LEARNING TO PROPOSE JOINT SENSING ACTIONS UNDER SHARED UNCERTAINTY

Anonymous authors
Paper under double-blind review

ABSTRACT

A sensing action can be valuable because of what it enables another action to reveal. Under shared receiver uncertainty, calibration can make a subsequent target observation informative even when calibration alone leaves target uncertainty unchanged. We study learning to propose executable sensing groups for a budget-limited physics-based evaluator. A two-slot Gaussian construction shows that exact singleton screening and sequential marginal greedy selection can recover an arbitrarily small fraction of the optimal variance reduction. An explicit calibration–repetition boundary identifies when resolving receiver ambiguity is more valuable than averaging extraction noise. Our recurrent proposer combines a joint target–receiver belief, learned control generation, executable projection, and group-cost prediction, with forecast feedback supporting within-decision refinement. On 192 simulated test episodes across eight scenarios, averaged over three training seeds, the full relational variant reduces mean GOSPA by 15.0% and RF energy by 7.0% relative to numerical search under matched forecast and construction budgets. Target recall improves by 4.37 percentage points, with both tracking metrics improving in every scenario mean. Fixed-checkpoint ablations identify control generation as the largest contributor: learning changes the executable choices available to the evaluator, as well as their ordering.

1 INTRODUCTION

A radar network allocates time to finding targets, maintaining tracks, and calibrating receivers. Each task also requires a choice of pointing direction, dwell time, and transmit power. These decisions determine the measurements acquired in the current cycle and the information available for subsequent tracking. The scheduler must therefore search over both discrete assignments and continuous controls, although it can evaluate only a small number of candidates before acting.

A target observation can mix position error with an unknown receiver delay. Repeated observations reduce measurement noise but do not resolve this ambiguity. Observing a surveyed calibration source helps distinguish the two errors. Calibration may thus improve a subsequent target estimate even when it provides no information about the target on its own. Screening actions by their individual gains can discard calibration before its value in a joint plan is considered.

This raises a question about the role of learning in sensing: should a model rank a fixed set of actions, or should it also generate their controls? Ranking directs computation toward promising candidates. Generating power, dwell, pointing, and timing changes the measurements those candidates can produce. We compare these two uses of learning in radar tracking with shared receiver errors, intermittent emissions, and delayed measurements.

Our method proposes sensing groups for evaluation with a shared measurement model. A recurrent encoder takes the current target–receiver belief and resource state, generates signal controls, and predicts group costs after the controls have been projected onto device and scheduling constraints. Candidate groups include calibration–tracking pairs and observations of the same target from different sensors, together with analytic alternatives. A two-cycle forecast evaluates these groups and selects one to execute. The second cycle allows the scheduler to account for calibration through its effect on a later target observation.

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.

We derive a condition under which calibration is preferable to another target observation and construct a case in which selection by exact individual gains recovers an arbitrarily small fraction of the optimal variance reduction. We then compare learned proposals under the same construction and evaluation budgets, using three encoders within a common scheduler. Across eight simulated scenarios, the full proposal procedure improves aggregate tracking for all three encoders. The relational model obtains the lowest mean GOSPA, a 15.0% reduction relative to numerical search. Switching control generation and feedback independently at inference attributes most of this improvement to control generation.

2 WHY CALIBRATION CAN CHANGE THE BEST SENSING PLAN

A target measurement mixes two unknowns. Consider a scalar target-position error x and a receiver bias b, expressed in the same distance coordinate. A receiver delay can be converted to this coordinate through propagation speed. Assume that x and b remain fixed over two observation slots. A target acquisition gives $y = x + b + \epsilon$, whereas a surveyed calibration source gives $z = b + \nu$. The four variables x, b, ϵ, ν are independent, zero-mean Gaussians with positive variances q, r, s, v, respectively. Here ϵ is target-measurement noise and ν is calibration noise.

Calibration reduces the bias variance from r to $r_z = rv/(r+v)$. Gaussian conditioning then gives

$$\mathrm{Var}(x \mid z) = q, \qquad \mathrm{Cov}(x, b \mid y) = -\frac{qr}{q+r+s}, \tag{1}$$

$$\mathrm{Var}(x \mid y) = \frac{q(r+s)}{q+r+s}, \qquad \mathrm{Var}(x \mid y, z) = \frac{q(r_z+s)}{q+r_z+s}. \tag{2}$$

Calibration alone leaves the target variance unchanged because the prior makes x and b independent. A target observation introduces the negative cross-covariance in Equation (1): a larger target error can be offset by a smaller receiver bias without changing their measured sum. Calibration then reduces uncertainty in both variables. During tracking, earlier measurements may already have introduced this dependence, allowing calibration to improve the current target estimate.

Calibration versus repetition. Suppose each acquisition uses one slot, with fixed measurement precision and equal resource cost. Two target acquisitions share the bias but have independent measurement noise. Their average has effective noise variance $r + s/2$. Calibration followed by a target acquisition instead has effective noise variance $r_z + s$. Since the posterior variance $q\eta/(q+\eta)$ increases with effective noise η, calibration gives the lower terminal target variance exactly when

$$\frac{r^2}{r+v} > \frac{s}{2}. \tag{3}$$

The left-hand side is the bias variance removed by calibration; the right-hand side is the measurement-noise variance removed by repetition. Their comparison determines which action gives the more precise target estimate. Appendix A.4 gives the derivation and limiting cases.

Accurate individual scores can still select the wrong pair. For a plan S drawn from a fixed observation family, let Y_S denote its observations and define the expected reduction in target squared error as

$$F(S) = \mathrm{tr}\,\mathrm{Cov}(x) - \mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(x \mid Y_S). \tag{4}$$

For scalar x, each trace reduces to a variance. The following construction shows why evaluating individual acquisitions accurately does not solve the combination problem.

Proposition 1 (Failure of individual-gain selection)**.** *Let $\ell > 0$ be a distance scale and $\rho \geq 2$ dimensionless. The available acquisitions are $y = x + b + \epsilon$, $z = b + \nu$, and $w = x + \eta$, with independent zero-mean Gaussian variables satisfying*

$$\mathrm{Var}(x) = \ell^2, \quad \mathrm{Var}(b) = \mathrm{Var}(\eta) = \rho\ell^2, \quad \mathrm{Var}(\epsilon) = \mathrm{Var}(\nu) = \ell^2/\rho.$$

Each acquisition can be selected once and consumes one of two slots. Selecting the two largest singleton gains, or choosing greedily by the exact conditional gain at each step, gives $\{w, y\}$. The optimal pair is $\{y, z\}$, and

$$\frac{F(\{y,w\})}{F(\{y,z\})} = \frac{2\rho^2+1}{\rho(\rho^2+1)} \longrightarrow 0 \quad \textit{as } \rho \to \infty. \tag{5}$$

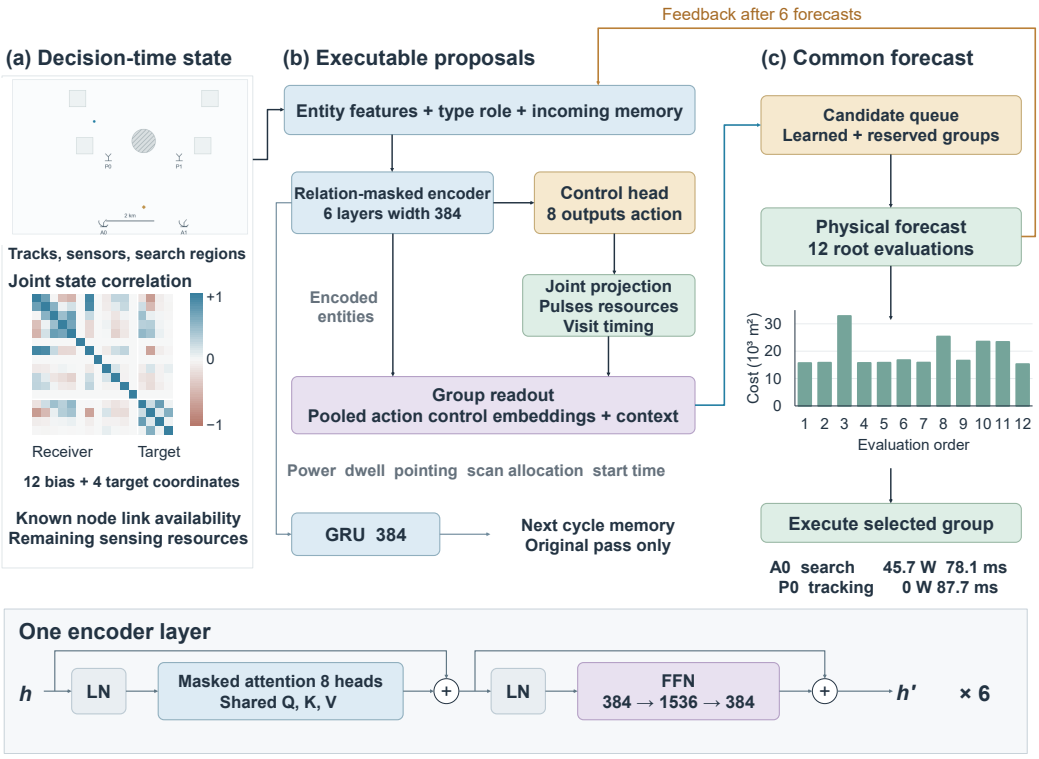


Figure 1: **Learned proposals within the radar scheduler.** (a) The current tracks, receiver uncertainty, and available resources form the input. (b) The recurrent encoder predicts eight controls per action and costs for projected action groups; the lower panel expands one of its six relation-masked blocks. (c) A physical forecast compares at most twelve groups and selects one to execute. After six evaluations, forecast feedback is used to refine the remaining proposals without advancing recurrent memory. The input map and correlation matrix are from the decision-time belief at 30 s; the bars give its twelve recorded forecast costs in evaluation order. LN, FFN, and GRU denote layer normalization, a feed-forward network, and a gated recurrent unit.

Observation w has the largest individual gain, but its large independent noise cannot be reduced by calibration. In contrast, much of the error in y comes from a shared bias that z can estimate precisely. Because z has zero gain both initially and after observing w, both selection rules exclude the optimal pair. Appendix A.5 compares all three pairs. Greedy methods admit information-gain guarantees under suitable conditional-independence assumptions (Krause & Guestrin, 2005). In this construction, shared receiver bias makes calibration and target acquisition complementary. Retaining such pairs allows their joint value to be evaluated before either action is discarded.

Shared errors also create redundancy. Repeated observations from one reference capture retain its common error, and repeated passive visits within a silent emission cycle cannot produce a target return. The forecast accounts for both common measurement errors and shared emission events, as derived in Appendix A.

3 Learning to Propose Radar Sensing Actions

At each cycle, the scheduler constructs candidate sensing groups, evaluates a subset with the two-cycle forecast, and executes the group with the lowest eligible cost (Figure 1). The model generates controls and prioritizes groups for evaluation; the forecast determines the final choice.

3.1 STATE, ACTIONS, AND THE PHYSICAL OBJECTIVE

At time t, the scheduler uses a belief β_t over receiver offsets b_t and tracked target states X_t, together with search history, node availability, and remaining resources. Its Gaussian moment representation retains the joint covariance

$$P_t = \begin{bmatrix} P_{bb} & P_{bX} \\ P_{Xb} & P_{XX} \end{bmatrix}, \tag{6}$$

where the off-diagonal blocks represent target–receiver dependence. The filter updates this belief using measurements that arrive within the acquisition cycle, accounting for their timestamps and shared error sources. When a measurement has several possible track assignments, the filter combines the association hypotheses by moment matching. The observation and inference models are specified in Appendix B.

A primitive action assigns a sensor to search, tracking, or calibration. Its signal role determines the measurement. Monostatic sensing provides range, bearing, and radial velocity. Direct passive sensing provides differences in arrival time and frequency between receivers, together with angle of arrival. Bistatic sensing uses excess delay, Doppler, and arrival angle relative to an illumination reference. A group A contains up to two compatible primitives executed in the same cycle. Sensors cannot be assigned twice, and a direct passive action reserves both participating receivers. A calibration followed by sensing on the same receiver occupies successive cycles, as in the two-slot example.

To compare groups, the forecast combines localization uncertainty, incomplete search, and track-maintenance costs:

$$C(\beta) = \sum_i \operatorname{tr} P_{pp}^i + \alpha_{\text{search}} \sum_c \lambda_c + \alpha_{\text{miss}} \sum_i p_{\text{miss},i} + \alpha_{\text{tent}} n_{\text{tent}}. \tag{7}$$

Here P_{pp}^i is track i's position covariance, λ_c is the remaining search-priority weight in region c, $p_{\text{miss},i}$ is the predicted miss probability for target i, and n_{tent} counts unconfirmed tracks. The search weights increase with time and decrease with predicted discovery. The three coefficients convert the dimensionless terms to squared-distance costs; their values are given in Appendix B.

For group A, let $\widehat{\beta}_A$ be the forecast belief after its observations and $\widehat{\beta}_A^-$ the prediction to the next cycle. We compare

$$\widehat{J}(\beta, A) = \tfrac{1}{2}\left[C(\widehat{\beta}_A) + \min_{A' \in \mathcal{T}(\widehat{\beta}_A^-)} C\big(\widehat{F}(\widehat{\beta}_A^-, A')\big) \right]. \tag{8}$$

The continuation set $\mathcal{T}$ contains waiting, up to two prioritized analytic actions, and their feasible pair. The map $\widehat{F}$ enumerates emission, delivery, reference-availability, and detection outcomes. Each measurement update uses the predicted observation, giving zero innovation, and the branch posteriors are combined by Gaussian moment matching. Between cycles, the belief is propagated through the motion and receiver-drift models. Evaluating the continuation accounts for the effect of calibration on a subsequent target observation.

3.2 GENERATING CONTROLS AND PREDICTING GROUP COST

The encoder represents sensors, receiver biases, tracks, search regions, resources, primitive actions, and covariance blocks as entities. Relations connect an action to the receiver and state it observes, and connect a covariance block to its two variables. To see why these links matter, consider a linear calibration $z_c = H_c b + \nu_c$, where H_c maps receiver offsets to measurements and independent noise ν_c has covariance R_c. In a joint Gaussian belief, the reduction in target covariance is

$$\Delta P_{XX} = P_{Xb} H_c^\top (H_c P_{bb} H_c^\top + R_c)^{-1} H_c P_{bX}. \tag{9}$$

The target–receiver cross-covariance determines how a calibration update affects the target estimate. We therefore encode each covariance block together with its endpoint variables. Six attention blocks process the entities, and a gated recurrent unit maintains an entity-specific memory across cycles.

For each primitive, a control head predicts peak power, dwell, pointing offset, scan half-span, three scan weights, and requested start time. These eight outputs are normalized to device ranges. Projection rounds them to admissible values, allocates pulses to scan directions, and resolves scheduling conflicts; infeasible groups are rejected. The controls change both measurement quality and

availability. More dwell can improve precision but delay delivery, while distributing dwell across directions trades per-direction precision for search coverage. For N pulse intervals of duration T_r and pulse width τ_p, the receive dwell is NT_r, whereas transmit time is $N\tau_p$. Thus RF energy is $P_{\text{peak}} N\tau_p$, where P_{peak} denotes peak transmit power.

The group head predicts costs from the controls after projection, so it scores the values that would actually be executed. Let h_a be action a's embedding, $\tilde{u}_a$ its projected normalized controls, and m_a a binary mask identifying applicable control coordinates. With learned parameters θ, concatenation denoted by brackets, and elementwise multiplication by $\odot$, the readout is

$$z_a = \phi_\theta([h_a, m_a \odot \tilde{u}_a, m_a]), \tag{10}$$

$$v_\theta(\beta, A) = \text{softplus}\left(\rho_\theta\left[\sum_{a\in A} z_a, \frac{\sum_{a\in A} z_a \odot z_a}{\max(1, |A|)}, \bar{h}, |A|/2\right]\right). \tag{11}$$

The functions ϕ_θ and ρ_θ are feed-forward networks; $\bar{h}$ is the mean embedding of valid entities, and $|A|$ is group size. Pooling the embeddings and their second moments allows the head to predict interactions between actions. Its two outputs estimate the log-transformed forecast cost and the subsequent tracking cost along a teacher trajectory. These estimates determine the order of learned candidates in the evaluation queue (Appendix D).

3.3 Candidate construction and online selection

The initial pool contains waiting, individual analytic actions, and selected pairs. Construction first attempts to project up to six calibration–tracking or same-object pairs and then fills the pool to at most forty groups. The model generates state-dependent controls for its preferred primitives. Each method has a budget of 192 candidate-construction and projection attempts, including rejected attempts, and twelve forecast evaluations per cycle. Each evaluation includes a current-cycle group and its continuations in Equation (8); we refer to this as a root evaluation.

The queue gives priority to overdue search, a reserved analytic candidate, and due calibration before considering learned proposals and remaining alternatives. After six evaluations, the model receives feedback containing the best costs found and the reasons for rejecting candidates. It uses this feedback to refine the remaining proposals, with the recurrent memory held at its value at the start of the cycle. Failed evaluations also count toward the budget. Among the evaluated groups eligible under the search-revisit and resource rules, the scheduler executes the one with the lowest forecast cost. Algorithm 1 gives the full procedure.

Control generation and group-cost prediction address different parts of the search. The former changes the available measurements; the latter allocates evaluations among proposed groups. Appendix A.8 formalizes this distinction by decomposing forecast suboptimality into candidate-coverage and evaluation-allocation terms. Our experiments compare ranking with the full proposal procedure and then vary control generation and feedback separately.

3.4 Learning from executed trajectories

Training uses two teachers with common constraints and forecast budgets: one with prescribed analytic controls, and one with scrambled Halton proposals and local repair. Each trajectory records the belief, evaluated groups and their costs, and the group actually executed. Evaluated costs supervise group ranking and cost prediction. Executed, post-projection controls supervise the control head. The discounted mean squared GOSPA along the teacher trajectory provides targets for the state and executed group over the remainder of the episode. GOSPA measures multi-target tracking error and is defined in Section 4.

The training objective is

$$\mathcal{L} = D_{\text{KL}}(q\|p) + \mathcal{L}_{\text{group}} + \mathcal{L}_{\text{control}} + \tfrac{1}{4}(\mathcal{L}_{\text{atom}} + \mathcal{L}_{\text{state}} + \mathcal{L}_{\text{future}}). \tag{12}$$

Here q assigns greater probability to groups with lower evaluated costs, and p is the distribution predicted by the model. The group and control losses regress log-transformed physical costs and executed controls, respectively. The auxiliary primitive loss predicts the best recorded cost of a group containing each primitive; the state and future losses predict the teacher's subsequent tracking

cost from the state and executed group. The regression terms use Huber loss over valid targets and applicable controls. Appendix D gives the target transforms and averaging rules. Training averages the losses from the initial and feedback passes, which use the same targets and memory input. The memory produced by the initial pass is carried to the next cycle.

4 EXPERIMENTS

We evaluate tracking accuracy and resource consumption, then isolate the effects of control generation and forecast feedback.

Setup. We use a planar measurement-level radar simulator with four sensors and up to three targets. Eight scenarios vary range, emission, illumination, target arrivals, motion, node outages, and transport. Training comprises 12,800 episodes of 64 one-second cycles, balanced across the two teachers and eight scenarios; validation uses 128 disjoint episodes. Each encoder is trained from three initializations for ten epochs (32,000 updates). For each run, we select the checkpoint with the lowest validation loss and evaluate it on 192 new episodes, with 24 per scenario. All methods share the same test-world realizations, drawn from the eight scenarios used in training. Appendices B and D give the simulator and training settings.

Comparisons. The analytic baseline uses prescribed controls and priorities. Numerical search uses twenty scrambled Halton proposals with local repair. Deep Sets, dense Self-attention, and relation-masked attention are compared as encoders within the same recurrent proposer, with common inputs, output heads, projection, and physical evaluation. The attention models each have 13.57 million parameters; Deep Sets has 11.79 million. *Rank* uses network priorities with analytic controls and no feedback. *Full* also generates controls and uses forecast feedback. The relational ablation also tests controls without feedback and feedback without controls, using the same selected checkpoints. Every method uses the construction budget and twelve forecast evaluations per cycle defined in Section 3.

Metrics. For true positions X and confirmed-track positions $\widehat{X}$, generalized optimal sub-pattern assignment (GOSPA) (Rahmathullah et al., 2017) combines localization error, missed targets, and false tracks:

$$d_G = \left[\min_{\Gamma} \left\{ \sum_{(i,j)\in\Gamma} \|x_i - \hat{x}_j\|^2 + \tfrac{c^2}{2}(|X| + |\widehat{X}| - 2|\Gamma|) \right\} \right]^{1/2}. \tag{13}$$

Here Γ is a partial one-to-one matching and $c = 250\,\mathrm{m}$. Each unmatched target or track contributes $c^2/2$ to the squared distance. We average GOSPA over each episode. Recall is the fraction of true target-time pairs matched to confirmed tracks; position root mean square error (RMSE) is computed over these matches. For learned methods, we first average the three training runs within each test episode and then average across scenarios with equal weights. Appendix F defines resource metrics and the episodes included in each estimate; Appendix G reports individual runs and paired bootstrap intervals.

Learned proposals improve target coverage. The relational Full model reduces mean GOSPA from 67.91 to 57.74 m relative to numerical search, a 15.0% improvement, and increases recall from 76.45% to 80.82% (Table 1). Both metrics improve in every scenario mean. Full also improves on Rank for all three encoders: GOSPA decreases by 11.07 m for Deep Sets, 10.86 m for Self-attention, and 12.95 m for the relational encoder. All three Rank variants, which use analytic controls, have higher GOSPA than numerical search.

The results show improved target coverage with little change in matched-position accuracy. Position RMSE changes from 17.19 to 16.49 m, and the paired interval for the difference includes zero (Table A6). Figure 2 illustrates how coverage changes during one transport/occlusion episode. The relational model confirms the first target by 6 s, before the other two displayed methods, but matches only two of three targets at 44 s, when both alternatives match all three. The curves show these changes over the full episode. Appendix H compares four methods at common times in each scenario and includes the analytic baseline in the episode curves.

Table 1: **Closed-loop tracking across eight sensing scenarios.** Results average 24 test episodes per scenario and three training seeds for learned methods.

Method		Estimation accuracy		Target tracking	
Proposer	Variant	GOSPA m ↓	Position RMSE m ↓	Target recall % ↑	False tracks per cycle ↓
Analytic baseline	—	74.63	18.09	74.37	0.0127
Numerical search	—	67.91	17.19	76.45	**0.0050**
Deep Sets	Rank	71.53	17.95	75.06	0.0103
Deep Sets	Full	60.46	17.05	79.97	**0.0050**
Self-attention	Rank	70.37	18.22	75.62	0.0081
Self-attention	Full	59.51	16.57	80.10	0.0080
Relational (ours)	Rank	70.69	17.79	75.41	0.0090
Relational (ours)	Full	**57.74**	**16.49**	**80.82**	0.0079

Rank uses analytic controls with feedback disabled; Full enables learned controls and forecast feedback. Bold marks the best value at the reported precision. Position errors use matched tracks; Table A5 lists sample counts. Appendix G reports seed variation and confidence intervals.

Tracking improves with lower RF expenditure. The relational Full model uses 2.493 J of RF energy per episode, compared with 2.681 J for numerical search, and reduces transmitter on-time from 54.93 to 51.51 ms (Table 2). Incident energy density at the fixed observation point falls from 1.892 to 1.773 nJ/m^2. Receiver energy remains similar, at 90.87 versus 90.25 J. The RF savings therefore occur with similar receiver energy consumption. Among the three Full models, Self-attention uses the least RF energy and the relational model has the lowest mean GOSPA.

Table 2: **Energy consumption and RF transmission** over 64-s test episodes.

Method		Energy consumption		Transmission	
Proposer	Variant	RF energy J ↓	Receiver energy J ↓	Tx on-time ms ↓	Incident energy density nJ/m^2 ↓
Analytic baseline	—	2.708	74.10	58.81	1.925
Numerical search	—	2.681	90.25	54.93	1.892
Deep Sets	Rank	2.682	72.64	58.24	1.908
Deep Sets	Full	2.439	88.39	51.48	1.744
Self-attention	Rank	2.683	72.51	58.25	1.904
Self-attention	Full	**2.352**	91.02	**49.21**	**1.686**
Relational (ours)	Rank	2.682	**72.50**	58.23	1.906
Relational (ours)	Full	2.493	90.87	51.51	1.773

Tx on-time sums pulse durations across nodes. Incident energy density is evaluated at a fixed observation point using isotropic propagation. Bold marks the lowest value at the reported precision.

Control generation accounts for the main ablation effect. With feedback disabled, enabling learned controls reduces GOSPA from 70.69 to 59.02 m and increases recall from 75.41% to 80.65% (Table 3). Adding feedback further lowers mean GOSPA to 57.74 m. The paired change is -1.28 m (95% interval $[-4.15, 1.45]$ m). Feedback alone leaves the fixed-control result nearly unchanged. Control generation accounts for most of the improvement; the additional feedback effect is small relative to variation across test episodes. Feedback reorders fixed candidates or refines generated controls within a decision, with model weights held fixed.

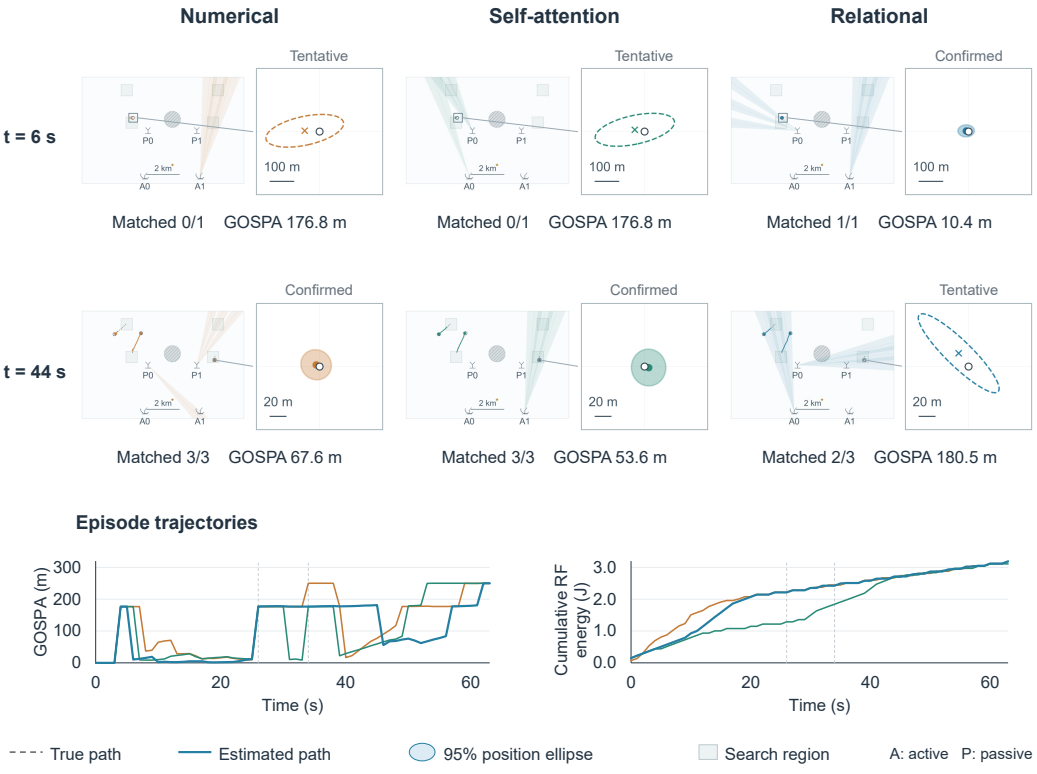


Figure 2: **Confirmation and loss of tracks occur at different times.** Columns compare numerical search, Self-attention, and the relational proposer in one transport/occlusion episode. Rows show 6 and 44 s; counts give matched targets over targets present. Dashed paths and open circles denote truth, solid paths and filled circles confirmed tracks, pale wedges scan directions, and the hatched disc an occluder. Detail views span 540 m at 6 s and 160 m at 44 s around the first and third targets, respectively. Crosses and dashed ellipses denote tentative tracks; solid ellipses denote confirmed tracks. The lower panels show full-episode GOSPA and cumulative RF energy; vertical lines mark target arrivals at 26 and 34 s. Episode selection follows Appendix H.

Table 3: **Inference-time ablation of control generation and feedback.** All settings use the same trained relational models and a budget of at most 12 forecast evaluations per cycle.

Learned controls	Feedback	GOSPA m ↓	Target recall % ↑	RF energy J ↓	Incident energy density nJ/m^2 ↓
—	—	70.69	75.41	2.682	1.906
—	✓	70.88	75.28	2.683	1.905
✓	—	59.02	80.65	2.494	1.787
✓	✓	**57.74**	**80.82**	**2.493**	**1.773**

With fixed controls, feedback reranks unscored candidates. With learned controls, it also refines action proposals. Bold marks the best value at the reported precision.

The best encoder varies with sensing conditions. The relational Full model has the lowest mean GOSPA in near range, far/weak return, maneuver/node outage, and transport/occlusion (Table A3). Its largest reduction relative to numerical search occurs with transport and occlusion, from 121.47 to 98.59 m. In this setting, a useful observation must also arrive in time to enter the tracking update. Self-attention leads in transition range, intermittent emission, and weak illumination; Deep Sets leads for late arrivals with a tight budget.

5 Related work

Sensor management and experimental design. Cognitive radar adapts sensing to an evolving environment through a perception–action cycle (Gurbuz et al., 2019). Sensor selection and non-myopic management study how measurements reduce uncertainty and tracking error under resource limits (Joshi & Boyd, 2009; Hernandez et al., 2024; Jones & García-Fernández, 2024; 2026). Conditional-independence assumptions yield greedy guarantees for information gain (Krause & Guestrin, 2005). Amortized experimental design learns an adaptive policy before deployment (Foster et al., 2021); recent approaches optimize information gain through contrastive diffusion or evaluate designs through downstream decisions (Iollo et al., 2025; Huang et al., 2024). Our analysis concerns dependence induced by shared receiver errors and how it affects candidate selection under a limited evaluation budget.

Learning within planning and search. Learned representations can reduce the space searched, change its expansion scale, or construct diverse solutions (Jiang et al., 2023; Zawalski et al., 2023; Hottung et al., 2025; Sun & Yang, 2023). Search also provides targets for amortized prediction (Ruoss et al., 2024). Recurrent models and decision-time planning connect these predictions to sequential control (Samsami et al., 2024; Hansen et al., 2024; Sokota et al., 2024). Our proposer learns sensing assignments and continuous controls from teacher trajectories. An explicit measurement model then evaluates the proposed groups online.

Representations and decision quality. Deep Sets and attention models offer different ways to encode interactions among entities (Zaheer et al., 2017; Lee et al., 2019; Diao & Loynd, 2023). We compare these encoders within a common recurrent scheduler. Predict-then-optimize studies distinguish prediction accuracy from the quality of the resulting decisions (Geng et al., 2024). We measure decision quality through closed-loop tracking and resource consumption, and compare the encoders across sensing scenarios.

6 Conclusion

Shared receiver errors make calibration and target acquisition complementary, so screening actions by individual gain can discard an effective sensing plan. We addressed this problem by learning to generate executable action groups for a fixed-budget forecast. Across eight simulated scenarios, the full proposal procedure improves tracking for all three encoders. The relational model reduces mean GOSPA and RF energy relative to numerical search, with control generation accounting for most of the improvement in the inference-time ablation. These results support using learning to shape the candidate actions available to a sensing scheduler.

AI USE STATEMENT

AI tools were used for language editing and figure preparation.

REPRODUCIBILITY STATEMENT

The supplementary material contains the implementation, evaluation configurations, episode-level results, training histories for all nine fits, and a selected checkpoint of the relational model. It includes scripts for recomputing the aggregate results and running individual evaluation episodes. Appendix A provides the proofs, Appendix B specifies the physical model and scenarios, and Appendix D details the architecture, optimization, and data partitions.

REFERENCES

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron Courville, and Marc G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29304–29320, 2021.

Y. T. Chan and K. C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, 42(8):1905–1915, 1994. doi: 10.1109/78.301830.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Cameron Diao and Ricky Loynd. Relational attention: Generalizing transformers for graph-structured tasks. In *International Conference on Learning Representations*, 2023.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. In *International Conference on Machine Learning*, pp. 3384–3395, 2021. URL https://proceedings.mlr.press/v139/foster21a.html.

Haoyu Geng, Hang Ruan, Runzhong Wang, Yang Li, Yang Wang, Lei Chen, and Junchi Yan. Benchmarking PtO and PnO Methods in the Predictive Combinatorial Optimization Regime. In *Advances in Neural Information Processing Systems*, 2024.

Sevgi Zubeyde Gurbuz, Hugh D. Griffiths, Alexander Charlish, Muralidhar Rangaswamy, Maria Sabrina Greco, and Kristine Bell. An overview of cognitive radar: Past, present, and future. *IEEE Aerospace and Electronic Systems Magazine*, 34(12):6–18, 2019. doi: 10.1109/MAES.2019.2953762.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations*, 2024.

Marcel Hernandez, Ángel F. García-Fernández, and Simon Maskell. Non-myopic sensor control for target search and track using a sample-based GOSPA implementation. *IEEE Transactions on Aerospace and Electronic Systems*, 60(1):387–404, 2024. doi: 10.1109/TAES.2023.3324908.

André Hottung, Mridul Mahajan, and Kevin Tierney. PolyNet: Learning diverse solution strategies for neural combinatorial optimization. In *International Conference on Learning Representations*, 2025.

Daolang Huang, Yujia Guo, Luigi Acerbi, and Samuel Kaski. Amortized Bayesian Experimental Design for Decision-Making. In *Advances in Neural Information Processing Systems*, 2024.

Jacopo Iollo, Christophe Heinkelé, Pierre Alliez, and Florence Forbes. Bayesian experimental design via contrastive diffusions. In *International Conference on Learning Representations*, 2025.

Zhengyao Jiang, Tianjun Zhang, Michael Janner, Yueying Li, Tim Rocktäschel, Edward Grefenstette, and Yuandong Tian. Efficient planning in a compact latent action space. In *International Conference on Learning Representations*, 2023.

George Jones and Ángel F. García-Fernández. GOSPA-driven multi-bernoulli gaussian sensor management. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 1–8, 2024. URL https://livrepository.liverpool.ac.uk/3190229/.

George Jones and Ángel F. García-Fernández. GOSPA-driven non-myopic multi-sensor management with multi-bernoulli filtering. arXiv:2511.01045v2, 2026. URL https://arxiv.org/abs/2511.01045v2. Preprint, version 2, May 2026.

Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009. doi: 10.1109/TSP.2008.2007095.

Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 324–331, 2005. URL https://www.cs.cmu.edu/~guestrin/Publications/UAI2005/uai2005.pdf.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3744–3753, 2019.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Abu Sajana Rahmathullah, Ángel F. García-Fernández, and Lennart Svensson. Generalized optimal sub-pattern assignment metric. In *20th International Conference on Information Fusion*, pp. 1–8, 2017. doi: 10.23919/ICIF.2017.8009645.

Anian Ruoss, Grégoire Delétang, Sourabh Medapati, Jordi Grau-Moya, Li K. Wenliang, Elliot Catt, John Reid, Cannada A. Lewis, Joel Veness, and Tim Genewein. Amortized Planning with Large-Scale Transformers: A Case Study on Chess. In *Advances in Neural Information Processing Systems*, 2024.

Mohammad Reza Samsami, Artem Zholus, Janarthanan Rajendran, and Sarath Chandar. Mastering memory tasks with world models. In *International Conference on Learning Representations*, 2024.

Samuel Sokota, Gabriele Farina, David J. Wu, Hengyuan Hu, Kevin A. Wang, J. Zico Kolter, and Noam Brown. The update-equivalence framework for decision-time planning. In *International Conference on Learning Representations*, 2024.

Zhiqing Sun and Yiming Yang. DIFUSCO: Graph-based Diffusion Solvers for Combinatorial Optimization. In *Advances in Neural Information Processing Systems*, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.03762.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Michał Zawalski, Michał Tyrolski, Konrad Czechowski, Tomasz Odrzygóźdź, Damian Stachura, Piotr Piękos, Yuhuai Wu, Łukasz Kuciński, and Piotr Miłoś. Fast and precise: Adjusting planning horizon with adaptive subgoal search. In *International Conference on Learning Representations*, 2023.

A CONDITIONAL INFORMATION AND JOINT ACTION SELECTION

We begin with the expected reduction in squared error from an additional observation. We then derive the calibration gain and prove Proposition 1. The remaining results describe shared measurement errors, intermittent emissions, and the effect of a limited candidate set on the forecast objective.

A.1 CONDITIONAL OBSERVATION VALUE

Let $x \in \mathbb{R}^d$ contain the target position coordinates, with $\mathbb{E}\|x\|^2 < \infty$. Assume a fixed joint distribution for the observations, so adding Y_a does not change Y_S. The information before and after this addition is represented by the sigma-algebras $\mathcal{H} = \sigma(Y_S)$ and $\mathcal{H}^+ = \sigma(Y_S, Y_a)$. Define the corresponding conditional means as $m = \mathbb{E}[x \mid \mathcal{H}]$ and $m^+ = \mathbb{E}[x \mid \mathcal{H}^+]$. The tower property gives $\mathbb{E}[m^+ \mid \mathcal{H}] = m$. Since $x - m^+$ is orthogonal to every square-integrable $\mathcal{H}^+$-measurable vector,

$$x - m = (x - m^+) + (m^+ - m),$$
$$\mathrm{Cov}(x \mid \mathcal{H}) = \mathbb{E}[\mathrm{Cov}(x \mid \mathcal{H}^+) \mid \mathcal{H}] + \mathrm{Cov}(m^+ \mid \mathcal{H}).$$

Taking expected traces establishes

$$F(S \cup \{a\}) - F(S) = \mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(m^+ \mid \mathcal{H}) = \mathbb{E}\|m^+ - m\|^2 \geq 0.$$

The gain equals the expected squared change in the conditional mean. It need not decrease as observations are added: in the example below, calibration has zero gain before a target observation and positive gain afterward. This argument assumes a fixed observation family; changing the time of a sensing visit can change its distribution.

A.2 INFORMATION TRANSFER THROUGH A SHARED RECEIVER STATE

Consider the scalar model $y = x + b + \epsilon$ and $z = b + \nu$, where x, b, ϵ, ν are independent zero-mean Gaussians with variances $q, r, s, v > 0$. Conditioning on calibration gives

$$r_z = \mathrm{Var}(b \mid z) = r - \frac{r^2}{r+v} = \frac{rv}{r+v}, \qquad \mathrm{Var}(x \mid z) = q.$$

Calibration reduces receiver uncertainty but initially leaves the target variance unchanged. A target observation instead creates target–receiver dependence:

$$\mathrm{Var}(x \mid y) = \frac{q(r+s)}{q+r+s}, \qquad \mathrm{Cov}(x, b \mid y) = -\frac{qr}{q+r+s}.$$

The negative covariance describes the ambiguity in the measured sum: an increase in target position can be offset by a decrease in receiver bias. After calibration, subtracting $\mathbb{E}[b \mid z]$ leaves residual noise variance $r_z + s$. Hence

$$\mathrm{Var}(x \mid y, z) = \frac{q(r_z+s)}{q+r_z+s},$$
$$\Delta_{\mathrm{cal}} = q^2\left(\frac{1}{q+r_z+s} - \frac{1}{q+r+s}\right) = \frac{q^2(r-r_z)}{(q+r+s)(q+r_z+s)} > 0.$$

Multivariate form. Let (x, b) have a joint Gaussian posterior, with x restricted to the target coordinates of interest. A calibration observes $z_c = H_c b + \nu_c$, where ν_c is independent and has covariance $R_c \succ 0$. With $S_c = H_c P_{bb} H_c^\top + R_c$, the target covariance reduction is

$$\Delta P_{XX} = P_{Xb} H_c^\top S_c^{-1} H_c P_{bX} \succeq 0, \qquad \mathrm{tr}\,\Delta P_{XX} = \|S_c^{-1/2} H_c P_{bX}\|_F^2.$$

The gain is zero if and only if $H_c P_{bX} = 0$. Thus calibration reduces target uncertainty only through receiver directions correlated with the target. Because x contains position coordinates, the trace is measured in square meters. For a joint position–velocity state, a position selector or an appropriate task-weight matrix is needed.

Nuisance marginalization. For joint precision $J = P^{-1} \succ 0$, completing the square yields

$$x^\top J_{XX} x + 2x^\top J_{Xb} b + b^\top J_{bb} b = \left(b + J_{bb}^{-1} J_{bX} x\right)^\top J_{bb} \left(b + J_{bb}^{-1} J_{bX} x\right) + x^\top \left(J_{XX} - J_{Xb} J_{bb}^{-1} J_{bX}\right) x.$$

Integrating over b gives the marginal target precision

$$P_{XX}^{-1} = J_{XX} - J_{Xb} J_{bb}^{-1} J_{bX}. \tag{14}$$

The second term removes precision that depends on knowing the receiver state. With uncertain association, marginalization also includes variation between association-conditioned means:

$$\bar{P} = \sum_h w_h \left[P_h + (\mu_h - \bar{\mu})(\mu_h - \bar{\mu})^\top\right], \qquad \bar{\mu} = \sum_h w_h \mu_h. \tag{15}$$

Here w_h, μ_h, and P_h are the posterior weight, mean, and covariance of hypothesis h.

A.3 Conditioning order

For $Y = (y, z)^\top$, the observation covariance is

$$S_Y = \begin{bmatrix} q + r + s & r \\ r & r + v \end{bmatrix}, \qquad D = \det S_Y = (q + s)(r + v) + rv.$$

The off-diagonal entry is the shared receiver variance. Since $\mathrm{Cov}(x, Y) = [q\ 0]$, joint conditioning gives

$$\mathrm{Var}(x \mid Y) = q - \frac{q^2(r + v)}{D} = \frac{q\{s + rv/(r + v)\}}{q + s + rv/(r + v)},$$

which agrees with calibration-first conditioning. In the opposite order, define $r_y = r(q + s)/(q + r + s)$ and $k_y = -qr/(q + r + s)$. The calibration residual has variance $r_y + v$ and target covariance k_y, so its target gain is

$$\frac{k_y^2}{r_y + v} = \frac{q^2 r^2}{(q + r + s)\{(q + s)(r + v) + rv\}} = \Delta_{\mathrm{cal}}.$$

The two conditioning orders give the same final posterior but different intermediate gains. In a sensing schedule, changing the visit order can also change geometry, process uncertainty, and delivery time. The forecast accounts for these quantities at the visit times determined by projection.

A.4 Calibration versus a repeated target visit

For two observations $y_i = x + b + \epsilon_i$ with independent extraction variance s, the mean contains $x + b$ with extraction variance $s/2$. Their difference contains only extraction noise and is independent of the mean. The effective noise of the pair is therefore $r + s/2$. Calibration with one target visit has effective noise $r_z + s$. Since $g(\eta) = q\eta/(q + \eta)$ satisfies $g'(\eta) = q^2/(q + \eta)^2 > 0$,

$$\begin{aligned} \mathrm{Var}(x \mid y, z) < \mathrm{Var}(x \mid y_1, y_2) &\iff r_z + s < r + s/2 \\ &\iff \frac{r^2}{r + v} > s/2. \end{aligned}$$

Calibration is preferred when the receiver variance it removes exceeds the extraction-noise reduction from repetition. The target prior controls the size of the improvement but cancels from the decision boundary. A known receiver ($r \to 0$) favors repetition; precise calibration ($v \to 0$) is preferred when $r > s/2$. This comparison fixes two acquisition slots and their precisions. Changing power or dwell changes s or v and requires a new feasibility calculation.

A.5 Proof of the singleton-screening counterexample

Use the family in Proposition 1 and normalize by ℓ, so the target prior variance is one. For $\rho \geq 2$, the singleton gains are

$$F(\{z\})/\ell^2 = 0, \quad F(\{y\})/\ell^2 = \frac{1}{1 + \rho + \rho^{-1}}, \quad F(\{w\})/\ell^2 = \frac{1}{1 + \rho}.$$

Singleton screening retains y, w. Their independent effective noises give precision increment

$$I_{yw} = \frac{1}{\rho + \rho^{-1}} + \frac{1}{\rho} = \frac{2\rho^2 + 1}{\rho(\rho^2 + 1)} =: \alpha_\rho.$$

For y, z, the residual receiver variance is $\rho/(\rho^2 + 1)$ and the extraction variance is $1/\rho$. Their sum is α_ρ, giving $I_{yz} = \alpha_\rho^{-1}$. The third pair has $I_{zw} = 1/\rho$. Because

$$\rho(\rho^2 + 1) - (2\rho^2 + 1) = \rho^2(\rho - 2) + \rho - 1 > 0,$$

we have $0 < \alpha_\rho < 1$ and $I_{yz} > I_{yw} > I_{zw}$. Thus y, z is the optimal pair. Adding the unit prior precision gives

$$\mathrm{Var}(x \mid y, w)/\ell^2 = \frac{1}{1 + \alpha_\rho}, \qquad F(\{y, w\})/\ell^2 = \frac{\alpha_\rho}{1 + \alpha_\rho},$$
$$\mathrm{Var}(x \mid y, z)/\ell^2 = \frac{\alpha_\rho}{1 + \alpha_\rho}, \qquad F(\{y, z\})/\ell^2 = \frac{1}{1 + \alpha_\rho}.$$

The ratio of gains is $\alpha_\rho \sim 2/\rho \to 0$, while the ratio of residual variances is α_ρ^{-1}. Exact sequential greedy selection also chooses w first. This leaves x independent of b, so calibration still has zero immediate target gain and the second choice is y. Both rules therefore miss the same complementary pair. The absolute variance remains bounded by ℓ^2; the ratio between the two residual variances diverges.

A.6 Common-reference uncertainty

Suppose n observations of a scalar parameter share one reference error of variance σ_b^2 and have independent extraction variance σ^2. Their covariance is $R = \sigma^2 I + \sigma_b^2 \mathbf{1}\mathbf{1}^\top$. The matrix inversion lemma gives

$$R^{-1} = \sigma^{-2} I - \frac{\sigma_b^2}{\sigma^2(\sigma^2 + n\sigma_b^2)} \mathbf{1}\mathbf{1}^\top, \qquad I_n = \frac{n}{\sigma^2 + n\sigma_b^2}.$$

Thus $I_n = \mathbf{1}^\top R^{-1} \mathbf{1} \to 1/\sigma_b^2$. With independent prior variance q, the posterior variance approaches $q\sigma_b^2/(q + \sigma_b^2)$. Repetition estimates the sum of target and reference error more precisely without separating its two components. More generally, in $Y = Hx + Gb + \epsilon$, a target displacement u is confounded whenever $Hu = Gv$ for a nuisance displacement v. Geometric diversity helps when additional rows distinguish these effects.

A.7 Shared emission and joint misses

Let a common emission event have probability π. Conditional on emission and the other availability gates, suppose visits detect the target independently with probabilities $d_1, \ldots, d_n$. Total probability gives

$$p_{\mathrm{miss}} = (1 - \pi) + \pi \prod_{j=1}^{n} (1 - d_j).$$

For two visits, replacing the shared event by independent emission draws changes the miss probability by

$$p_{\mathrm{miss}} - (1 - \pi d_1)(1 - \pi d_2) = \pi(1 - \pi) d_1 d_2 \geq 0.$$

All visits in a silent cycle miss the target because of the same emission event; a later cycle has a new emission draw. The forecast also conditions on shared reference and transport events when combining detection outcomes.

A.8 CANDIDATE COVERAGE AND EVALUATION ALLOCATION

Fix the decision-time belief, forecast objective, and eligibility rule. Let $\mathcal{C}$ be the finite nonempty set of constructed executable groups and $\mathcal{S} \subseteq \mathcal{C}$ the nonempty subset successfully evaluated. With $\mathcal{C} \subseteq \mathcal{F}$, adding and subtracting the best constructed cost gives

$$\min_{A \in \mathcal{S}} \widehat{J}(A) - \inf_{A \in \mathcal{F}} \widehat{J}(A) = \left[\min_{A \in \mathcal{S}} \widehat{J}(A) - \min_{A \in \mathcal{C}} \widehat{J}(A) \right] + \left[\min_{A \in \mathcal{C}} \widehat{J}(A) - \inf_{A \in \mathcal{F}} \widehat{J}(A) \right]. \tag{16}$$

Both terms are nonnegative. The first is the cost of leaving constructed candidates unevaluated; the second is the cost of restricting the search to the constructed set. The set $\mathcal{C}$ contains the distinct projected groups from both proposal passes, since different continuous proposals can map to the same executable group. Failed evaluations consume budget but do not add a group to $\mathcal{S}$. Equation (16) separates these two sources of suboptimality for the forecast objective in Equation (8).

A.9 CONTINUATION AFTER FUTURE OBSERVATIONS

For continuation cost $V(\beta'(Y), A')$, allowing the next action to depend on observation Y gives

$$\mathbb{E}_Y \left[\min_{A'} V(\beta'(Y), A') \right] \leq \min_{A'} \mathbb{E}_Y [V(\beta'(Y), A')].$$

For any fixed continuation, the pointwise minimum is no larger than its cost. Taking expectations and then minimizing over fixed continuations gives the inequality. Equality holds if one continuation is optimal for almost every observation. This distinction motivates nonmyopic sensing (Hernandez et al., 2024). The implemented forecast branches over shared availability and detection, applies zero-innovation Gaussian updates, and searches a short analytic continuation set. The learned future-value head uses realized teacher-trajectory costs to allocate evaluations within this common forecast model.

B PHYSICAL MODEL AND INFERENCE

The stage-cost weights are $\alpha_{\text{search}} = 9 \times 10^4\,\text{m}^2$, $\alpha_{\text{miss}} = 10^3\,\text{m}^2$, and $\alpha_{\text{tent}} = 10^2\,\text{m}^2$. These weights express search priority, miss probability, and tentative-track count on the same squared-distance scale as position variance.

B.1 GEOMETRY, MOTION, AND TARGET ARRIVALS

The model is planar. Two monostatic sensors occupy $(-1700, -1200)$ and $(1700, -1200)\,\text{m}$; two passive receivers occupy $(-1500, 1600)$ and $(1500, 1600)\,\text{m}$. The surveyed calibration source is at $(0, -500)\,\text{m}$, the illuminator at $(0, -6000)\,\text{m}$, and the fixed observation point at $(8000, 6000)\,\text{m}$.

Each target state is $x = (p_x, p_y, v_x, v_y)^\top$. With cycle duration $\Delta = 1\,\text{s}$, prediction uses a constant-velocity transition and white acceleration spectral density $q_a = 0.8\,\text{m}^2/\text{s}^3$:

$$F = \begin{bmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{bmatrix}, \qquad Q = q_a \begin{bmatrix} \Delta^3 I_2/3 & \Delta^2 I_2/2 \\ \Delta^2 I_2/2 & \Delta I_2 \end{bmatrix}. \tag{17}$$

The maneuver scenario adds a prescribed acceleration interval to one target. Up to three targets enter at successive geometric waiting times, with a default per-cycle arrival probability 0.07. Their initial regions are distinct, their nominal speeds are uniform between 15 and $65\,\text{m/s}$, and their directions are uniform on the circle. Position and velocity perturbations have standard deviations of $30\,\text{m}$ and $2\,\text{m/s}$, respectively. Radar cross sections are uniform on $[0.7, 1.3]\,\text{m}^2$.

Receiver delay, bearing, and frequency states follow independent random walks before coupling through observations. Initial estimator standard deviations are $25\,\text{m}$, $0.6°$, and $12\,\text{Hz}$ per receiver. Per-cycle drift standard deviations are $0.15\,\text{m}$, $0.005°$, and $0.05\,\text{Hz}$. The simulated receiver offsets start from prescribed nonzero values with small random perturbations, while the estimator starts at zero mean. Calibration therefore corrects the initial offset and tracks subsequent drift.

B.2 MEASUREMENT EQUATIONS

Let p and v be target position and velocity at the visit midpoint, s_i a stationary receiver, $r_i = \|p - s_i\|$, $u_i = (p - s_i)/r_i$, and $\theta_i = \text{atan2}(p_y - s_{iy}, p_x - s_{ix})$. Write effective receiver measurement biases as b_i^r in meters, b_i^θ in radians, and b_i^f in hertz. The wavelength is $\lambda = c_0/f_c$, where c_0 is the speed of light.

The monostatic mean observation is

$$h_{\text{mono},i}(x, b) = \begin{bmatrix} r_i + b_i^r \\ \theta_i + b_i^\theta \\ u_i^\top v - \lambda b_i^f/2 \end{bmatrix}, \qquad \text{units } (\text{m}, \text{rad}, \text{m/s}). \tag{18}$$

For direct reception of a target emitter at receivers i and j, the mean is

$$h_{\text{dir},ij}(x, b) = \begin{bmatrix} (r_i - r_j + b_i^r - b_j^r)/c_0 \\ -(u_i - u_j)^\top v/\lambda + b_i^f - b_j^f \\ \theta_i + b_i^\theta \\ \theta_j + b_j^\theta \end{bmatrix}, \quad \text{units } (\text{s}, \text{Hz}, \text{rad}, \text{rad}). \tag{19}$$

The range-difference constraint is the classical hyperbolic localization geometry (Chan & Ho, 1994). Both delay and Doppler are measured relative to the second receiver. Its angle of arrival supplies an additional geometric constraint.

For an illuminator at s_L, define $r_L = \|p - s_L\|$ and $u_L = (p - s_L)/r_L$. The bistatic mean relative to its direct reference path is

$$h_{\text{bi},i}(x, b) = \begin{bmatrix} (r_i + r_L - \|s_i - s_L\| + b_i^r)/c_0 \\ -(u_i + u_L)^\top v/\lambda + b_i^f \\ \theta_i + b_i^\theta \end{bmatrix}, \qquad \text{units } (\text{s}, \text{Hz}, \text{rad}). \tag{20}$$

The distance-bias state enters monostatic range additively; direct reception uses the difference of two receiver states divided by c_0; bistatic reception uses a residual surveillance–reference delay bias

b_i^r/c_0. Frequency bias enters through the role-specific coefficients above. These states parameterize persistent residual measurement offsets. The independent reference-extraction draw for each capture enters the observation covariance separately.

Calibration subtracts the known source geometry and produces noisy residuals for delay, frequency, and bearing offset. Active calibration uses a surveyed reflector and consumes transmit energy. Passive calibration receives the surveyed beacon. Observation Jacobians include the motion from cycle time to the actual visit midpoint.

B.3 Signal strength, detection, and measurement uncertainty

The device parameters in Table A1 determine the signal model. Let D be aperture diameter, η aperture efficiency, $G = \eta(\pi D/\lambda)^2$ boresight gain, and $\vartheta = 0.886\lambda/D$ the beamwidth. The Gaussian power pattern at angular offset $\delta\theta$ is $g = \exp[-4\log(2)(\delta\theta/\vartheta)^2]$. The modeled measurement support is restricted to the half-power beam. Write $N_0 = k_B T F_n L$ for receiver noise power spectral density in $\mathrm{W/Hz}$, including noise factor and system loss. Per-visit signal-to-noise ratios are

$$\gamma_{\mathrm{mono}} = \frac{E_{\mathrm{RF}} G^2 g^2 \lambda^2 \sigma}{(4\pi)^3 r_i^4 N_0}, \tag{21}$$

$$\gamma_{\mathrm{dir}} = \frac{P_e T_{\mathrm{dwell}} G g \lambda^2}{(4\pi r_i)^2 N_0}, \tag{22}$$

$$\gamma_{\mathrm{bi}} = \frac{P_L T_{\mathrm{dwell}} G g \lambda^2 \sigma}{(4\pi)^3 r_L^2 r_i^2 N_0}. \tag{23}$$

Here σ is target cross section and P_e, P_L are external emitter and illuminator equivalent isotropic radiated powers. Monostatic SNR decays as r_i^{-4}, direct-path SNR as r_i^{-2}, and bistatic SNR as $(r_L r_i)^{-2}$.

Conditional on available propagation and a signal, detection uses $p_D = p_{\mathrm{FA}}^{1/(1+\gamma)}$, with per-resolution-cell false alarm probability $p_{\mathrm{FA}} = 10^{-7}$. Occlusion, node outage, and absent illumination remove the corresponding signal. Direct passive observations also require the common target emission event. Bistatic reference acquisition is shared by returns in the same capture. The reference channel has its own signal-to-noise ratio and detection event.

Conditional on detection, the measurement variances follow the local information scaling

$$\sigma_\tau^2 = \frac{1}{8\pi^2\beta^2\gamma}, \quad \sigma_f^2 = \frac{1}{8\pi^2(T_{\mathrm{dwell}}^2/12)\gamma}, \quad \sigma_\theta^2 = \frac{\vartheta^2}{8\gamma}, \qquad \beta = B/\sqrt{12}. \tag{24}$$

For monostatic range and radial velocity, delay and frequency variances are multiplied by $c_0^2/4$ and $\lambda^2/4$. Reference extraction errors are added to passive delay and frequency variance. A bistatic capture shares the same reference error across its target returns, producing off-diagonal covariance blocks. Direct references are target-emitter specific; persistent effective receiver biases remain in the augmented state. These expressions specify the measurement noise in the simulator.

The pulse timing admits monostatic ranges between $c_0\tau_p/2 \simeq 300\,\mathrm{m}$ and $c_0(T_r - \tau_p)/2 \simeq 14.69\,\mathrm{km}$. These limits describe receive blanking and ambiguity support. Detection probability inside that interval still varies with energy, aperture, range, target cross section, and pointing. Passive delay processing uses a $200\,\mu\mathrm{s}$ window. External illumination and emitter powers are environmental variables rather than outputs of the scheduler.

B.4 Association, birth, and observation arrival

Clutter count is Poisson with mean equal to the number of resolution cells times the per-cell false alarm probability; clutter values are uniform over the measurement support. Returns unresolved within the same measurement cell retain one noisy observation. Association hypotheses use the fixed squared normalized-residual threshold 16.27 for both three- and four-dimensional observations and are capped at 2048. Above this cap, the filter uses a global nearest assignment. Candidate new tracks use monostatic inverse geometry or passive weighted nonlinear fitting with a velocity prior. Their covariance includes the sensitivity to receiver nuisance states. New-track fits outside signal or geometric support are rejected; an AOA-only observation does not initialize an absolute range.

Table A1: Device and resource parameters. Dwell is receive/integration time; transmit time counts the pulse widths only. Default resource budgets apply to the full 64-cycle episode unless marked per cycle.

Parameter	Value	Parameter	Value
Carrier	$3\,\mathrm{GHz}$	Bandwidth	$10\,\mathrm{MHz}$
Pulse width	$2\,\mu\mathrm{s}$	Repetition interval	$100\,\mu\mathrm{s}$
Aperture	$0.6\,\mathrm{m}$	Aperture efficiency	0.6
Noise factor	4	Propagation/system loss	4
Temperature	$290\,\mathrm{K}$	Beamwidth	$0.1476\,\mathrm{rad}$
Peak power	2–100 W	Dwell	6–180 ms
Slew rate	$25\,\mathrm{rad/s}$	Settling time	$1\,\mathrm{ms}$
RF energy	$3.2\,\mathrm{J}$	Transmit time	$80\,\mathrm{ms}$
Reception/cycle	$1.2\,\mathrm{s}$	Receive electrical power	$5\,\mathrm{W}$
Operations/cycle	10^8	Communication/cycle	$4096\,\mathrm{bytes}$
Processing rate	$2\times10^9\,\mathrm{s}^{-1}$	Default link rate	$1\,\mathrm{Mbit/s}$

Table A2: Sensing conditions in the eight scenarios. Each scenario includes stochastic arrivals and up to three targets.

Scenario	Sensing conditions
Near range	Stronger links; resource allocation and calibration near the network.
Transition range	Weaker links and geometry-dependent active/passive choices.
Far range	Search regions extend to $7\,\mathrm{km}$ in the northward coordinate.
Intermittent emission	Emission probability 0.2 and periodic illumination gaps.
Late births, scarce energy	Later arrivals, half the RF budget, and tighter incident-energy-density limit.
Maneuver and outage	A turning target and scheduled unavailability of two nodes.
Weak illumination	Far-range targets with weaker emitter and illuminator powers.
Transport and occlusion	An opaque region, 15% packet loss, and a $20\,\mathrm{kbit/s}$ link.

Two detections in distinct cycles confirm a track. A track is removed after more than nine cycles without a hit. Gaussian updates use a common predicted linearization anchor within a cycle, extended when a track is initialized. Each measurement record is used once. The filter updates existing tracks and initializes new ones as usable packets arrive.

Each beam visit consumes steering time, settling time, and integration time. Same-band active transmissions are serialized. Processing is queued per node, and communication uses a shared first-in, first-out link. A packet carries 512 bytes per visit; reception requires eight processing operations per bandwidth sample. A packet is discarded if it arrives after its acquisition cycle. The forecast uses the same resource and arrival-time calculations to evaluate candidate groups.

The default incident energy density at observation point s_M is

$$\mathcal{E}_t = \sum_{a\in A_t:\,\mathrm{active}} \frac{E_a}{4\pi\|s_{i(a)} - s_M\|^2}, \qquad [\mathcal{E}_t] = \mathrm{J/m^2}. \tag{25}$$

Its per-cycle limit is $5\times10^{-10}\,\mathrm{J/m^2}$, tightened to $10^{-11}\,\mathrm{J/m^2}$ in the late-birth regime. This quantity measures radiated energy per unit area at the fixed observation point under isotropic propagation.

B.5 Scenario specification

Search-region half-widths are $450\,\mathrm{m}$ in transition range, $600\,\mathrm{m}$ in far/weak return, and $350\,\mathrm{m}$ in the other six scenarios. Scenarios that share a geometry use the same region centers.

Each regime has four square search regions, arranged in two symmetric pairs. In near geometry, both pairs use eastward coordinates $\pm800\,\mathrm{m}$; their northward coordinates are $-500\,\mathrm{m}$ and $700\,\mathrm{m}$. The region half-width is $350\,\mathrm{m}$. Transition geometry places the lower pair at eastward coordinates $\pm2500\,\mathrm{m}$ and northward coordinate $2100\,\mathrm{m}$, and the upper pair at $\pm2800\,\mathrm{m}$ and $4100\,\mathrm{m}$, respectively. Its half-width is $450\,\mathrm{m}$. The corresponding far-range coordinates are $\pm3800\,\mathrm{m}$ and $4200\,\mathrm{m}$

for the lower pair, and $\pm 4200\,\mathrm{m}$ and $7000\,\mathrm{m}$ for the upper pair, with half-width $600\,\mathrm{m}$. The remaining regimes modify these geometries as follows.

Intermittent emission uses transition geometry, target emission probability 0.2, and illumination gaps at cycles whose remainder modulo 20 is 10, 11, or 12. Late births use transition geometry and arrival probability 0.035, with $1.6\,\mathrm{J}$ RF energy and $40\,\mathrm{ms}$ transmit-time budgets. Maneuver and outage use near geometry, with one active node absent in cycles 20–31 and one passive node absent in cycles 36–42; the first target's northward velocity increases by $2\,\mathrm{m/s}$ per cycle over cycles 16–20. Weak illumination uses far geometry, $1\,\mathrm{kW}$ illuminator power and $1\,\mathrm{mW}$ target emitter power, plus illumination gaps. Transport and occlusion use transition geometry, packet loss 0.15, a $20\,\mathrm{kbit/s}$ link, and an opaque disc centered at $(0, 2300)\,\mathrm{m}$ with radius $500\,\mathrm{m}$. Default target emitter power is $10\,\mathrm{mW}$, illuminator power is $10\,\mathrm{kW}$, emission probability is 0.7, and packet loss is 0.03.

Search-priority mass starts at 0.4 per region and increases by 0.02 per cycle before reduction by discovery. Search priority evolves separately from the geometric arrival process that generates at most three targets. Search becomes overdue after six cycles. Due calibration proposals are inserted at a twelve-cycle interval. A search opportunity meets the revisit criterion when its forecast discovery probability reaches 0.5. Target detection is determined by the observations generated after execution.

C Geometry, signal precision, and usable information

The distinction between receiver uncertainty and measurement noise in the scalar example extends to radar observations. Geometry determines which state directions are observable; bandwidth, integration time, and received energy determine measurement precision. Receiver offsets and common references correlate measurement errors, while visibility, emission, and transport determine whether observations are available. Below we relate these quantities to the controls in Section B.

C.1 TDOA, FDOA, and AOA constrain different parts of the state

Let $p \in \mathbb{R}^2$ be target position, v target velocity, and s_i a known stationary receiver position. Define

$$r_i = \|p - s_i\|, \qquad u_i = \frac{p - s_i}{r_i}, \qquad B_i = \frac{I - u_i u_i^\top}{r_i}.$$

The line-of-sight unit vector u_i is dimensionless; B_i has units of inverse meters. Differentiating the range and direction gives

$$\begin{aligned} \mathrm{d}r_i &= \frac{(p - s_i)^\top \mathrm{d}p}{\|p - s_i\|} = u_i^\top \mathrm{d}p, \\ \mathrm{d}u_i &= \frac{\mathrm{d}p}{r_i} - \frac{p - s_i}{r_i^2}\,\mathrm{d}r_i \\ &= \frac{I - u_i u_i^\top}{r_i}\,\mathrm{d}p = B_i\,\mathrm{d}p. \end{aligned}$$

The matrix $I - u_i u_i^\top$ removes a radial displacement. A small motion along the line of sight changes range to first order, while a transverse motion changes bearing.

A time difference is one range-difference constraint. For direct emission received at i and j, with range-equivalent receiver offsets b_i^r, b_j^r, the TDOA is

$$\begin{aligned} \tau_{ij} &= \frac{r_i - r_j + b_i^r - b_j^r}{c_0}, \\ \mathrm{d}\tau_{ij} &= \frac{(u_i - u_j)^\top \mathrm{d}p + \mathrm{d}b_i^r - \mathrm{d}b_j^r}{c_0}. \end{aligned}$$

Here c_0 is propagation speed in meters per second, so τ_{ij} is in seconds. A fixed delay defines a hyperbola in position space when offsets are known. One scalar delay cannot by itself determine both planar coordinates. When the two line-of-sight vectors nearly coincide, $u_i - u_j$ is small and the delay changes little with target displacement. Even a precise delay estimate then provides little information about position.

A frequency difference couples geometry and motion. For wavelength λ and frequency offsets b_i^f, b_j^f, the direct-emission FDOA is

$$\begin{aligned} f_{ij} &= -\frac{(u_i - u_j)^\top v}{\lambda} + b_i^f - b_j^f, \\ \mathrm{d}f_{ij} &= -\frac{(u_i - u_j)^\top \mathrm{d}v}{\lambda} - \frac{v^\top(\mathrm{d}u_i - \mathrm{d}u_j)}{\lambda} + \mathrm{d}b_i^f - \mathrm{d}b_j^f \\ &= -\frac{(u_i - u_j)^\top \mathrm{d}v}{\lambda} - \frac{v^\top(B_i - B_j)\mathrm{d}p}{\lambda} + \mathrm{d}b_i^f - \mathrm{d}b_j^f. \end{aligned}$$

The position sensitivity depends on velocity, while the velocity sensitivity depends on geometry. FDOA therefore constrains the joint position–velocity state. The tracker estimates position and velocity jointly with stationary receivers.

A bearing constrains a transverse direction. For $\theta_i = \mathrm{atan2}(p_y - s_{i,y}, p_x - s_{i,x}) + b_i^\theta$, define $u_i^\perp = (-u_{i,y}, u_{i,x})$. Then

$$\mathrm{d}\theta_i = \frac{(u_i^\perp)^\top}{r_i}\,\mathrm{d}p + \mathrm{d}b_i^\theta.$$

One AOA measurement gives a ray. Two nonparallel rays can constrain a planar position; nearly parallel rays give a weak intersection. Unknown bearing offsets rotate these rays. Calibration helps distinguish receiver bearing error from target displacement.

C.2 Observability and receiver ambiguity

Linearize a collection of observations at a common belief anchor:

$$\delta Y = H_x \delta x + H_b \delta b + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, R).$$

The whitened sensitivity is $R^{-1/2}[H_x\ H_b]$. A displacement (u, w) satisfying $H_x u + H_b w = 0$ leaves the linearized mean unchanged. Improving signal strength rescales the measurement precision but does not remove this null direction. Adding a calibration or a different geometry can add a row that distinguishes u from w.

For $y = x + b + \epsilon$, the row is $[1\ 1]$ and the ambiguity direction is $[1\ -1]^\top$. The calibration row $[0\ 1]$ has nonzero projection onto that direction, so the two rows together have full rank. Repeating $[1\ 1]$ with a smaller noise variance does not change the rank. This rank difference explains how calibration can be preferable to a repeated target observation.

Difference measurements also have gauge directions. Adding the same constant to every receiver delay offset leaves all TDOAs unchanged; the same holds for common frequency offset in FDOA. A surveyed reference or a prior anchors these common offsets. A positive-definite prior can yield a nonsingular posterior even when the measurements alone do not identify every state direction. The tracker combines the prior and measurement information while estimating receiver offsets jointly with target states.

After nuisance marginalization, the weighted sensitivity describes which target directions the observations resolve. A second receiver can add a constraint along a previously ambiguous direction. The proposer receives these geometric and covariance relations when forming groups of sensing actions.

C.3 How controls change measurement precision

The propagation laws in Section B distinguish a monostatic round trip, a direct path, and the two path lengths of a bistatic echo. Peak power increases SNR and reduces extraction noise. Dwell increases integrated energy and separately improves frequency precision through the T_d^2 factor; bandwidth improves delay precision through β^2. Calibration changes the receiver-bias posterior. Power, dwell, bandwidth, and calibration therefore reduce different components of estimation error.

The same separation explains detection saturation. The chosen energy-detector model has normalized noise-only energy $Z \sim \mathrm{Exp}(1)$ and signal-present energy $Z \sim \mathrm{Exp}(1+\gamma)$, with the exponential distribution parameterized by its mean. For a normalized threshold h,

$$\begin{aligned} p_{\mathrm{FA}} &= \Pr(Z > h \mid \text{noise}) = e^{-h}, \\ h &= -\log p_{\mathrm{FA}}, \\ p_D &= \Pr(Z > h \mid \text{signal}) = e^{-h/(1+\gamma)} \\ &= p_{\mathrm{FA}}^{1/(1+\gamma)}. \end{aligned}$$

The simulator uses this detection probability for a fluctuating-energy signal. At high SNR, detection approaches one, so additional power mainly improves measurement precision. A silent emitter or a target outside the beam support produces no target detection.

C.4 Measurement precision is different from availability

Let V be the event that the signal, reference, processing, and delivery needed for an observation are all available in time. Hold the local sensitivity H, conditional covariance R, and availability probabilities fixed. The expected information contribution from an available measurement is

$$\mathbb{E}[I_{\mathrm{admitted}} \mid A] = \Pr(V \mid A)\, H^{\top} R^{-1} H.$$

This expression separates measurement availability from conditional precision. It gives the contribution of successful measurements; availability events can provide additional information when their probabilities depend on the target state. Shared emission, reference, or transport events correlate availability across visits and must be included when computing joint probabilities.

For one sensing visit with first-sample time t_0, the observation timestamp is the visit midpoint, while processing waits for the end of acquisition:

$$\begin{aligned} t_{\mathrm{obs}} &= t_0 + T_d/2, \\ t_{\mathrm{acq,end}} &= t_0 + T_d, \\ t_{\mathrm{process}} &= \max(t_{\mathrm{acq,end}}, t_{\text{processor available}}) + t_{\mathrm{processing}}, \\ t_{\mathrm{arrive}} &= \max(t_{\mathrm{process}}, t_{\text{link available}}) + t_{\mathrm{transmission}}. \end{aligned}$$

Steering and settling occur before t_0. Processing and communication queues, together with the specified delay faults, determine packet arrival. A measurement is available to the filter only if it arrives within its acquisition cycle, regardless of SNR. Projection therefore checks timing and shared-resource constraints before the group head scores the projected controls.

C.5 Example: allocating pulses in a three-direction search

Consider a three-direction active search using $P_{\mathrm{peak}} = 40\,\mathrm{W}$, $T_r = 100\,\mu\mathrm{s}$, $\tau_p = 2\,\mu\mathrm{s}$, and $N = 900$ pulse intervals. Receive dwell, transmit time, and RF energy are

$$\begin{aligned} T_d &= 900 \times 100\,\mu\mathrm{s} = 90\,\mathrm{ms}, \\ T_{\mathrm{tx}} &= 900 \times 2\,\mu\mathrm{s} = 1.8\,\mathrm{ms}, \\ E_{\mathrm{RF}} &= 40\,\mathrm{W} \times 1.8\,\mathrm{ms} = 0.072\,\mathrm{J}. \end{aligned}$$

For requested scan weights $(1, 2, 1)$, the implemented integer allocation first gives two intervals to each direction and apportions the remaining 894 by largest remainder. This yields $(226, 449, 225)$ intervals, summing to 900. The central direction receives more energy and integration time, while the side directions preserve search coverage. Steering and settlement add to elapsed time without adding to this RF energy. The integer allocation preserves the total number of pulse intervals while implementing the requested scan proportions.

An action with the same N but twice the power has twice the RF energy and a different SNR. An action with twice N changes both elapsed time and integration precision, and may miss its communication window. Replacing active transmission by passive listening sets controlled RF power to zero but uses receiver time and relies on an external source. Replacing target acquisition by calibration changes which latent variable is observed. The group forecast accounts for these changes in precision, availability, and the state being observed.

C.6 From an action to an available observation

Figure 3 expands the spatial assignments into visit and transport times. Receive dwell determines the acquisition window; processing and communication queues determine whether the measurement reaches the filter in time. In this example, the last report arrives after the one-second acquisition cycle and is discarded.

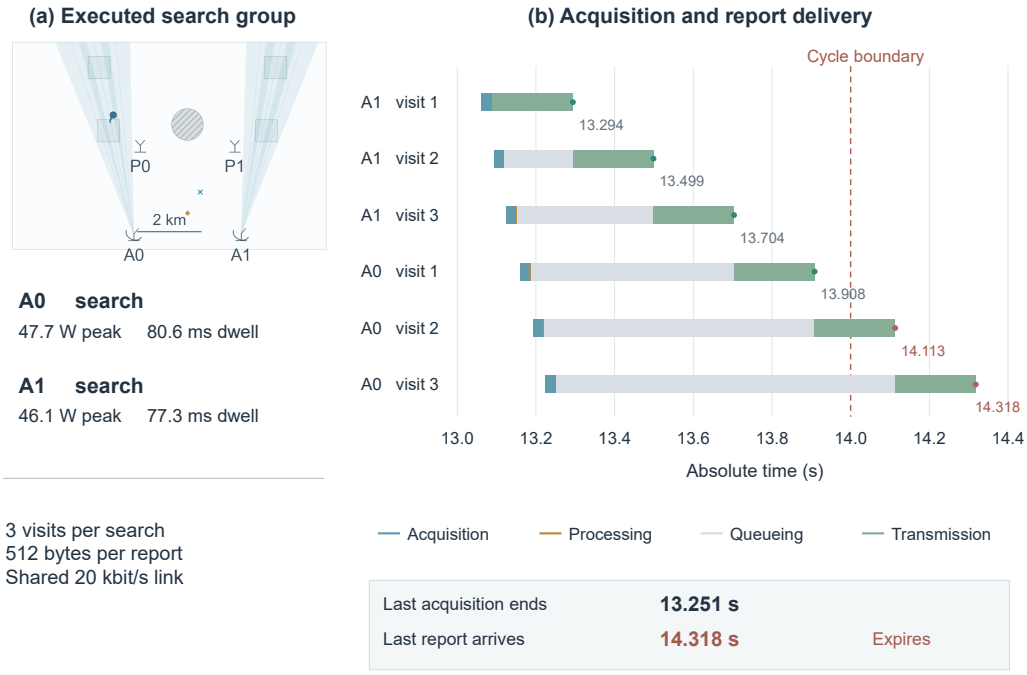


Figure 3: **A shared communication link delays observation arrival.** The selected group and all reports from cycle 13 are shown for the relational proposer in the transport/occlusion episode. Each horizontal bar follows a visit through acquisition, processing, queueing, and transmission. The last acquisition ends at 13.251 s, while its report arrives at 14.318 s, after the acquisition cycle has ended, so the filter discards it. W denotes peak RF power and ms denotes receive dwell; a passive task has zero transmit power.

D ARCHITECTURE, PROJECTION, AND TRAINING DETAILS

Multi-head attention uses scaled dot products (Vaswani et al., 2017); recurrent memory uses gated recurrent units (Cho et al., 2014). Figure 1 shows the control, group-readout, and memory branches. Algorithm 1 gives the complete decision cycle.

D.1 ENCODERS AND PREDICTION HEADS

The encoder has six residual layers of width 384, with eight attention heads and feed-forward width 1536. Let h_i be a normalized entity representation. The relation-masked and dense-attention models use the same multi-head transformation but different attention masks. Deep Sets instead forms a global mean $\bar{h}$ and an entity update $\mathrm{GELU}(W_{\mathrm{local}} h_i + W_{\mathrm{context}} \bar{h})$. All three models use identically structured residual feed-forward layers, role and type embeddings, recurrent memory, and heads for controls, primitive costs, state cost, and group costs. Padding is masked from attention, pooling, and loss. Both attention encoders compute a dense attention matrix; the relational model masks entries to restrict information exchange.

The control and state-value heads use one hidden layer of width 384. The primitive–control encoder maps its 400 inputs through two layers of width 384. The group head maps the pooled 1153-dimensional vector through widths 768 and 384 to two positive outputs. The recurrent update is a gated recurrent unit with input and hidden width 384. Group and state heads predict log-transformed costs.

D.2 ENTITY FEATURES AND THE RELATION MASK

Each entity has a zero-padded feature vector, role and type embeddings, and a recurrent memory indexed by its persistent identity. Bias entities contain normalized means and variances. Track entities contain normalized position and velocity, their 4×4 covariance, confirmation status, and time since the last hit. Sensor entities encode location, signal role, carrier, bandwidth, pointing, and availability. Search entities encode region centers, priority mass, and revisit age; resource and geometry entities encode remaining budgets, surveyed sources, and occluders. Action entities additionally encode task identity, nominal controls, relevant target or region state, and device limits.

For state blocks i, j with component scales s_i, s_j, each joint-covariance entity contains the block dimensions, the zero-padded arrays

$$\widetilde{P}_{ij} = \mathrm{diag}(s_i)^{-1} P_{ij} \,\mathrm{diag}(s_j)^{-1}, \qquad [C_{ij}]_{kl} = \frac{[P_{ij}]_{kl}}{\sqrt{[P_{ii}]_{kk}[P_{jj}]_{ll}}},$$

and both endpoints' normalized means and marginal variances. Each array occupies 4×4 entries, with a scalar bias occupying one row or column. All block pairs, including zero blocks, are present. Position and distance-bias scales are $1000\,\mathrm{m}$, velocity and frequency-bias scales are $50\,\mathrm{m/s}$ and $50\,\mathrm{Hz}$, and angle-bias scale is $0.05\,\mathrm{rad}$. The same numerical block and endpoint information is supplied to every encoder.

Sensor, action, and resource entities also receive device specifications and the current queue state. Waiting times use seconds; processing and communication rates are divided by 2×10^9 operations/s and 10^6 bit/s. Resource scales are 10 J, 0.1 s transmit time, 1 s reception, 10^8 operations, and 4096 bytes. Pulse width and interval use $1\,\mu$s and $100\,\mu$s, respectively. SNR and source powers use $\log_{10}(1 + x)$ in their stated base units; probabilities and availability indicators remain dimensionless. Presence indicators distinguish missing reference, queue, and geometry features from observed zeros. Known availability schedules cover the current and next three cycles. Arrival-time estimates use the queues present at decision time.

Let $M_{ij} \in \{0, 1\}$ mark an allowed attention pair. The mask contains self-loops and undirected versions of the graph relations: sensor ownership, action–resource use, action–track or action–region observation, and calibration–bias observation. Each covariance entity is connected to both of its endpoint variables. Hyperedges form cliques among a sensor and its owned actions, and among a covariance entity and its endpoints. For one attention head of dimension $d_h = 48$,

$$a_{ij} = \frac{M_{ij} \exp(q_i^\top k_j / \sqrt{d_h})}{\sum_{k \in \mathcal{V}} M_{ik} \exp(q_i^\top k_k / \sqrt{d_h})}, \qquad o_i = \sum_{j \in \mathcal{V}} a_{ij} v_j,$$

where $\mathcal{V}$ is the set of valid entities and q_i, k_i, v_i are learned query, key, and value projections of layer-normalized embeddings. Dense attention sets $M_{ij} = 1$ for valid pairs. Relations determine connectivity; the query, key, and value transformations are shared across relations. Each block has an attention residual followed by a layer-normalized feed-forward residual, as expanded in Figure 1.

D.3 Queue priority and forecast feedback

Let $v_{A,1}, v_{A,2}$ denote the two positive log-cost outputs for executable group A. The learned candidates in each queue segment are ordered by increasing predicted cost:

$$Q_\theta(A) = s_0\big(e^{\min(v_{A,1},12)} - 1\big) + \tfrac{1}{4} s_0\big(e^{\min(v_{A,2},12)} - 1\big), \qquad s_0 = 10^4\,\mathrm{m}^2. \tag{26}$$

Ties preserve the incoming group order. The global context $\bar{h}$ in Equation (11) averages all valid entity embeddings, including action and covariance entities. The first output estimates the two-cycle forecast cost. The second estimates discounted tracking cost and is supervised on executed teacher groups.

The feedback vector has 24 components: a presence bit; seven indicators for RF, timing, processing, reference, geometry, risk, and scoring outcomes; an affected-action bit; transformed value, limit, and relative deficit; remaining forecast and construction allowances; used and failed forecast counts; best evaluated total cost and its four components; two outcome-origin indicators; and the refinement index. Action entities receive their matching outcome records; other entities receive the aggregate records. Values and cost components use $\log_{10}(1 + x)$. For a recorded value x and limit l, relative deficit is $(x - l)/(1 + |x| + |l|)$. Forecast counts are divided by 64 and construction allowance by 256. Refinement leaves the belief unchanged. Both forward passes use the memory from the start of the cycle, and the first pass supplies the memory for the next cycle.

Construction inserts waiting and considers pairs assigned to distinct sensors. Calibration–tracking pairs and pairs with the same task and object are ordered by the sum of their analytic primitive priorities; up to six are projected first. Individual actions fill the remaining slots in the initial pool of at most forty groups. All compared proposers use these reservations. The runtime queue starts with overdue-search coverage, the best analytic nonempty reserve, and due calibration, followed by learned or numerical proposals, waiting, and remaining analytic candidates. Projection rejects infeasible groups and removes duplicates. Each distinct projected group is evaluated at most once. Final selection minimizes forecast cost among groups satisfying the revisit rule, with ties broken by a deterministic ordering of the executable schedules.

D.4 Projection and numerical search

The power control maps to a $0.1\,\mathrm{W}$ grid between 2 and $100\,\mathrm{W}$. Dwell is an integer multiple of the $100\,\mu\mathrm{s}$ pulse repetition interval, bounded by the device limits. Pointing selects the nearest admissible 0.05° grid point within 12° of the nominal bearing. Search half-span lies between 0 and 15°. Each of its three visits receives two pulse intervals first; the remaining intervals are allocated in proportion to the three generated weights by largest remainders. All-zero weights map to equal proportions. The start control spans the one-second decision cycle. Projection delays conflicting starts while preserving their requested lower bounds and rejects a group if the completed visits exceed the cycle.

Section C.5 works through the pulse allocation, elapsed dwell, actual transmit time, and RF energy for a three-direction search. A companion passive action adds reception and transport costs without controlled RF transmission. Feasibility depends on both actions' steering, reference ownership, and arrival times.

The nonlearned numerical reference draws twenty scrambled Halton points in sixteen dimensions. The first eight coordinates parameterize one primitive; alternating proposals use the remaining eight coordinates for a compatible second primitive. Primitive identities follow analytic priorities. The learned variant generates up to six initial numerical proposals from its preferred primitives. Numerical variants also repair proposals using feedback after the first six evaluations. The fixed-control variants rank the initial analytic groups without this numerical regeneration pass. The Full–Rank comparison therefore includes both control generation and subsequent refinement.

The construction budget counts attempts to create and project candidates, including rejected attempts; pair enumeration is recorded separately. Each root evaluation includes the continuation

trials and matrix operations required by Equation (8). Evaluation stops after twelve attempts or when no unevaluated group remains.

Algorithm 1 Joint proposal and evaluation in one sensing cycle

Require: Decision-time belief β_t, incoming memory m_{t-1}, trained model f_θ.
Require: Construction budget $M = 192$, initial pool size $K = 40$, evaluation budget $B = 12$.
Ensure: Executed group A_t and recurrent memory m_t.

```
Construct analytic groups, reserving calibration–tracking and same-object pairs, within M and K.
(h_t, u_t, m_t) ← f_θ(β_t, m_{t−1}); project numerical proposals jointly.
Form queue Q: overdue search, reserved analytic group, due calibration, learned proposals, and other groups.
E ← ∅; evaluation history H ← ∅; b ← 0; feedback unused.
while b < B do
    if b ≥ B/2, feedback unused, and construction allowance remains then
        Summarize H and projection outcomes as feedback η_t.
        Recompute proposals from f_θ(β_t, η_t, m_{t−1}); retain the original m_t.
        Prepend jointly projected repairs and refined proposals to Q; mark feedback used.
    end if
    Take the next unevaluated group A from Q; stop if none remains.
    b ← b + 1; forecast A from a fresh copy of β_t using Equation (8).
    Record the outcome in H and each successful score in E.
end while
If search is overdue and an evaluated group meets the coverage rule, retain those groups in E.
Select A_t ∈ arg min_{A∈E} Ĵ(β_t, A); use waiting if E is empty.
Execute A_t, generate its observations, and advance the world by one cycle.
```

D.5 Supervision and loss computation

For evaluated costs $c_j = \widehat{J}(\beta, A_j)$, use the dimensionless transform $\psi(c) = \log(1+c/s_0)$ with $s_0 = 10^4\,\mathrm{m}^2$. The ranking temperature is $\tau = \frac{1}{2}\max(\mathrm{IQR}(c), 100\,\mathrm{m}^2)$, where IQR is the interquartile range over valid group costs. The target and predicted distributions are

$$q_j = \frac{e^{-(c_j - \min_k c_k)/\tau}}{\sum_k e^{-(c_k - \min_l c_l)/\tau}}, \qquad p_j = \frac{e^{-s_0(\exp(\min(v_{j,1},12))-1)/\tau}}{\sum_k e^{-s_0(\exp(\min(v_{k,1},12))-1)/\tau}}. \tag{27}$$

The minimum subtraction improves numerical conditioning without changing q. The temperature assigns similar probabilities to groups with nearly equal costs. Only evaluated valid groups enter either distribution.

Let d_k be realized GOSPA along the teacher trajectory. Its discounted mean squared value is

$$R_t = \frac{\sum_{k=t}^{T-1} \gamma^{k-t} d_k^2}{\sum_{k=t}^{T-1} \gamma^{k-t}}, \qquad \gamma = 0.95. \tag{28}$$

Dividing by the sum of discount weights averages the remaining trajectory costs, including near the end of an episode. The target remains in square meters and supervises the state value and the future-cost output for the executed group. Each primitive-cost target is the minimum recorded cost among groups containing that primitive. Control targets are the teacher's projected values, with scan weights normalized by their maximum. Future-cost regression is applied only to the executed group, for which the subsequent trajectory is observed.

For residual e, the unit-transition Huber loss used here is

$$\mathrm{Huber}(e) = \begin{cases} e^2/2, & |e| \le 1, \\ |e| - 1/2, & |e| > 1. \end{cases} \tag{29}$$

For a batch, $\mathcal{L}_{\mathrm{group}}$ averages $\mathrm{Huber}(v_{j,1} - \psi(c_j))$ over all valid evaluated groups. The control term averages $\mathrm{Huber}(\hat{u}_{a,d} - \tilde{u}_{a,d})$ over applicable coordinates of primitives in executed teacher groups.

The primitive-cost term averages its log-cost regression over primitives appearing in at least one evaluated group. The state and future terms average $\mathrm{Huber}(\hat{v} - \psi(R_t))$ over worlds, with the group future value selected at the executed group. The ranking divergence is first summed across valid groups per world and then averaged across worlds. Padding is excluded from every reduction.

The initial and feedback passes use the same recorded targets and memory input, and their losses are averaged. Only the recurrent state from the initial pass is carried forward, so within-cycle refinement does not advance the temporal state.

D.6 Optimization and partitioning

The optimizer is AdamW (Loshchilov & Hutter, 2019) with learning rate 2×10^{-4} and weight decay 10^{-4}. The learning rate warms up over 500 updates and follows a cosine schedule that reaches 10% of its initial value at update 64,000. Training stops after the first 32,000 updates, corresponding to ten epochs. Gradients are clipped at Euclidean norm 5. Computation uses bfloat16 autocasting on supported GPUs, with losses computed in float32. Validation runs after each epoch, and the checkpoint with the lowest total validation loss is selected, with the earlier epoch winning a tie.

Each batch contains 32 complete trajectories, with 64 cycles per trajectory and truncated backpropagation over eight consecutive cycles. An epoch therefore contains $(12{,}800/32)(64/8) = 3200$ updates, giving 32,000 updates over ten epochs. The $819{,}200$ distinct training contexts are revisited over epochs. Ten epochs give $8{,}192{,}000$ context visits per model. Half of the trajectories come from each teacher, with 800 trajectories per teacher in each regime. Validation is balanced in the same way, using 128 disjoint worlds. Development testing and the 192-episode confirmation test are also balanced across the eight regimes. Checkpoint selection uses validation loss; the reported tables use confirmation episodes. Separate random processes govern motion, birth, emission, detection, clutter, and transport. Methods are paired using common world seeds.

D.7 Sequence learning

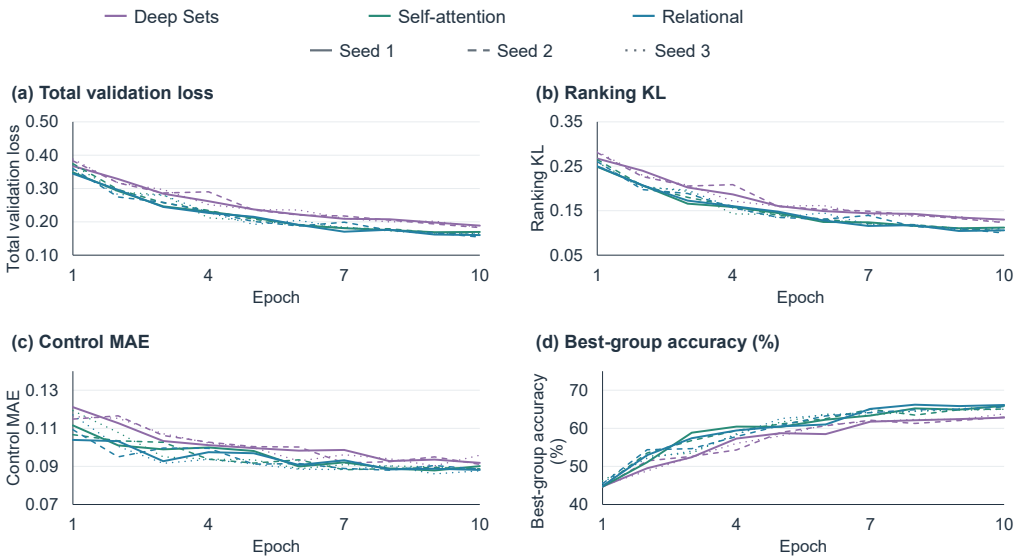
Algorithm 2 Trajectory supervision with one memory update per cycle

Require: Teacher trajectories of decision-time beliefs, evaluated groups, executed controls, and realized errors.
Require: Model f_θ, world batch size 32, truncation length 8, training epochs 10.
Ensure: Model parameters from the epoch with the lowest validation loss.

```
Compute discounted realized targets R_t by Equation (28).
Form group costs, ranking targets, and applicability masks from the teacher archive.
for each training epoch do
    Shuffle whole trajectories and partition them into batches of 32 worlds.
    for each batch of worlds do
        Initialize entity memories at the beginning of each trajectory.
        for each consecutive block of eight cycles do
            for each cycle t in the block do
                Align incoming memory m_{t-1} with persistent entity identities.
                Evaluate the initial pass and retain its outgoing memory m_t.
                Evaluate recorded feedback with the same incoming memory m_{t-1}.
                Average the two masked losses in Equation (12).
                Apply realized future supervision to the executed group and state value.
            end for
            Average valid losses; backpropagate through the block; clip gradients at norm 5.
            Take one AdamW update; detach the carried memory at the block boundary.
        end for
    end for
    Measure validation loss and retain the lowest-loss completed epoch; break ties by earliest epoch.
end for
```

All evaluated teacher groups provide forecast labels, and the executed group also provides an observed future-error target. The initial and feedback passes share the memory at the start of the cycle; recurrent time advances only when the cycle advances.

D.8 LEARNING TRAJECTORIES



Selected checkpoints	Deep Sets		Self-attention		Relational	
Seed 1	Epoch 10	0.1887	Epoch 9	0.1689	Epoch 10	0.1603
Seed 2	Epoch 10	0.1829	Epoch 10	0.1599	Epoch 10	0.1543
Seed 3	Epoch 10	0.1833	Epoch 9	0.1640	Epoch 10	0.1661

Figure 4: **Validation trajectories across encoders and initializations.** Each curve contains ten evaluations on the same 128 held-out episodes. Colors identify encoders, and line styles identify their three training seeds. The total objective and ranking divergence decrease over training in all nine runs. Control mean absolute error (MAE) measures error in applicable normalized controls; best-group accuracy measures agreement with the lowest-cost evaluated teacher group. The total validation objective selects the checkpoint used in closed-loop evaluation.

For a validation batch $\mathcal{B}$, let $\hat{u}_{bak}$ be a predicted normalized control, $u^{\star}_{bak}$ its executed teacher target, and $m_{bak} \in \{0, 1\}$ the applicability mask for control k of action a in episode b. The control error is

$$\mathrm{MAE}_u(\mathcal{B}) = \frac{\sum_{b\in\mathcal{B},a,k} m_{bak}|\hat{u}_{bak} - u^{\star}_{bak}|}{\max(1, \sum_{b\in\mathcal{B},a,k} m_{bak})}. \tag{30}$$

Let c_{bj} be the physical cost of teacher-evaluated group j in state b, and $\hat{j}_b$ the network's predicted minimum-cost group. Best-group accuracy is

$$\mathrm{Acc}(\mathcal{B}) = \frac{1}{|\mathcal{B}|}\sum_{b\in\mathcal{B}} \mathbf{1}\{c_{b\hat{j}_b} - \min_j c_{bj} \le \varepsilon\}. \tag{31}$$

Here $\mathbf{1}\{\cdot\}$ is the indicator function, and $\varepsilon = 10^{-6}\,\mathrm{m}^2$ resolves numerical ties. Validation averages the initial and feedback passes when both are supervised, then averages batches and cycles weighted by episode count.

E EVALUATION SETTINGS

E.1 METHODS AND CONTROLLED COMPARISONS

All methods receive the same state features and use the same filter, measurement model, projection procedure, resource limits, and budget of twelve forecast evaluations per cycle. The analytic baseline uses prescribed controls and priorities; numerical search uses twenty scrambled Halton proposals and local repair. Each encoder is tested with Rank and Full settings. The relational encoder is also evaluated with fixed controls and feedback, and with generated controls but no feedback. The eight learned settings each use three training seeds; together with the two nonlearned baselines, this gives 26 runs per test world and 4,992 completed evaluations over 192 worlds.

The encoder comparison holds entity width, depth, recurrence, and output heads fixed. The inference-time ablation holds trained weights fixed while switching control generation and feedback independently. During closed-loop evaluation, each method chooses its own sensing actions and updates its belief from the resulting measurements. Pairing methods on the same world captures the effects of these choices on later decisions.

F METRICS AND STATISTICAL ANALYSIS

Multi-target estimation. Generalized optimal sub-pattern assignment (GOSPA) combines localization error with missed-target and false-track penalties. At cycle t, let $x_{i,t} \in \mathbb{R}^2$ be the true position of target i and $\widehat{x}_{j,t}$ the position of confirmed track j. Let $\mathcal{A}_t$ be the minimum-cost one-to-one assignment with matched distances below $c = 250\,\mathrm{m}$. Write N_t^{miss} and N_t^{false} for the numbers of unmatched true targets and unmatched confirmed tracks. With order $p = 2$ and cardinality parameter $\alpha = 2$, GOSPA is

$$G_t = \left[\sum_{(i,j)\in\mathcal{A}_t} \|x_{i,t} - \widehat{x}_{j,t}\|_2^2 + \frac{c^2}{2}\left(N_t^{\mathrm{miss}} + N_t^{\mathrm{false}}\right) \right]^{1/2}. \tag{32}$$

The episode score is $G = T^{-1}\sum_{t=1}^{T} G_t$ for $T = 64$ cycles. False tracks are $T^{-1}\sum_t N_t^{\mathrm{false}}$.

Position error and target recall. Let $M_t = |\mathcal{A}_t|$ and let N_t be the number of true targets at cycle t. The episode position root mean square error (RMSE) and target recall are

$$R_x = \left[\frac{\sum_t \sum_{(i,j)\in\mathcal{A}_t} \|x_{i,t} - \widehat{x}_{j,t}\|_2^2}{\sum_t M_t} \right]^{1/2}, \tag{33}$$

$$C = 100\,\frac{\sum_t M_t}{\sum_t N_t} \quad (\%). \tag{34}$$

Velocity RMSE uses the corresponding velocity errors. The position-error quantile is computed within each episode and then averaged over episodes. Recall is defined for episodes containing at least one true target. Conditional error metrics require at least one confirmed match in every trained model evaluated on that episode.

Discovery and continuity. The unconfirmed-target count is the number of targets that appear but are never matched to a confirmed track before the episode ends. ID switches count changes in the confirmed-track identity assigned to the same true target. The maximum tracking gap is the largest run of unmatched cycles after a first confirmation, multiplied by the one-second cycle duration. It is evaluated on episodes with at least one confirmed match.

Receiver bias. For a bias type b with d_b state components, its final error is

$$R_b = \left[\frac{1}{d_b} \sum_{\ell=1}^{d_b} \left(\widehat{b}_{\ell,T} - b_{\ell,T}\right)^2 \right]^{1/2}. \tag{35}$$

Range bias uses meters, frequency bias uses hertz, and angular bias is converted from radians to degrees. All three groups contain four effective receiver-state components in this experiment.

F.1 RESOURCE ACCOUNTING AND AGGREGATION

Resource accounting. For a transmitted action a, let P_a be peak RF power, n_a the number of transmitted pulses and τ_a pulse width. Let $\mathcal{E}$ be the set of executed actions over an episode; passive actions have $n_a = 0$. Then

$$E_{\mathrm{RF}} = \sum_{a \in \mathcal{E}} P_a n_a \tau_a, \qquad T_{\mathrm{tx}} = \sum_{a \in \mathcal{E}} n_a \tau_a, \tag{36}$$

$$E_{\mathrm{rx}} = \sum_{a \in \mathcal{E}} P_{\mathrm{rx},a} L_a, \qquad D = \sum_{a \in \mathcal{E}} \frac{P_a n_a \tau_a}{4\pi r_a^2}. \tag{37}$$

Here $P_{\mathrm{rx},a}$ is receiver electrical power, L_a is accounted receiver listening time, and r_a is the distance from the transmitting node to the fixed observation point. D is the incident energy density under isotropic propagation at that observation point. The tables use $10^3 T_{\mathrm{tx}}$ in milliseconds and $10^9 D$ in nJ/m^2.

Scenario-balanced estimates. Let y_{mswk} denote an episode metric for method m, scenario s, test episode w and training run k. Let $K_m = 3$ for learned methods and $K_m = 1$ for a nonlearned baseline. A test episode is included for a given metric only when that metric is defined for every trained model of the method. Its score is

$$\overline{y}_{msw} = \frac{1}{K_m} \sum_{k=1}^{K_m} y_{mswk}, \tag{38}$$

$$\widehat{\mu}_m = \frac{1}{8} \sum_{s=1}^{8} \frac{1}{|\mathcal{W}_{ms}|} \sum_{w \in \mathcal{W}_{ms}} \overline{y}_{msw}, \tag{39}$$

where $\mathcal{W}_{ms}$ is the set of valid test episodes for that method and scenario. Thus each scenario has equal weight. For a difference between methods a and b, the mean is computed on their common valid test episodes:

$$\widehat{\Delta}_{a,b} = \frac{1}{8} \sum_{s=1}^{8} \frac{1}{|\mathcal{W}_{as} \cap \mathcal{W}_{bs}|} \sum_{w \in \mathcal{W}_{as} \cap \mathcal{W}_{bs}} \left(\overline{y}_{asw} - \overline{y}_{bsw} \right). \tag{40}$$

For conditional metrics, the paired difference uses the intersection of valid episodes and can therefore differ from the difference between the separately reported means.

Intervals and training variation. The 95% intervals are the 2.5th and 97.5th percentiles of 10,000 bootstrap draws. Each draw resamples test episodes within each scenario, preserving all paired methods and the three trained models. The primary comparison is GOSPA for the relational Full model versus numerical search. Secondary contrasts report marginal 95% intervals without multiplicity adjustment. In Table A4, the standard deviation is instead computed across the three training-run means. It describes variation across training runs; the bootstrap intervals describe variation across sampled test episodes.

Inference-time ablations. Let μ_{nf} denote the mean with learned action parameters enabled when $n = 1$ and evaluation feedback enabled when $f = 1$. The interaction contrast is

$$\Delta_{\mathrm{int}} = (\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00}). \tag{41}$$

Feedback acts within a scheduling decision. With fixed action parameters it changes the order of unscored candidates; with learned parameters it also refines proposals. All settings use the same belief update and two-cycle forecast.

G Additional experimental results

G.1 Results across sensing scenarios

The aggregate scores in Section 4 give equal weight to eight sensing scenarios. Table A3 reports GOSPA and target recall for each scenario, using the same methods and checkpoints as the main comparison. Range and source strength affect measurement precision, while emission, outages and transport affect the availability of information.

Table A3: **Results by sensing scenario.** Each scenario has 24 test episodes; learned-method results average three training seeds.

Method		Near range		Transition range		Far / weak return		Intermittent emission	
Proposer	Variant	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑
Analytic baseline	—	61.78	81.89	83.08	69.44	73.31	78.54	76.99	76.02
Numerical search	—	49.97	84.52	76.64	72.94	61.81	83.64	67.44	77.93
Deep Sets	Rank	62.26	81.43	78.14	71.93	69.15	81.24	68.61	78.32
Deep Sets	Full	54.02	84.71	62.57	78.89	61.26	84.84	58.53	81.56
Self-attention	Rank	60.72	82.01	77.85	71.97	68.75	81.23	66.32	79.22
Self-attention	Full	55.86	84.34	**58.13**	**80.82**	56.90	85.79	**53.49**	**83.18**
Relational (ours)	Rank	61.09	81.59	74.93	72.89	67.36	81.63	67.68	78.57
Relational (ours)	Full	**46.52**	**86.80**	62.55	78.60	**53.70**	**87.02**	56.09	82.75
Episodes with defined metric		24	24	24	23	24	24	24	24

Method		Late arrival / tight budget		Maneuver / node outage		Weak illumination		Transport / occlusion	
Proposer	Variant	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑	GOSPA m ↓	Recall % ↑
Analytic baseline	—	43.35	69.83	66.49	82.13	79.18	71.70	112.83	65.43
Numerical search	—	43.23	71.18	59.97	81.63	62.75	78.89	121.47	60.87
Deep Sets	Rank	43.50	69.42	62.59	81.76	72.28	73.27	115.72	63.15
Deep Sets	Full	**34.42**	**76.95**	53.55	83.41	60.41	79.84	98.92	69.52
Self-attention	Rank	44.17	69.25	60.23	82.44	69.46	75.69	115.43	63.17
Self-attention	Full	39.33	74.66	54.51	82.79	**56.00**	**81.00**	101.88	68.26
Relational (ours)	Rank	44.16	69.25	63.07	81.60	71.52	74.70	115.71	63.03
Relational (ours)	Full	37.88	76.42	**50.50**	**84.40**	56.11	80.73	**98.59**	**69.81**
Episodes with defined metric		24	20	24	23	24	24	24	24

Recall excludes six episodes containing no targets. Bold marks the best value at the reported precision for each scenario and metric.

The relational Full model improves on numerical search in both metrics in every scenario mean. The largest absolute GOSPA difference is 22.88 m with transport and occlusion; target recall in that scenario increases from 60.87% to 69.81%. Weak illumination produces a smaller GOSPA difference, from 62.75 to 56.11 m. Comparing encoders reveals a different pattern: Self-attention has the lowest mean GOSPA in transition range, intermittent emission and weak illumination, and Deep Sets has the lowest value for late arrivals with a tight energy budget. Learned proposals improve on numerical search across scenarios, while the best-performing encoder depends on the sensing conditions.

G.2 VARIATION ACROSS TRAINING SEEDS

Table A4 evaluates each training run separately using its lowest-validation-loss checkpoint. The run-level rows show the nine models underlying the encoder averages. The summary rows report the sample standard deviation across three run means.

Table A4: **Variation across training seeds.** Each run uses its lowest-validation-loss checkpoint and the Full variant on 192 test episodes.

Encoder	Run	Epoch	GOSPA m ↓	RF energy J ↓	Incident energy density nJ/m^2 ↓
Deep Sets	1	10	60.74	2.465	1.783
	2	10	59.76	2.427	1.728
	3	10	60.88	2.425	1.721
Mean ± SD			60.46 ± 0.61	2.439 ± 0.023	1.744 ± 0.034
Self-attention	1	9	57.50	2.264	1.609
	2	10	60.86	2.385	1.715
	3	9	60.17	2.406	1.732
Mean ± SD			59.51 ± 1.78	2.352 ± 0.076	1.686 ± 0.067
Relational (ours)	1	10	56.93	2.501	1.769
	2	10	57.08	2.498	1.774
	3	10	59.22	2.480	1.777
Mean ± SD			57.74 ± 1.29	2.493 ± 0.011	1.773 ± 0.004

SD is the sample standard deviation across the three run means. Checkpoints are selected by validation loss over epochs 1–10.

The three relational runs obtain mean GOSPA values of 56.93, 57.08 and 59.22 m, giving 57.74 ± 1.29 m across training runs. The corresponding summaries are 60.46 ± 0.61 m for Deep Sets and 59.51 ± 1.78 m for Self-attention. All three relational runs improve on the numerical-search mean of 67.91 m. Self-attention's lower RF expenditure is also present in every run: its three means range from 2.264 to 2.406 J, compared with 2.480 to 2.501 J for the relational model.

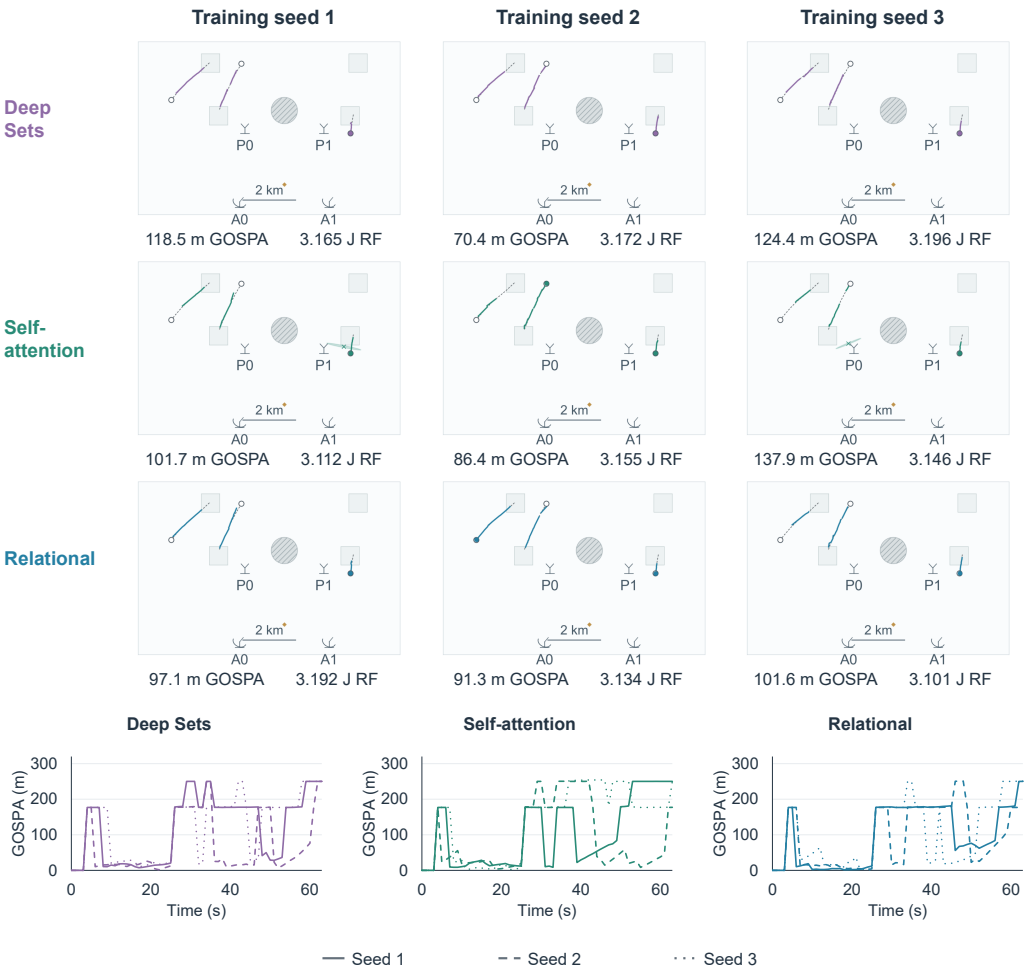


Figure 5: **Training initialization changes the executed trajectory.** Three encoders and their three training seeds act in the same transport/occlusion episode. Each panel shows all trajectories through 63 s and reports the episode-mean GOSPA and total RF energy. Each run uses its checkpoint with the lowest validation loss. Seeds 1, 2, and 3 correspond to the three initializations used throughout the tables.

G.3 TRACK CONTINUITY AND RECEIVER-STATE ESTIMATION

Table A5 complements GOSPA and recall with errors for matched tracks, continuity measures, and final receiver-bias error. These metrics distinguish track accuracy and persistence from target discovery.

Table A5: **Tracking continuity and receiver calibration.** Values are test-episode means.

A. Localization and track continuity

Proposer	Variant	n	Position error 95th percentile m ↓	Velocity RMSE m/s ↓	Max. tracking gap s ↓
Analytic baseline	—	185	34.98	2.89	7.25
Numerical search	—	185	32.88	3.06	6.07
Deep Sets	Rank	184	35.16	2.85	7.36
Deep Sets	Full	184	33.32	2.95	6.51
Self-attention	Rank	185	35.74	**2.83**	7.15
Self-attention	Full	184	32.31	2.85	5.56
Relational (ours)	Rank	185	34.73	2.85	7.18
Relational (ours)	Full	184	**31.91**	2.85	**5.42**

B. Confirmation, identity and final bias estimates

Method		Target tracking		Final bias RMSE		
Proposer	Variant	Unconfirmed per episode ↓	ID switches per episode ↓	Range m ↓	Frequency Hz ↓	Angle deg ↓
Analytic baseline	—	0.151	0.359	6.64	3.10	0.158
Numerical search	—	0.135	0.391	6.23	3.08	0.145
Deep Sets	Rank	0.134	0.262	6.12	3.03	0.150
Deep Sets	Full	**0.087**	0.325	6.48	**2.92**	0.140
Self-attention	Rank	0.122	0.286	**6.09**	2.94	0.151
Self-attention	Full	0.123	0.286	6.76	3.00	**0.139**
Relational (ours)	Rank	0.127	0.283	6.10	2.99	0.151
Relational (ours)	Full	0.118	**0.234**	6.34	**2.92**	0.141

n counts episodes with defined conditional errors and also applies to position RMSE in Table 1. The maximum tracking gap starts after a target is first confirmed. Panel B uses all 192 episodes. Bold marks the lowest value at the reported precision.

The relational Full model has a mean maximum tracking gap of 5.42 s and 0.234 identity switches per episode, compared with 6.07 s and 0.391 for numerical search. Deep Sets yields the fewest unconfirmed targets, at 0.087 per episode. Final bias errors are similar across methods: range-bias RMSE spans 6.09–6.76 m, frequency-bias RMSE 2.92–3.10 Hz, and angular-bias RMSE 0.139–0.158 degrees. The differences in discovery and track continuity occur alongside similar final receiver-bias errors.

G.4 PAIRED EFFECTS AND INFERENCE-TIME ABLATIONS

Table A6 compares methods on their common valid test episodes. Panel A reports the relational Full model minus numerical search. Panel B reports the paired effects of control generation and feedback. The resampling unit is a test episode within a scenario; the three trained models remain fixed and paired during resampling (Agarwal et al., 2021).

Table A6: **Paired comparisons and ablation effects.** Intervals are obtained by stratified bootstrap resampling of test episodes.

A. Full relational proposer minus numerical search

Metric	Difference	95% CI	n
GOSPA (m)	-10.17	[-14.30, -6.05]	192
Position RMSE (m)	-0.73	[-2.37, +0.85]	184
Target recall (pp)	+4.37	[+2.58, +6.17]	186
False tracks (per cycle)	+0.0028	[-0.0031, +0.0086]	192
RF energy (J)	-0.188	[-0.215, -0.159]	192
Receiver energy (J)	+0.61	[-0.06, +1.30]	192
Tx on-time (ms)	-3.42	[-4.20, -2.66]	192
Incident energy density (nJ/m^2)	-0.119	[-0.141, -0.095]	192
Position error 95th percentile (m)	-1.06	[-3.84, +1.72]	184
Velocity RMSE (m/s)	-0.15	[-0.36, +0.04]	184
Max. tracking gap (s)	-0.67	[-1.56, +0.24]	184
Unconfirmed targets (per episode)	-0.017	[-0.062, +0.028]	192
ID switches (per episode)	-0.156	[-0.238, -0.076]	192
Range bias RMSE (m)	+0.10	[-0.67, +0.82]	192
Frequency bias RMSE (Hz)	-0.15	[-0.48, +0.17]	192
Angle bias RMSE (deg)	-0.004	[-0.027, +0.017]	192

B. Interventions within the relational proposer

Intervention	Δ GOSPA m	Δ recall pp	Δ RF energy J
Learned parameters; feedback off	-11.67 [-15.35, -8.03]	+5.24 [+3.73, +6.77]	-0.188 [-0.219, -0.157]
Learned parameters; feedback on	-13.14 [-17.25, -9.13]	+5.54 [+3.93, +7.13]	-0.190 [-0.218, -0.161]
Feedback; fixed parameters	+0.19 [-0.72, +1.15]	-0.13 [-0.69, +0.39]	+0.001 [-0.003, +0.006]
Feedback; learned parameters	-1.28 [-4.15, +1.45]	+0.17 [-0.93, +1.23]	-0.001 [-0.021, +0.020]
Parameters × feedback	-1.47 [-4.34, +1.32]	+0.30 [-0.83, +1.40]	-0.003 [-0.023, +0.018]

CI: confidence interval; pp: percentage points. Panel B gives each difference with its 95% CI below. Differences use common valid episodes: $n = 192$ for GOSPA and energy, and $n = 186$ for recall. The resampling procedure and interaction contrast are defined in Appendix F.

The primary GOSPA difference is -10.17 m, with a 95% interval of $[-14.30, -6.05]$ m. Recall increases by 4.37 percentage points, with an interval of $[2.58, 6.17]$. In the inference-time ablation, control generation reduces GOSPA by 11.67 m without feedback and by 13.14 m with feedback. The incremental feedback effect with generated controls is -1.28 m, with an interval of $[-4.15, 1.45]$ m; the interaction interval also spans zero. Control generation accounts for most of the GOSPA reduction; the estimated feedback effect is smaller and its interval includes zero.

H CLOSED-LOOP SCENE SEQUENCES

Each sequence follows one complete test episode. Within each scenario, we sort the 24 episodes by the numerical baseline's mean GOSPA and select rank 12, breaking ties by world seed. The displayed learned policies use training seed 1 and their validation-selected checkpoints. All policies share target motion and scene geometry but acquire observations through their own sensing actions.

Rows show numerical search and the three Full encoders. Columns show common times: two seconds after each target arrival and the final time, 63 s. The single-target maneuver/outage episode instead includes outage and recovery. Dashed paths are truth, solid paths are confirmed estimates, and ellipses are 95% marginal position contours at their physical scale. Crosses denote tentative tracks. Pale wedges show recorded beam directions; squares are search regions, the hatched disc is an occluder, and the diamond is the surveyed calibration source. A0/A1 are active sensors and P0/P1 passive receivers. Each panel has a distance scale and gives the number of matched targets over targets present.

Below each panel, S, T, and C denote search, tracking, and calibration; the values are executed peak power in watts and receive dwell in milliseconds. Row summaries give episode-mean GOSPA and energy. Both lower plots cover all 64 cycles and include the analytic baseline. Energy is controlled RF expenditure except in the late-arrival episode, where all methods use passive reception and receiver electrical energy is shown instead.

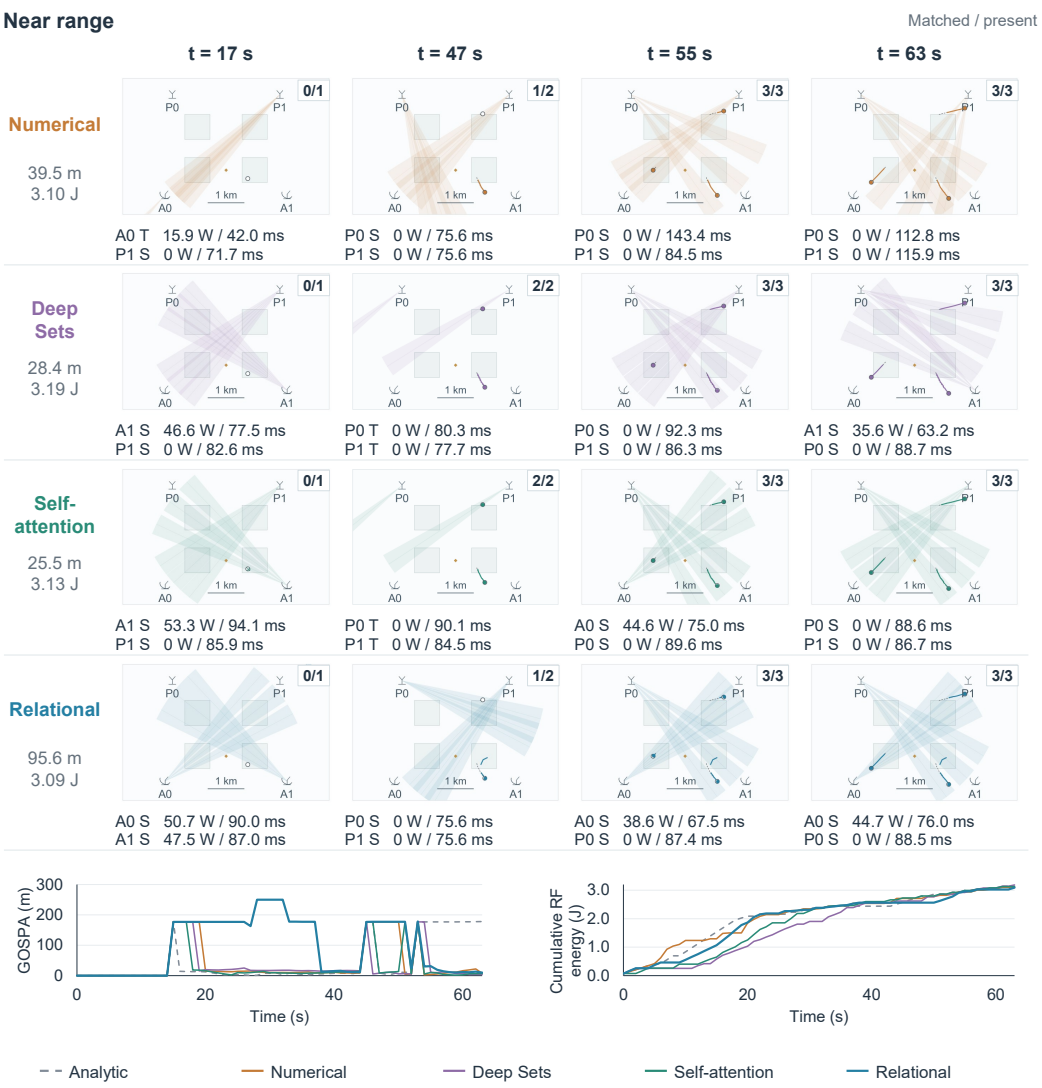


Figure 6: **Near-range acquisition and recovery.** All four displayed methods match the three targets by 55 s, although their earlier tracking histories differ. The relational run has an episode-mean GOSPA of 95.6 m; Self-attention has 25.5 m. The episode curves show the earlier periods of missed or inaccurate tracking.

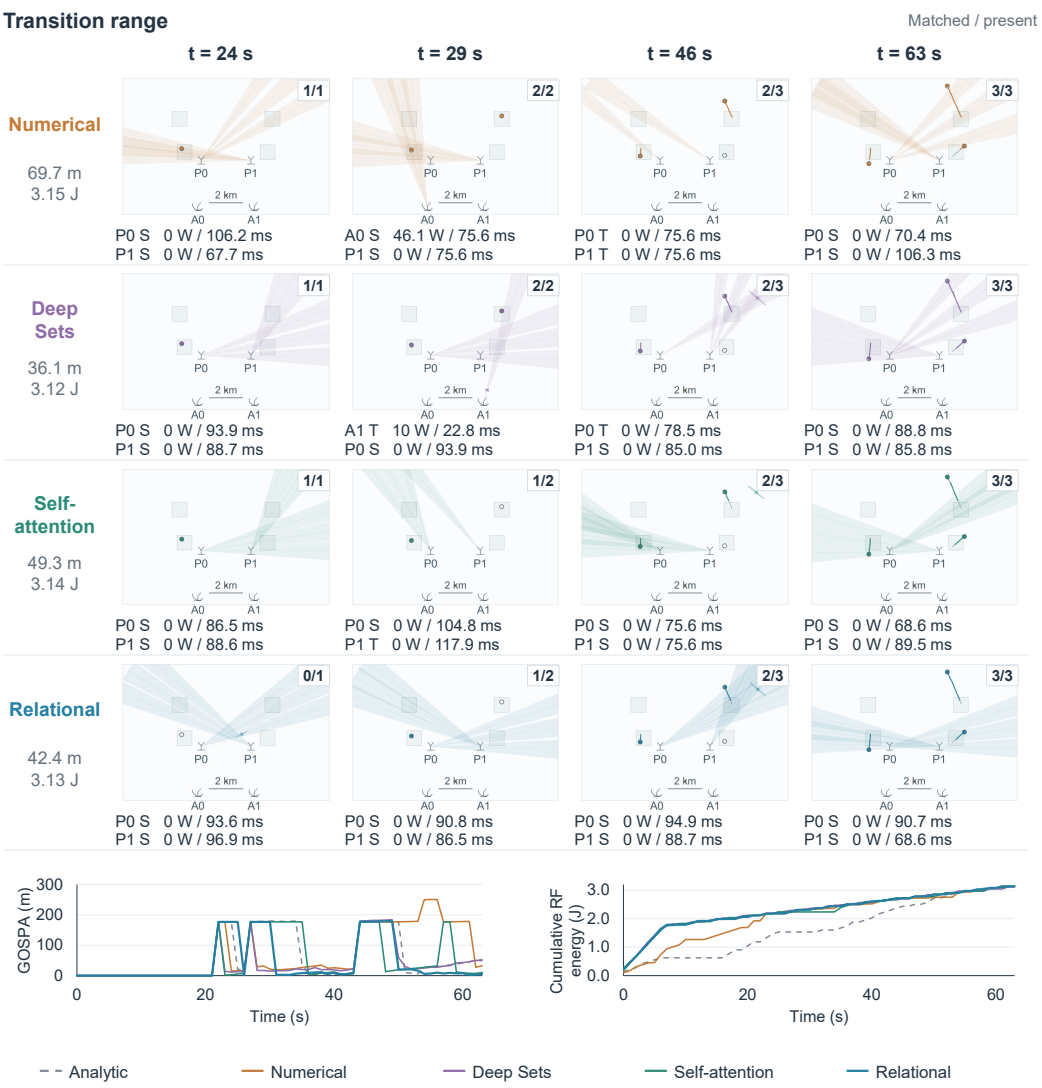


Figure 7: **Successive arrivals at transition range.** The methods differ in how quickly they establish tracks after the first two arrivals. All match three targets at 63 s. Episode-mean GOSPA ranges from 36.1 m for Deep Sets to 69.7 m for numerical search.

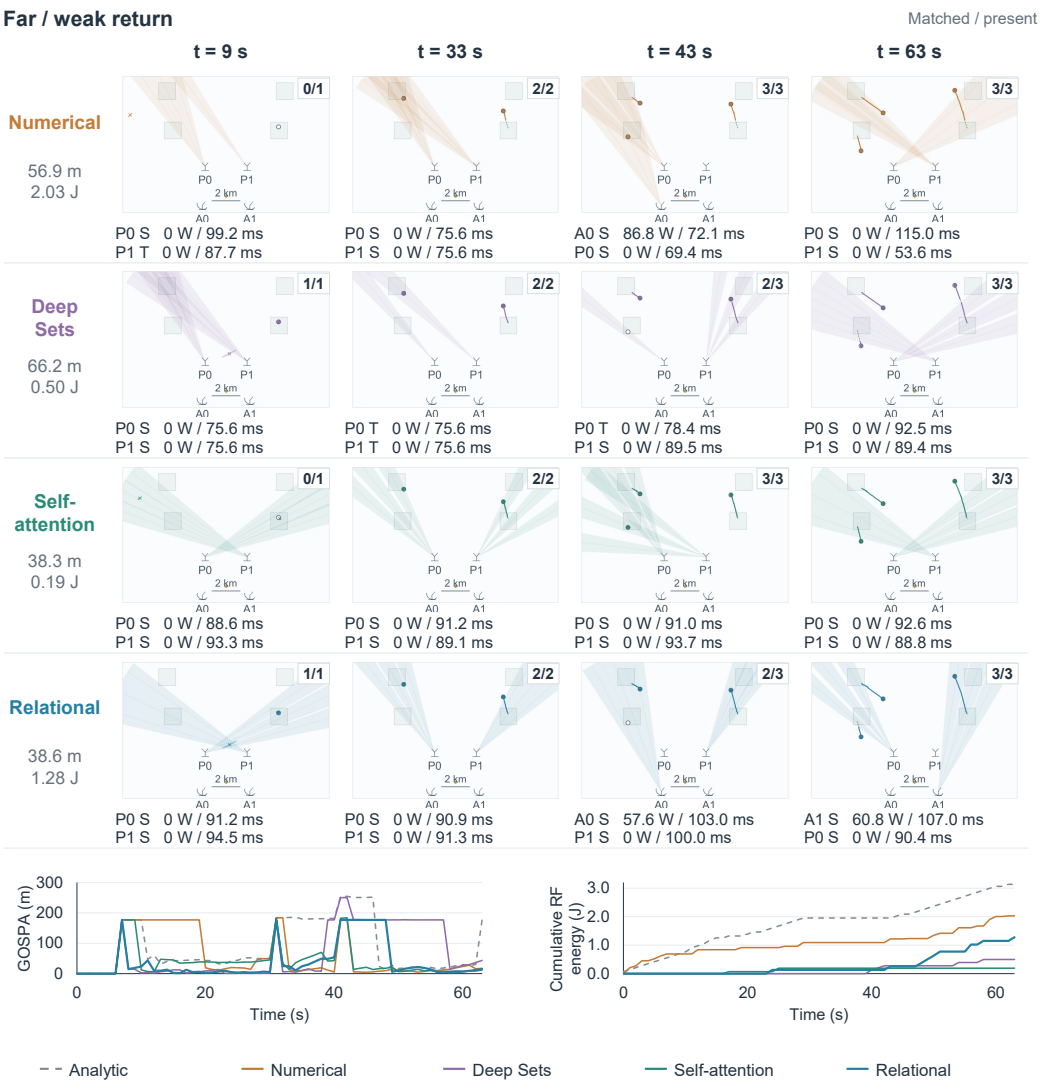


Figure 8: **Tracking with weak distant returns.** The displayed policies combine active search and passive reception in different proportions. Self-attention and the relational model obtain mean GOSPA of 38.3 and 38.6 m, with RF expenditures of 0.191 and 1.278 J; numerical search uses 2.029 J and obtains 56.9 m.

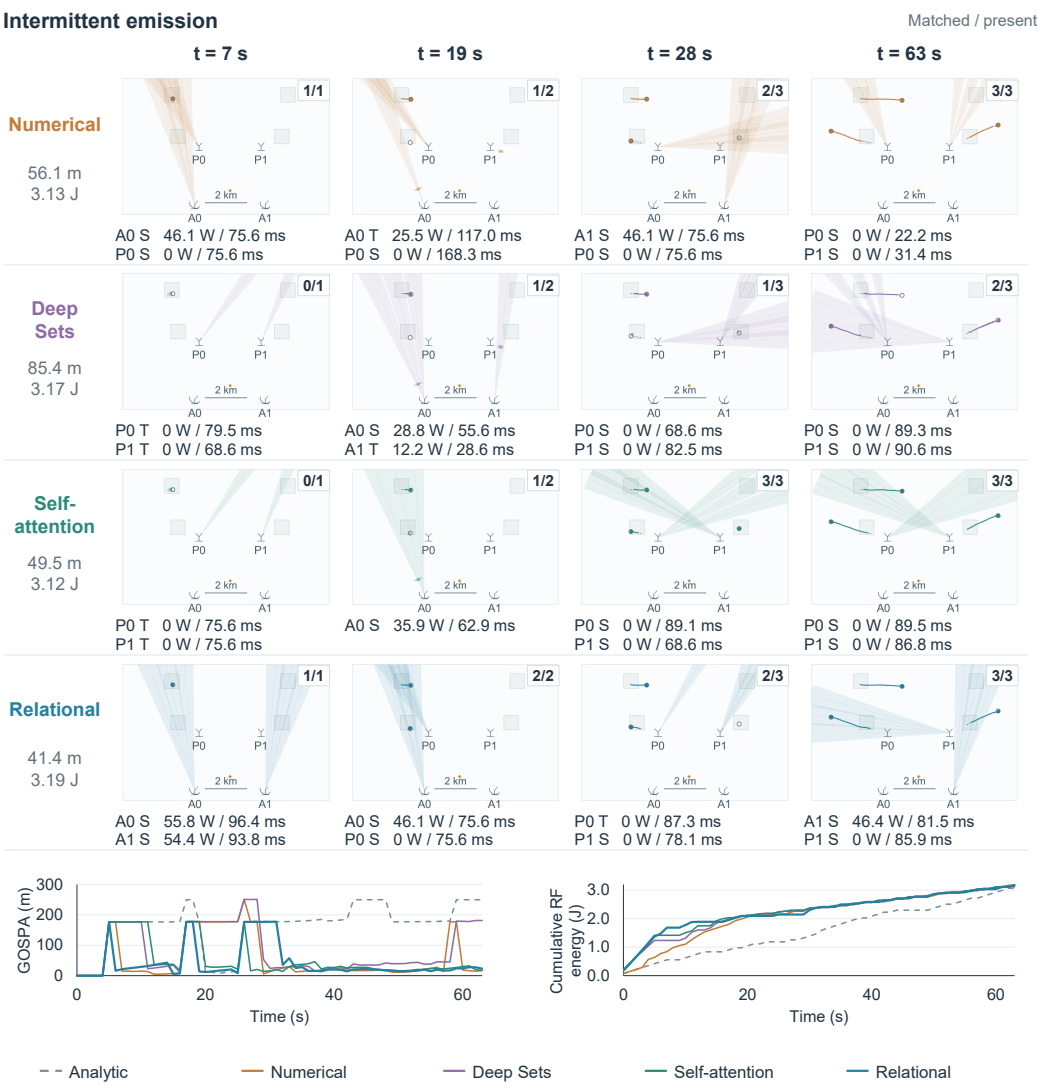


Figure 9: **Track formation under intermittent emission.** At 19 s the relational run matches both present targets, while the other displayed methods match one. At the end of the episode, Deep Sets matches two of three targets and the other methods match all three. The time series locate these changes between the selected frames.

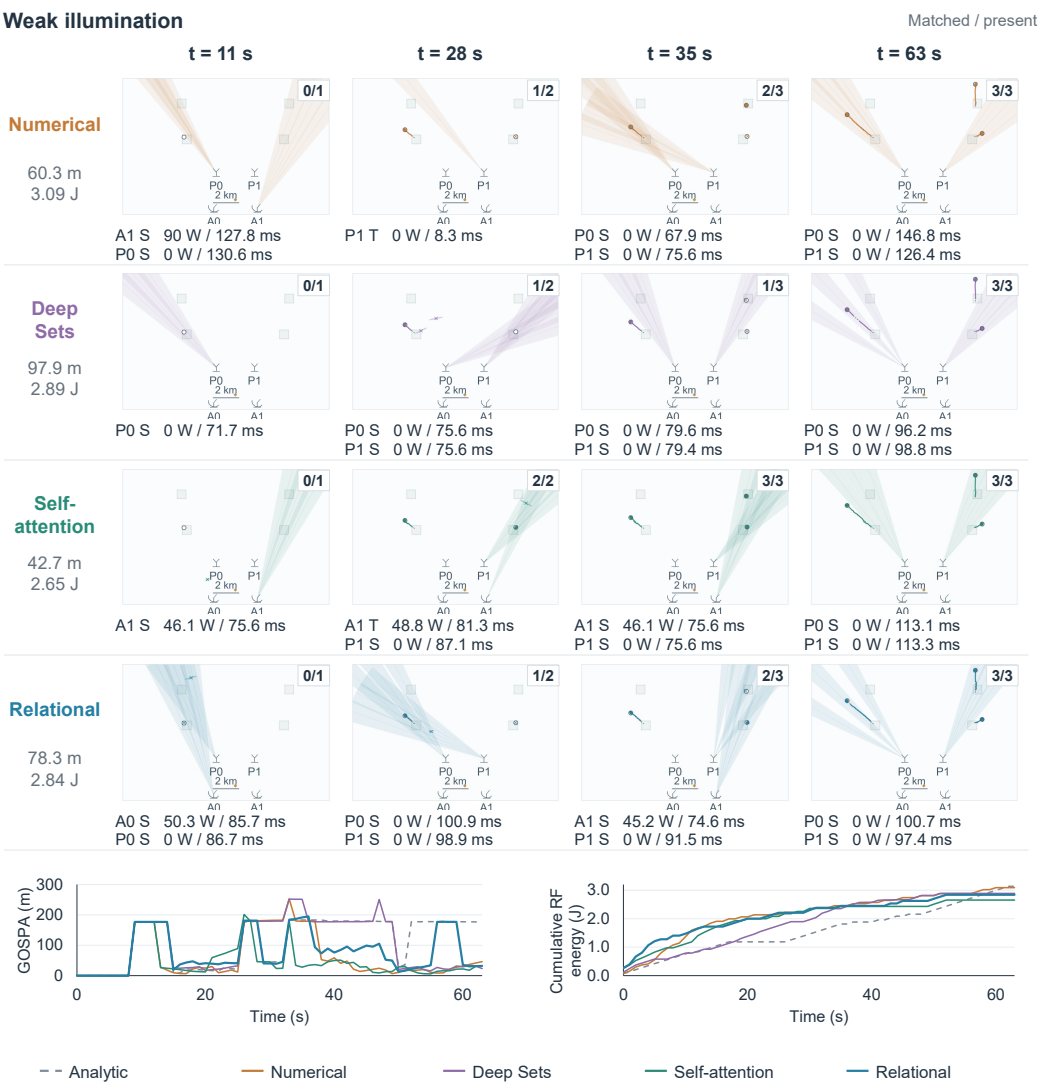


Figure 10: **Track acquisition with weak illumination.** Self-attention matches all three targets by 35 s, while the other displayed methods still miss at least one. All four match the three targets at 63 s. The episode curves show how long each method leaves targets unmatched.

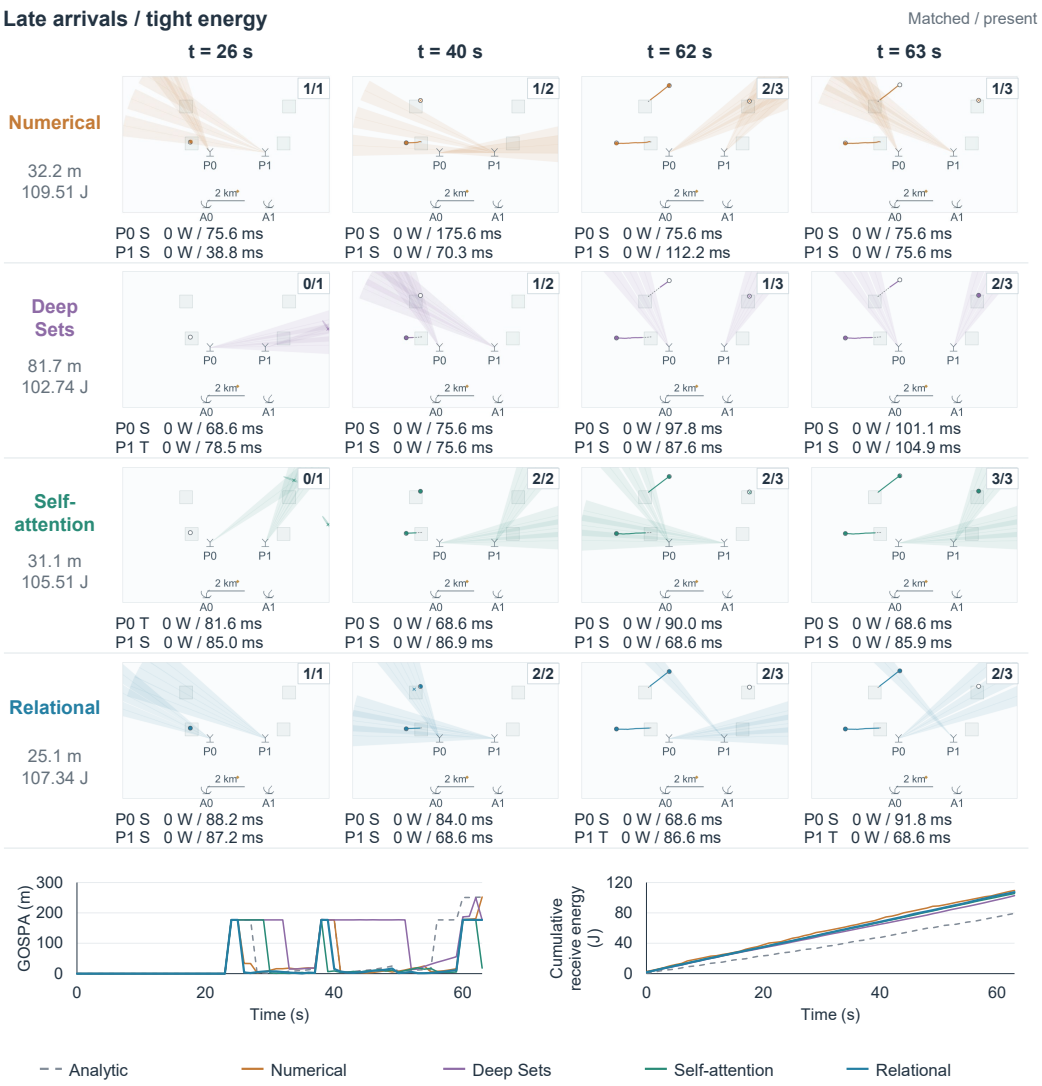


Figure 11: **Late arrivals with tight resource limits.** All five recorded policies use passive reception throughout this episode, so controlled RF expenditure is zero. The right plot therefore shows cumulative receiver electrical energy. The final two columns resolve the response to the third target, which enters near the end of the episode.

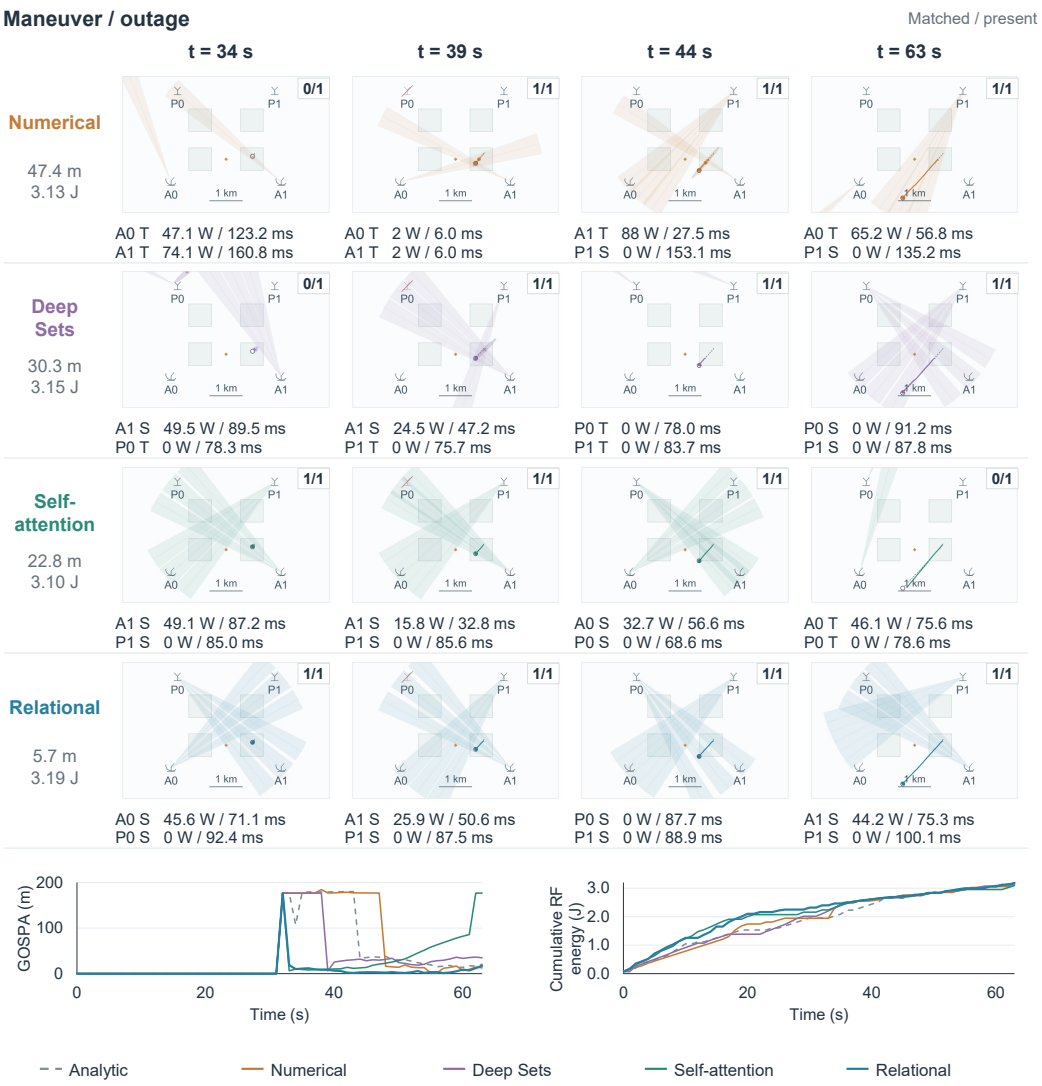


Figure 12: **Continuity through receiver outages.** This episode contains one target. The first displayed frame follows the active-node outage; the middle frames cover the passive-node outage and recovery. The relational run maintains a matched track at all four displayed times and obtains an episode-mean GOSPA of 5.7 m.

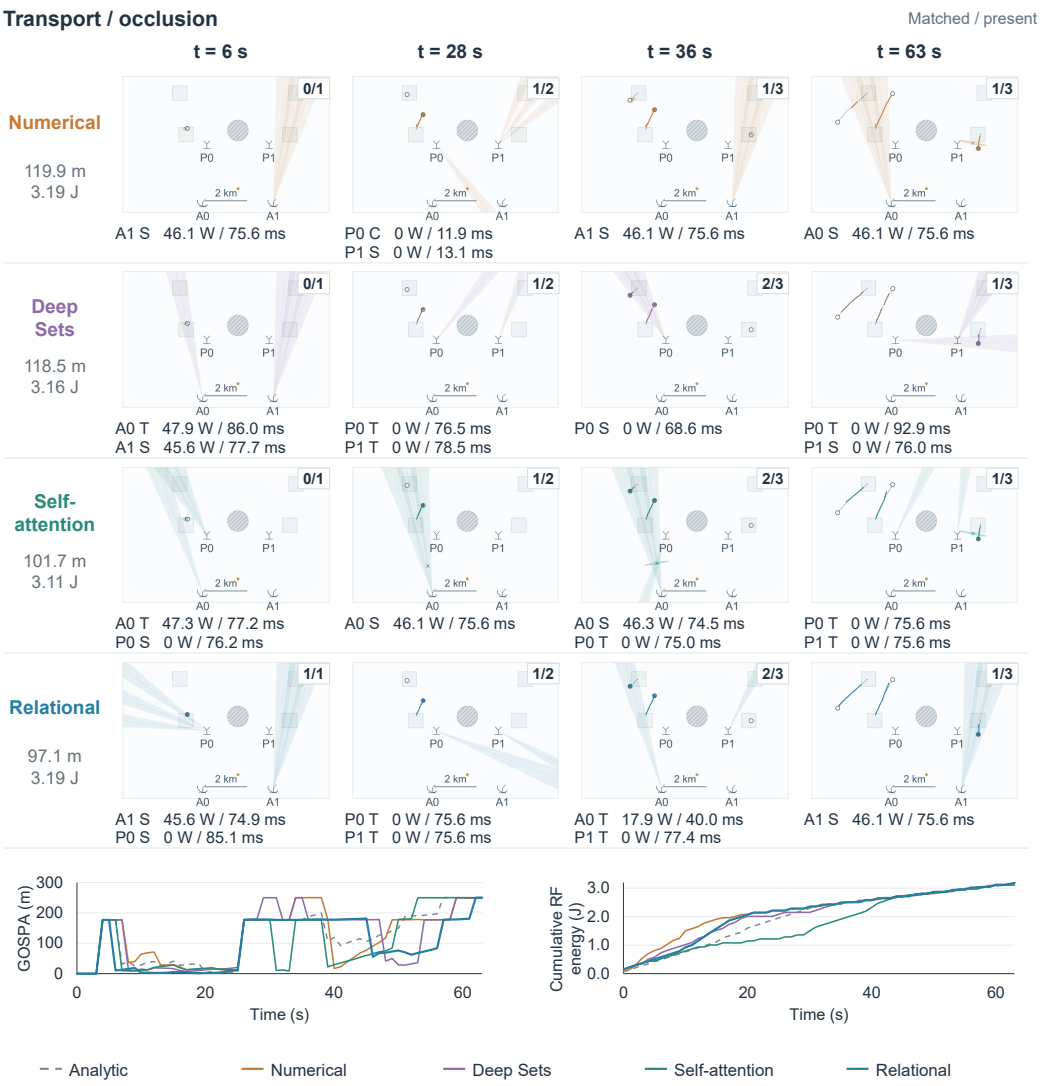


Figure 13: **Different tracking histories under transport and occlusion.** The relational model confirms the first target at 6 s, earlier than the other displayed methods. At 63 s, all four match only one of three targets. Their equal final counts conceal differences in earlier discovery and track continuity, which contribute to their different episode-mean GOSPA values.