate fault-free evaluations on the same source set. Family rows use the intensities in Figure 3, and report $s2$ and $s4$ separately.

Map	Clean source	Clean with SIR-MARL
8m	90.62 ± 12.20	85.94 ± 17.61
3s5z	77.60 ± 19.78	77.60 ± 19.78
2s3z	83.85 ± 14.85	83.85 ± 14.85

		$s2$		$s4$	
Map	Family	Corrupted	Recovered	Corrupted	Recovered
8m	F0	11.98 ± 5.92	50.00 ± 24.56	2.08 ± 0.90	29.17 ± 4.51
	F1	9.90 ± 11.09	53.65 ± 19.66	7.29 ± 12.63	29.17 ± 15.01
	F2	38.54 ± 12.73	44.27 ± 10.97	6.25 ± 3.12	5.73 ± 3.61
	F3	38.54 ± 13.29	35.94 ± 8.27	18.23 ± 9.92	20.31 ± 10.94
	F4	13.02 ± 6.31	11.46 ± 6.31	0.00 ± 0.00	0.00 ± 0.00
	F5	4.69 ± 4.69	6.25 ± 3.12	0.00 ± 0.00	0.00 ± 0.00
3s5z	F0	11.98 ± 11.30	19.27 ± 9.92	2.08 ± 3.61	4.69 ± 4.69
	F1	13.54 ± 20.81	21.88 ± 19.95	5.21 ± 9.02	3.65 ± 3.25
	F2	63.02 ± 29.08	56.25 ± 26.01	41.15 ± 40.21	46.35 ± 31.88
	F3	52.60 ± 26.58	46.88 ± 24.26	30.73 ± 30.55	26.56 ± 27.24
	F4	29.69 ± 14.32	31.25 ± 11.80	7.29 ± 7.38	9.38 ± 8.27
	F5	7.29 ± 5.92	3.65 ± 3.93	0.00 ± 0.00	0.00 ± 0.00
2s3z	F0	6.25 ± 4.69	27.08 ± 13.29	0.00 ± 0.00	0.52 ± 0.90
	F1	19.79 ± 8.61	11.98 ± 5.49	0.00 ± 0.00	0.00 ± 0.00
	F2	51.04 ± 8.61	53.65 ± 7.71	22.40 ± 3.93	19.27 ± 6.31
	F3	36.98 ± 11.73	42.71 ± 6.51	21.88 ± 5.63	21.35 ± 9.55
	F4	31.77 ± 14.77	30.21 ± 14.52	3.12 ± 2.71	1.56 ± 1.56
	F5	7.81 ± 8.70	8.33 ± 7.38	0.00 ± 0.00	0.52 ± 0.90

G Additional Comparisons and Paired Results

G.1 Policy-training baselines

EIR and MIR3 use policy training, while SIR-MARL uses clean-only recovery on a frozen source. The following table repeats their full fixed-cardinality outcomes from Table 1 for the same five-policy benchmark. All values are percentages. Appendix C.2 specifies the corresponding MAPPO implementations and comparison protocols.

Table 29: Additional policy baselines, mean and standard deviation over five policy seeds.

Map	Method	Clean	$s1$	$s2$	$s3$	$s4$
8m	EIR	59.06 ± 15.44	21.56 ± 12.71	5.31 ± 5.25	0.63 ± 1.40	0.00 ± 0.00
8m	MIR3	82.19 ± 24.20	43.44 ± 26.32	10.63 ± 9.40	3.75 ± 4.50	0.31 ± 0.70
3s5z	EIR	70.94 ± 19.19	25.00 ± 15.15	8.75 ± 6.50	2.19 ± 2.37	0.31 ± 0.70
3s5z	MIR3	96.25 ± 3.92	66.56 ± 20.69	54.38 ± 13.68	24.69 ± 16.18	3.44 ± 3.01
2s3z	EIR	92.50 ± 4.87	48.13 ± 9.14	6.88 ± 7.04	0.94 ± 2.10	0.00 ± 0.00
2s3z	MIR3	83.44 ± 6.01	53.44 ± 10.09	9.38 ± 4.56	0.31 ± 0.70	0.00 ± 0.00

G.2 One-stage neural recovery comparison

SIR-Direct is a separately constructed one-stage raw-context belief network. It estimates a clean candidate from the available input, then uses residual-based support inference and blending to recover observations. It shares SIR-MARL's frozen source policies, clean-only training interface, and unknown-support/unknown-cardinality protocol. Its architecture produces recovery in one stage; SIR-MARL feeds the initial correction into a second stage with raw and corrected LOO contexts.

SIR-Direct and SIR-MARL use the same frozen source actors for decision supervision. This comparison evaluates the alternative recovery architectures. The *w/o second stage* ablation in Table 2 instead retains SIR-MARL's first-stage candidate and support for recovery and retrains that variant. The two comparisons therefore evaluate architecture choice and module removal, respectively.

Table 30: One-stage neural recovery comparison on SMAC (win rate, %). Five recovery seeds are averaged within each of five source-policy seeds; entries report mean±std over the policy aggregates. Both methods use the same benchmark fault protocol.

Method	Clean	$s1$	$s2$	$s3$	$s4$
`8m`					
SIR-Direct	83.44 ± 10.76	72.38 ± 13.74	57.25 ± 21.91	42.00 ± 22.56	23.75 ± 14.66
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
`3s5z`					
SIR-Direct	82.25 ± 12.69	76.00 ± 14.96	65.19 ± 19.45	53.44 ± 18.87	31.50 ± 14.35
SIR-MARL	82.19 ± 12.33	73.44 ± 14.53	68.25 ± 18.88	56.06 ± 23.63	36.13 ± 25.12
`2s3z`					
SIR-Direct	83.50 ± 12.95	80.25 ± 15.28	69.50 ± 21.66	17.38 ± 7.37	0.56 ± 0.56
SIR-MARL	83.75 ± 12.01	79.75 ± 16.48	69.50 ± 24.08	13.19 ± 3.56	0.38 ± 0.34

SIR-MARL has higher mean win rate over $s1$–$s4$ than SIR-Direct by 7.50 percentage points on `8m` and 1.94 points on `3s5z`. On `2s3z`, SIR-Direct is higher by 1.22 points. Both architectures improve mean win rate averaged over $s1$–$s4$ relative to the unprotected source, with task-dependent differences between them. Appendix G.4 reports source-paired differences and source-normalized performance.

G.3 Observation-augmentation policy-training control

SparseAug-Mix trains a recurrent MAPPO policy with Gaussian augmentation of sampled agents' local observations during rollout collection. Its augmentation routine scales noise using the selected

local observation. This supplementary comparison evaluates policy adaptation with observation-noise exposure during training. SIR-MARL uses the frozen nominal source policy and trains observation recovery on nominal traces with generic masking. Both are evaluated under the team-level benchmark fault protocol in Appendix C.1.2; model-selection rules are specified in Appendix C.2.

Table 31: Observation-augmentation policy-training control on SMAC (win rate, %). Entries are mean±std over five policy seeds. For SIR-MARL, five recovery seeds are first averaged within each source-policy seed. Evaluation uses the same per-step group-sparse Gaussian fault protocol as Table 1.

Method	Clean	$s1$	$s2$	$s3$	$s4$
8m					
SparseAug-Mix	77.19±14.09	72.50±13.51	74.38±13.05	75.63±12.91	60.31±9.67
SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
3s5z					
SparseAug-Mix	70.31±24.83	74.06±19.91	68.44±21.06	66.25±13.60	54.06±18.07
SIR-MARL	82.19 ± 12.33	73.44 ± 14.53	68.25 ± 18.88	56.06 ± 23.63	36.13 ± 25.12
2s3z					
SparseAug-Mix	87.19±12.32	84.69±11.34	81.56±14.12	72.50±15.05	45.94±11.51
SIR-MARL	83.75 ± 12.01	79.75 ± 16.48	69.50 ± 24.08	13.19 ± 3.56	0.38 ± 0.34

Table 31 shows higher mean win rates for SparseAug-Mix at $s3$ and $s4$ on all three maps. These results compare policy adaptation through observation augmentation with post-training recovery of a frozen nominal policy under extensive fault coverage.

G.4 Source-normalized performance and paired effects

Policy-training baselines retain the recurrent actor–critic implementations and method-specific objectives described in Appendix C.2. Their absolute win rates reflect both clean checkpoint quality and fault response. SIR-MARL, SIR-Direct, and PCA-SIR share the same frozen source policies and clean-only recovery interface.

Source-paired differences. For each source policy, we first average its recovery seeds, then average the selected severities, and subtract the corresponding baseline score. The table reports mean differences across the five sources and percentile intervals from all $5^5 = 3,125$ resamples of these source-level pairs. These exploratory percentile intervals describe source-level resampling conditional on the fitted recovery models and the observed source-policy results. Recovery models share pooled training traces, and the source policies have different training and validation roles; the intervals do not represent uncertainty from retraining on independent source-policy populations. The resampling and pairing unit is the source checkpoint; episodes are aggregated within their respective evaluation batches. The source-only reference uses the evaluation batches in Appendix G.5.

Normalized degradation and gap recovery. For policy-training methods we report $\mathrm{ND}_s = (R_0 - R_s)/R_0$ using each method's own clean score R_0. This expresses degradation relative to each method's initial performance; source policy quality remains method-dependent. For recovery methods sharing a frozen actor, GR_s uses the corresponding MAPPO clean and corrupted scores as in Appendix C. Below we average the four severity-specific ratios, each formed from policy-aggregate means. The order of aggregation is source averaging, severity-specific normalization, then severity averaging. Values are unclipped, and a negative degradation means the faulted score exceeded the clean score in that evaluation.

Table 32: SIR-MARL minus baseline. Interior uses $s1$–$s3$ for eight-agent maps and $s1$–$s2$ for 2s3z; All uses $s1$–$s4$.

Map	Baseline	Regime	Mean gain (pp)	95% source CI
8m	Corrupted	Interior	+44.17	[+31.69, +56.65]
8m	Corrupted	All	+41.73	[+29.69, +54.22]
8m	SIR-Direct	Interior	+6.13	[+0.52, +12.65]
8m	SIR-Direct	All	+7.50	[+3.63, +12.02]
3s5z	Corrupted	Interior	+38.21	[+26.90, +47.71]
3s5z	Corrupted	All	+36.91	[+25.41, +47.78]
3s5z	SIR-Direct	Interior	+1.04	[-0.69, +2.85]
3s5z	SIR-Direct	All	+1.94	[-1.28, +5.84]
2s3z	Corrupted	Interior	+42.59	[+28.91, +52.63]
2s3z	Corrupted	All	+23.91	[+16.36, +29.05]
2s3z	SIR-Direct	Interior	-0.25	[-2.22, +2.03]
2s3z	SIR-Direct	All	-1.22	[-2.11, -0.17]

Table 33: Source-normalized performance. Lower ND and higher GR indicate smaller source-relative degradation and larger recovery, respectively. ND describes within-method degradation; GR describes recovery of the shared source-policy gap.

Map	Method	Metric	Mean $s1$–$s4$ (%)
8m	MAPPO	ND	82.69
8m	ERNIE	ND	39.19
8m	ROMANCE	ND	75.08
8m	WALL	ND	82.36
8m	PCA-SIR	GR	-11.71
8m	SIR-Direct	GR	52.29
8m	SIR-MARL	GR	62.29
3s5z	MAPPO	ND	73.86
3s5z	ERNIE	ND	39.14
3s5z	ROMANCE	ND	66.58
3s5z	WALL	ND	67.79
3s5z	PCA-SIR	GR	-5.97
3s5z	SIR-Direct	GR	61.64
3s5z	SIR-MARL	GR	63.35
2s3z	MAPPO	ND	79.64
2s3z	ERNIE	ND	65.03
2s3z	ROMANCE	ND	78.66
2s3z	WALL	ND	75.90
2s3z	PCA-SIR	GR	-23.49
2s3z	SIR-Direct	GR	48.08
2s3z	SIR-MARL	GR	46.36

G.5 NOMINAL-RANGE CLIPPING

For each map, nominal-range clipping computes the empirical minimum and maximum of every agent-row observation coordinate from the pooled nominal training traces of source policies 1–4. Source policy 5 is excluded from fitting the bounds. At deployment, each incoming observation coordinate is clipped to its fitted interval. The physical SMAC agent-row and feature ordering are retained; available-action masks and recurrent states are handled separately by the frozen actor. The transformation uses the incoming observation and fixed bounds. The intervals describe empirical nominal ranges rather than environment-wide legal limits. Exactly constant coordinates have identical lower and upper bounds: $152/1632$ on `8m`, $232/2016$ on `3s5z`, and $100/720$ on `2s3z`.

Source, clipping, and SIR-MARL use the same five frozen source actors, deterministic action selection, and per-step resampled, fixed-scale Gaussian fault protocol. Evaluation uses one environment, environment seed $50{,}000q$, fault seed $700{,}000 + q$, and 64 completed episodes per source and condition, where q is the source-policy seed. SIR-MARL averages five recovery seeds within each source before source-level aggregation. Table 34 reports the same Source and SIR-MARL results as Table 1. All win rates are computed from integer wins divided by completed episodes before aggregation and formatting.

Table 34: Nominal-range clipping under the benchmark fault protocol (win rate, %). Entries are mean±sample SD across five source policies; SIR-MARL first averages five recovery seeds within each source.

Map	Method	Clean	$s1$	$s2$	$s3$	$s4$
8m	Source	84.38 ± 11.43	42.50 ± 4.34	12.19 ± 6.09	2.81 ± 4.74	0.94 ± 1.40
	Clipping	83.44 ± 11.46	20.94 ± 9.80	2.19 ± 3.42	0.31 ± 0.70	0.00 ± 0.00
	SIR-MARL	83.25 ± 10.59	74.38 ± 12.65	64.88 ± 14.35	50.75 ± 19.00	35.38 ± 17.41
3s5z	Source	82.50 ± 13.32	49.69 ± 16.44	24.06 ± 13.83	9.38 ± 10.07	3.13 ± 6.99
	Clipping	85.63 ± 9.72	55.00 ± 24.22	24.69 ± 26.41	12.81 ± 18.13	6.25 ± 10.83
	SIR-MARL	82.19 ± 12.33	73.44 ± 14.53	68.25 ± 18.88	56.06 ± 23.63	36.13 ± 25.12
2s3z	Source	82.50 ± 13.68	49.38 ± 15.13	14.69 ± 5.13	1.56 ± 1.10	1.56 ± 2.71
	Clipping	84.06 ± 11.02	59.69 ± 14.88	18.44 ± 9.91	3.13 ± 2.21	0.63 ± 1.40
	SIR-MARL	83.75 ± 12.01	79.75 ± 16.48	69.50 ± 24.08	13.19 ± 3.56	0.38 ± 0.34

Clipping preserves near-source clean performance but has task-dependent effects under faults. It lowers faulted win rates on `8m` and improves several conditions on the heterogeneous maps. SIR-MARL attains higher mean win rates than clipping in eleven of the twelve map–fault-count conditions. The exception is `2s3z` at $s4$, where all three methods have near-zero win rates. These results show that empirical range clipping alone does not reproduce the recovery gains under the evaluated faults.