ECHOBELIEF: EVIDENCE-ANCHORED SPATIAL MEMORY FOR ACTIVE ACOUSTIC MAPPING

Anonymous authors
Paper under double-blind review

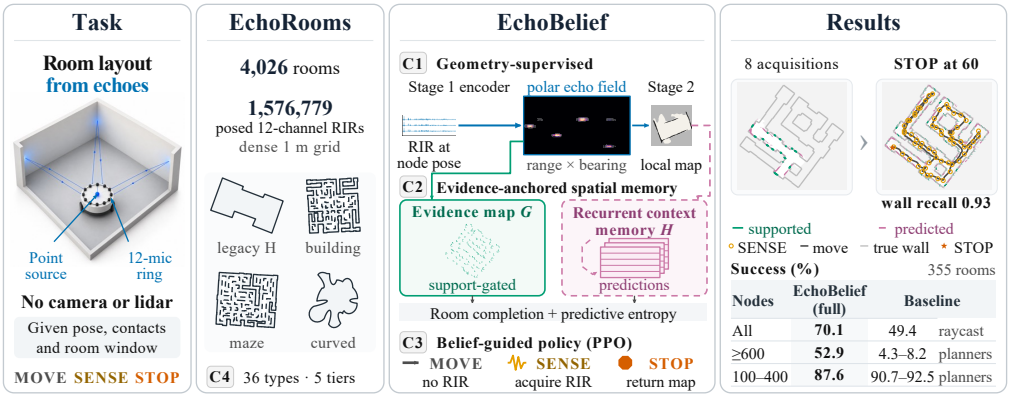

Figure 1: Overview of our proposed system, EchoBelief, which maps rooms from echoes alone, without cameras or lidar. A robot must decide where to move, when to sense and when to stop. We build EchoRooms, a benchmark of 4,026 simulated rooms, to train and test this skill (C4). EchoBelief encodes each echo into a polar map of nearby walls (C1), stores echo-supported evidence G apart from learned predictions H (C2), and uses this memory to choose its next action (C3). On 355 test rooms, it maps 70.1 % of episodes to 90 % wall recall, compared with 49.4 % for a tuned raycast planner, while sensing half as often. The example episode and the renderings of G and H are illustrative.

ABSTRACT

Echoes encode room geometry, and learned models can infer room layouts from room impulse responses (RIRs). However, one RIR mainly reveals nearby walls in direct view, so a robot mapping a whole room must decide where to move, when to measure and when to stop, as every measurement has a cost. Existing acoustic methods take responses as given, and active systems rely on vision; none chooses measurements from a persistent acoustic map. Such a map faces a dilemma: treating learned predictions as measurements erases observed walls, while discarding them loses spatial context. We propose EchoBelief, a system that resolves this dilemma with *evidence-anchored spatial memory*: an evidence map stores only structure that echoes or executed motion support, and a separate recurrent memory holds learned predictions. A geometry-supervised polar encoder turns each RIR into a map of wall reflections, and a belief-guided PPO policy reads the memory to move, sense or stop under an acquisition budget. We also introduce EchoRooms, a benchmark of 4,026 simulated rooms across 36 room types and five difficulty tiers, with 1.58M posed 12-channel RIRs on a dense 1 m grid, which we will release. On 355 test rooms, with known poses, EchoBelief reaches 90 % wall recall in 70.1 % of episodes, against 49.4–52.0 % for three tuned planners, while sensing half as often. The gain is largest in large rooms, where the planners rarely succeed (52.9 % vs. at most 8.2 %); in smaller rooms, the planners match or exceed it but sense far more. On 60 rooms with geometry disjoint from EchoRooms, evaluated only once, EchoBelief still outperforms raycast planning, and retraining without the memory pathway costs 22.8 success points.

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.

1 Introduction

Robots map indoor spaces mostly with cameras and lidar, but these sensors degrade in darkness, in smoke and dust, and on glass or mirrored surfaces. Sound offers a complementary route. Bats navigate in the dark by echolocation, and in the same way a room impulse response (RIR) recorded by a microphone array encodes the delays and directions of wall reflections (Allen & Berkley, 1979; Dokmanić et al., 2013). Learned models already use this signal to infer room layouts from a single RIR (Christensen et al., 2020; Yeon & Choi, 2024; Yeon et al., 2024), and model-based echolocation runs on small robots (Dümbgen et al., 2023). Yet one RIR mainly reveals nearby walls in direct view, so no single measurement can map a whole room. A robot that maps by sound must therefore explore actively: it must decide where to move, when to sense and when to stop, because each acquisition costs time and energy and the robot can afford only a limited number.

Yet no existing method explores in this way. Acoustic geometry methods infer structure from the responses they are given, but do not choose where to sense (Bicer et al., 2024; Yeon et al., 2024). Methods that do act on sound still depend on vision: audio-visual systems use sound to support cameras (Purushwalkam et al., 2021; Majumder et al., 2023), and active acoustic agents learn from audio-visual input to move and emit sound that refines visual maps (Hu et al., 2023) or to sample RIRs for acoustic rendering rather than geometry (Somayazulu et al., 2024). Prediction-guided exploration decides where to go next from anticipated maps, but builds those maps from depth, lidar or images (Ramakrishnan et al., 2020; Ho et al., 2025). We therefore study *active acoustic room mapping*, in which a robot with a point source and a 12-microphone ring, and no camera or lidar, maps a room from echoes under an acquisition budget.

The central challenge is how to remember what the robot has heard. Sparse echoes leave most of the room unobserved, so a useful map must also predict walls the robot has not yet reached. These predictions, however, become unreliable away from direct reflections. If the map writes them in as measurements, unsupported free-space predictions overwrite walls the robot has already observed, a failure we call *prediction-induced wall erasure*. If the map discards them, it loses the spatial context that guides exploration. The map must therefore use learned predictions without letting them erase what the robot has measured.

We propose EchoBelief, a system that meets this requirement by keeping measured evidence and learned predictions in separate stores (Fig. 3). First, a geometry-supervised encoder turns each RIR into a polar echo field, a range–bearing map of first-order reflections, and a local mapper converts this field into a wall/free/unknown map around the robot. Second, an *evidence-anchored spatial memory* splits these local maps into two stores: an evidence map records structure only where an echo or an executed move supports it, while a recurrent context memory integrates all learned predictions. Decoders read both stores to complete the room and estimate its uncertainty. Third, a belief-guided policy trained with proximal policy optimisation (PPO) reads this memory and chooses to MOVE, SENSE or STOP, which makes each acquisition an explicit decision. To train and evaluate such policies, we build EchoRooms, a benchmark of 4,026 simulated rooms across 36 room types and five difficulty tiers, with 1.58M posed 12-channel RIRs on a dense 1 m grid.

We evaluate EchoBelief on the 355 test rooms of EchoRooms under a budget of 128 acquisitions. It reaches 90 % wall recall in 70.1 % of episodes, against 49.4–52.0 % for three tuned planners, while sensing half as often. The gain is largest in large rooms, where EchoBelief succeeds in 52.9 % of episodes and the planners in at most 8.2 %; in smaller rooms, the planners match or exceed its success but sense far more. The advantage over raycast planning also holds on 60 rooms with unseen geometry. Ablations confirm the two key ingredients: retraining without the memory pathway costs 22.8 success points, and polar supervision raises the near-wall F_1 of a mapping probe from 42.2 % with masked reconstruction to 71.3 %. In summary, our contributions are:

- **Geometry-supervised echo encoding (C1)**, a polar representation that locates first-order reflections in range and bearing and gives downstream mapping an explicit geometric signal.
- **Evidence-anchored spatial memory (C2)**, a two-store belief that separates support-gated evidence from recurrent predictive context, reducing prediction-induced wall erasure.
- **Belief-guided acoustic exploration (C3)**, a PPO policy that jointly decides where to move, when to sense and when to stop, without vision.
- **EchoRooms (C4)**, the first benchmark for active acoustic mapping, with dense posed RIRs and geometric labels that enable policy training without online acoustic simulation.

2 Related work

Acoustic geometry inference. Echo sorting relates first-order echoes to room geometry (Dokmanić et al., 2013), acoustic SLAM jointly estimates the sensor trajectory and reflector positions from echoes (Kreković et al., 2016; Evers & Naylor, 2018), and model-based echolocation runs on small robots (Dümbgen et al., 2023). Learned methods detect reflectors and refine maps (Bicer et al., 2024; Saqib et al., 2024), predict layouts (Christensen et al., 2020), infer non-convex geometry from multi-order reflections (Yeon & Choi, 2024) and reconstruct floorplans of curved layouts (EchoScan; Yeon et al., 2024), while physics-regularised acoustic fields fit sparse RIRs (Wang et al., 2026). All of these methods infer geometry from the responses they are given. We also learn echo geometry, but our polar encoder feeds a persistent memory that decides where to sense next.

Audio-visual mapping and active acoustic sensing. Several systems pair sound with vision. SoundSpaces renders acoustics for audio-visual navigation (Chen et al., 2020; 2022), and VisualEchoes learns spatial image representations through echolocation (Gao et al., 2020). AV-Map reconstructs floorplans from audio-visual sequences (Purushwalkam et al., 2021), and Chat2Map uses audio and pose to select visual observations (Majumder et al., 2023). Active agents also learn where to act: Hu et al. (2023) learn navigation and emission to refine visual maps acoustically, and ActiveRIR learns navigation and selective RIR sampling for acoustic rendering (Somayazulu et al., 2024). Costly-observation formulations study when to acquire state or observations (Nam et al., 2021; Zhou et al., 2024). Like these works, we treat acquisition as a costly decision; unlike them, our camera-free agent reconstructs room geometry from posed multichannel RIRs.

Prediction-guided exploration and stopping. Predicted maps guide exploration through anticipated occupancy (Ramakrishnan et al., 2020), map uncertainty (MapEx; Ho et al., 2025), generated layouts (Wang et al., 2025), path coverage (Baek et al., 2025), or context and budget (Harutyunyan et al., 2025). Learned policies explore with images or lidar (Chen et al., 2025; Cao et al., 2024), imagined Gaussians plan reconstruction (Li et al., 2026), and completeness estimates decide when to stop (Luperto et al., 2026). These methods observe the room through depth, lidar or images. With sparse echoes, predictions are far less reliable, so our belief keeps supported evidence apart from learned completion, and our policy also decides when to sense.

Belief representations and world models. Neural Map (Parisotto & Salakhutdinov, 2018) and Active Neural SLAM (Chaplot et al., 2020) learn spatial memory and mapping for visual navigation, and world models connect prediction and control (Ha & Schmidhuber, 2018; Hafner et al., 2020; 2025). We bring learned memory to acoustic sensing but keep it apart from measured evidence. By replaying identical observations through beliefs with and without memory, we separate its benefits for reconstruction and for decisions, and we test whether forecasting helps planning.

3 EchoRooms: a benchmark for active acoustic mapping

Training and evaluating an active acoustic mapper requires an RIR at every place the robot might sense and geometric labels to score every intermediate map. Existing visual-acoustic benchmarks (Chen et al., 2022; Brunetto et al., 2023) target audio-visual perception and do not provide both. We therefore build EchoRooms, a benchmark of 4,026 simulated rooms with 1,576,779 posed 12-channel RIRs on a dense 1 m lattice (2.71 TB). Because every RIR is precomputed, a SENSE action reads the nearest cached response, so policies train without online acoustic simulation. Table 3 summarizes the statistics, Fig. 2 shows example rooms, and Appendix A gives full details.

Room generation. Three builder families generate 36 room types in five difficulty tiers, ranging from simple corner-chain polygons to building, maze and curved layouts with interior holes. Each wall receives an octave-band absorption class, such as brick, glass or curtain.

Acoustic simulation. At every lattice node, a point source sits at the centre of a 0.3 m ring of twelve omnidirectional microphones, with 0.45 m clearance from walls. We render each RIR with pyroomacoustics 0.10.1 (Scheibler et al., 2018), combining image sources and ray tracing in the extruded room (16 kHz, 3 s, noiseless), and apply quality gates to reject faulty renders.

Annotations. Each room comes with labels at three scales: the first-order echoes of each RIR on a polar grid, a ray-cast wall/free/unknown view within 8 m of each pose, and a raster of the whole room. These labels supervise the echo encoder, the local mapper and room completion, respectively, and shared metrics score both sequential reconstruction and budgeted exploration (Appendix F).

Splits and statistics. We split rooms by geometry into 3,447 training, 224 validation and 355 test rooms, so no test room shares its layout with a training room. The test split is deliberately harder

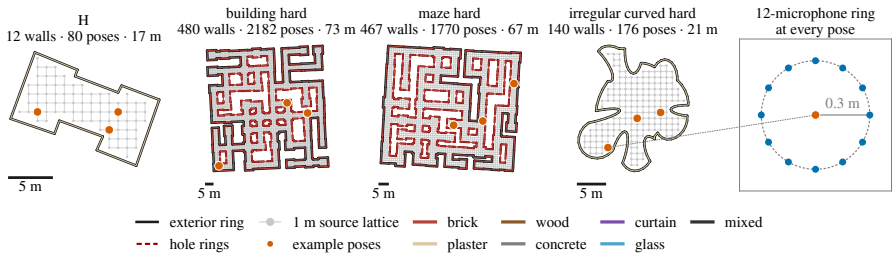

Figure 2: EchoRooms spans simple polygons, buildings, mazes and curved rooms with holes. Each panel shows one room with a 5 m scale bar: a legacy polygon and examples of the building, maze and irregular curved families. The drawings mark exterior and hole rings, per-wall material classes, the 1 m source lattice and three example poses. The last panel shows the 12-microphone ring (0.3 m radius) placed at every pose. Fig. 6 shows more rooms.

than training: its rooms are larger and have more walls (median extent 42.0 m and 32 walls, against 29.8 m and 19), and it includes 113 large building, maze and curved rooms generated only for testing. Section 5.1 reports when we added these rooms and how often we read the test split.

4 Method

EchoBelief has four stages (Fig. 3). First, a polar echo encoder turns each RIR into a *polar echo field*: a grid over range and bearing around the robot, where each cell gives the probability that a wall reflects the pulse exactly once at that distance and direction. The echo's delay determines the range, and the microphone array determines the bearing. Second, a local geometry mapper converts this field into a wall/free/unknown map of the robot's surroundings. Third, an evidence-anchored spatial memory merges these local maps into two stores: an evidence map that records only structure supported by an echo or an executed move, and a recurrent context memory that accumulates learned predictions. Support-gated fusion writes local wall and free-space estimates into the evidence map only where echo-derived support or executed motion supports them. The recurrent context memory retains the full local predictions, including their unsupported spatial context. Finally, a belief-guided policy reads this memory and chooses to move, sense or stop.

4.1 Problem setting

A room is a polygon $\mathcal{P} \subset \mathbb{R}^2$ bounded by an exterior ring and, in complex layouts, interior hole rings. The robot may stand only at admissible poses, the nodes of a 1 m lattice $\mathcal{L}$ at which EchoRooms stores an RIR. An *episode* is one mapping run from a starting pose. At each decision step t, the robot chooses to MOVE, SENSE or STOP. Each SENSE acquires an RIR window $s_k \in \mathbb{R}^{12\times 1600}$ (100 ms at 16 kHz, one row per microphone) at pose $p_k \in \mathcal{L}$. From the acquisitions so far, $\{(p_i, s_i)\}_{i\le k}$, the system maintains a belief $\mathcal{B}_t$, which a decoder turns into a map of the whole room on a 0.125 m grid, the *room completion*. The episode ends when the robot stops or after K acquisitions or D decisions; our primary budget is $K/D = 128/1{,}024$, and our secondary budget is $64/512$. We score the final completion by wall recall, the fraction of true wall length recovered within 0.30 m, together with wall F_1 and the IoU of the room interior. An episode succeeds if its final wall recall reaches 90 %.

4.2 Geometry-supervised echo encoding and local mapping

Polar echo encoder (Stage 1). The encoder maps each RIR window to polar echo field logits $z \in \mathbb{R}^{60\times 72}$ (0.2 m range and 5° bearing bins). A temporal CNN with microphone attention turns the waveform into 1,200 *delay tokens*, one per microphone and delay step, which a polar head decodes into the field (Appendix B). *Soft oracle targets* $y \in [0,1]^{60\times 72}$, computed from the true geometry, mark where visible walls produce first-order echoes, not how loud they are (Fig. 9). Since echoes occupy only 0.15 % of cells, we use binary cross-entropy with positive weight $w = 8$, plus an auxiliary waveform reconstruction from the delay tokens (Eqs. (1) and (2)).

Local geometry mapper (Stage 2). The mapper outputs an egocentric wall/free/unknown map of 128×128 cells at 0.125 m (8 m radius). A four-level U-Net (Ronneberger et al., 2015) reads the

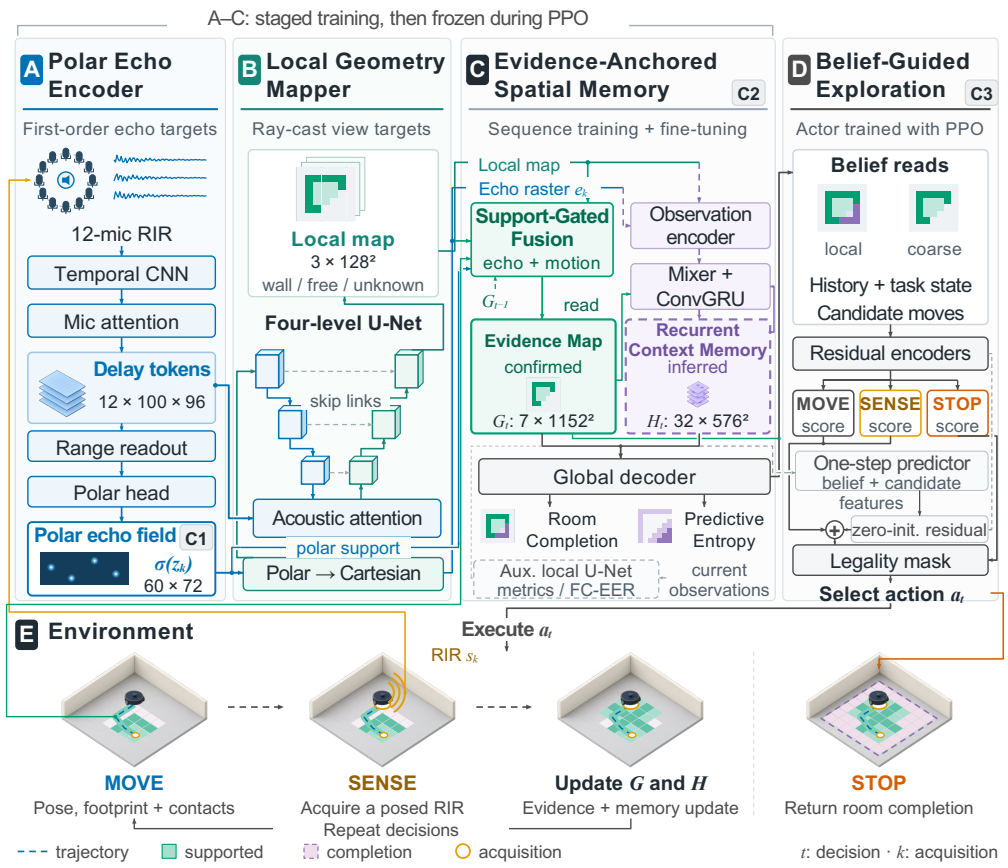

Figure 3: EchoBelief keeps measured evidence apart from learned predictions. (A) The polar echo encoder turns each RIR into a polar echo field. (B) The local geometry mapper converts the field into a wall/free/unknown map around the robot. (C) Support-gated fusion writes wall and free cells to the evidence map G only where an echo or an executed move supports them, while a ConvGRU accumulates local predictions in the recurrent context memory H. (D) The policy reads both stores and chooses MOVE, SENSE or STOP. (E) Each action returns an RIR or contact feedback, which updates the stores. Stages A–C are trained before PPO and then frozen (layer details: Fig. 7).

resampled polar probabilities, and at its bottleneck, *acoustic attention* lets each cell attend to all delay tokens (Vaswani et al., 2017); we therefore call it the *hybrid mapper*. We train it on ray-cast views, the walls and free space visible within 8 m, with class-weighted cross-entropy and a wall Dice loss (Milletari et al., 2016; Eq. (3), Appendix C). The deployed hybrid mapper processes each RIR independently. For analysis only, a causal *history refiner* adds a *view* head (visible now) and a *knowledge* head (observed so far).

4.3 Evidence-anchored spatial memory

Two stores. Writing learned predictions into the map as measurements causes prediction-induced wall erasure. With such writes, 64.5 % of the wall length the map once predicted reads as free after 64 acquisitions on 8 curved selection rooms, and 12.2 % on 58 polygonal rooms (Appendix G). Our evidence-anchored spatial memory $\mathcal{B}_t = (G, H)$ therefore separates the two: the evidence map G holds only structure that echoes or executed moves support, and the recurrent context memory H holds local predictions. *Anchored* refers to this write rule: local estimates enter G only where echoes or executed motion support them.

Evidence, memory and decoding. G and H share a 144 m canvas. G stores seven channels at 0.125 m: wall and free probabilities, support, echo and taper channels, and stamps for visited poses and collisions. *Support-gated fusion* writes wall or free evidence into a cell only where a first-order echo or the swept footprint of an executed move supports it; for comparison, *legacy fusion* writes

Stage 2 predictions directly into G wherever the mapper considers a cell known ($p_{\text{wall}} + p_{\text{free}} \geq 0.5$), which permits the erasure above (Appendix D.6). For H, an observation encoder embeds each Stage 2 prediction, and a ConvGRU (Ballas et al., 2016; Cho et al., 2014) at 0.25 m uses this embedding and the evidence to update a 26 m patch around the current pose (Eq. (4)). A global decoder reads both stores over the visited-pose bounding box plus a 10 m margin and outputs the room completion; the normalised entropy of its temperature-calibrated logits (Guo et al., 2017), set to zero on observed cells, forms the *predictive-entropy map*, which marks where the completion remains uncertain (Juergens et al., 2024; Appendix D). An auxiliary local U-Net with *view*, *knowledge*, *completion* and *forecast* heads supports training, local metrics and the FC-EER planner.

Losses and training. We train the memory and decoders on episode windows with $\mathcal{L} = \mathcal{L}_{\text{view}} + \mathcal{L}_{\text{know}} + \mathcal{L}_{\text{comp}}^{\text{local}} + \mathcal{L}_{\text{comp}}^{\text{global}} + \frac{1}{2}\,\mathcal{L}_{\text{fc}}$, one term per decoder head (Eq. (3); targets: Appendix D), with the forecast weight fixed before training. Episodes come from depth-first, random-walk and coverage-sweep explorers, and training runs for 20,000 updates with truncated back-propagation through time over at most 16 steps; we then fit four temperatures on calibration rooms. The full system fine-tunes this belief with support-gated fusion and *hole-aware* completion targets, which label hole interiors as outside the room, and refits its global temperature (Appendices F and D.6). Two ablation beliefs follow from this design. The *legacy-fusion belief* keeps H but uses legacy fusion without fine-tuning. The *geometric-only belief* removes H. Under legacy fusion, past Stage 2 predictions remain available through G; under support-gated fusion, only their supported components remain, so the *without-memory-pathway* arm removes the retained context beyond this evidence.

FC-EER planner. As a planning baseline built on the same belief, FC-EER (*forecast-conditioned entropy reduction*) scores each candidate pose by how much one imagined observation there would reduce uncertainty. It forms the imagined observation in one of three ways: the forecast head's prediction, the nearest stored observation (*copy*) or a knowledge-head read at the candidate pose (*knowledge read*). It writes this observation into a copy of the evidence and memory, decodes the same crop before and after, and computes $U = \sum_c \max(\mathcal{H}_c^{\text{before}} - \mathcal{H}_c^{\text{after}}, 0)\, a$ over cells c of area a, minus path-length, collision-risk and revisit costs. This clipped entropy reduction is a heuristic, not an expectation over possible observations (Appendix D.8).

4.4 Belief-guided acoustic exploration

Environment. The robot moves continuously, off the lattice, and chooses among 34 actions: SENSE, STOP and 32 MOVE actions that combine 16 headings, $2\pi(i \bmod 16)/16$, with step lengths of 0.5 or 1.0 m. Separating MOVE from SENSE makes acquisition an explicit decision rather than a side effect of movement (Nam et al., 2021; Zhou et al., 2024). A move succeeds only if its swept path keeps 0.3 m clearance from every wall. To prevent early training from collapsing into premature stops, STOP is masked until the robot has made a minimum number of acquisitions, the *STOP floor*; we anneal this floor from 24 to 4 over 0.5–1.5 M transitions and give its current value to the policy as an input. An episode ends at STOP, at the acquisition cap or at the decision limit.

Information contract and policy inputs. All methods receive the same declared priors before mapping (Table 24): exact poses, contact feedback, and a square *room window* that bounds the room, computed from the admissible nodes and polygon corners. The window centres the canvas, bounds moves and normalises the policy's scalar inputs (rooms span up to about 116 m). A failed move stamps its contact point into G, and a *collision memory* masks that move from that pose. SENSE is legal within 1 m of a lattice node and returns the node's cached RIR, registered at the node's true pose; beyond this, the policy never sees the lattice directly. The policy reads local and coarse windows of the belief, including a *conflict* channel that marks cells where wall and free evidence disagree, together with 65 scalars on trajectory, history and task state and 15 features per move (position, heading, contacts, SENSE/STOP legality). Residual encoders process these inputs, and separate heads score MOVE, SENSE and STOP (Appendix E). An online *one-step predictor* adds features describing how sensing would change the belief at the current pose and at the most promising move destinations, without imagined rollouts.

Reward and training. The reward measures map improvement. After every SENSE that writes new evidence, every new collision and at termination, we decode the belief and pay $10\,\Delta\text{Recall}_{0.30} + 5\,\Delta F_1^{0.30} + 2\,\Delta\text{Known}_{\text{reach}}$, where $\text{Known}_{\text{reach}}$ is the resolved fraction of reachable cells, plus interaction costs and potential-based frontier shaping (Table 21, Eq. (6)). A *terminal block* is paid once

Table 1: On 355 test rooms, EchoBelief (full) outperforms every planner and walker: $+18.1$ to $+20.8$ success points over the three tuned planners at 128 acquisitions and $+5.9$ to $+10.0$ at 64. The three-seed mean averages seeds 42–44 within each room. Raycast IG: raycast information gain (policy motion: with the policy's move set); imagined planners: FC-EER with copy or knowledge read; walkers: belief frontier or random; three starts per room. Success, F_1, hole-aware IoU, STOP in %; senses exclude the start; Acq. 90 is failure-penalised. Δ = main method minus row (success points, paired 95 % CI). The 64-acquisition rows re-run the same checkpoint.

Policy	Succ.	F_1	IoU	Acq. 90	Senses	Coll.	Dist.	Stop	Δ succ. [CI]
All 355 test rooms, 128 acquisitions (1,024 decisions)									
EchoBelief (full, seed 42, floor 4)	70.1	90.9	92.2	66.6	59.3	5.6	347.9	78.8	
EchoBelief (full), seed 43	71.8	91.1	92.1	65.3	63.5	7.1	392.1	73.6	−1.7 [−4.7, 1.3]
EchoBelief (full), seed 44	70.5	91.2	92.6	65.9	57.9	5.1	305.0	84.0	−0.4 [−3.2, 2.3]
EchoBelief (full), three-seed mean	70.8	91.1	92.3	65.9	60.2	5.9	348.4	78.8	
Raycast IG planner	49.4	83.9	80.9	84.0	124.3	22.3	429.6	0.0	+20.8 [16.3, 25.2]
Raycast IG planner, policy motion	49.5	83.7	80.4	84.5	124.4	32.1	350.7	0.0	+20.7 [16.2, 25.2]
Imagined planner, copy	50.0	80.6	76.6	79.9	124.4	15.1	626.7	0.0	+20.1 [15.6, 24.8]
Imagined planner, read	52.0	85.5	84.0	79.8	124.4	18.9	510.3	0.0	+18.1 [13.8, 22.6]
Belief-frontier walker	21.2	64.1	60.3	114.0	108.5	0.0	397.1	0.0	+48.9 [44.6, 53.3]
Random walker	7.3	43.1	42.7	120.6	23.7	76.3	507.6	92.2	+62.8 [58.7, 66.8]
All 355 test rooms, 64 acquisitions (512 decisions)									
EchoBelief (full, seed 42, floor 4)	49.4	84.6	84.5	42.5	40.9	3.5	204.0	55.3	
EchoBelief (full), seed 43	50.1	84.5	83.8	42.3	44.4	4.4	212.4	50.7	−0.8 [−2.8, 1.3]
Raycast IG planner	39.3	75.5	71.8	48.6	62.6	14.3	229.7	0.0	+10.0 [7.1, 13.1]
Raycast IG planner, policy motion	38.5	75.2	71.2	49.1	62.9	17.8	188.7	0.0	+10.9 [8.0, 13.9]
Imagined planner, copy	43.5	72.1	68.1	46.5	62.8	8.8	326.4	0.0	+5.9 [3.3, 8.6]
Imagined planner, read	42.8	78.4	76.5	46.7	62.6	12.3	276.0	0.0	+6.6 [3.9, 9.2]
Belief-frontier walker	12.8	53.3	51.8	61.1	60.0	0.0	163.3	0.0	+36.6 [32.4, 41.0]
Random walker	3.7	37.8	38.5	63.0	15.5	48.9	330.9	2.0	+45.7 [41.2, 50.2]

on the final recall R, rewarding success and penalising shortfalls in proportion to their size:

$$r_T = 1.5\,\mathbf{1}[R \geq 0.9] - 1.0\,\mathbf{1}[R < 0.9] - 2.0\max(0.9 - R, 0).$$

We train the actor with PPO (Schulman et al., 2017; 2016) and a *privileged critic* (Pinto et al., 2018), which sees ground-truth information unavailable to the actor, for 15 M steps on all 3,447 training rooms under the primary budget (Appendix E.7). The belief stays frozen, and rewards and the critic use hole-aware ground truth. We select the checkpoint on 75 validation rooms by success, then wall F_1, then acquisitions; this chose update 3,600 of the seed-42 run, which we froze before testing. This checkpoint is the full system, EchoBelief (full) in the tables.

Reference recipe. To measure how much our training changes matter, we also report the *reference recipe*, our earlier PPO policy. It uses a single network trunk without per-move features, the one-step predictor or collision memory; additional terminal reward terms; a fixed STOP floor of 24; legacy fusion with *hole-blind* ground truth, which ignores interior holes; and a canvas anchored at the start pose, which limits rooms to a 52 m span (Appendices E.1–E.6). Its ablation arm (Table 2) retrains this recipe in the full system's setting.

5 Experiments

5.1 Experimental setup

We evaluate on the 355 test rooms of EchoRooms and use three disjoint sets of validation rooms for checkpoint selection, temperature calibration and reporting (75, 74 and 75 rooms); every policy checkpoint is frozen before testing. As an additional check, we froze a 60-room confirmation cohort with geometry disjoint from all 4,026 corpus rooms before selecting the full system's checkpoint, and evaluated it exactly once, as declared in advance (Appendix E.10). Besides success (Sec. 4.1), we report *acquisitions to 90 % recall*, which counts episodes that never reach 90 % as the cap plus one (65 or 129); this metric combines speed and reliability, so we do not compare it across budgets, and it differs from one minus success because an episode can cross 90 % recall and still end below it. The lattice planner study uses its own criterion: wall $F_1 \geq 90$ % at the last scored point within 128 motion attempts. We compare methods with paired bootstraps over rooms (10,000 resamples, 95 % percentile intervals; Patterson et al., 2024) and call a difference *resolved* when its interval excludes zero, applying Holm correction within Stage 1 comparison families but not to policy contrasts (Appendix F). Finally, we compare against three planners and two walkers that share the full system's belief and information contract. The *raycast* planner scores candidate poses by ray-cast information gain over unresolved cells, and two *imagined* planners use FC-EER with the copy or knowledge-read observation (Sec. 4.3); we tune all three on validation rooms (Appendix E.7), and they sense until the acquisition cap without stopping early. The *belief-frontier* walker moves toward the nearest frontier of the belief, and the *random* walker picks legal actions at random.

5.2 Does the complete system map better under a sensing budget?

EchoBelief maps more rooms to 90 % wall recall than every planner, while sensing about half as often. At 128 acquisitions, it succeeds in 70.1 % of test episodes, against 49.4–52.0 % for the three tuned planners, a gain of 18.1 to 20.8 points (Table 1; example episodes in Fig. 4). It uses 59.3 senses instead of 124.3–124.4, because it learns when to stop, whereas the planners sense until the cap. It also reaches 90 % recall faster: 66.6 failure-penalised acquisitions, against 79.8–84.0. The advantage holds on the 60 confirmation rooms, which we read only once: EchoBelief succeeds in 66.7 % of episodes, against 51.7 % for raycast (+15.0 [1.7, 28.3]; Table 32).

The gain comes from large rooms. On the 170 test rooms with at least 600 lattice nodes, EchoBelief succeeds in 52.9 % of episodes, while the planners almost never do (4.3–8.2 %; Table 29). In smaller rooms, the planners succeed as often or slightly more often at 128 acquisitions (for example, 90.7–92.5 % against 87.6 % with 100–400 nodes), but they need about 125 senses where EchoBelief needs about 27. Under the tighter 64-acquisition budget, EchoBelief leads in these rooms as well. The confirmation cohort shows the same pattern: a 40.7-point gain on its 27 large rooms and no resolved difference on its 22 small ones.

Controls rule out simpler explanations (Table 36). *Stopping:* the gain does not come from learned stopping alone, since EchoBelief still leads by 11.0 to 15.1 points with STOP disabled at 64 acquisitions, and giving raycast a tuned stopping rule leaves it 23.4 points behind. *Motion:* it does not come from a richer action set, since raycast with the policy's 32 moves shows no resolved change. *Priors:* without acoustic input, the retrained policy never succeeds; without the room window, success shows no resolved change on the 246 rooms within the control's 52 m span limit, but the policy senses 11.4 more times, so the window mainly buys efficiency. *Seeds and simulation:* seeds 43 and 44 reach 71.8 and 70.5 %, and simulating each RIR live at the true pose instead of reading the cache gives 69.0 %; neither differs resolvably from the reported run (Table 31).

Larger budgets extend the operating range. Large layouts need longer paths, so we also raise the budget to 256 and 512 acquisitions on 500 fresh rooms generated beyond the corpus. With these budgets and extra starts for rooms that fail, EchoBelief reaches 90 % recall on 78.0 % of rooms (Appendix E.8).

5.3 Which ingredients matter?

The memory pathway matters most. Retraining the belief and policy without context memory H yields the largest ablation drop: success falls from 70.1 to 47.3 % (+22.8 [18.8, 27.0] for the full system; Table 2). A second seed reproduces this gap (+22.0), which persists at 64 acquisitions and with STOP disabled. Without memory, 51.9 % of episodes never reach 90 % recall, versus 29.5 % with memory, despite 14.3 more sensing actions (Table 30). Support-gated fusion keeps only supported predictions in G, so removing H removes the predictive context retained beyond it; the 22.8-point gap thus measures the whole memory pathway.

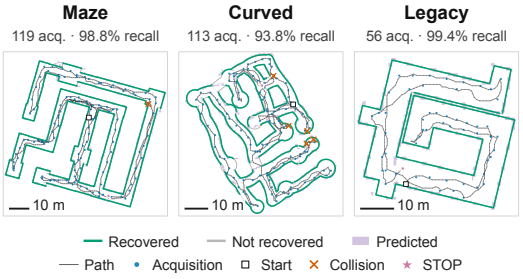

Figure 4: Three successful test episodes of the main method: the largest room per layout family (128 acquisitions, STOP floor 4). Headers: acquisitions and final wall recall at 0.30 m; More examples: Fig. 17.

Memory improves reconstruction and decisions. To separate these effects, we replay planner runs through beliefs with and without memory on identical observations and collisions (Table 19). Memory adds 1.5–1.7 wall F_1 and 2.5–3.1 IoU points at 64 acquisitions. Using the memory belief for action selection further improves trajectories by 2.1 points for raycast and 3.3 for FC-EER.

Evidence anchoring makes the policy safer. Legacy fusion, which writes predictions into G, collides more than the full system under every protocol. Success also trails for the second seed (+6.6 [3.1, 10.0]; Table 35) and at 64 acquisitions with STOP disabled (+3.1 [1.6, 4.7]).

Training choices also matter. Retraining our earlier reference recipe in the full system's setting reaches only 57.8 % (+12.3), and removing collision memory lowers success to 66.6 % (+3.6). Removing the one-step predictor shows no resolved success difference (Table 2).

Forecasting does not help planning. In FC-EER, forecasting the next observation yields 76.1 % wall F_1 at 64 acquisitions, below copying the nearest stored observation (+1.1) and reading the knowledge head (+1.6); both match raycast (77.6 %; Table 17).

Table 2: The memory pathway (recurrence and, under support-gated fusion, the route for retaining past local predictions beyond the support-gated evidence) is the largest policy ablation: retraining belief and policy without it costs 22.8 success points at 128 acquisitions. 355 rooms, three starts each; 15 M training steps; ablation arms are seed 42 (seed-43 replicates of the legacy-fusion and without-memory-pathway arms: Table 35). Δ = main method minus arm (paired 95 % CI). The reference-recipe arm retrains the earlier recipe (Sec. 4.4) in the main method's setting (Appendix E.12).

Policy (all 355 test rooms)	Success (%)	Stop (%)	Main − arm: Δ success (pts)	Main − arm: Δ acq. to 90%	Main − arm: Δ collisions	Main − arm: Δ senses
STOP floor 4, 128-acquisition cap, 1,024 decisions (the main method's own protocol)						
EchoBelief (full, seed 42)	70.1	78.8				
Seed 43	71.8	73.6	−1.7 [−4.7, 1.3]	+1.3 [−0.7, 3.5]	−1.5 [−2.0, −0.9]	−4.2 [−5.4, −3.1]
Seed 44	70.5	84.0	−0.4 [−3.2, 2.3]	+0.8 [−1.4, 3.0]	+0.6 [0.2, 0.9]	+1.5 [0.4, 2.5]
Legacy fusion	67.6	67.9	+2.5 [−1.0, 6.1]	−2.2 [−4.9, 0.5]	−1.1 [−1.8, −0.5]	−0.5 [−2.0, 0.9]
Reference-recipe arm (floor 24)	57.8	87.5	+12.3 [8.7, 16.0]	−5.0 [−7.5, −2.5]	−15.0 [−18.5, −11.8]	+1.5 [−0.4, 3.5]
No predictor	68.8	77.8	+1.3 [−1.7, 4.3]	−0.5 [−2.7, 1.6]	−0.1 [−0.6, 0.4]	−0.8 [−2.1, 0.4]
No collision memory	66.6	76.2	+3.6 [0.4, 6.8]	−4.9 [−7.4, −2.3]	−2.0 [−7.5, 1.0]	−5.0 [−6.6, −3.5]
Without memory pathway (belief and policy retrained)	47.3	64.9	+22.8 [18.8, 27.0]	−19.6 [−22.3, −16.8]	−0.9 [−1.4, −0.4]	−14.3 [−16.1, −12.5]
STOP floor 4, 64-acquisition cap, 512 decisions						
EchoBelief (full, seed 42)	49.4	55.3				
Seed 43	50.1	50.7	−0.8 [−2.8, 1.3]	+0.3 [−0.4, 1.0]	−0.9 [−1.3, −0.6]	−3.4 [−4.2, −2.6]
Seed 44	51.2	57.4	−1.8 [−4.0, 0.5]	+0.6 [−0.2, 1.3]	+0.28 [0.05, 0.50]	−0.5 [−1.0, 0.1]
Legacy fusion	47.7	51.9	+1.7 [−0.7, 3.9]	+0.1 [−0.8, 1.1]	−0.6 [−1.1, −0.2]	+1.8 [1.3, 2.4]
Reference-recipe arm (floor 24)	48.3	61.1	+1.1 [−1.1, 3.4]	+0.5 [−0.3, 1.3]	−10.7 [−13.1, −8.5]	−1.7 [−2.6, −0.7]
No predictor	48.9	55.6	+0.5 [−1.8, 2.7]	+0.04 [−0.80, 0.88]	−0.2 [−0.5, 0.1]	+0.1 [−0.4, 0.6]
No collision memory	47.7	53.7	+1.7 [−0.7, 3.9]	−0.80 [−1.65, 0.02]	−1.1 [−3.7, 0.4]	−1.2 [−1.8, −0.6]
Without memory pathway (belief and policy retrained)	39.8	51.0	+9.6 [6.9, 12.3]	−7.1 [−8.2, −6.0]	+0.1 [−0.1, 0.4]	−5.5 [−6.4, −4.7]
STOP disabled (floor 64), 64-acquisition cap, 512 decisions						
EchoBelief (full, seed 42)	54.5	0.0				
Seed 43	53.1	0.0	+1.4 [−0.1, 2.9]	−0.8 [−1.2, −0.3]	−1.4 [−1.9, −0.9]	−2.9 [−3.4, −2.3]
Seed 44	54.6	0.0	−0.1 [−1.6, 1.4]	+0.1 [−0.3, 0.5]	+0.4 [0.1, 0.6]	−3.4 [−3.9, −2.9]
Legacy fusion	51.4	0.0	+3.1 [1.6, 4.7]	+0.45 [0.00, 0.88]	−1.6 [−2.2, −1.2]	+7.4 [6.4, 8.4]
Reference-recipe arm	51.5	0.0	+3.0 [1.3, 4.8]	−0.2 [−0.7, 0.3]	−15.6 [−18.9, −12.6]	−3.8 [−4.3, −3.3]
No predictor	53.5	0.0	+0.9 [−0.7, 2.6]	−0.2 [−0.6, 0.3]	−0.6 [−0.9, −0.3]	+1.8 [1.2, 2.4]
No collision memory	52.3	0.0	+2.2 [0.5, 3.8]	−0.9 [−1.4, −0.4]	+0.42 [0.00, 0.82]	−3.4 [−3.9, −2.8]
Without memory pathway (belief and policy retrained)	48.8	0.0	+5.6 [3.6, 7.9]	−6.4 [−7.2, −5.6]	+0.25 [−0.05, 0.56]	−3.9 [−4.4, −3.4]

5.4 Do polar supervision and memory improve mapping?

The polar encoder locates echoes accurately. The Stage 1 encoder reaches 97.0 % near-range F_1 on test RIRs and 87.1–88.6 % on the harder curved families (Table 6).

Polar supervision makes acoustic features useful for mapping. A probe that maps from the delay tokens alone reaches 71.3 % near-wall F_1 when the encoder is trained with polar supervision, against 42.2 % with masked reconstruction and 35.2 % with waveform reconstruction (Table 9). The deployed mapper, in turn, relies on the polar field: zeroing the field drops near-wall F_1 from 80.5 to 1.6 %, whereas zeroing the tokens leaves it at 79.8 % (Table 10).

Acoustic history and memory improve the map. Adding past acquisitions improves the history refiner by 9.3 points (Tables 11 and 12). Removing the context memory lowers whole-room completion F_1 in both belief seeds, by 1.4 points on reporting rooms and 1.2 on test rooms, while it slightly raises the local knowledge read (Table 15). Memory therefore serves what it was designed for: completing the whole room rather than reading local structure.

6 Limitations and conclusion

Our claims concern simulated acoustic mapping with exact poses, shared contact feedback and spatial priors, and idealized acoustics; measured RIRs remain untested. The start-anchored control covers only a subset of large rooms, and randomly translated windows remain untested (Appendix E.12). Without a rate-matched periodic-sensing control, results support the joint exploration strategy but do not isolate adaptive timing. No external learned comparator is evaluated; only raycast is tested with matched moves. The main policy has three seeds; ablations have one or two. Large, intricate layouts remain sensitive to budget and starting pose (Table 29). Appendix E.8 details sensing limitations.

Conclusion. EchoBelief maps rooms from echoes, without cameras or lidar, using evidence-anchored spatial memory: an evidence map of echo-supported structure beside a recurrent context memory of learned predictions. Removing that memory pathway lowers success on rooms with at least 600 nodes from 52.9 to 15.9 %, polar supervision raises a token-only probe's near-wall F_1 from 42.2 % (masked) to 71.3 %, and the belief-guided policy succeeds on 70.1 % of test episodes against 49.4–52.0 % for three tuned planners; releasing EchoRooms in full will make the task reproducible.

AI use statement

Large-language-model assistance was used to draft and edit the paper, write and refactor experiment, figure and table scripts, and summarise training and evaluation logs. The authors executed all AI-assisted code and verified its outputs against the logged artefacts. Table and figure scripts read numerical results from those artefacts; values quoted in the prose were transcribed from them. The authors take full responsibility for the content.

Ethics statement

The work uses simulated rooms only, with no human subjects, personal data or recordings of people. Incorrect maps could cause collisions when robots operate near people. We therefore report false-free predictions on true walls and collision counts. The models have not been validated on measured data.

Reproducibility statement

Appendix A specifies the room generators, materials, simulator settings, quality gates, labels and seeded split algorithm. Appendices B–E detail the architectures, losses, hyperparameters, seeds, selection rules and baselines; Appendix F defines metrics, censoring and bootstrap procedures. Section 5.1 and Appendices E.7. Frozen checkpoints, temperatures, thresholds and the dataset manifest are identified by SHA-256 in the provenance accompanying the tables and figures. The public release will comprise the full RIR archives of the training, validation and 355-room test splits, the geometric labels, the generator, the training and evaluation code and the EchoRooms manifest with split lists (Appendix A).

References

Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979. doi: 10.1121/1.382599. URL https://jontalle.web.engr.illinois.edu/Public/AllenBerkley79.pdf.

Seungjae Baek, Brady Moon, Seungchan Kim, Muqing Cao, Cherie Ho, Sebastian Scherer, and Jeong Hwan Jeon. PIPE planner: Pathwise information gain with map predictions for indoor robot exploration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7684–7691, 2025. doi: 10.1109/IROS60139.2025.11246190. URL https://arxiv.org/abs/2503.07504.

Nicolas Ballas, Li Yao, Chris Pal, and Aaron Courville. Delving deeper into convolutional networks for learning video representations. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.06432.

H. Nazim Bicer, Cagdas Tuna, Andreas Walther, and Emanuël A. P. Habets. Data-driven joint detection and localization of acoustic reflectors. In *IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pp. 745–749, 2024. doi: 10.1109/ICASSPW62465.2024.10626611. URL https://arxiv.org/abs/2402.06246.

Amandine Brunetto, Sascha Hornauer, Stella X. Yu, and Fabien Moutarde. The audio-visual BatVision dataset for research on sight and sound. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–8, 2023. doi: 10.1109/IROS55552.2023.10341715. URL https://doi.org/10.1109/IROS55552.2023.10341715.

Yuhong Cao, Rui Zhao, Yizhuo Wang, Bairan Xiang, and Guillaume Sartoretti. Deep reinforcement learning-based large-scale robot exploration. *IEEE Robotics and Automation Letters*, 9(5):4631–4638, 2024. doi: 10.1109/LRA.2024.3379804. URL https://doi.org/10.1109/LRA.2024.3379804.

Devendra Singh Chaplot, Dhiraj Gandhi, Saurabh Gupta, Abhinav Gupta, and Ruslan Salakhutdinov. Learning to explore using active neural SLAM. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HklXn1BKDH.

Changan Chen, Unnat Jain, Carl Schissler, Sebastia Vicenc Amengual Gari, Ziad Al-Halah, Vamsi Krishna Ithapu, Philip Robinson, and Kristen Grauman. SoundSpaces: Audio-visual navigation in 3D environments. In *European Conference on Computer Vision*, pp. 17–36. Springer, 2020.

Changan Chen, Carl Schissler, Sanchit Garg, Philip Kobernik, Alexander Clegg, Paul Calamia, Dhruv Batra, Philip Robinson, and Kristen Grauman. SoundSpaces 2.0: A Simulation Platform for Visual-Acoustic Learning. In *Advances in Neural Information Processing Systems (Datasets and Benchmarks Track)*, volume 35, pp. 8896–8911, 2022. doi: 10.52202/068431-0647. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/3a48b0eaba26ba862220a307a9edb0bb-Abstract-Datasets_and_Benchmarks.html.

Xiao Chen, Tai Wang, Quanyi Li, Tao Huang, Jiangmiao Pang, and Tianfan Xue. GLEAM: Learning generalizable exploration policy for active mapping in complex 3D indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5558–5568, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Chen_GLEAM_Learning_Generalizable_Exploration_Policy_for_Active_Mapping_in_Complex_ICCV_2025_paper.html.

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 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

Jesper Haahr Christensen, Sascha Hornauer, and Stella X. Yu. BatVision: Learning to See 3D Spatial Layout with Two Ears. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1581–1587, 2020. doi: 10.1109/ICRA40945.2020.9196934. URL https://doi.org/10.1109/ICRA40945.2020.9196934.

Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M. Lu, and Martin Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, 2013. doi: 10.1073/pnas.1221464110. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3725100/.

Frederike Dümbgen, Adrien Hoffet, Mihailo Kolundžija, Adam Scholefield, and Martin Vetterli. Blind as a bat: Audible echolocation on small robots. *IEEE Robotics and Automation Letters*, 8(3):1271–1278, 2023. doi: 10.1109/LRA.2022.3194669. URL https://doi.org/10.1109/LRA.2022.3194669.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, 1989. doi: 10.1109/2.30720. URL https://www.cs.cmu.edu/~motionplanning/papers/sbp_papers/integrated4/elfes_occup_grids.pdf.

Christine Evers and Patrick A Naylor. Acoustic SLAM. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(9):1484–1498, 2018.

Ruohan Gao, Changan Chen, Ziad Al-Halah, Carl Schissler, and Kristen Grauman. VisualEchoes: Spatial image representation learning through echolocation. In *European Conference on Computer Vision*, pp. 658–676. Springer, 2020.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330, 2017.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018. URL https://arxiv.org/abs/1803.10122.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025. doi: 10.1038/s41586-025-08744-2. URL https://www.nature.com/articles/s41586-025-08744-2.

Narek Harutyunyan, Brady Moon, Seungchan Kim, Cherie Ho, Adam Hung, and Sebastian Scherer. MapExRL: Human-inspired indoor exploration with predicted environment context and reinforcement learning. In *22nd International Conference on Advanced Robotics (ICAR)*, 2025. URL https://arxiv.org/abs/2503.01548.

Cherie Ho, Seungchan Kim, Brady Moon, Aditya Parandekar, Narek Harutyunyan, Chen Wang, Katia Sycara, Graeme Best, and Sebastian Scherer. MapEx: Indoor structure exploration with probabilistic information gain from global map predictions. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13074–13080, 2025. doi: 10.1109/ICRA55743.2025.11128862. URL https://arxiv.org/abs/2409.15590.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Xixi Hu, Senthil Purushwalkam, David Harwath, and Kristen Grauman. Learning to map efficiently by active echolocation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1505–1510, 2023. doi: 10.1109/IROS55552.2023.10341664. URL https://ieeexplore.ieee.org/document/10341664/.

Mira Juergens, Nis Meinert, Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Is epistemic uncertainty faithfully represented by evidential deep learning methods? In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22624–22642, 2024. URL https://proceedings.mlr.press/v235/juergens24a.html.

Miranda Kreković, Ivan Dokmanić, and Martin Vetterli. EchoSLAM: Simultaneous localization and mapping with acoustic echoes. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11–15. IEEE, 2016.

Shiyao Li, Antoine Guédon, Shizhe Chen, and Vincent Lepetit. MAGICIAN: Efficient long-term planning with imagined Gaussians for active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21606–21615, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Li_MAGICIAN_Efficient_Long-Term_Planning_with_Imagined_Gaussians_for_Active_Mapping_CVPR_2026_paper.html.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

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

Matteo Luperto, Marco Maria Ferrara, Matteo Princisgh, Giacomo Boracchi, and Francesco Amigoni. Estimating map completeness in robot exploration. *Autonomous Robots*, 50:6, 2026. doi: 10.1007/s10514-025-10221-8. URL https://link.springer.com/article/10.1007/s10514-025-10221-8.

Sagnik Majumder, Hao Jiang, Pierre Moulon, Ethan Henderson, Paul Calamia, Kristen Grauman, and Vamsi Krishna Ithapu. Chat2Map: Efficient scene mapping from multi-ego conversations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10554–10564, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Majumder_Chat2Map_Efficient_Scene_Mapping_From_Multi-Ego_Conversations_CVPR_2023_paper.html.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision (3DV)*, pp. 565–571, 2016. doi: 10.1109/3DV.2016.79.

HyunJi Alex Nam, Scott Fleming, and Emma Brunskill. Reinforcement learning with state observation costs in action-contingent noiselessly observable Markov decision processes. In *Advances in Neural Information Processing Systems*, volume 34, pp. 15650–15666, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/83e8fe6279ad25f15b23c6298c6a3584-Abstract.html.

Andrew Y. Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the 16th International Conference on Machine Learning*, pp. 278–287, 1999.

Emilio Parisotto and Ruslan Salakhutdinov. Neural map: Structured memory for deep reinforcement learning. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Bk9zbyZCZ.

Andrew Patterson, Samuel Neumann, Martha White, and Adam White. Empirical design in reinforcement learning. *Journal of Machine Learning Research*, 25(318):1–63, 2024. URL https://jmlr.org/papers/v25/23-0183.html.

Lerrel Pinto, Marcin Andrychowicz, Peter Welinder, Wojciech Zaremba, and Pieter Abbeel. Asymmetric actor critic for image-based robot learning. In *Robotics: Science and Systems*, 2018. doi: 10.15607/RSS.2018.XIV.008.

Senthil Purushwalkam, Sebastià Vicenc Amengual Garí, Vamsi Krishna Ithapu, Carl Schissler, Philip Robinson, Abhinav Gupta, and Kristen Grauman. Audio-visual floorplan reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1183–1192, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Purushwalkam_Audio-Visual_Floorplan_Reconstruction_ICCV_2021_paper.html.

Santhosh K. Ramakrishnan, Ziad Al-Halah, and Kristen Grauman. Occupancy anticipation for efficient exploration and navigation. In *European Conference on Computer Vision*, pp. 400–418, 2020. doi: 10.1007/978-3-030-58558-7_24. URL https://doi.org/10.1007/978-3-030-58558-7_24.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. URL https://arxiv.org/abs/1505.04597.

Usama Saqib, Letizia Marchegiani, and Jesper Rindom Jensen. A machine learning framework for acoustic reflector mapping. *arXiv preprint arXiv:2409.12094*, 2024. URL https://arxiv.org/abs/2409.12094.

Robin Scheibler, Eric Bezzam, and Ivan Dokmanić. Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 351–355, 2018. doi: 10.1109/ICASSP.2018.8461310. URL https://arxiv.org/abs/1710.04196.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-Dimensional Continuous Control Using Generalized Advantage Estimation. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1506.02438.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Arjun Somayazulu, Sagnik Majumder, Changan Chen, and Kristen Grauman. ActiveRIR: Active audio-visual exploration for acoustic environment modeling. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13830–13836, 2024. doi: 10.1109/IROS58592.2024.10801510. URL https://doi.org/10.1109/IROS58592.2024.10801510.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Haoran Wang, Shengzhi Qiang, Qi Wang, Qing Lu, and Yuanqing Wang. Physics-regularized neural acoustic fields for spatial layout inference from sparse room impulse responses. *The Journal of the Acoustical Society of America*, 160(1):493–506, 2026. doi: 10.1121/10.0044365. URL https://pubmed.ncbi.nlm.nih.gov/42455750/.

Yizhuo Wang, Haodong He, Jingsong Liang, Yuhong Cao, Ritabrata Chakraborty, and Guillaume Adrien Sartoretti. CogniPlan: Uncertainty-guided path planning with conditional generative layout prediction. In *Proceedings of the 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 1382–1396, 2025. URL https://proceedings.mlr.press/v305/wang25d.html.

Yuxin Wu and Kaiming He. Group Normalization. In *European Conference on Computer Vision*, 2018. URL https://arxiv.org/abs/1803.08494.

Brian Yamauchi. A frontier-based approach for autonomous exploration. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp. 146–151, 1997. doi: 10.1109/CIRA.1997.613851. URL https://robotfrontier.com/papers/cira97.pdf.

Inmo Yeon and Jung-Woo Choi. RGI-Net: 3D room geometry inference from room impulse responses with hidden first-order reflections. In *18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 439–443, 2024. doi: 10.1109/IWAENC61483.2024.10694653. URL https://arxiv.org/abs/2309.01513.

Inmo Yeon, Iljoo Jeong, Seungchul Lee, and Jung-Woo Choi. EchoScan: Scanning complex room geometries via acoustic echoes. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:4768–4782, 2024. doi: 10.1109/TASLP.2024.3485516. URL https://arxiv.org/abs/2310.11728.

Helen Zhou, Audrey Huang, Kamyar Azizzadenesheli, David Childers, and Zachary Lipton. Timing as an action: Learning when to observe and act. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3979–3987, 2024. URL https://proceedings.mlr.press/v238/zhou24c.html.

A EchoRooms: statistics, splits, simulation and labels

This appendix documents the EchoRooms benchmark of Sec. 3: the manifest and public release, the corpus statistics, the split rule and test-read history, the simulator settings and the label construction. It supports the benchmark contribution and the split disclosures of Sec. 5.1.

A.1 Manifest, storage and release

A single manifest describes all 4,026 rooms. Geometric fields record the global identifier, room and catalogue types, difficulty tier, layout family and complexity, scale stratum, extent, and counts of corners, holes and sources. Acoustic fields record wall-material classes and their length fractions, material profile, RT_{60} estimate and class, and echo-consistency score. Acquisition and split fields specify the ring radius, clearance, split, split group and validation role. Each room file is identified by a hash, and every geometry identity is unique.

Each room file stores raw RIRs as a float32 array of shape (sources, 12, 48,000), the polygon with hole rings, per-wall materials and the lattice. The corpus occupies 2.71 TB (training 2.15 TB, validation 0.13 TB, test 0.42 TB).

Benchmark protocol. The manifest fixes the splits and validation roles for the tasks of Sec. 4.1 (visual-acoustic benchmark context: Chen et al., 2022; Brunetto et al., 2023). Shared metrics and uncertainty estimates are defined in Appendix F.

Public release. The public release will cover the full corpus: the RIR archives of the training, validation and 355-room test splits, the geometric labels, the room generator, the training and evaluation code and this manifest with its split lists. Checkpoints and calibrated settings are identified by hash in the artefact provenance. EchoRooms supports echo detection, local and whole-room mapping, and active acquisition; use for speech, source localisation or measured-RIR transfer requires further validation.

Corpus statistics Table 3 and Figs. 5 and 6 summarise the corpus; the paragraphs below give the counts behind them for each split.

Table 3: EchoRooms at a glance: rooms and acquisition poses by split, layout family and difficulty tier, with per-room extent, wall count and RT_{60}; the test split over-represents large, many-walled rooms. *Legacy* denotes the corner-chain room types. Validation roles selection / calibration / reporting = 75 / 74 / 75. The Schroeder fit behind the RT_{60} medians is flagged unreliable for 3,067 of 3,447 training rooms (paragraph *Training acoustics*).

	Train	Validation	Test	All
Rooms	3,447	224	355	4,026
Acquisition poses	1,250,664	79,711	246,404	1,576,779
Layout family				
legacy	2,737	183	180	3,100
building	121	8	61	190
maze	139	11	45	195
curved	77	7	57	141
irregular curved	373	15	12	400
Difficulty tier				
simple	180	20	0	200
medium	180	20	0	200
complex	2,607	159	319	3,085
very complex	449	22	26	497
extreme	31	3	10	44
Geometry and acoustics				
Extent (m), median [min–max]	29.8 [4.1–79.3]	28.9 [7.2–79.6]	42.0 [14.6–78.8]	30.4 [4.1–79.6]
Walls per room, median	19	17	32	20
RT60 (s), median	1.61	1.57	1.49	1.59

Training rooms. The training split (3,447 rooms, 1,250,664 poses) contains 180 simple, 180 medium, 2,607 complex, 449 very complex and 31 extreme rooms. By scale stratum, it contains 63 small, 267 medium, 1,584 large and 1,533 very large rooms; by layout family, 2,737 legacy (corner-chain), 121 building, 139 maze, 77 curved and 373 irregular curved rooms. Room extent has a median of 29.8 m (5th and 95th percentiles 13.2 and 61.2 m, maximum 79.3 m). Rooms have a median of 19 walls (95th percentile 259, maximum 3,231) and 285 poses (minimum 4, 95th percentile 1,181, maximum 2,453); 337 rooms have holes (up to 35).

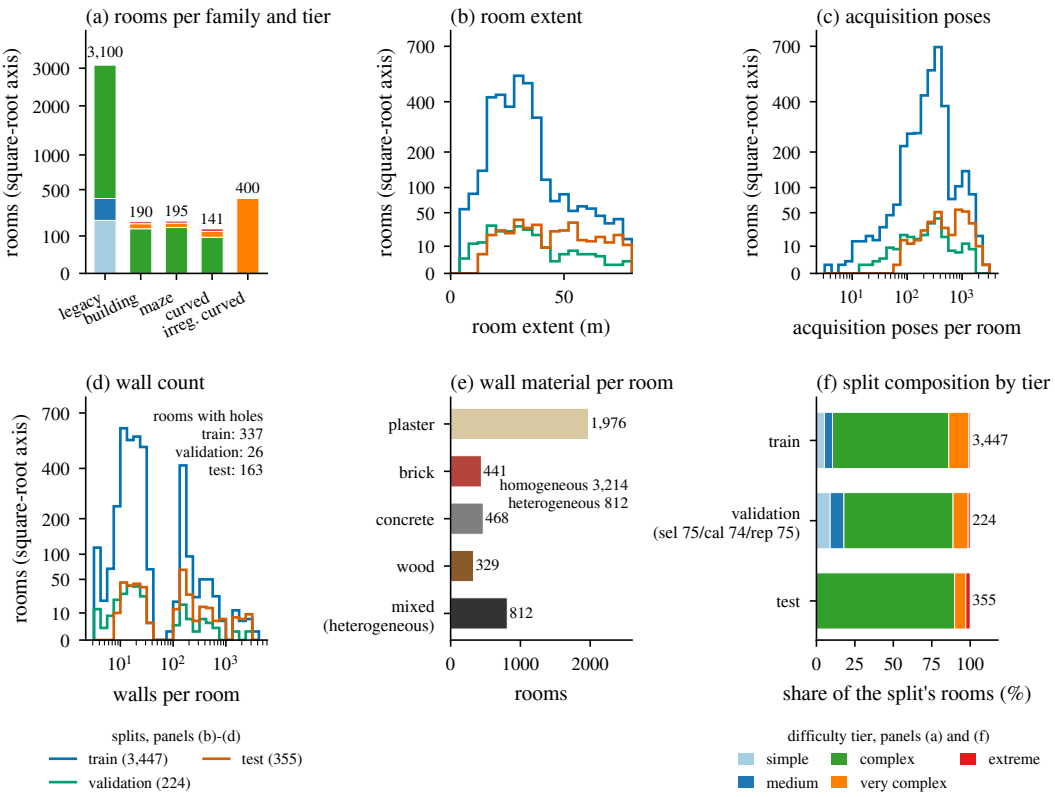

Figure 5: EchoRooms statistics: (a) rooms per layout family and difficulty tier; (b–d) per-split distributions of room extent, acquisition poses per room and walls per room, with the rooms with holes per split in (d); (e) rooms per wall-material class, with the homogeneous / heterogeneous split; (f) split composition by tier for the training, validation and test splits.

Training acoustics. Room RT_{60} medians range from 0.50 to 7.67 s, with a median of 1.61 s across rooms. The Schroeder fit is reliable for 380 rooms and flagged *unreliable* for 3,067, mostly the largest rooms whose decay is not exponential. Wall-material classes comprise plaster (1,680 rooms), mixed (690), concrete (409), brick (380) and wood (288). Material profiles comprise office (1,236), residential (871), industrial (505), masonry (477) and timber (358).

Validation and test rooms. The validation split (224 rooms, 79,711 poses) contains 20 simple, 20 medium, 159 complex, 22 very complex and 3 extreme rooms from all 36 types (183 legacy, 8 building, 11 maze, 7 curved, 15 irregular curved). The test split (355 rooms, 246,404 poses) has a median extent of 42.0 m (quartiles 30.1 and 56.8), a median of 32 walls (95th percentile 1,527, maximum 3,131), a median of 505 poses (minimum 72, maximum 2,749), and 163 rooms with holes. It contains 180 legacy, 61 building, 45 maze, 57 curved and 12 irregular curved rooms, spanning 319 complex, 26 very complex and 10 extreme rooms.

A.2 Split algorithm and test-read history

Rooms are assigned by a fixed-quota rule for each difficulty tier and room type (seed 42). Within each type, rooms are ordered by the SHA-256 hash of the seed and geometry identity, and the first q rooms fill the current split, where q is that split's quota for the type. Split groups are geometry identities, so no geometry can appear in two splits.

Test allocation and read history. The test split takes 10 rooms per complex type, 12 per very-complex type and 8 per extreme type, drawn only from rooms not already assigned to training and capped by the number of such rooms: no building-hard or maze-hard room qualified, and only one building-extreme, one maze-extreme and six curved-extreme rooms did. The test split also includes 113 rooms generated exclusively for testing after the original quota rooms had been read once: 50 building, 34 maze and 29 curved rooms (median extent 58.1 m, 131,139 poses). Their per-type counts

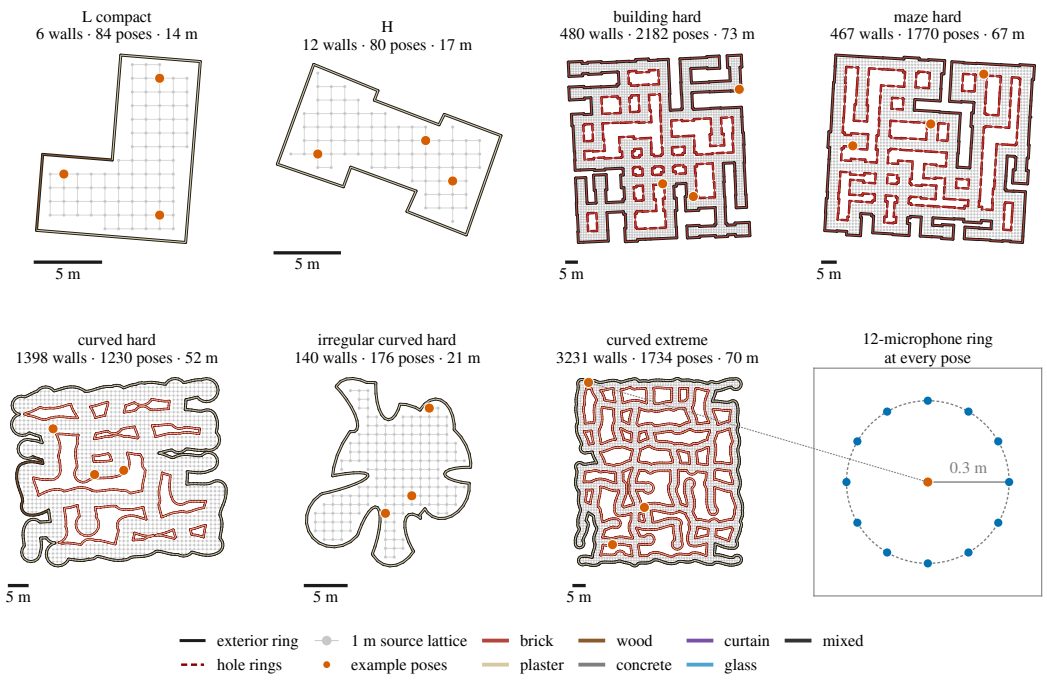

Figure 6: Seven EchoRooms rooms with 5 m scale bars (two legacy types, one per procedural family, one extreme-tier room), drawn as in Fig. 2: exterior ring, hole rings, per-wall material classes, the 1 m source lattice and three example poses; the last cell is the 12-microphone ring at every pose.

are included in Table 4; they add two curved-hard and two curved-extreme rooms to the quota rooms. Simple and medium types are absent from the test split.

Table 5 shows how adding the large rooms changed the component results and those of the reference recipe's policy: each model was read on the 242 original rooms and again on all 355 rooms, without recipe changes. Completion and policy scores fall with the added large rooms, whose interiors remain mostly unmeasured within the 64-acquisition budget. The occlusion and STOP-floor analyses also use the 355-room split (Appendices D.7 and E.6). Appendix E.7 gives the full test-read history.

Validation allocation. Difficulty quotas set the validation tier counts given above, retaining at least one room per type in both validation and training. Strictly improving swaps within each type balance the material-profile quotas. The selection / calibration / reporting roles (75 / 74 / 75) are assigned by per-type largest remainder with a seeded generator.

Simulator settings and quality gates Three builder families produce the 36 room types in five difficulty tiers. Corner chains give four simple, four medium and eighteen complex types (the legacy layout family); procedural graph layouts with interior holes give nine building, maze and curved types (footprints up to 80 m); and the irregular curved family is bounded by ellipse arcs. Each wall receives one of six material classes with octave-band absorption from five profiles. The platform is a point source at the center of a 0.3 m ring of twelve omnidirectional microphones at a height drawn from $\mathcal{U}(1.0, 1.5)$ m; sources sit at every node of a 1 m lattice inside the polygon eroded by the 0.45 m platform clearance.

Acoustic simulation. We use pyroomacoustics 0.10.1 at 16,000 Hz with 48,000-sample RIRs (3.0 s). Image-source order is capped at 6/4/3 for rooms with at most 8/12/more walls, and at first order above 80 faces. Ray tracing uses 50,000 rays, receiver radius 0.15 m, histogram bins of 4 ms, energy threshold 10^{-7} and time threshold 3.0 s. Air absorption is enabled; temperature and relative humidity are drawn per room from $\mathcal{U}(18, 26)$°C and $\mathcal{U}(30, 70)$ %. For late-tail Poisson arrivals, we replace the library's sampler with exact inverse-integrated-intensity sampling with rate $\lambda(t) = 4\pi c^3 t^2 / V$ (speed of sound c, room volume V). Rooms are extruded floor plans with earcut-triangulated floors and ceilings; ceiling heights follow the scale stratum (2.4–7.0 m).

Materials. Wall classes are brick (rough brick, brickwork; scattering 0.10–0.30), plaster (plasterboard; 0.03–0.12), wood (lining, 16 mm wood, thin plywood; 0.05–0.20), curtain, glass and concrete

Table 4: Rooms per room type in the training, validation and test splits, grouped by difficulty tier.

Room type	Train	Validation	Test	All
simple (4 types)				
convex	45	5	0	50
near rect	45	5	0	50
rectangle	45	5	0	50
trapezoid	45	5	0	50
medium (4 types)				
L compact	45	5	0	50
T compact	45	5	0	50
U notch	45	5	0	50
plus compact	45	5	0	50
complex (21 types)				
E	132	8	10	150
H	132	8	10	150
S	132	8	10	150
U	132	8	10	150
angled	132	8	10	150
apartment	132	8	10	150
building	132	8	10	150
courtyard	132	8	10	150
cross	132	8	10	150
culdesac	132	8	10	150
double corr	132	8	10	150
gallery	132	8	10	150
hospital	132	8	10	150
labyrinth	132	8	10	150
maze	132	8	10	150
multi wing	132	8	10	150
office	132	8	10	150
ring	133	7	10	150
building standard	78	4	60	142
maze standard	99	7	44	150
curved standard	53	5	35	93
very complex (4 types)				
building hard	31	3	0	34
maze hard	27	3	0	30
curved hard	18	1	14	33
irregular curved hard	373	15	12	400
extreme (3 types)				
building extreme	12	1	1	14
maze extreme	13	1	1	15
curved extreme	6	1	8	15
Total	3,447	224	355	4,026

Table 5: Component and reference-recipe results before and after adding 113 large rooms (242 vs. 355; one test read per stage, unchanged recipes). Primary memory uses the seed-42 legacy-fusion fit before fine-tuning. Mapper/refiner results show mean $\pm$ SD over four seeds; policy results show room-bootstrap 95% CIs. Policy acquisitions are failure-penalised; success requires final wall recall $\geq 90\%$. Evaluation details are in Appendices D and F.

Model and metric	Earlier test (242 rooms)	Test (355 rooms)
Polar echo encoder (Stage 1), near-range echo F1	97.1	97.0
Local geometry mapper (Stage 2), current-view near-wall F1 (4 seeds)	83.2 ± 0.2	80.6 ± 0.2
History refiner (Stage 2), prefix-knowledge near-wall F1 (4 seeds)	92.1 ± 0.1	91.1 ± 0.1
Spatial memory (primary), prefix-knowledge near-wall F1	88.9	87.5
Spatial memory (primary), whole-room completion F1 (0.30 m) at 64	56.1	45.2
Spatial memory (primary), interior IoU at 64	58.5	45.1
Trained-STOP policy, usable rooms (node span $\leq$ 52 m)	211	246
Trained-STOP policy, acquisitions to 90% [95% CI]	28.6 [26.1, 31.5]	33.8 [30.9, 36.8]
Trained-STOP policy, success within 64 (%) [95% CI]	73.9 [68.6, 79.1]	63.7 [58.0, 69.2]

(rough and unpainted; 0.10–0.25). Floors use thin carpet, soft 10 mm carpet, linoleum on concrete, closed-cell foam carpet or concrete; ceilings use plasterboard, perforated gypsum or fissured tile. Profile priors are office 0.35, residential 0.25, masonry 0.15, industrial 0.15 and timber 0.10. Accent materials cover at most 30 % of the boundary and are assigned per tangent-continuous surface group (maximum turn 30°).

Quality gates. RIRs must have finite samples, a direct arrival on every channel with amplitude at least 0.1 of the peak, first-order echo consistency of at least 0.85 on four probe poses, and last-50 ms tail energy at most 10^{-3}. A Schroeder T_{30} fit with $R^2 < 0.8$ marks the RT_{60} label as unreliable; the room is retained.

A.3 Label construction

Three label types serve the pipeline: polar echoes for the polar echo encoder (Stage 1), local view rasters for the local geometry mapper (Stage 2), and whole-room rasters for completion targets and interior IoU.

Polar echoes. The Stage 1 target for a (pose, heading) pair is computed in the device frame rotated by the heading yaw. For every wall segment of the exterior and hole rings, the source is reflected across the segment line. The image is retained if the perpendicular foot lies within the segment (corner tolerance 0.01 m), is not corner-adjacent and has an unblocked line of sight from the source. Echo range is half the image distance and bearing is the image direction. Each echo is rendered as a Gaussian with $\sigma_r = 0.1\,\mathrm{m}$ and $\sigma_\varphi = 5^\circ$ on the 60×72 grid; the sum is clipped to $[0, 1]$. Per-pair metadata records counts of walls, emitted and kept echoes, and dropped images by reason (foot outside, corner adjacent, occluded or degenerate).

Local and whole-room rasters. The Stage 2 view target casts one ray per cell bearing from the pose over the 0.125 m grid within 8 m. Cells before the first hit are free; cells whose nearest hit on the selected segment lies within 0.125 m are wall; all other cells, including those outside the circle, are unknown. The whole-room raster labels wall (a 0.125 m band around every segment), interior and exterior over the polygon with a 2 m margin. Wall takes precedence over interior.

B Polar echo encoder (Stage 1)

This appendix specifies the polar echo encoder of Sec. 4.2: its architecture, echo loss, training recipe and echo metric. It also reports the representation control behind the polar-supervision result of Sec. 5.4. It opens with a layer-level view of the whole deployed pipeline, to which later appendices refer.

B.1 The deployed pipeline at layer level

Fig. 7 expands the deployed pipeline of Fig. 3, showing each layer and interface from acquisition through the environment step.

B.2 Encoder architecture

Input and temporal encoder. The input comprises a float32 RIR window of 1,600 samples per microphone (100 ms), microphone positions u_i in metres in the device frame, and a boolean microphone mask. All observed channels are divided by their shared maximum absolute amplitude, or set to zero if it is below 10^{-12}. A shared per-microphone temporal CNN starts with Conv1d($1 \to 32$, $k = 9$) and masked GroupNorm, followed by four residual blocks ($32 \to 48 \to 64 \to 96 \to 96$, $k = 5$, stride 2 each, masked GroupNorm with 8 groups, GELU). The total stride of 16 yields 100 delay tokens of 96 channels per microphone.

Microphone mixing and range readout. At each delay, 4-head attention of width 96 mixes microphones using additive bias $-\|u_i - u_j\|^2$, masked keys, a residual connection and LayerNorm. In the ingredient study (Table 8), *without mic attention* removes this block and *capacity control* replaces it with a feed-forward block $96 \to 192 \to 96$. A fixed linear interpolation matrix maps delay tokens to 60 target ranges d at sample index $n = f_s \cdot 2d/c + 40$, with sampling rate f_s and $c = 343\,\mathrm{m/s}$; the 40-sample offset is the simulator's fractional-delay filter centre. The interpolated tokens provide queries and the residual for a *global readout*, 4-head attention over all 100 delay tokens with 8-octave Fourier time positions. *Without global readout* removes this attention, and *global capacity control* replaces it with an equally sized feed-forward block.

Polar head and microphone subsets. The coordinate polar head steers actual microphone azimuths onto 12 bearing taps (kernel $12 \times 96 \times 64$), followed by GroupNorm, two circular convolutions over bearing, and a 1×1 logit convolution with bias initialised at -4.0. It outputs 60×72 logits. The encoder has 471,457 parameters for every microphone count. The *legacy encoder*, a baseline with 842,081 parameters, uses a ring-convolution trunk with a steered polar head at fixed

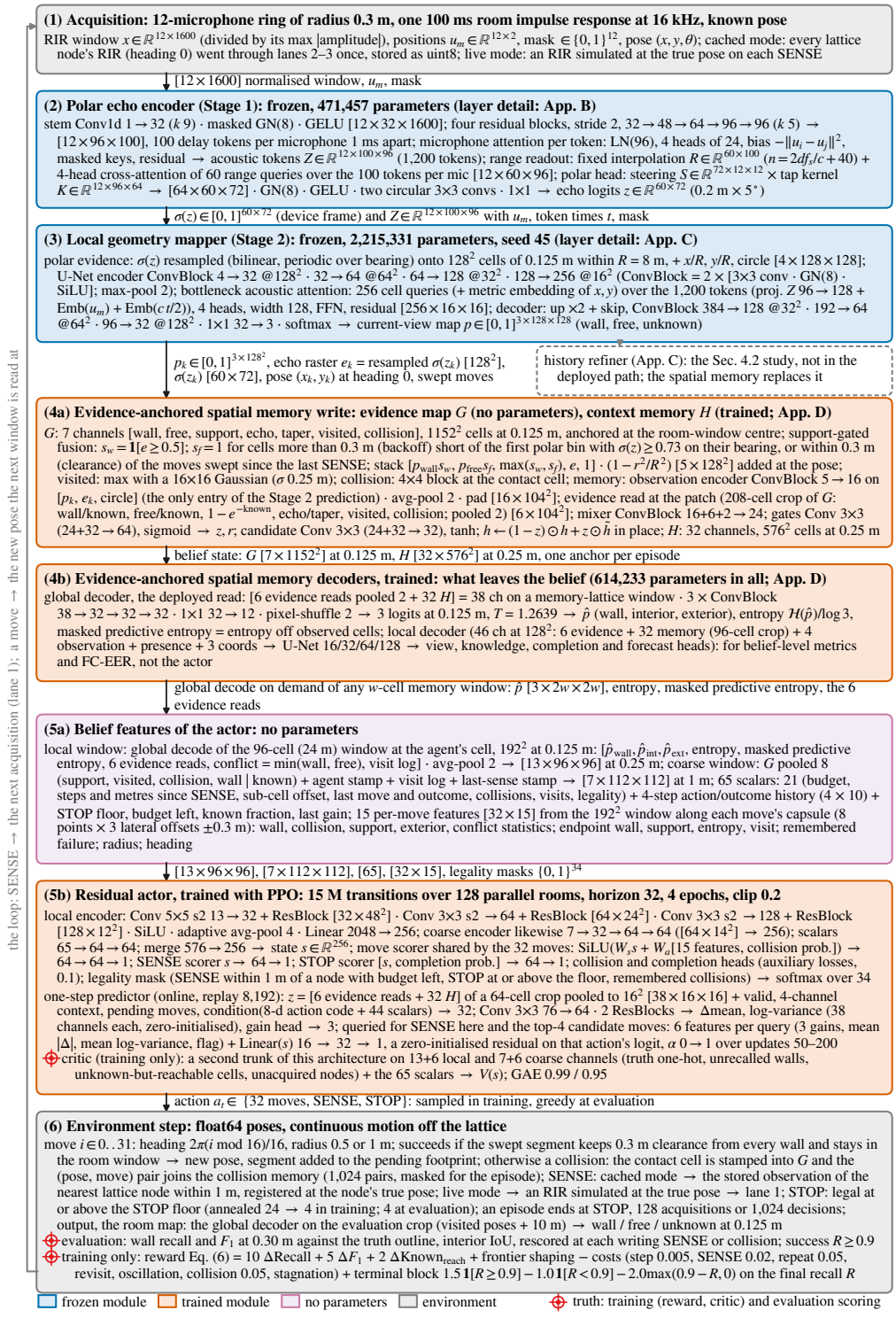

Figure 7: The deployed pipeline at layer level, six lanes top to bottom: (1) acquisition, (2) the frozen polar echo encoder (Stage 1), (3) the frozen local geometry mapper (Stage 2), (4) the write and decoders of the evidence-anchored spatial memory, (5) the actor with its belief features, one-step predictor and privileged critic, (6) the environment step and its outputs; the grey arrow closes the loop (a SENSE yields the next acquisition, a move the next pose). Colours as in Fig. 3: frozen blue, trained orange, parameter-free violet, environment grey; target icons mark where ground truth enters: training (reward, critic) and evaluation scoring. Bracketed shapes are per environment as [channels × height × width]; w memory cells (0.25 m) decode to $2w$ evidence cells (0.125 m). The history refiner of Sec. 4.2 is a side branch, not in the deployed path.

six-microphone angles. The 3-microphone subset uses channels (0, 4, 8); the 6-microphone subset uses even channels; the 9-microphone subset uses (0, 1, 2, 4, 5, 6, 8, 9, 10); the 12-microphone model uses all channels.

B.3 Echo loss and loss weights

The echo loss of Sec. 4.2 is the positively weighted binary cross-entropy

$$\mathcal{L}_{\text{echo}} = \frac{1}{60 \cdot 72} \sum_{r,\varphi} \Big[w\, y_{r\varphi}\, \text{softplus}(-z_{r\varphi}) + (1 - y_{r\varphi})\, \text{softplus}(z_{r\varphi}) \Big], \qquad w = 8. \tag{1}$$

Setting the derivative of the per-element term to zero at a cell with expected target μ gives $\sigma(z^*) = w\mu/(w\mu + 1 - \mu)$, equivalently, logit $q^* = \text{logit}\,\mu + \log w$ for the optimal probability $q^* = \sigma(z^*)$. On EchoRooms the binary positive rate is 0.00148 and the soft-target mass 0.00212 (inverse 471.4). The seed-0 screen (training recipe below) compared $w \in \{1, 8, 32, 128, 619\}$, the binary and soft inverse ratios and a focal loss (Lin et al., 2017) (selection F_1: 93.11, 93.13, 92.76, 91.80, 89.73, 89.14, 90.44, 92.98 at 5,000 updates); $w = 8$ was selected. In Table 8, *unweighted BCE* uses $w = 1$ and *historical positive-weight BCE* the historical weight; the *legacy encoder* keeps its historical loss, and *legacy encoder, selected loss* trains it with $w = 8$. The total Stage 1 loss adds a reconstruction auxiliary on the observed microphones M,

$$\mathcal{L}_1 = \alpha_{\text{echo}}\, \mathcal{L}_{\text{echo}} + \alpha_{\text{rec}}\, \frac{1}{|M|} \sum_{m \in M} \big\| \hat{s}_m - s_m \big\|_2^2, \tag{2}$$

with fixed coefficients $\alpha_{\text{echo}} = 89.1565$ and $\alpha_{\text{rec}} = 16.2737$, the latter chosen so that the reconstruction gradient is a fraction 0.3 of the encoder gradient at initialisation; *without reconstruction* (Table 8) sets $\alpha_{\text{rec}} = 0$. The reconstruction decoder is Conv1d($96 \to 64$, $k = 3$) with GELU and a linear map $64 \to 16$ per token, producing a 1,600-sample waveform per microphone. A window-length screen (100, 200, 400 ms: 93.30, 93.10, 92.86 selection F_1) chose 100 ms.

B.4 Training recipe and checkpoint selection

Training uses AdamW (Loshchilov & Hutter, 2019) with learning rate 3×10^{-4}, weight decay 0.01, 500 warm-up updates, cosine decay to 10^{-5} over 20,000 updates, and gradient clipping at 1.0. A room-chunk sampler forms batches of 64 samples as 8 rooms $\times$ 8 poses, with random heading augmentation. Training uses FP32 with TF32 disabled, saves checkpoints every 250 updates and evaluates every 1,000.

A seed-0 screen trains each recipe for 5,000 updates. Checkpoints are selected by room-macro average precision on the selection rooms. Decision thresholds are fitted on the calibration rooms by room-macro near-range F_1, searching logits $-12 \ldots 12$ in steps of 0.25. Recipes are ranked by calibrated selection F_1 and then confirmed by a seed-42 fit of 20,000 updates. Frozen-feature probes use 2,000 AdamW updates. The report was frozen before the reporting rooms were opened. The selected checkpoint is update 20,000 with threshold logit 1.0.

B.5 Peak decoding, echo metric and ablations

The echo metric decodes discrete peaks from the polar map and matches them to oracle echoes. Peaks are 3×3 local maxima (circular in bearing) of the sigmoid map above the threshold, pruned by greedy non-maximum suppression with a radius of 2 range bins and 2 bearing bins, at most 64 per map. Predicted peaks and oracle echoes are matched by maximum-cardinality bipartite matching under a gate of 0.5 m and 10°. True positives, false positives and false negatives are summed within a room over all poses and the 12 headings; F_1 is computed per room, and rooms are averaged with equal weight. Near range covers the first 44 range bins (8.8 m).

Table 6 reports the selected 12-microphone encoder on every validation role and on the test split, by difficulty tier and layout family, and Fig. 9 shows its decoded peaks at one test pose. On the reporting rooms it has near-range AP 95.4, range MAE 0.050 m and bearing MAE 1.45°. At heading 0 it scores 64,747 true positives, 866 false positives and 1,435 false negatives over 66,182 oracle echoes. Batch-1 latency is 6.99 ms median on an A40. Table 7 and Fig. 8 compare microphone counts, and Table 8 compares the encoder ingredients at six microphones. There, the legacy encoder trained with the selected loss scores 1.8 points above the six-microphone reference (97.0 against 95.3 near-range F_1) but has 842,081 parameters against 471,457. The deployed encoder keeps the smaller architecture with 12 microphones and reaches 97.5 (Table 7); the study has no 12-microphone legacy encoder.

Table 6: Polar echo encoder (Stage 1; 12 microphones, seed 42, calibrated threshold logit 1.0): near-range F_1 (%; first 8.8 m), precision, recall, average precision, range and bearing errors of matched echoes, and full-range F_1. Validation roles and the test split (355 rooms) are reported with equal room weights; test results are stratified by difficulty tier and layout family.

Partition	Rooms	Near F1	Precision	Recall	AP	Range MAE (m)	Bearing MAE (°)	Full F1
Validation (role)								
selection	75	97.5	98.2	97.0	95.5	0.051	1.39	97.2
calibration	74	97.3	98.3	96.5	95.4	0.050	1.42	97.1
reporting	75	97.5	98.0	97.1	95.4	0.050	1.45	97.2
combined	224	97.4	98.1	96.9	95.5	0.050	1.42	97.2
Held-out test (355 rooms)								
all rooms	355	97.0	98.4	95.9	96.6	0.050	1.45	97.0
Test by difficulty								
complex	319	98.1	98.7	97.5	97.1	0.050	1.40	98.0
very complex	26	87.0	95.6	79.8	91.4	0.050	1.88	87.2
extreme	10	90.5	96.0	85.8	93.4	0.049	1.67	90.5
Test by layout family								
legacy	180	98.9	98.6	99.3	96.8	0.051	1.40	98.6
building	61	99.4	99.1	99.8	99.0	0.050	1.16	99.6
maze	45	99.5	99.2	99.8	99.0	0.050	1.43	99.6
curved	57	88.6	97.0	81.5	92.6	0.049	1.79	88.8
irregular curved	12	87.1	95.8	80.0	91.4	0.050	1.99	87.7

Table 7: Microphone-count study (75 reporting rooms, seed 42): near/full-range F_1 (%), AP, matched-echo errors and batch-1 A40 latency; 471,457 parameters per model. Δ = subset minus 12 microphones (10,000 paired room-bootstrap resamples, 95 % CI; Holm-corrected p). Last column: the same architecture trained and evaluated on the earlier 10,507-room development corpus.

Microphones	Near F1	Full F1	AP	Range MAE (m)	Bearing MAE (°)	Latency (ms)	Δ vs 12 mics [95% CI]	Holm p	Earlier-corpus F1
3	90.8	89.8	84.5	0.051	1.86	6.92	−6.6 [−8.4, −5.1]	<0.001	86.2
6	95.3	94.5	91.3	0.050	1.62	6.78	−2.2 [−3.0, −1.5]	<0.001	96.3
9	96.8	96.5	93.6	0.050	1.53	6.90	−0.6 [−0.9, −0.5]	<0.001	97.5
12	97.5	97.2	95.4	0.050	1.45	6.99	–	–	98.1

B.6 Representation control: what polar supervision adds to the tokens

This control asks how the training objective shapes the delay tokens, separately from the polar field. The acoustic controls of Table 10 already show that the deployed local geometry mapper (the *hybrid mapper*, which reads both the polar field and the delay tokens) depends on the polar field. Table 9 therefore compares the same encoder trained with one of three objectives: reconstruction alone ($\alpha_{\text{echo}} = 0$, leaving the clean-waveform auxiliary, which the encoder solves almost trivially), masked reconstruction (25 % of 64-sample input blocks zeroed; this single setting was not tuned), or polar supervision (Table 6). Each frozen encoder feeds a token-only mapper with zero polar input and acoustic attention retained, trained with the Stage 2 recipe and budget (one seed, 42, per arm).

Polar-supervised tokens yield near-wall F_1 of 72.9 % on the 75 reporting rooms and 71.3 % on the 355-room test, versus 36.8 and 35.2 % for reconstruction-only tokens (paired room-bootstrap differences $+36.1$ [34.5, 37.6] and $+36.1$ [35.3, 36.9] points). Masked reconstruction improves these scores to 44.2 and 42.2 %, but remains 28.7 [27.4, 30.0] and 29.1 [28.4, 29.8] points below polar supervision. Echo supervision thus improves the geometry available to the token-only mapper beyond either reconstruction objective as configured here; a tuned masking setting or further seeds could narrow the masked-reconstruction gap.

The seed-42 hybrid mapper, the control, also reads the explicit echo field and reaches 85.9 and 80.7 % ($+13.1$ [11.7, 14.4] and $+9.4$ [8.8, 9.9] points over the polar token-only arm). Appendix C describes this hybrid mapper.

C Local geometry mapper (Stage 2) and history refiner

This appendix specifies the local geometry mapper of Sec. 4.2, the deployed single-shot model that turns one encoded RIR into a wall/free/unknown map around the robot, and the diagnostic history

Table 8: Ingredients of the polar echo encoder (Stage 1; 75 reporting rooms, 6-microphone reference, seed 42): near-range F_1 (%), parameters and reference-minus-variant differences (10,000 paired room-bootstrap resamples, 95 % CI; p Holm-corrected within family). The variants are defined in the architecture and echo-loss paragraphs above.

Recipe (vs. 6 microphones)	Family	Near F1	Params	Δ (6 mics − recipe) [95% CI]	Holm p
6 microphones (reference)	reference	95.3	471,457	–	–
global capacity control	ingredients	94.4	471,457	+0.9 [0.2, 1.5]	0.031
without global readout	ingredients	94.2	430,945	+1.1 [0.4, 1.8]	0.011
unweighted BCE	ingredients	95.0	471,457	+0.30 [0.00, 0.68]	0.132
capacity control	ingredients	94.4	471,457	+0.9 [0.1, 1.5]	0.037
without reconstruction	ingredients	95.0	471,457	+0.23 [0.01, 0.50]	0.132
without mic attention	ingredients	94.3	434,017	+1.0 [0.2, 1.6]	0.029
legacy encoder, selected loss	ingredients	97.0	842,081	−1.8 [−2.5, −1.1]	<0.001
historical positive-weight BCE	ingredients	93.2	471,457	+2.1 [1.7, 2.6]	<0.001
legacy encoder	primary	95.7	842,081	−0.5 [−1.1, 0.1]	0.134

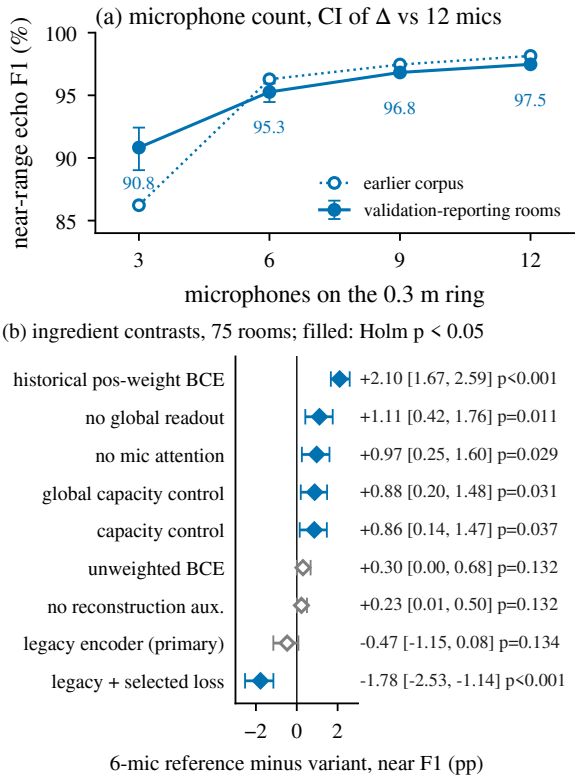

Figure 8: (a) Near-range echo F_1 against microphone count on the validation reporting rooms (75), seed 42, with the paired-bootstrap interval of each count minus 12 microphones, and the same architecture on the earlier development corpus (hollow); (b) forest plot of the ingredient contrasts of Table 8 (6-microphone reference minus variant), filled where Holm $p < 0.05$.

refiner. The mapper is called the *hybrid mapper* because it reads both Stage 1 outputs, the polar echo field and the delay tokens. The controls below support two statements of Sec. 5.4: the mapper draws its geometry mainly from the polar echo field, and acoustic history improves the refiner.

C.1 Mapper loss, architecture and training

The mapping loss of Sec. 4.2 is

$$\mathcal{L}_{\text{map}}(p, y) = \frac{\sum_{c \in \mathcal{C}} w_{y_c} \, \text{CE}(p_c, y_c)}{\sum_{c \in \mathcal{C}} w_{y_c}} + \tfrac{1}{2}\Big(1 - \frac{2 \sum_c p_c^{\text{wall}} \, \mathbf{1}[y_c = \text{wall}] + 1}{\sum_c p_c^{\text{wall}} + \sum_c \mathbf{1}[y_c = \text{wall}] + 1}\Big), \tag{3}$$

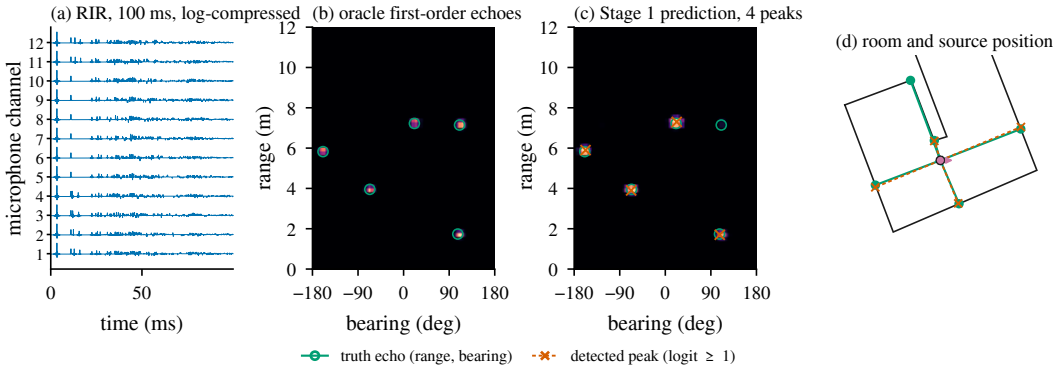

Figure 9: Qualitative Stage 1 example at one pose of a held-out test-split room: (a) the 12-channel 100 ms RIR window (log-compressed), (b) the oracle first-order echoes in range and bearing, (c) the predicted polar map with the detected peaks (logit ≥ 1) against the truth echoes, and (d) the room with the source position.

Table 9: Representation control (seed 42): near-wall and full-radius F_1 (%) for token-only local geometry mappers (Stage 2; polar input zeroed) with each encoder objective, and the deployed hybrid mapper as control. Reporting: 75 rooms; test: 355 rooms. Near-wall differences use 10,000 paired room-bootstrap resamples (95 % CI).

	Validation reporting (75)		Held-out test (355)	
Encoder objective × mapper (seed 42)	Near-wall F1	Full-radius F1	Near-wall F1	Full-radius F1
Polar-supervised encoder, hybrid mapper (deployed)	85.9	72.0	80.7	65.0
Polar-supervised encoder, token-only mapper	72.9	51.4	71.3	49.9
Reconstruction-only encoder, token-only mapper	36.8	19.7	35.2	18.0
Masked-reconstruction encoder, token-only mapper	44.2	27.1	42.2	24.2
Paired room bootstrap, near-wall F1 (10,000 resamples, seed 42), difference as labelled [95% CI]				
hybrid − token-only (polar)	+13.1 [11.7, 14.4]		+9.4 [8.8, 9.9]	
token-only (polar) − token-only (reconstruction)	+36.1 [34.5, 37.6]		+36.1 [35.3, 36.9]	
token-only (polar) − token-only (masked)	+28.7 [27.4, 30.0]		+29.1 [28.4, 29.8]	

with predicted class probabilities p_c and label y_c at cell c, class weights $w = (4, 1, 1)$ for wall, free and unknown, and $\mathcal{C}$ the cells inside the circle. The output grid is 128×128 at 0.125 m resolution (radius 8 m), with a 0.125 m wall band; cells outside the circle are forced to (wall, free, unknown) = (0, 0, 1).

Inputs and architecture. The frozen 12-microphone polar echo encoder (Stage 1) supplies polar probabilities (60×72) and delay tokens ($12 \times 100 \times 96$). Polar probabilities are resampled bilinearly onto the grid with periodic angular padding, masked by the circle and concatenated with the cell coordinates x/R and y/R (circle radius R) and the mask (4 channels). A four-level U-Net has widths 32, 64, 128 and 256; each block contains two Conv3 × 3–GroupNorm (Wu & He, 2018)–SiLU layers. It uses max-pool downsampling, bilinear upsampling with skip connections, and a 1×1 output layer producing three logits.

Acoustic attention operates at the $16 \times 16 \times 256$ bottleneck. Delay tokens pass through LayerNorm and a linear map from 96 to 128 channels, then receive a metric embedding of microphone position and token time, scaled by 171.5 m/s (half the speed of sound). Queries combine a 1×1 bottleneck convolution with a metric embedding of cell position. The block uses 4-head attention of width 128, a feed-forward layer $128 \to 256 \to 128$, and a residual 1×1 projection back to 256 channels. The mapper has 2,215,331 parameters.

Training and calibration. Training runs for 20,000 updates with effective batch size 32 (micro-batch 8, FP32), AdamW learning rate 10^{-4}, weight decay 10^{-4}, 500 warm-up updates, cosine decay to 10^{-6} and gradient clipping at 1.0. Evaluation occurs every 1,000 updates and checkpoints are saved every 250. Rooms are sampled uniformly from the training split, with at most 16 observations

per room in a batch. Contiguous windows come from seeded blind depth-first episodes: cardinal first-visit search, translation cost 1 m, turn cost $\pi/6$, at most 256 attempts and 64 acquisitions. Each fit takes about 78.5 ks on one A40, with 714 MB peak memory. For seed 45, selection by near-wall F_1 chooses update 20,000 with a score of 0.8591. Across seeds 42–45, peak scores range from 0.8582 to 0.8608. Seed 45, selected together with its history refiner by prefix-knowledge near-wall F_1 (below), is the deployed local geometry mapper that feeds the spatial memory and the policy. One temperature per model is fitted on the 74 calibration rooms (1,183 measurements), giving 0.990 for seed 45; it serves the mapper's standalone reports only, and the spatial memory reads the mapper at temperature 1 (Appendix D).

C.2 Diagnostic history refiner

The causal history refiner is diagnostic and not in the deployed path; it tests how much acoustic history can add to a local map. The refiner writes five rasters per acquisition in its own frame (at most 64 observations) and reads them back warped into the query frame. It feeds 13 conditioning channels (current view, memory read-out, coordinates and budget) to a four-level U-Net (widths 16, 32, 64 and 128) with two zero-initialised 1×1 residual heads: a current-view head and a prefix-knowledge head, whose target is the union of wall and free labels in the causal prefix. The refiner has 489,526 trainable parameters and is trained for 20,000 updates on fixed training-split trajectories (blind depth-first search, at most 64 acquisitions). Seed 45 is selected by prefix-knowledge near-wall F_1 (knowledge temperature 1.132, view temperature 1.076).

C.3 Local-mapping results and controls

The results below compare the single-shot mapper with the refiner and then test which inputs each relies on. On the 355-room test, near-wall F_1 across four seeds is 80.6 ± 0.2 % for the hybrid mapper, 90.5 ± 0.1 % for the history current view and 91.1 ± 0.1 % for prefix knowledge. Table 11 gives every partition and the per-seed paired differences between the refiner and its matched foundation; Fig. 10 shows a qualitative example of the hybrid mapper.

History controls. Table 12 evaluates the refiner on reporting and test rooms. Disabling history or using the frozen local foundation (the refiner's own frozen single-shot mapper) returns predictions to the single-shot level. On reporting rooms, zeroing the stored acoustics leaves only the pose trace (near-wall F_1 83.2 %), shuffling them across poses reduces near-wall F_1 to 34.2 %, and Gaussian pose noise (0.05 m translation, 2° yaw) costs 2.0 points.

Acoustic controls. Table 10 tests the single-shot hybrid mapper (seed 45) without retraining. At read time, one or both Stage 1 outputs are zeroed or replaced by outputs from donor poses, assigned through a seeded derangement of the evaluation records. Removing the polar echo field collapses mapping performance; perturbing the delay tokens costs 0.7–3.1 near-wall F_1 points. These controls identify the polar field as the mapper's main source of acoustic geometry. Appendix B separately tests how supervision shapes the tokens.

Inference cost. Table 13 reports parameter counts, latencies, memory and calibrated temperatures of the frozen modules. Its last row times one full decision of the FC-EER planner (forecast-conditioned entropy reduction, Sec. 4.3), which copies the belief for up to 32 candidates and decodes each on the covering crop; Appendix F gives the decision and candidate counts of this timing.

D Evidence-anchored spatial memory and the lattice planner study

This appendix specifies the evidence-anchored spatial memory of Sec. 4.3: its dimensions, losses, training, support-gated fusion and the fine-tuned belief. It then reports the belief-level evaluations behind Sec. 5.4 (whole-room completion, retrained ablations, visible and occluded walls) and the lattice planner study with the crossed replay of Sec. 5.3. Unless stated otherwise, *primary* denotes the EchoRooms fit with the recurrent context memory H under legacy fusion (seed 42 unless another seed is named), before the fine-tuning of Appendix D.6. The seed-42 primary is the legacy-fusion arm's belief, not the main method's fine-tuned belief.

D.1 Module dimensions, writes and reads

This subsection gives the sizes and data flow behind the description in Sec. 4.3; Fig. 11 shows the write path with its tensor shapes. The recurrent context memory H holds a 32-channel state per cell.

Table 10: Acoustic controls of the local geometry mapper (Stage 2; single-shot hybrid mapper, seed 45), near-wall and full-radius F_1 (%) on the validation reporting rooms (75) and the held-out test split (355): the frozen Stage 1 polar echo field or the 1,200 delay tokens are zeroed or replaced by those of a donor pose at read time; *zero RIR* passes a zero waveform through Stage 1 and *shuffled RIR* passes the donor pose's waveform.

	Validation reporting (75)		Held-out test (355)	
Condition (single-shot hybrid mapper, seed 45)	Near-wall F1	Full-radius F1	Near-wall F1	Full-radius F1
Full model	85.9	72.2	80.5	65.1
Zero echo field	0.9	2.6	1.6	1.8
Shuffled echo field (donor pose)	8.5	6.5	8.4	6.6
Zero delay tokens	84.9	68.0	79.8	61.3
Shuffled delay tokens (donor pose)	82.8	68.2	78.4	62.4
Zero echo field and tokens	0.0	0.9	0.0	1.1
Zero RIR through Stage 1	11.0	6.7	10.8	6.4
Shuffled RIR (donor pose, both outputs)	8.4	6.5	8.3	6.5

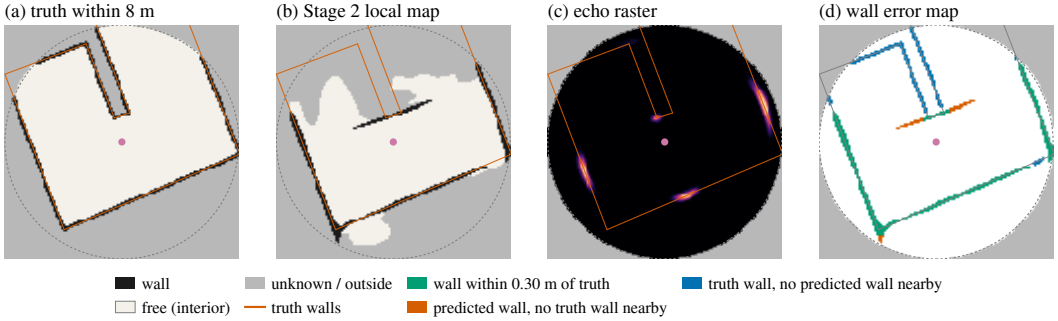

Figure 10: Qualitative Stage 2 example at one pose of a held-out test-split room: (a) the ray-cast truth within 8 m, (b) the single-shot hybrid local map, (c) the polar echo evidence rasterised onto the grid, and (d) the wall error map at a 0.30 m tolerance (predicted walls near a true wall, true walls without a prediction, and predictions without a true wall).

Within the write patch, the state h_t is updated from the mixer output x_t (encoded observation and evidence read; below) by

$$\begin{aligned} z_t &= \sigma\big(W_z * [x_t, h_{t-1}]\big), & r_t &= \sigma\big(W_r * [x_t, h_{t-1}]\big), \\ \tilde{h}_t &= \tanh\big(W_h * [x_t, r_t \odot h_{t-1}]\big), & h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \tag{4}$$

where z_t and r_t are the update and reset gates, σ is the logistic sigmoid, $*$ is a convolution with 3×3 kernels W_z, W_r, W_h, and $\odot$ is the elementwise product. The EchoRooms fit and the earlier-corpus fit use the same dimensions. The evidence map G is an occupancy-style grid (Elfes, 1989) of 1152^2 cells at 0.125 m resolution (144 m); H is a 576^2, 32-channel grid at 0.25 m resolution. Both are indexed by the world offset from a shared anchor (for the main method, the room-window centre; Sec. 4.4). Memory writes cover 104 cells (26 m); reads cover 96 cells (24 m). The global decoder's margin around the visited-pose bounding box is 10 m, the completion-loss radius 12 m and the maximum pose offset from the anchor 56 m. The observation encoder has width 16, the local U-Net decoder has widths 16, 32, 64 and 128, and evidence dropout (zeroing the evidence read of a training sequence) has probability 0.25.

Evidence and memory writes. The legacy-fusion writer rotates the five-channel observation stack of Fig. 11 onto a 184-cell world-aligned window, using exact quarter-turn rotations for multiples of 90° and bilinear grid sampling otherwise, then adds it to G at the pose's lattice offset. The visited stamp is a 16×16 Gaussian with $\sigma = 0.25$ m; the collision stamp is a 4×4 block. The support-gated writer is specified in Appendix D.6. For the memory update, the observation encoder, a ConvBlock $(5 \to 16)$, encodes the rotated Stage 2 probabilities, echo raster and sensing-circle mask; the encoding is average-pooled by 2 to 92 cells and zero-padded to the 104-cell patch. A mixer ConvBlock $(16 + 6 + 2 \to 24)$ combines this encoding with the pooled evidence read, a distance map and a presence flag into x_t. The gate convolution $(3 \times 3,\ 24 + 32 \to 64)$ supplies the sigmoid update and reset gates;

the candidate convolution (3×3, $24 + 32 \to 32$) uses tanh. The new state is written back into H in place.

Reads and decoders. A 96-cell memory patch is warped to a 64^2 egocentric grid and upsampled by $\times 2$ with taps 0.25/0.75 to 128^2. The local decoder receives 46 channels: 6 evidence, 32 memory, 4 current-observation channels (wall, free, unknown, echo), 1 presence flag and 3 coordinates. The global decoder applies three ConvBlocks ($38 \to 32 \to 32 \to 32$) to the pooled evidence and memory crop, a 1×1 convolution to 12 channels, and pixel-shuffle 2 to produce three classes at 0.125 m. Parameter counts are 614,233 for the primary, 552,201 for the geometric-only arm (the same belief trained without H) and 494,397 for the stateless arm (no G or H; current observation only). The policy reads the belief through the local and coarse windows of Appendix E.1, which include the predictive-entropy map defined in Sec. 4.3.

D.2 Forecast and completion losses

This subsection gives the loss terms of the window loss $\mathcal{L}$ in Sec. 4.3. The forecast head predicts the class and echo logits of the observation the agent would obtain at a query pose. With target probabilities t, forecast logits ℓ, echo raster e and sensing-circle cells $\mathcal{C}$, its term is

$$\mathcal{L}_{\text{fc}} = \frac{\sum_c m_c \,\mathrm{KL}\big(t_c \,\|\, \mathrm{softmax}(\ell_c)\big)}{\sum_c m_c} + \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \mathrm{BCE}\big(\ell_c^{\text{echo}}, e_c\big), \quad m_c = (t_c^{\text{wall}} + t_c^{\text{free}})\, \mathbf{1}[c \in \mathcal{C}]\,. \tag{5}$$

Here KL is the Kullback–Leibler divergence, BCE the binary cross-entropy of the echo logit ℓ_c^{echo}, and m_c weights each sensing-circle cell by its target wall-plus-free probability. The forecast is queried before each write and at one pose 2–8 steps ahead, using the masked Stage 2 probabilities and echo raster as targets. The view and knowledge heads use Eq. (3) with class weights (4, 1, 1) for wall, free and unknown and Dice weight 0.5. The softmax completion heads use $\sum_c w_{y_c}\,\mathrm{CE}_c\, m_c / \sum_c w_{y_c}\, m_c + 0.5\,\mathrm{Dice}_{\text{wall}}$, with class weight w_{y_c} of the true class y_c, per-cell cross-entropy CE_c and a mask m that differs from the forecast mask above. For the local head, m is the sensing circle, weighted by 1 inside and 0.25 outside the true room raster. The global term uses the evaluation crop with $m = \mathbf{1}[\text{within 12 m of a visited pose}] \times (1 \text{ or } 0.25)$, the same inside/outside weights. It is evaluated at two window steps per sequence: the last and one randomly chosen step. Sequence losses are averaged over window steps and weighted by (sequence length)/32.

D.3 Training, selection and calibration

This subsection details the training and temperature calibration summarised in Sec. 4.3. Both EchoRooms fits (seeds 42 and 43) use the same schedule. Each fit runs for 20,000 updates of 32 observations with AdamW: learning rate 10^{-4}, weight decay 10^{-4}, 500 warm-up updates, minimum learning rate 10^{-6} and gradient clipping at 1. The window length is 16, with two sequences per micro-batch, two global reads per sequence and one forecast-ahead query. Evaluation occurs every 1,000 updates and checkpoints are saved every 250.

Rooms are sampled uniformly. Training episodes come from a mixture of blind depth-first search (weight 0.4), causal random walk (0.3) and a privileged coverage sweep with 2 m radius (0.3). Episodes have at most 64 acquisitions and 256 motion attempts; the first sequence of each batch comes from an episode with at least 32 acquisitions. Training uses contiguous windows and random $k \times 90^\circ$ rotations about the anchor. Frozen inputs are the local geometry mapper (Stage 2) at temperature 1 and the echo raster (channel 0 of the polar evidence), cached as 8-bit unsigned maps.

The seed-42 EchoRooms fit uses 3,447 training rooms and 1,250,664 poses, taking 18,657 s with 4.18 GB peak memory. Its selection score, computed on the 75 selection rooms, is $0.5\times$ knowledge near-wall F_1 $+0.5\times$ completion F_1 at 0.30 m at the end of the depth-first episode. The score rises from 0.460 at update 1 to 0.760 at 1,000 and 0.790 at 6,000, then stays between 0.770 and its peak of 0.7930 at 19,000 (Fig. 12c). Seed 43 takes 19,815 s and peaks at 0.8007, also at 19,000.

On the 74 EchoRooms calibration rooms, the seed-42 fitted temperatures are 1.515 for knowledge (negative log-likelihood, NLL, $0.228 \to 0.206$), 1.096 for view, 1.035 for local completion and 1.311 for global completion (NLL $0.352 \to 0.337$ over 2.94 M cells). Temperature calibration adjusts the global head's predictive probabilities; it does not separate model uncertainty from observation ambiguity.

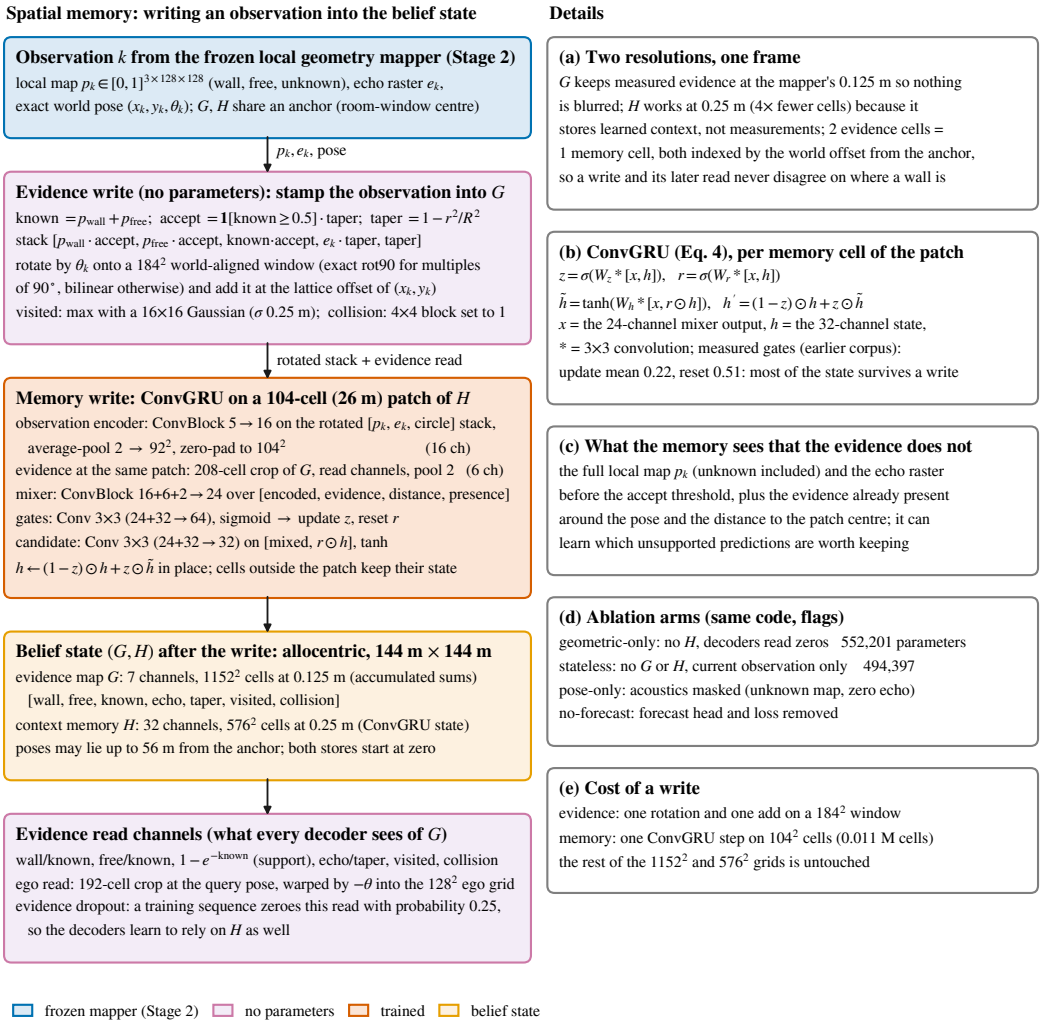

Figure 11: Write path of the evidence-anchored spatial memory, with tensor shapes. The evidence write shown is legacy fusion; support-gated fusion (Appendix D.6) instead writes local wall and free-space estimates into the evidence map G only where a first-order echo or an executed move's swept footprint supports them, while the full local predictions, including their unsupported spatial context, reach the recurrent context memory H through the observation encoder. A ConvGRU (Eq. (4)) updates a 26 m patch of the 32-channel H at 0.25 m resolution. All decoders read G through six fixed channels. Right: (a) resolutions, (b) ConvGRU equations, (c) memory inputs, (d) ablation arms and (e) write cost.

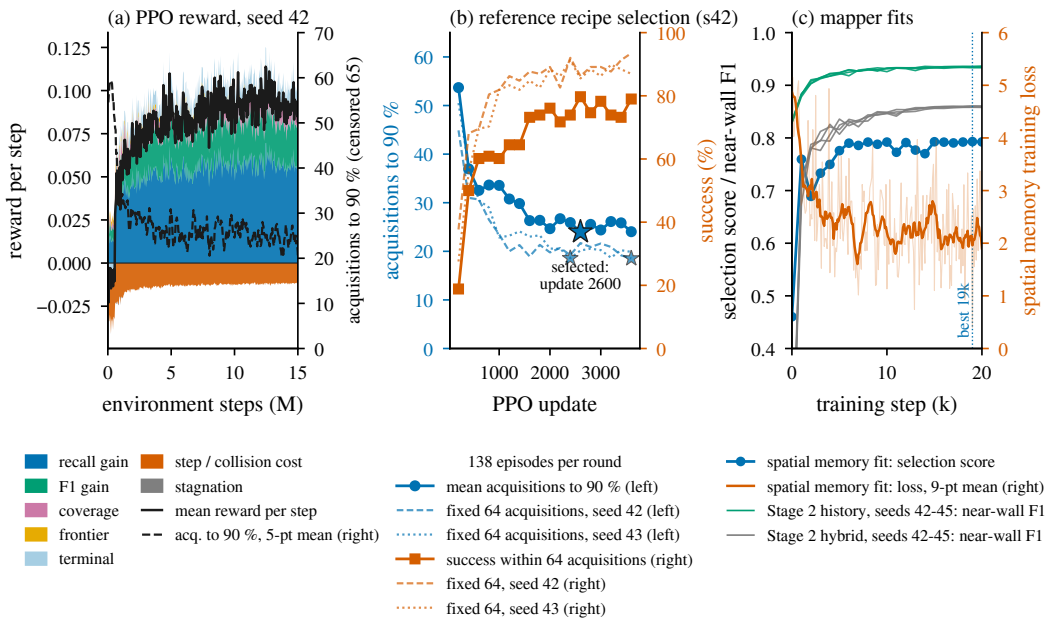

Figure 12: Training curves of the reference recipe (our earlier PPO recipe) and the mapping modules. (a) PPO reward components per environment step (positive terms stacked upwards, costs downwards), with the mean reward and the 5-round mean of acquisitions to 90 % recall on training rollouts; seed 42, 15 M steps. (b) Selection rounds of the same policy on the 69 usable selection rooms × 2 starts (138 episodes): acquisitions to 90 % recall and success, with the selected checkpoint (update 2,600) starred. Dashed lines are the selection rounds of the reference recipe retrained with STOP disabled (STOP floor 64; seeds 42 and 43), each small star marking its selected checkpoint. (c) Selection score and training loss of the seed-42 EchoRooms fit of the evidence-anchored spatial memory (best update, 19,000, marked), and the selection near-wall F_1 of the Stage 2 history refiner and hybrid mapper (the deployed local geometry mapper; Sec. 4.2) for seeds 42–45.

D.4 Whole-room completion on EchoRooms

This subsection reports the belief-level results of the EchoRooms fits: completion against the number of acquisitions, seed replication, read-time controls and forecast fidelity.

Fig. 13 illustrates belief evolution during a test-split episode of the reference recipe. Table 14 reports both EchoRooms seeds alongside the seed-42 fit on the earlier development corpus. Its columns use different rooms and corpora and are not a like-for-like comparison.

Completion over acquisitions. On the 75 validation reporting rooms, seed 42's completion wall F_1 at 0.30 m is 34.2, 36.2, 39.0, 42.4, 48.1, 57.1 and 67.7 % at 1, 2, 4, 8, 16, 32 and 64 acquisitions, respectively. These means include all 75 rooms; eight depth-first trajectories end before 64 acquisitions. Over the same range, the wall F_1 of evidence fusion (fusion of the stored local maps, the baseline of Table 16) rises from 33.7 to 54.7 %, completion wall recall from 25.7 to 59.2 %, and interior IoU from 39.3 to 68.6 %. On the 355 test rooms, the corresponding F_1 curve is 17.6, 18.9, 20.9, 24.2, 28.6, 35.1 and 45.2 % (fusion: 17.0 to 37.4 %), with recall rising from 11.3 to 35.3 % and interior IoU from 23.2 to 45.1 %. Depth-first episodes stay within 56 m of their start, and completion is scored inside the evaluation crop; the large rooms added to the test split (Sec. 5.1) therefore contribute mostly unmeasured interior.

Seeds and controls. Paired on the same records, seed 43 improves knowledge near-wall F_1 and completion F_1 by $+1.5$ [1.1, 2.0] and $+0.5$ [0.1, 0.8] points on the reporting rooms. The test differences are $+2.7$ [2.4, 3.0] and $+0.06$ [-0.02, 0.14], with no resolved difference in test completion. Under read-time controls on the reporting rooms, seed-42 knowledge near-wall F_1 is 83.3 % for memory knockout (H zeroed at read time), 82.5 for stateless read, 82.8 for frozen foundation, 90.2 for pose noise, 91.6 for shuffled history, 15.6 for shuffled acoustics and 0.0 for zero acoustics. The cor-

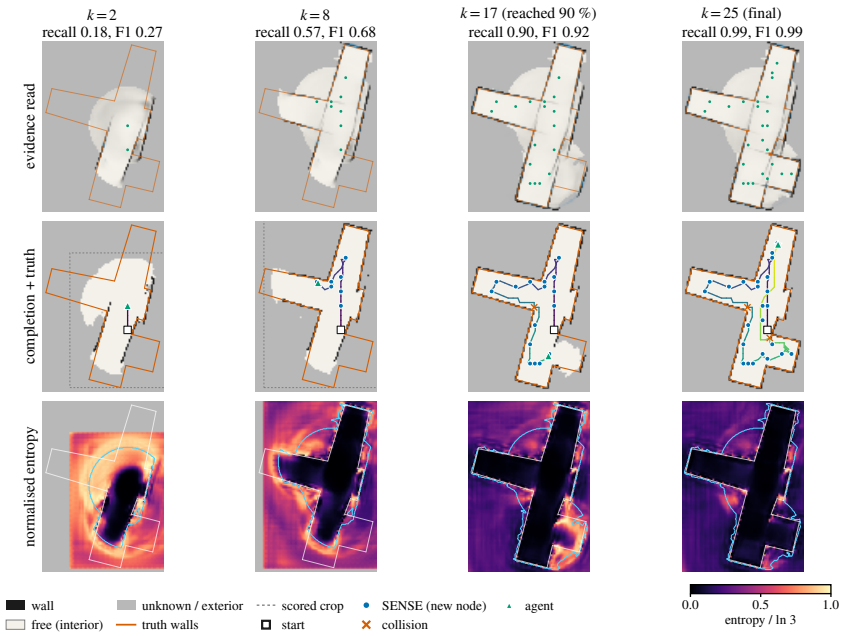

Figure 13: Belief evolution during a test episode of the reference recipe's seed-42 policy, which chooses when to STOP (EchoRooms test room 3080, start 167, 282 lattice nodes). The episode first crosses 90 % wall recall at 17 acquisitions, the median acquisitions to 90 % recall among rooms with $\geq$100 nodes; this first crossing is distinct from final-episode success. Columns show $k = 2$, 8, 17 and 25 (final) acquisitions. Rows show the evidence read (SENSE nodes in green), the completion with true walls and trajectory (dashed box: scored crop), and the normalised completion entropy. Panel titles report wall recall and F_1 at 0.30 m.

responding test values are 79.2, 75.7, 77.2, 85.6, 87.1, 16.4 and 0.0 %. Seed 43's memory-knockout scores are 86.4 on reporting and 81.9 on test rooms.

Forecast fidelity. At unmeasured poses, the reporting-room near-wall F_1 of the predicted observation is 78.5 % for the forecast head, 78.6 for the nearest stored observation, 78.3 for the knowledge read, 75.7 for the evidence read and 14.8 for the class prior. The corresponding test scores are 74.7, 72.6, 75.0, 71.8 and 14.5 %. These fidelity results do not establish a benefit from forecasting; the planner study in Appendix D.10 likewise finds no acquisition improvement from it.

D.5 Retrained belief ablations

Retraining without a component shows what the belief loses when it can never rely on that component; these arms support the recurrent-context-memory result of Sec. 5.4. Table 15 compares three retrained arms with the primary on identical depth-first records (the same trajectory seed, rooms and scored acquisitions): geometric-only (trained without H), no forecast head (head and loss removed) and pose-only (acoustic inputs masked). The primary fits (seeds 42 and 43) and the geometric-only arms compared with them all use legacy fusion. The same holds in the crossed replay (Tables 18 and 19), the occlusion analysis (Table 16) and the planner study (Appendix D.10). Stage 2 predictions therefore remain readable in the evidence on both sides of these comparisons.

For seed 42 on the reporting rooms, the geometric-only arm changes knowledge near-wall F_1 by $+1.7$ [1.2, 2.2] points, completion F_1 by -1.1 [-1.8, -0.5] and interior IoU by -2.7 [-3.9, -1.4] at 64 acquisitions. Removing the forecast head changes these metrics by $+0.4$ [0.1, 0.7], $+0.5$ [0.2, 0.9] and $+0.30$ [0.01, 0.62] points, respectively: small but resolved improvements. The head is kept only because the forecast arm of the FC-EER study (Appendix D.10) queries it. The policy never queries the head; the policy's one-step predictor (Sec. 4.4) is a separate module. The pose-only changes are -51.8, -31.1 and -22.3 points.

Memory knockout and retraining answer different questions. Knockout (-8.6 knowledge points for the primary) removes H at read time from a model trained with it; retraining instead lets the model

learn to rely on the evidence map. The seed-42 geometric-only arm improves local knowledge but reduces whole-room completion.

Replication across seeds. Paired with the seed-43 primary, the seed-43 geometric-only arm changes reporting-room completion F_1 by -1.6 [-2.2, -1.1] and interior IoU by -3.5 [-4.5, -2.5] points. The two-seed per-room mean differences are -1.4 [-1.9, -0.8] and -3.1 [-4.1, -2.1]. Reporting-room knowledge has no resolved difference for seed 43 ($+0.04$ [-0.20, 0.26], versus $+1.7$ for seed 42; two-seed mean $+0.9$ [0.5, 1.2]). On the 355 test rooms, the completion differences for seeds 42 and 43 are -1.1 [-1.3, -1.0] and -1.2 [-1.4, -1.1] points (two-seed mean -1.2 [-1.4, -1.0]; interior IoU differences -0.1 and -0.8). Test knowledge improves by $+3.1$ [2.7, 3.5] and $+0.4$ [0.2, 0.5] points. Thus, removing H consistently reduces completion, while its effect on local knowledge depends on the seed.

D.6 Support-gated fusion and the fine-tuned belief

This subsection specifies the write rule and the fine-tuning that produce the main method's belief (Sec. 4.3). The main method's belief uses the support-gated writer: local wall and free-space estimates enter the measured wall and free channels of G only where a first-order echo or an executed move's swept footprint supports them; unsupported estimates are not written to these channels. The echo threshold is 0.73, backoff and clearance are both 0.3 m, and the wall threshold is 0.5; free space without an echo has weight 0, and redundancy filtering is disabled. The legacy-fusion writer instead records Stage 2 predictions wherever $p_{\text{wall}} + p_{\text{free}} \geq 0.5$. Under support-gated fusion, the full Stage 2 predictions, including their unsupported spatial context, reach the recurrent context memory H through the observation encoder. The evidence map also retains separate predicted channels, but the six-channel decoder read does not access them. Under legacy fusion, by contrast, the read's wall, free and support channels are built from the Stage 2 predictions. Support-gated fusion reduces, but does not eliminate, prediction-induced wall erasure: on curved selection rooms the erased wall length falls from 64.5 % to 7.4 % (Appendix G).

The main method's *fine-tuned belief* (arm C of Appendix G) starts from the seed-42 primary at update 19,000. It is fine-tuned under support-gated fusion with hole-aware completion targets (hole interiors labelled exterior; Appendix F) for 5,000 updates, with learning rate 2.5×10^{-5}, 200 warm-up updates and seed 42. The checkpoint is selected by mean completion F_1 over 8, 16, 32 and 64 acquisitions on selection replays. The global completion temperature is then refitted on the 74 calibration rooms from 1.3112 to $T = 1.2639$, which lowers the fine-tuned belief's calibration NLL from 0.3288 to 0.3180; this temperature is paired with the selected checkpoint.

Here *polygonal* denotes the legacy layout family and *curved* the curved and irregular-curved families. On identical selection-room depth-first replays at 64 acquisitions, the fine-tuned belief's completion wall F_1 has no resolved difference from the primary: $+0.2$ [-0.3, 0.7] points on the 62 polygonal rooms and -0.3 [-1.0, 0.4] on the 8 curved rooms. Interior IoU decreases on polygonal rooms (-0.6 [-1.0, -0.2]) and has no resolved difference on curved rooms ($+0.8$ [-2.2, 3.6]). Support-gated fusion changes the meaning of the measured channels and reduces false-free predictions on curved walls at eight acquisitions (-4.1 [-6.9, -2.2] points). Appendix G gives the supporting selection experiments.

D.7 Recall of visible and occluded walls

This analysis asks where completion gains come from: walls visible from the trajectory, or walls hidden behind other walls. It supports the statement in Sec. 5.4 that gains concentrate on visible walls.

Visibility and matching. Each ground-truth wall sample is classified from the visited poses of a fixed blind depth-first trajectory. A sample is *visible* if a pose within the 8 m sensing radius sees it without crossing another wall, *occluded in range* if it lies within 8 m of a pose but behind a wall, and *beyond* otherwise. Evaluation uses the same predicted wall points as the overall completion metric: known wall cells of the completion crop for belief maps, and stored local rasters for evidence fusion (*Fusion of local maps* in Table 16). Matching at 0.30 m reproduces overall recall exactly from the visibility classes.

A prediction on a visible face can fall within the matching tolerance of an occluded sample. To control for this, the $\geq 0.6\,m$ column retains only occluded-in-range samples at least 0.6 m from every visible sample; *from occluded points* matches occluded samples only to occluded predicted points. Ratios

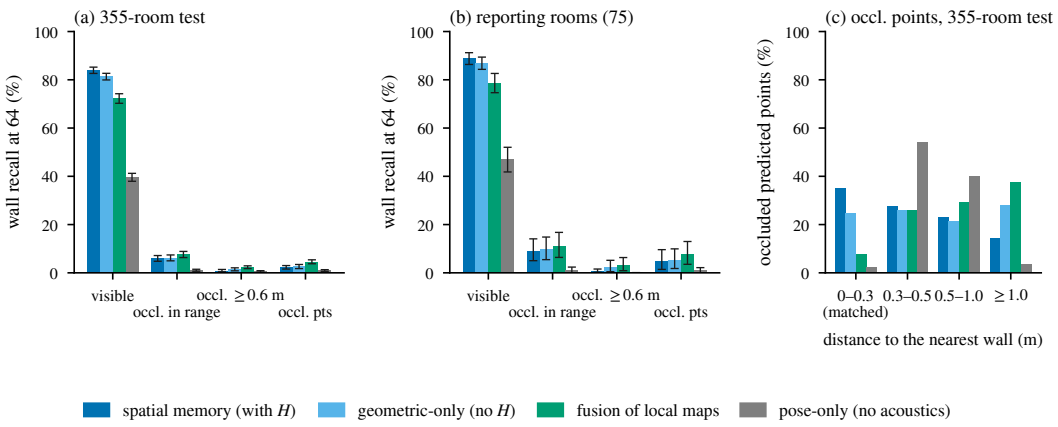

Figure 14: Occlusion analysis after 64 acquisitions on fixed blind depth-first trajectories, for the completion of the evidence-anchored spatial memory (primary, seed 42, with H), the geometric-only arm, evidence fusion and the pose-only arm: wall recall by visibility class (a) on the 355 test rooms and (b) on the 75 reporting rooms, with room-bootstrap 95 % CIs over room means (occl. pts: recall from occluded predicted points only), and (c) the distance of each map's occluded predicted points to the nearest true wall on the test rooms. Values: Table 16.

are computed per room and averaged over rooms with a non-empty denominator; Table 16 lists those counts. Paired differences use room bootstraps with 10,000 resamples.

Completion versus evidence fusion. Table 16 and Fig. 14 report the 355 test rooms and 75 reporting rooms after 64 acquisitions, using the primary's completion (seed 42). On the test rooms, 40.0 % of wall length is visible, 6.2 % occluded in range and 53.7 % beyond. At 64 acquisitions on the test rooms, completion exceeds fusion's visible recall by $+11.7$ [10.7, 12.7] points but recalls fewer separated occluded walls (≥ 0.6 m column: -1.5 [-2.1, -0.9]). Both differences also hold at 16 and 32 acquisitions: visible recall changes by $+11.8$ and $+11.8$ points, and ≥ 0.6 m recall by -1.6 and -1.7, with all intervals excluding zero. On the reporting rooms, the ≥ 0.6 m difference is resolved only at 64 acquisitions (-2.2 [-5.4, -0.2]).

Occluded predictions of the evidence-anchored spatial memory are closer to true walls than fusion's. Of all its occluded predicted points on the test rooms, within or beyond 8 m of a pose (Fig. 14c), 35.1 % lie within 0.3 m of a true wall and 14.1 % farther than 1 m, versus 7.8 and 37.3 % for fusion. The precision column of Table 16 (35.0 and 7.3 %) is the within-0.3 m share restricted to occluded predictions within 8 m of a pose. Fusion's occluded predictions represent low-support back sides of visible walls. The pose-only arm (acoustics zeroed, seed 42) recalls 39.6 % of visible and 1.0 % of occluded-in-range samples, with 2.3 % precision on occluded predictions. The behind-wall signal in the local maps therefore depends on acoustics and is not amplified by completion.

Contribution of the recurrent context memory. The retrained geometric-only arm (seed 42, its own calibrated temperatures; Table 15) recalls 81.3 % of visible test samples, versus 83.9 for the primary with H: a paired gain from H of $+2.6$ [2.3, 3.0] points at 64 acquisitions and $+3.4$ at 16 and 32. Occluded-in-range recall has no resolved difference (6.1 without memory versus 5.9 with it; primary minus geometric only -0.2 [-0.5, 0.1]). For samples at least 0.6 m from a visible sample, recall is slightly lower with memory (0.8 with versus 1.3 % without; -0.5 [-0.9, -0.2]). Memory does, however, raise occluded-prediction precision (21.7 without versus 35.0 % with; paired $+12.5$ [9.8, 15.2]) and lower the share of occluded predictions farther than 1 m from any wall (28.1 versus 14.1 %). On the reporting rooms, the visible-recall gain is $+1.9$ [1.4, 2.6], while the ≥ 0.6 m column shows no resolved difference (-1.4 [-4.2, 0.1]). Memory thus improves completion of walls visible from the trajectory and suppresses spurious predictions behind walls; it does not resolve an improvement in occluded-in-range test recall.

D.8 Forecast queries and the FC-EER crop

This subsection specifies how the FC-EER planner of Sec. 4.3 queries the belief and scores a candidate. A forecast reads the belief at a query pose without an attached observation. The local decoder receives the six evidence channels and 32 memory channels warped to the query, plus the three coordinate channels. The four current-observation channels (wall, free, unknown and echo) and the presence flag are zeroed. The forecast head therefore has no access to the future observation; the knowledge baseline reduces to evidence support, and the view head is not evaluated. The forecast head's class softmax and echo-logit sigmoid form the imagined observation, written at heading 0 into copies of G and H. Candidates at already acquired nodes are left unwritten.

The entropies before and after the imagined write ($\mathcal{H}_c^{\text{before}}$ and $\mathcal{H}_c^{\text{after}}$ in the utility U) use the same world-aligned crop, fixed before the decision to cover the current evaluation crop and every candidate's sensing footprint. The implementation rejects any pair that differs in shape, origin or cell size.

D.9 Lattice planner definitions

The lattice planners below are compared in Appendix D.10. Raycast information gain and the copy and knowledge-read variants of FC-EER, re-implemented inside the policy environment, are the planner baselines of Sec. 5.2 (Appendix E.4). The lattice-planner settings were selected on the earlier development corpus and transferred unchanged to the EchoRooms lattice study (Appendix D.10). The planners used in the continuous-motion policy comparisons were separately tuned in the policy environment (Appendix E.4). The belief-based planners share a candidate generator and cost term. Candidates are unacquired lattice points within 8 m, reachable by unit cardinal-step routes over the belief's passable raster, with path length at most 16 m and at most 32 candidates. On the earlier development corpus, clearance is 1.15 m plus a 0.177 m margin. The cost is $\lambda_d \cdot \text{length} + \lambda_c \cdot \text{risk} + \lambda_r \cdot \text{revisits}$, with hysteresis 1.5; a planner stops after two quiet decisions, whose utility falls below the threshold τ.

FC-EER uses the clipped entropy-reduction utility U of Sec. 4.3, with observation-free queries and a common crop as specified in Appendix D.8. It deterministically writes soft probabilities, without sampling or taking an expectation over observations. The signed, unclipped utility is also recorded. *Uncertainty coverage* sums completion entropy over the unseen part of a candidate's footprint. *Raycast information gain* casts 72 rays and sums entropy along each until $p_{\text{wall}} \geq 0.5$. *Approximate EIG* (expected information gain) uses a binary sensor model with accuracy 0.8, and *nearest frontier* chooses the closest passable cell adjacent to unknown. These information-gain, coverage and approximate-EIG planners are our re-implementations on the shared candidate generator. *Blind DFS* (depth-first search) and *random walk* ignore the belief. The privileged references are the *FC-EER oracle* (*true future observation* in Table 17), which writes the cached true future observation, and the *privileged sweep* (*privileged true-graph sweep* in the table), which covers the true lattice.

Default costs are $(\lambda_d, \lambda_c, \lambda_r) = (3, 9, 3)$ and $\tau = 5$. Each planner's *selected* variant is tuned on the earlier development corpus's 176 selection rooms; FC-EER selects $(1, 3, 1)$ and $\tau = 2$. Planner success is wall F_1 at 0.30 m $\geq 90\,\%$ at the last scored point within 128 attempts, distinct from the policy's final-wall-recall criterion (Sec. 5.1; Appendix F).

D.10 Lattice planner study and crossed replay on EchoRooms

This study tests whether imagined observations help a belief-based planner (the forecasting result of Sec. 5.3) and, through crossed replay, separates what the recurrent context memory adds to reconstruction from what it adds to planning.

Protocol. Nineteen planner arms are evaluated on the seed-42 primary belief over 75 validation reporting rooms with three starts each: 225 episodes per arm, with budgets of 64 acquisitions and 256 motion attempts. All arms share the checkpoint, temperatures and evaluation implementation. Only 67 rooms have at least 64 lattice nodes. Attempts to 90 % recall are censored in more than half of every non-privileged planner's episodes, which limits that metric's informativeness here. Conventional planners use the definitions above. Selected settings are carried over from the earlier corpus's selection rooms; none is re-selected on EchoRooms. The forecasting negative result of Sec. 5.3 therefore rests on transferred settings: with their default settings, neither copy ($+0.5$ [-0.6, 1.5]) nor knowledge read ($+0.05$ [-0.99, 1.11]) has a resolved wall-F_1 difference from the selected forecast arm.

FC-EER's observation-source variants write the *forecast head*'s prediction, the *nearest stored observation* copied to the candidate (*copy*), the observation-free *knowledge read*, or the privileged *true future observation*. The knowledge loss does not supervise observation-free queries, limiting what this baseline establishes about read-based imagination. The *signed* variants instead select on the unclipped utility. Table 17 reports means and paired room bootstraps, averaging starts within each room before 10,000 resamples and reporting 95 % percentile intervals without multiplicity correction.

Utility fidelity. The forecast arm selects goals within 2 m of an acquired pose in 76.9 % of decisions, versus 57.8 for copy and 66.1 % for read. Its predicted signed utility is anti-correlated with the realised signed utility (rank correlation -0.196). Across scored FC-EER variants, 19.9–30.8 % of realised signed utilities are negative, including when selection uses signed utility. Clipped utility is therefore a heuristic score, and using the signed sum does not remove this limitation. Signed utility also changes hysteresis and stopping, so those arms' stop metrics are not directly comparable (‡ in Table 17).

Imagined updates without the recurrent context memory. The copy, read and forecast-head arms and raycast information gain are also evaluated with the retrained geometric-only belief (seed 42, no H, its own calibrated temperatures) on the same 75 rooms, starts, budgets and evaluation implementation. Relative to the same planner on the primary belief, every selected planner loses wall F_1 at 64 acquisitions. Raycast information gain changes by -3.7 [-4.7, -2.7] points, copy by -2.1 [-3.3, -1.0], read by -4.6 [-5.8, -3.4] and forecast by -4.9 [-6.4, -3.5] (the total column of Table 18 with the sign reversed). On the geometric-only belief, selected copy exceeds selected raycast by $+1.12$ [0.03, 2.22] points, so a useful imagined update can operate through the evidence map and its effect on completion entropy.

Crossed replay: reconstruction and trajectory quality. Crossed replay separates what H adds to the reconstruction of a fixed observation stream from what it adds to trajectory quality. Every recorded episode of the four selected planners is replayed through both beliefs, the primary and the geometric-only arm, giving a 2×2 design per planner: trajectory source (the belief the planner ran on) by evaluator (the belief that scores the replay). Goal routes come from recorded candidate tables, random steps use the evaluation random stream, and counters are checked at every decision. The same cached observations and new-collision events are then replayed through a fresh belief state with its own calibrated temperatures and scored after every event.

The diagonal cells, where the evaluator is the belief that generated the trajectory, are the recorded runs. Identity replay through that belief reproduces every scored point on the first start of each room: 75 episodes per diagonal cell, with drift 0.0 against a tolerance of 10^{-4}. Off-diagonal replay covers every start. Table 19 gives all four cells per planner (225 episodes each), with paired room bootstraps using 10,000 resamples; Table 18 summarises the decomposition used in Sec. 5.3. Each planner's third row in Table 19 reports within-evaluator trajectory differences, with total recorded-run differences in the Δ columns. Wall F_1 is scored at 0.30 m. The score at 64 acquisitions equals final F_1 in every episode: 553 of the 2,400 replayed episode rows reach the 256-attempt cap with fewer acquisitions and retain their last score.

Holding the trajectory fixed, the primary belief improves wall F_1 by $+1.5$ to $+1.7$ points and interior IoU by $+2.5$ to $+3.1$ for every planner and both trajectory sources, with all intervals excluding zero. Holding the primary evaluator fixed, trajectories planned on the primary belief improve F_1 by $+3.3$ [1.9, 4.6] points for forecast, $+2.9$ [1.8, 4.0] for read and $+2.1$ [1.2, 3.0] for raycast. Copy has no resolved trajectory difference ($+0.5$ [-0.6, 1.6]), so only its reconstruction gain is resolved. The two effects account for the recorded totals, with wall-F_1 interactions within 0.2 points; trajectory quality accounts for 56–67 % of the total gains for forecast, read and raycast. On fixed streams, the primary also reduces failure-penalised attempts to 90 % recall in six of eight planner–trajectory pairs, by 1.6 to 5.1 attempts, with no resolved difference in the other two.

The belief-frontier walker of the reference recipe provides a fixed-trajectory comparison on the test split. Its 738 trajectories across the 246 test rooms within the reference recipe's span cap (Sec. 4.4) have identical motion records under both beliefs, because frontier selection reads only the evidence map G; only the belief-dependent reward components differ. On these streams, the primary improves final F_1 by $+1.3$ [1.1, 1.4] points (62.1 versus 60.8 %) and interior IoU by $+2.3$ [1.8, 2.8].

Table 11: Local mapping (Stage 2) by the hybrid mapper and the history refiner on reporting (75), all validation (224) and test (355) rooms: near-wall F_1 (walls within 4 m, 0.125 m tolerance), full-radius F_1, false-free wall rate (all in %), NLL and Brier score; mean $\pm$ sd over seeds 42–45. The *matched foundation* is the seed-45 history refiner's frozen single-shot mapper evaluated on the refiner's acquisition records, which differ from the hybrid-mapper rows' records. Bottom: per-seed refiner current view minus matched foundation on identical records (10,000 paired room-bootstrap resamples, 95 % CI).

Model	Partition	Near-wall F1	Full-radius F1	False-free wall	NLL	Brier
Hybrid mapper, current view						
mean $\pm$ sd (4 seeds)	validation reporting (75)	85.9 ± 0.1	72.2 ± 0.2	15.5 ± 0.2	0.260 ± 0.004	0.140 ± 0.002
mean $\pm$ sd (4 seeds)	all validation (224)	85.9 ± 0.1	72.0 ± 0.2	15.4 ± 0.2	0.264 ± 0.004	0.143 ± 0.002
mean $\pm$ sd (4 seeds)	held-out test (355)	80.6 ± 0.2	65.1 ± 0.2	20.2 ± 0.2	0.323 ± 0.004	0.178 ± 0.002
History refiner, current view						
mean $\pm$ sd (4 seeds)	validation reporting (75)	93.9 ± 0.1	87.0 ± 0.1	6.3 ± 0.3	0.120 ± 0.002	0.064 ± 0.001
mean $\pm$ sd (4 seeds)	all validation (224)	93.5 ± 0.0	86.8 ± 0.1	6.7 ± 0.2	0.123 ± 0.002	0.066 ± 0.001
mean $\pm$ sd (4 seeds)	held-out test (355)	90.5 ± 0.1	82.2 ± 0.1	8.5 ± 0.2	0.163 ± 0.002	0.087 ± 0.001
History refiner, prefix knowledge						
mean $\pm$ sd (4 seeds)	validation reporting (75)	94.1 ± 0.1	88.4 ± 0.1	5.9 ± 0.1	0.138 ± 0.002	0.073 ± 0.001
mean $\pm$ sd (4 seeds)	all validation (224)	93.7 ± 0.1	88.0 ± 0.1	6.2 ± 0.1	0.142 ± 0.002	0.075 ± 0.001
mean $\pm$ sd (4 seeds)	held-out test (355)	91.1 ± 0.1	83.8 ± 0.1	7.5 ± 0.1	0.190 ± 0.002	0.101 ± 0.001
Matched foundation, current view (seed 45; the history refiner's frozen single-shot mapper on the identical acquisition records)						
seed 45	validation reporting (75)	84.7	70.2	16.9	0.279	0.151
seed 45	held-out test (355)	78.9	63.6	22.6	0.333	0.183
Paired Δ (refiner current view − matched foundation, identical acquisition records) [95% CI]						
seed 42	validation reporting (75)	+9.3 [8.0, 10.5]	+17.0 [15.4, 18.5]	−10.4 [−12.0, −8.9]	−0.162 [−0.175, −0.149]	−0.089 [−0.098, −0.081]
seed 42	held-out test (355)	+11.4 [10.6, 12.2]	+18.6 [17.8, 19.4]	−14.2 [−15.3, −13.2]	−0.172 [−0.177, −0.166]	−0.097 [−0.100, −0.093]
seed 43	validation reporting (75)	+9.2 [7.9, 10.5]	+16.9 [15.3, 18.5]	−10.7 [−12.2, −9.2]	−0.161 [−0.173, −0.148]	−0.089 [−0.097, −0.080]
seed 43	held-out test (355)	+11.6 [10.8, 12.4]	+18.6 [17.8, 19.5]	−14.1 [−15.1, −13.1]	−0.171 [−0.177, −0.165]	−0.096 [−0.100, −0.092]
seed 44	validation reporting (75)	+8.9 [7.7, 10.2]	+16.6 [15.0, 18.1]	−10.1 [−11.6, −8.7]	−0.156 [−0.169, −0.143]	−0.087 [−0.095, −0.078]
seed 44	held-out test (355)	+11.4 [10.6, 12.2]	+18.4 [17.6, 19.2]	−13.8 [−14.8, −12.8]	−0.167 [−0.173, −0.161]	−0.094 [−0.098, −0.090]
seed 45	validation reporting (75)	+9.3 [8.0, 10.6]	+17.0 [15.4, 18.5]	−10.9 [−12.6, −9.3]	−0.158 [−0.171, −0.145]	−0.088 [−0.096, −0.079]
seed 45	held-out test (355)	+11.5 [10.7, 12.3]	+18.6 [17.8, 19.4]	−14.3 [−15.3, −13.3]	−0.169 [−0.175, −0.163]	−0.095 [−0.099, −0.091]

Table 12: Stage 2 history-refiner controls (seed 45, prefix-knowledge head), near-wall and full-radius F_1 (%) on the validation reporting rooms (75) and the held-out test split (355). Conditions are described in the paragraph *History controls*; the frozen local foundation is the refiner's own frozen single-shot mapper.

	Validation reporting (75)		Held-out test (355)	
Condition (history refiner, prefix knowledge, seed 45)	Near-wall F1	Full-radius F1	Near-wall F1	Full-radius F1
Full model	94.3	88.5	91.2	83.8
History disabled	82.9	66.9	77.3	60.2
Zero acoustic history	83.2	67.3	77.6	60.6
Shuffled acoustic history	34.2	28.9	32.9	26.9
Zero acoustics	0.0	0.0	0.0	0.0
Shuffled acoustics	10.8	8.8	10.9	9.0
Frozen local foundation	82.9	66.9	77.3	60.2
Pose-noise alignment	92.3	82.6	88.8	77.7

Table 13: Frozen-module inference cost: parameters, median and tail batch-1 latency on an otherwise idle A40 (tail: p90 for the polar echo encoder, p95 otherwise), peak CUDA memory and calibration. Primary calibration is the encoder's decision threshold, the local geometry mapper's temperature or the knowledge-head temperature; other temperatures are view (history refiner) and view/local/global completion (evidence-anchored spatial memory). The spatial-memory row is the primary fit (seed 42, legacy fusion, before fine-tuning; Appendix D) measured on EchoRooms. The full FC-EER decision uses the forecast head and is timed on a shared A40 over decisions with at least one candidate.

Module	Trainable parameters	Total parameters	Median latency (ms)	Tail latency (ms)	Peak CUDA (MiB)	Primary temperature	Other temperatures
Polar echo encoder (Stage 1, 12 mics)	471,457	471,457	6.99	7.05 (p90)	47	threshold 1.0	–
Local geometry mapper (Stage 2)	2,215,331	2,215,331	14.09	14.31 (p95)	76	0.990	–
History refiner (Stage 2)	489,526	2,704,857	19.80	20.14 (p95)	122	1.132	1.076 (view)
Spatial memory (primary, seed 42)	614,233	614,233	9.47	13.92 (p95)	317	1.515	1.096, 1.035, 1.311
Full FC-EER decision (EchoRooms, $\geq$1 candidate)	–	–	138.5	353.1 (p95)	–	–	–

Table 14: Fits of the evidence-anchored spatial memory (primary: legacy fusion with H) on EchoRooms (3,447 training rooms; seeds 42 and 43, mean $\pm$ sd). Blocks report selection (75 rooms), calibration (seed 42; 74 EchoRooms rooms; temperatures, with negative log-likelihood (NLL) before $\rightarrow$ after), reporting and held-out evaluation (355 test versus 840 sealed rooms). Corpora are not directly comparable. Prefix knowledge and current view are the knowledge and view heads; ECE is expected calibration error; memory-knockout rows zero H at read time (selection Δ: points lost under knockout). F_1, recall and IoU are in %. Completion is scored at 64 acquisitions on fixed depth-first (DFS) trajectories, and at the end of the DFS episode for selection. Paired seed-43-minus-seed-42 differences use identical records and paired room bootstraps (10,000 resamples, 95 % CI).

	EchoRooms seed 42	EchoRooms seed 43	Mean $\pm$ sd (2 seeds)	Earlier corpus seed 42
Selection rooms (75 validation-selection rooms; earlier corpus: not applicable)				
Prefix-knowledge near-wall F1	91.7	93.0	92.4 $\pm$ 0.9	–
Prefix-knowledge full-radius F1	85.9	87.4	86.6 $\pm$ 1.0	–
Prefix-knowledge false-free wall	11.4	10.3	10.9 $\pm$ 0.7	–
Prefix-knowledge NLL	0.238	0.171	0.205 $\pm$ 0.048	–
Prefix-knowledge ECE	0.025	0.016	0.021 $\pm$ 0.007	–
Current-view near-wall F1	92.4	92.5	92.5 $\pm$ 0.0	–
Whole-room completion F1 (0.30 m) at end of DFS	66.9	67.1	67.0 $\pm$ 0.2	–
Memory-knockout Δ	8.8	7.6	8.2 $\pm$ 0.8	–
Selection score	0.7930	0.8007	0.7968 $\pm$ 0.0054	–
Best checkpoint step	19,000	19,000	–	–
Calibration temperature, seed 42 (NLL before $\rightarrow$ after)				
Prefix knowledge	1.515 (0.228 $\rightarrow$ 0.206)	–	–	1.609 (0.122 $\rightarrow$ 0.105)
Current view	1.096 (0.145 $\rightarrow$ 0.144)	–	–	1.380 (0.096 $\rightarrow$ 0.090)
Local completion	1.035 (0.271 $\rightarrow$ 0.271)	–	–	1.038 (0.138 $\rightarrow$ 0.138)
Whole-room completion	1.311 (0.352 $\rightarrow$ 0.337)	–	–	1.161 (0.144 $\rightarrow$ 0.142)
Reporting evaluation (EchoRooms: 75 validation-reporting rooms; earlier corpus: 176 reporting rooms)				
Prefix-knowledge near-wall F1	92.0	93.5	92.7 $\pm$ 1.1	94.7
Current-view near-wall F1	92.9	93.4	93.1 $\pm$ 0.3	94.5
Local completion near-wall F1	91.5	91.6	91.5 $\pm$ 0.0	93.4
Prefix-knowledge full-radius F1	86.3	88.1	87.2 $\pm$ 1.3	91.3
Knowledge NLL	0.204	0.154	0.179 $\pm$ 0.036	0.107
Knowledge ECE	0.018	0.013	0.015 $\pm$ 0.004	0.010
Whole-room completion F1 (0.30 m) at 64	67.7	68.1	67.9 $\pm$ 0.3	83.5
Whole-room completion recall (0.30 m) at 64	59.2	59.3	59.2 $\pm$ 0.1	78.6
Interior IoU at 64	68.6	69.0	68.8 $\pm$ 0.3	83.6
Memory-knockout control, knowledge near-wall F1	83.3	86.4	84.9 $\pm$ 2.1	91.7
Δ (43 $-$ 42) knowledge F1 [95% CI]	–	+1.5 [1.1, 2.0]	–	–
Δ (43 $-$ 42) completion F1 at 64 [95% CI]	–	+0.5 [0.1, 0.8]	–	–
Δ (43 $-$ 42) interior IoU at 64 [95% CI]	–	+0.4 [$-$0.1, 1.0]	–	–
Held-out evaluation (EchoRooms: 355 test rooms; earlier corpus: 840 sealed held-out rooms)				
Prefix-knowledge near-wall F1	87.5	90.2	88.8 $\pm$ 1.9	94.6
Current-view near-wall F1	89.6	89.5	89.5 $\pm$ 0.0	–
Local completion near-wall F1	87.6	87.5	87.5 $\pm$ 0.0	–
Prefix-knowledge full-radius F1	80.3	83.0	81.7 $\pm$ 1.9	90.9
Knowledge NLL	0.262	0.209	0.236 $\pm$ 0.037	0.112
Knowledge ECE	0.023	0.017	0.020 $\pm$ 0.004	–
Whole-room completion F1 (0.30 m) at 64	45.2	45.3	45.3 $\pm$ 0.0	83.0
Whole-room completion recall (0.30 m) at 64	35.3	35.2	35.3 $\pm$ 0.0	78.0
Interior IoU at 64	45.1	45.6	45.4 $\pm$ 0.4	83.7
Memory-knockout control, knowledge near-wall F1	79.2	81.9	80.6 $\pm$ 1.9	–
Δ (43 $-$ 42) knowledge F1 [95% CI]	–	+2.7 [2.4, 3.0]	–	–
Δ (43 $-$ 42) completion F1 at 64 [95% CI]	–	+0.06 [$-$0.02, 0.14]	–	–
Δ (43 $-$ 42) interior IoU at 64 [95% CI]	–	+0.5 [0.3, 0.7]	–	–

Table 15: Retrained ablation arms of the evidence-anchored spatial memory on EchoRooms, against the primary (legacy fusion with H). Arms: geometric only (trained without H), no forecast head (head and loss removed) and pose only (acoustic inputs masked). Blocks report seed-42 fits on 75 selection rooms (best step: selected update; elapsed: training hours), calibrated evaluation on 75 reporting rooms at 64 acquisitions, and paired arm-minus-primary differences on identical depth-first (DFS) records. Prefix knowledge is the knowledge head's near-wall F_1; whole-room completion uses F_1 at 0.30 m (at the end of the DFS episode for selection); F_1 and interior IoU are in %, differences in points. The memory-knockout Δ is the change in knowledge near-wall F_1 when H is zeroed at read time (a loss is negative; Table 14 reports it as points lost). The final block pairs geometric-only and primary fits by seed (42, 43) on reporting and 355 test rooms; two-seed means average each room's matched differences before resampling. All intervals use paired room bootstraps (10,000 resamples, 95 % CI).

Arm (seed 42)	Prefix-knowledge near-wall F1	Whole-room completion F1 (0.30 m) at end of DFS	Selection score	Best step	Parameters	Elapsed (h)
Fit on the training split; selection rooms (75) at the best checkpoint						
Primary	91.7	66.9	0.7930	19,000	614,233	5.2
Geometric only (no H)	93.2	65.7	0.7944	20,000	552,201	4.6
No forecast head	92.0	67.2	0.7960	20,000	614,165	5.3
Pose only (no acoustics)	38.8	35.3	0.3706	20,000	614,233	7.5

Arm (seed 42)	Prefix-knowledge near-wall F1	Whole-room completion F1 (0.30 m) at 64	Interior IoU at 64	Memory-knockout Δ (points)
Validation reporting rooms (75), calibrated temperatures				
Primary	92.0	67.7	68.6	−8.6
Geometric only (no H)	93.7	66.6	65.9	–
No forecast head	92.4	68.2	68.9	−6.3
Pose only (no acoustics)	40.2	36.6	46.3	−10.0

Arm (seed 42)	Δ knowledge near-wall F1 [95% CI]	Δ completion F1 at 64 [95% CI]	Δ interior IoU at 64 [95% CI]
Arm − primary on the reporting rooms, paired room bootstrap (10,000 resamples) over the same DFS records			
Primary	reference	reference	reference
Geometric only (no H)	+1.7 [1.2, 2.2]	−1.1 [−1.8, −0.5]	−2.7 [−3.9, −1.4]
No forecast head	+0.4 [0.1, 0.7]	+0.5 [0.2, 0.9]	+0.30 [0.01, 0.62]
Pose only (no acoustics)	−51.8 [−54.0, −49.6]	−31.1 [−34.4, −27.9]	−22.3 [−24.4, −20.2]

Arm and seed	Partition	Δ knowledge near-wall F1 [95% CI]	Δ completion F1 at 64 [95% CI]	Δ interior IoU at 64 [95% CI]
Geometric only (no H) − primary of the same seed, paired room bootstrap (10,000 resamples) over the same DFS records				
Seed 42	reporting (75)	+1.7 [1.2, 2.2]	−1.1 [−1.8, −0.5]	−2.7 [−3.9, −1.4]
Seed 43	reporting (75)	+0.04 [−0.20, 0.26]	−1.6 [−2.2, −1.1]	−3.5 [−4.5, −2.5]
Seeds 42 and 43 (per-room mean)	reporting (75)	+0.9 [0.5, 1.2]	−1.4 [−1.9, −0.8]	−3.1 [−4.1, −2.1]
Seed 42	test (355)	+3.1 [2.7, 3.5]	−1.1 [−1.3, −1.0]	−0.1 [−0.7, 0.6]
Seed 43	test (355)	+0.4 [0.2, 0.5]	−1.2 [−1.4, −1.1]	−0.8 [−1.4, −0.3]
Seeds 42 and 43 (per-room mean)	test (355)	+1.7 [1.5, 2.0]	−1.2 [−1.4, −1.0]	−0.5 [−1.0, 0.1]

Table 16: Wall recall by visibility after 64 acquisitions on fixed blind depth-first trajectories: 355 test rooms (top) and 75 reporting rooms (bottom). Columns give visible and occluded-in-range recall, recall on occluded samples at least 0.6 m from visible samples, recall using only occluded predictions, occluded-prediction precision (share of occluded predicted points within 8 m of a pose that lie within 0.3 m of a true wall) and the share of predictions that are occluded. Maps are the seed-42 completion of the evidence-anchored spatial memory with H (spatial-memory rows; the primary, legacy fusion), the retrained geometric-only (no H) and pose-only (no acoustics) arms, and evidence fusion (*Fusion of local maps*). Values are room means in %; differences are spatial memory minus control in points. Intervals use room bootstraps (10,000 resamples, 95 % CI), paired for differences. Predicted-point shares include all rooms; recall and precision use non-empty denominators, with counts listed in each block.

Map (64 acquisitions)	Visible	Occluded in range	Occluded ≥0.6 m from visible	Occluded from occluded points	Precision occluded in range	Predicted points occluded
355-room test: 6.2% of the wall length is occluded within 8 m of a pose (5.2% also ≥0.6 m from every visible sample)						
rooms with a non-empty denominator (recall columns): visible 355, in range 310, ≥0.6 m 304, occluded points 310; precision column, map rows in order: 338 / 346 / 350 / 355						
Spatial-memory completion (with H)	83.9 [82.6, 85.2]	5.9 [4.8, 7.2]	0.8 [0.4, 1.4]	2.2 [1.6, 3.0]	35.0 [32.0, 38.1]	6.2 [5.7, 6.8]
Geometric-only completion (no H)	81.3 [79.9, 82.7]	6.1 [5.0, 7.4]	1.3 [0.8, 2.1]	2.5 [1.8, 3.5]	21.7 [19.3, 24.3]	7.2 [6.6, 7.8]
Pose-only completion (no acoustics)	39.6 [37.9, 41.2]	1.0 [0.7, 1.5]	0.5 [0.3, 0.9]	0.9 [0.6, 1.3]	2.3 [1.6, 3.1]	22.9 [22.0, 23.9]
Fusion of local maps	72.3 [70.3, 74.2]	7.5 [6.3, 8.9]	2.3 [1.7, 2.9]	4.5 [3.8, 5.4]	7.3 [6.5, 8.2]	26.4 [25.3, 27.6]
Δ spatial memory − fusion [95% CI]	+11.7 [10.7, 12.7]	−1.6 [−2.3, −1.0]	−1.5 [−2.1, −0.9]	−2.3 [−3.0, −1.6]	+27.4 [24.4, 30.4]	−20.2 [−21.3, −19.1]
Δ spatial memory − geometric only [95% CI]	+2.6 [2.3, 3.0]	−0.2 [−0.5, 0.1]	−0.5 [−0.9, −0.2]	−0.3 [−0.7, 0.1]	+12.5 [9.8, 15.2]	−0.9 [−1.4, −0.5]
Δ spatial memory − pose only [95% CI]	+44.4 [42.8, 46.0]	+4.8 [3.7, 6.1]	+0.3 [−0.3, 0.9]	+1.3 [0.5, 2.2]	+32.7 [29.7, 35.9]	−16.7 [−17.8, −15.6]
75 validation-reporting rooms: 4.7% of the wall length is occluded within 8 m of a pose (3.7% also ≥0.6 m from every visible sample)						
rooms with a non-empty denominator (recall columns): visible 75, in range 50, ≥0.6 m 47, occluded points 50; precision column, map rows in order: 60 / 67 / 75 / 74						
Spatial-memory completion (with H)	88.8 [86.4, 91.2]	8.9 [5.0, 14.1]	0.8 [0.2, 1.6]	4.6 [1.4, 9.6]	37.2 [29.2, 45.2]	4.8 [3.9, 5.9]
Geometric-only completion (no H)	86.9 [84.3, 89.4]	9.5 [5.5, 14.8]	2.2 [0.5, 5.2]	5.0 [1.8, 9.9]	21.9 [15.9, 28.4]	5.8 [4.6, 7.1]
Pose-only completion (no acoustics)	46.8 [41.8, 52.0]	0.9 [0.0, 2.4]	0.0 [0.0, 0.1]	0.8 [0.0, 2.2]	1.1 [0.4, 2.1]	23.6 [21.8, 25.5]
Fusion of local maps	78.7 [74.7, 82.6]	11.1 [6.4, 16.8]	3.0 [0.9, 6.3]	7.5 [3.5, 13.0]	6.1 [4.5, 7.9]	27.2 [24.4, 30.0]
Δ spatial memory − fusion [95% CI]	+10.1 [8.1, 12.3]	−2.1 [−4.4, −0.4]	−2.2 [−5.4, −0.2]	−3.0 [−5.5, −0.9]	+30.1 [22.5, 38.1]	−22.4 [−25.4, −19.4]
Δ spatial memory − geometric only [95% CI]	+1.9 [1.4, 2.6]	−0.6 [−2.7, 1.0]	−1.4 [−4.2, 0.1]	−0.4 [−2.7, 1.5]	+13.1 [4.6, 21.8]	−0.9 [−2.3, 0.4]
Δ spatial memory − pose only [95% CI]	+42.0 [37.3, 46.8]	+8.0 [4.5, 13.0]	+0.8 [0.2, 1.5]	+3.7 [0.3, 8.9]	+35.9 [28.3, 43.7]	−18.8 [−20.9, −16.7]

Table 17: Lattice planners on the seed-42 primary belief (legacy fusion), EchoRooms: 75 reporting rooms × 3 starts. *Selected*: costs tuned on the earlier corpus; *default*: default costs. Top: wall F_1 at 0.30 m after 64 acquisitions (all rooms; the 67 with ≥64 nodes), recall at 64, motion attempts to 90 % recall (censored at 257) with censored share, success ($F_1 \geq 90$ % within 128 attempts), collisions and distance (m). Bottom: row minus the selected forecast-head arm, and for F_1 also minus selected raycast (paired, 10,000 room-bootstrap resamples, 95 % CI); *natural stop*: attempts at which the planner's stop rule first fires (the episode continues); *stop regret*: final minus natural-stop F_1. †Privileged (true geometry or future observations). ‡Unclipped signed utility; stop metrics not comparable.

Planner	Setting	F1 (0.30 m) at 64, all	F1 at 64, ≥64 nodes	Recall (0.30 m) at 64	Attempts to 90%	Censored (%)	Success at 128 (%)	Collisions	Distance (m)
FC-EER, forecast head	selected	76.1	73.6	70.2	166.9	60.4	40.0	6.87	122.8
FC-EER, forecast head	default	75.5	73.0	69.6	168.7	60.4	39.1	6.51	117.3
FC-EER, nearest-observation copy	selected	77.2	74.7	71.4	164.5	58.2	42.2	6.54	122.3
FC-EER, nearest-observation copy	default	76.5	74.1	71.1	165.5	58.7	41.8	6.60	121.1
FC-EER, knowledge read	selected	77.7	75.4	72.2	168.9	61.3	40.4	6.47	119.4
FC-EER, knowledge read	default	76.1	73.6	70.3	173.5	63.1	38.2	6.84	118.6
FC-EER, true future observation†	selected	77.7	75.4	72.2	161.1	57.3	44.0	6.78	123.4
FC-EER signed utility, forecast head‡	selected	72.0	69.0	65.4	182.6	66.7	32.4	6.61	136.1
FC-EER signed utility, forecast head‡	default	71.3	68.2	64.4	183.7	67.1	34.2	6.98	128.1
FC-EER signed utility, nearest-observation copy‡	selected	75.0	72.4	69.2	170.4	60.0	40.9	6.40	124.8
FC-EER signed utility, knowledge read‡	selected	76.3	73.9	70.5	172.9	63.1	40.0	6.77	123.0
Raycast information gain	selected	77.6	75.3	72.0	165.5	59.1	42.2	7.06	117.3
Raycast information gain	default	77.1	74.8	71.5	160.6	56.4	41.8	6.80	119.2
Uncertainty coverage	default	74.4	71.6	68.4	178.2	64.4	36.4	6.56	119.6
Approximate EIG	default	74.7	72.0	68.0	181.0	65.8	33.8	6.32	116.3
Nearest frontier	default	73.3	70.5	66.9	191.0	68.4	30.2	13.10	172.2
Blind DFS	default	66.8	63.2	58.4	195.6	71.6	28.9	22.28	92.5
Random walk	default	63.1	59.0	54.3	218.4	78.7	16.0	15.68	211.9
Privileged true-graph sweep†	default	83.7	82.0	80.0	140.1	47.6	53.8	0.00	108.4

Planner	Setting	ΔF1 at 64 vs forecast [95% CI]	ΔF1 at 64 vs raycast [95% CI]	Δ attempts vs forecast [95% CI]	Δ success vs forecast [95% CI]	Natural stop (attempts)	Stop regret (F1 points)
FC-EER, forecast head	selected	reference	−1.5 [−2.6, −0.5]	reference	reference	41.1	9.2
FC-EER, forecast head	default	−0.6 [−1.7, 0.5]	−2.1 [−3.3, −1.0]	+1.7 [−4.9, 8.5]	−0.9 [−4.0, 2.2]	18.7	20.9
FC-EER, nearest-observation copy	selected	+1.1 [0.1, 2.1]	−0.5 [−1.4, 0.5]	−2.5 [−10.0, 5.0]	+2.2 [−1.3, 5.8]	49.9	6.6
FC-EER, nearest-observation copy	default	+0.5 [−0.6, 1.5]	−1.08 [−2.17, 0.01]	−1.4 [−9.3, 6.1]	+1.8 [−3.1, 6.7]	27.9	14.8
FC-EER, knowledge read	selected	+1.6 [0.6, 2.7]	+0.1 [−0.8, 1.0]	+1.9 [−9.0, 12.9]	+0.4 [−5.3, 5.8]	44.2	1.2
FC-EER, knowledge read	default	+0.05 [−0.99, 1.11]	−1.5 [−2.5, −0.5]	+6.6 [−0.8, 14.9]	−1.8 [−5.8, 1.8]	37.2	6.4
FC-EER, true future observation†	selected	+1.7 [0.7, 2.6]	+0.1 [−0.8, 1.0]	−5.8 [−13.3, 1.1]	+4.0 [0.4, 8.0]	49.3	3.6
FC-EER signed utility, forecast head‡	selected	−4.1 [−5.4, −2.9]	−5.6 [−7.0, −4.3]	+15.6 [6.9, 25.5]	−7.6 [−13.3, −2.7]	12.3	24.1
FC-EER signed utility, forecast head‡	default	−4.8 [−6.3, −3.3]	−6.3 [−8.1, −4.6]	+16.7 [8.9, 25.5]	−5.8 [−10.2, −1.8]	9.8	25.5
FC-EER signed utility, nearest-observation copy‡	selected	−1.1 [−1.9, −0.2]	−2.6 [−3.8, −1.4]	+3.5 [−6.3, 13.4]	+0.9 [−3.1, 5.3]	35.4	12.9
FC-EER signed utility, knowledge read‡	selected	+0.2 [−0.8, 1.2]	−1.3 [−2.3, −0.4]	+5.9 [−4.6, 16.6]	+0.00 [−4.89, 4.89]	44.0	5.8
Raycast information gain	selected	+1.5 [0.5, 2.6]	reference	−1.5 [−9.8, 6.0]	+2.2 [−1.3, 6.7]	27.4	12.5
Raycast information gain	default	+1.01 [−0.03, 2.07]	−0.5 [−1.3, 0.3]	−6.3 [−15.1, 2.3]	+1.8 [−3.1, 6.7]	10.6	28.0
Uncertainty coverage	default	−1.7 [−2.9, −0.6]	−3.2 [−4.5, −2.0]	+11.3 [2.3, 21.1]	−3.6 [−8.0, 0.4]	12.9	22.0
Approximate EIG	default	−1.4 [−2.7, −0.1]	−2.9 [−4.1, −1.8]	+14.0 [5.7, 23.8]	−6.2 [−11.6, −1.8]	52.0	5.9
Nearest frontier	default	−2.7 [−4.0, −1.5]	−4.3 [−5.6, −3.0]	+24.0 [13.9, 35.4]	−9.8 [−15.6, −4.4]	18.3	16.2
Blind DFS	default	−9.2 [−11.9, −6.6]	−10.8 [−13.5, −8.0]	+28.6 [15.1, 42.6]	−11.1 [−18.7, −4.0]	112.5	0.0
Random walk	default	−13.0 [−15.5, −10.6]	−14.5 [−17.1, −12.0]	+51.5 [35.0, 69.2]	−24.0 [−32.9, −15.6]	–	–
Privileged true-graph sweep†	default	+7.6 [5.4, 10.0]	+6.1 [3.9, 8.4]	−26.8 [−40.5, −14.9]	+13.8 [7.6, 20.9]	39.6	0.8

Table 18: Crossed replay with legacy-fusion beliefs: H improves reconstruction of identical observation streams and, for three of four planners, the trajectories they choose. 75 reporting rooms $\times$ 3 starts; wall F_1 at 64 acquisitions (%) of recorded runs on the primary (with H) and geometric-only (no H) beliefs. Their difference (points) splits into *reconstruction* (both beliefs scoring the geometric-only trajectories), *trajectory* (the primary scoring both trajectory sources) and *total* (recorded runs); interactions are within 0.2 points. Paired room bootstraps (10,000 resamples, 95 % CI). Bottom: belief-frontier walker on 246 test rooms $\times$ 3 starts (64-acquisition budget), final F_1; its trajectories are identical under both beliefs. Details: Table 19, Appendix D.10.

Planner	F_1 at 64 (%)		ΔF_1 at 64, primary − geometric only (points, 95% CI)		
	primary belief	geometric only	reconstruction (fixed stream)	trajectory (fixed evaluator)	total (recorded runs)
FC-EER, copy	77.2	75.0	+1.6 [1.2, 2.1]	+0.5 [−0.6, 1.6]	+2.1 [1.0, 3.3]
FC-EER, knowledge read	77.7	73.1	+1.7 [1.2, 2.1]	+2.9 [1.8, 4.0]	+4.6 [3.4, 5.8]
FC-EER, forecast head	76.1	71.2	+1.6 [1.2, 2.1]	+3.3 [1.9, 4.6]	+4.9 [3.5, 6.4]
Raycast information gain	77.6	73.9	+1.6 [1.1, 2.1]	+2.1 [1.2, 3.0]	+3.7 [2.7, 4.7]
Belief-frontier walker, test split	62.1	60.8	+1.3 [1.1, 1.4]	identical	+1.3 [1.1, 1.4]

Table 19: Full crossed replay on EchoRooms with the primary (legacy fusion with H) and geometric-only (no H) beliefs: 75 reporting rooms $\times$ 3 starts for each selected planner. Rows fix the trajectory source (the belief the planner ran on) and columns the evaluator. Wall F_1 at 64 acquisitions (0.30 m tolerance) and final interior IoU are in %; differences are in points with paired room-bootstrap 95 % CIs (10,000 resamples). The third row per planner gives trajectory differences within each evaluator and total recorded-run differences in the Δ columns. Bottom: identical belief-frontier trajectories on 246 test rooms (738 episodes). Replay checks and scores carried forward at the attempt cap are described in Appendix D.10.

Planner	Trajectory	F1 at 64 primary belief	F1 at 64 geometric only	ΔF1 at 64, primary − geometric only [95% CI]	Final interior IoU primary belief	Final interior IoU geometric only	ΔFinal interior IoU, primary − geometric only [95% CI]
FC-EER, nearest-observation copy	planned on the primary belief	77.2	75.7	+1.5 [1.0, 1.9]	78.5	75.9	+2.6 [1.3, 3.7]
	planned on the geometric-only belief	76.7	75.0	+1.6 [1.2, 2.1]	78.2	75.7	+2.5 [1.0, 3.8]
	Δ trajectory, primary − geometric only [95% CI]	+0.5 [−0.6, 1.6]	+0.7 [−0.4, 1.8]	+2.1 [1.0, 3.3]	+0.3 [−0.9, 1.5]	+0.2 [−1.0, 1.5]	+2.8 [0.9, 4.6]
FC-EER, knowledge read	planned on the primary belief	77.7	76.2	+1.5 [1.0, 2.0]	79.1	76.4	+2.7 [1.5, 3.7]
	planned on the geometric-only belief	74.8	73.1	+1.7 [1.2, 2.1]	76.2	73.7	+2.5 [1.1, 3.7]
	Δ trajectory, primary − geometric only [95% CI]	+2.9 [1.8, 4.0]	+3.1 [2.0, 4.2]	+4.6 [3.4, 5.8]	+2.9 [1.7, 4.2]	+2.8 [1.3, 4.2]	+5.4 [3.4, 7.4]
FC-EER, forecast head	planned on the primary belief	76.1	74.4	+1.7 [1.2, 2.2]	76.9	74.0	+2.9 [1.8, 4.0]
	planned on the geometric-only belief	72.8	71.2	+1.6 [1.2, 2.1]	74.3	71.8	+2.5 [1.0, 3.8]
	Δ trajectory, primary − geometric only [95% CI]	+3.3 [1.9, 4.6]	+3.2 [1.9, 4.5]	+4.9 [3.5, 6.4]	+2.7 [1.1, 4.3]	+2.3 [0.5, 3.9]	+5.2 [2.7, 7.5]
Raycast information gain	planned on the primary belief	77.6	75.9	+1.7 [1.2, 2.2]	78.3	75.2	+3.1 [2.0, 4.3]
	planned on the geometric-only belief	75.5	73.9	+1.6 [1.1, 2.1]	76.7	73.6	+3.1 [1.5, 4.5]
	Δ trajectory, primary − geometric only [95% CI]	+2.1 [1.2, 3.0]	+2.0 [1.2, 2.9]	+3.7 [2.7, 4.7]	+1.6 [0.5, 2.8]	+1.6 [0.1, 2.9]	+4.7 [2.7, 6.6]

Walker	Trajectory	Final F1 primary belief	Final F1 geometric only	Δ final F1, primary − geometric only [95% CI]	Final interior IoU primary belief	Final interior IoU geometric only	Δ interior IoU, primary − geometric only [95% CI]
Belief frontier, 64-acquisition budget (246 test rooms, 738 episodes)	identical under both beliefs	62.1	60.8	+1.3 [1.1, 1.4]	61.4	59.1	+2.3 [1.8, 2.8]

E Belief-guided exploration policy

This appendix supports Secs. 4.4, 5.2 and 5.3. Appendices E.1–E.6 describe the *reference recipe*, our earlier PPO policy, with its environment, reward, walkers and planners. Appendices E.7–E.12 then cover the main method: training and freeze, planner and stopping comparisons, live-sensing, single-episode, confirmation and fresh-room evaluations, ablations and input controls. Appendix E.13 closes with a collision-write check on the reference recipe.

Names used in the policy appendix. The text and tables of this appendix use the short names below; each is also defined where it first appears.

Name	Meaning
Main method (EchoBelief (full) in tables)	The selected seed-42 PPO checkpoint of Sec. 4.4 (update 3,600), run at STOP floor 4 with 128 acquisitions and 1,024 decisions unless stated.
Main method's recipe	The PPO recipe that produced the main method (Sec. 4.4).
Fine-tuned belief	The seed-42 fit of the evidence-anchored spatial memory fine-tuned with support-gated fusion and hole-aware completion targets (Appendix D.6; selected in Appendix G); the main method's frozen belief.
Legacy belief, legacy fusion	The seed-42 fit without that fine-tuning, used by the reference recipe; legacy fusion writes Stage 2 predictions into G wherever $p_{\text{wall}} + p_{\text{free}} \geq 0.5$.
Legacy-fusion arm	An ablation arm: the main method's recipe retrained on the legacy belief (*Legacy fusion* in tables).
Geometric-only (memory-free) belief	The same belief trained without H. Under legacy fusion Stage 2 predictions stay in G; under support-gated fusion H is the route for retaining past local predictions beyond the support-gated evidence.
Without-memory-pathway arm	An ablation arm: the geometric-only belief given support-gated fusion and the fine-tuning, with the policy retrained on it (*Without memory pathway* in tables).
Reference recipe	Our earlier PPO recipe (Sec. 4.4): legacy belief, fixed STOP floor 24 and a start-anchored canvas (the global grid shared by G and H, Sec. 4.3) that loads only rooms with node spans up to 52 m.
Trained-STOP policy	The reference recipe's selected seed-42 checkpoint at its own STOP floor of 24.
Capped policy	The reference recipe retrained with STOP disabled (floor 64), seeds 42 and 43.
Reference checkpoint, floor 128	The trained-STOP checkpoint re-run with 128 acquisitions, 1,024 decisions and STOP disabled.
Reference-recipe arm	An ablation arm: the reference recipe's network, features, terminal block and STOP floor of 24 retrained in the all-rooms configuration.
64-trained runs	The main method's recipe trained at 64 acquisitions and 512 decisions on the 3,177 training rooms within 52 m (seeds 42 and 43).
All-rooms configuration	The main method's environment: fine-tuned belief and a canvas centred on the room window, which loads all 3,447 training and 355 test rooms.
Floor k (STOP floor)	STOP is legal only after k acquisitions; a floor equal to the acquisition cap disables STOP.
Protocols	*128 acquisitions*: STOP floor 4, 128-acquisition cap, 1,024 decisions (the main method's own); *64 acquisitions*: STOP floor 4, 64-acquisition cap, 512 decisions; *STOP disabled*: floor 64, 64-acquisition cap, 512 decisions (the reference recipe's capped protocol). In the STOP-on protocols the trained-STOP policy, the reference-recipe arm and the walkers keep floor 24.
Planners	Raycast information gain (*Raycast IG planner* in tables; *policy motion*: its routes walked with the policy's move set) and the two imagined-observation FC-EER planners, the copy planner and the knowledge-read planner (*Imagined planner, copy* and *Imagined planner, read* in tables; *imagined-update planner* in Table 22); they never choose STOP except in the stop-rule control (Appendix E.7).
Walkers	The belief-frontier and random walkers (Appendix E.3).

Usable rooms	Rooms within the 52 m span cap, the only ones the reference recipe and the 64-trained runs can load: 246 of the 355 test rooms and 69 of the 75 selection rooms.
Sealed holdout, validation rooms, added rooms	The 355-room test split comprises 120 original rooms held out from policy development, 122 original rooms used during policy development, and 113 large rooms added after the first test read. The separate 224-room validation split contains 75 selection, 74 calibration, and 75 reporting rooms.
Success	Final wall recall $\geq 90\,\%$ at 0.30 m; the lattice planner studies of Appendices D.9 and D.10 instead use wall $F_1 \geq 90\,\%$.
Acquisitions-to-90 (*Acq. 90* in tables)	Acquisitions until wall recall first reaches 90 %, failure-penalised: an episode that never crosses is censored at the cap plus one (65 or 129).
Senses	SENSE actions per episode, excluding the start acquisition.

E.1 Reference-recipe environment and observations

This subsection and the next five describe the reference recipe of Sec. 4.4 and its evaluation. The main method's differences (fine-tuned belief, window-centre anchor, hole-aware truth, collision memory, terminal block) are specified in Sec. 4.4 and Appendix E.7. The environment runs 128 rooms in parallel and represents positions in float64.

The frozen belief of the reference recipe is the EchoRooms fit of the evidence-anchored spatial memory (best checkpoint at update 19,000, global completion temperature 1.311). Observations come from caches of the Stage 2 outputs at every lattice node at heading 0. Rooms whose lattice span exceeds 52 m or whose environment window exceeds 320 cells are excluded as unusable. This leaves 3,177 of the 3,447 training rooms, 69 of the 75 selection rooms and 246 of the 355 test rooms (211 of the 242 original rooms and 35 of the 113 added rooms).

The clearance rule for moves is 0.3 m against every wall segment (the lattice itself was generated with a 0.45 m platform clearance). Every episode starts with one acquisition at the start node. SENSE snaps to the lattice node nearest to the agent by Euclidean distance, is legal only within the 1 m snap radius, and writes that node's cached observation registered at the node's true pose. There is no line-of-sight test between the agent and the node, and the rule is shared by the learned policy and both walkers (Appendix E.3).

An audit of every SENSE in the test trajectories of the trained-STOP policy and the two walkers (Table 20) checks whether the segment from the agent position to the sensed node crosses any wall segment of the room. It finds 0 of 23,084 learned, 0 of 11,351 random and 3 of 39,705 belief-frontier snaps that do, at mean snap distances of 0.39, 0.41 and 0.49 m. The three crossing snaps (room 3428, two episodes, snap distances 0.96–0.98 m) each cross the two faces of one thin partition, two wall segments about 0.09 m apart.

Observations. *Reference recipe.* The local window has 12 channels at 0.25 m (96×96, centred on the agent's belief cell): p_{wall}, p_{interior}, p_{exterior}, $\mathcal{H}/\ln 3$, the predictive-entropy map (the same quantity set to zero on observed cells), the six evidence channels (wall ratio, free ratio, saturating support, mean echo, visited, collision) and a visit channel $\log(1 + \text{count})/\log 6$. The coarse window has 7 channels at 1 m (112×112, anchored at the episode anchor): the belief's four coarse channels, an agent stamp, the visit channel and a last-sense stamp.

The privileged critic receives 6 local and 6 coarse channels: the true wall/interior/exterior one-hot, the unrecalled-wall mask, the unknown-but-reachable mask and the unacquired-node stamp.

The main method. The actor reads the same windows with a thirteenth local channel (conflict: the minimum of the wall and free evidence), 65 scalars and 15 per-move features sampled along the swept path (Sec. 4.4). The scalars extend the reference inputs with recent action/outcome history, STOP floor, remaining acquisitions, measured-support fraction and last support gain. Residual encoders of widths 32/64/128 feed a shared MOVE scorer, separate SENSE and STOP scorers, and collision and completion heads. The online one-step predictor is queried at most 5 times per decision and feeds a zero-initialised residual; its weight ramps from 0 to 1 over updates 50–200.

Table 20: SENSE snaps of the reference recipe's trained-STOP policy and the two walkers on the held-out test split (246 usable rooms × 3 starts, 738 episodes per policy): SENSE actions, those with the agent off the node (displaced), mean and maximum snap distance, snaps whose agent-to-node segment crosses or touches a wall segment of the room, their share of all SENSEs, and the episodes containing one. Room geometry from the corpus room files, verified against the policy environment's corpus manifest.

Policy	Episodes	SENSEs	Displaced SENSEs	Mean snap (m)	Max snap (m)	Wall-crossing snaps	Crossing (% of SENSEs)	Episodes with a crossing
Trained-STOP policy (seed 42)	738	23,084	22,905	0.388	1.000	0	0.00	0
Random walker	738	11,351	11,323	0.406	0.998	0	0.00	0
Belief-frontier walker	738	39,705	39,695	0.494	1.000	3	0.01	2

E.2 Reward and PPO training

This subsection specifies the reward and PPO optimisation summarised in Sec. 4.4. The reference recipe uses the following reward:

$$\begin{aligned} r_t = {} & 10\,\Delta\text{Recall}_{0.30} + 5\,\Delta F_1^{0.30} + 2\,\Delta\text{Known}_{\text{reach}} + \big(\gamma\,\phi(s_{t+1})\,(1-d_t) - \phi(s_t)\big) \\ & + \mathbf{1}[\text{recall first} \geq 0.9] + \tfrac{1}{2}\,\mathbf{1}[\text{STOP after success}] - 0.005 - 0.02\,\mathbf{1}[\text{SENSE}] - 0.05\,\mathbf{1}[\text{repeat}] \\ & - 0.01\,\min(v_t, 5) - 0.02\,\mathbf{1}[\text{oscillation}] - 0.05\,\mathbf{1}[\text{collision}] - 0.01\,\mathbf{1}[\text{stagnation}] \\ & - 2\,(1 - \text{Known}_{\text{reach}})\,d_t, \end{aligned} \tag{6}$$

where v_t is the number of prior visits of the current 0.5 m cell, d_t marks the terminal step, and the ϕ term is potential-based frontier shaping (Ng et al., 1999); Table 21 lists every term with its coefficient and trigger. The main method retains the dense terms and interaction costs, replacing the first-crossing success bonus, STOP bonus and terminal cost by the task-aligned terminal block of Sec. 4.4, paid once per episode.

Optimisation. PPO uses horizon 32, 4 epochs, 4 minibatches, learning rate 3×10^{-4} decayed linearly to 0 over 15 M steps, clip 0.2, entropy coefficient $0.01 \rightarrow 0.001$, value coefficient 0.5, gradient clipping 0.5, $\gamma = 0.99$, GAE $\lambda = 0.95$, 4,096 environment steps per update, checkpoints every 50 updates, no teacher.

Reference recipe. Its actor has two convolutional window encoders and a scalar MLP (21 scalars) merged into a 34-way masked softmax head, with a critic on the same trunk over the actor and privileged channels. Every 200 updates it is evaluated on the 69 usable selection rooms with two fixed starts (seed 1000), scored lexicographically by mean failure-penalised acquisitions-to-90, then success fraction, then mean final F_1; the selected seed-42 checkpoint is update 2,600 (10.65 M steps).

E.3 Walkers and reference-recipe test results

This subsection defines the two walkers, simple baselines used throughout this appendix, and reports the reference recipe's own test results, against which Table 27 compares the main method. One walker ignores the belief and one follows it. *Random* draws uniformly over the legal actions. *Belief frontier* pools the belief to 0.5 m and declares a cell passable if it is known and its interior share is at least half its known share; the frontier is the set of passable cells adjacent to unknown cells. It plans to the nearest frontier by breadth-first search with a move cost of distance plus 0.5 m per prior visit. It senses when standing on a frontier cell or every 2 m of travel, provided the nearest node is unacquired, and stops when no frontier remains and STOP is legal.

Evaluated greedily on the 246 usable test rooms with three starts (seed 2000; 64 acquisitions, 512 decisions), the reference recipe's trained-STOP policy reaches 33.8 failure-penalised acquisitions-to-90 with 63.7 % success. The capped policies (STOP floor 64, seeds 42 and 43; Fig. 12b) score 30.2 [27.3, 33.0] and 30.9 [28.0, 33.9] failure-penalised acquisitions-to-90 with 73.0 and 69.6 % success and 14.0 and 19.5 collisions. Under the same cap the belief-frontier walker scores 59.3 failure-penalised acquisitions-to-90 with 14.1 % success, and random actions 62.2 with 5.1 %. Table 22 pairs the capped policy with the planners run inside the policy environment (Appendix E.4).

Table 21: Shared dense reward terms and reference-recipe terminal terms (Eq. (6)). The first-crossing success bonus, STOP bonus and terminal coverage cost apply only to the reference recipe; the main method replaces them with the terminal block in Sec. 4.4.

Term	Coefficient	Trigger
Δ wall recall at 0.30 m	10	every writing SENSE, new collision stamp, episode end
Δ wall F_1 at 0.30 m	5	same
Δ known reachable fraction	2	same
Frontier shaping $\gamma\phi(s')(1-d)-\phi(s)$	1	every decision; $\phi = -0.02\times$ geodesic distance (m, cap 64) to the nearest reachable unknown cell, 0 if none
Success bonus	1	first time recall ≥ 0.9
Stop bonus	0.5	STOP after success
Step cost	-0.005	every decision
Sense cost	-0.02	every SENSE
Repeat-sense cost	-0.05	SENSE on an acquired node (in addition)
Revisit cost	$-0.01\min(v, 5)$	entering a previously visited 0.5 m cell
Oscillation cost	-0.02	within 0.6 m of a position at $t-2 \ldots t-8$
Collision cost	-0.05	blocked move
Stagnation cost	-0.01	each decision after 16 without a new cell or coverage gain
Terminal cost	$-2\,(1-K)$	episode end, K = known reachable fraction

E.4 Planners inside the policy environment

This subsection specifies how the lattice planners of Appendix D.10 run inside the policy environment, first against the reference recipe (Table 22) and then as the planner rows of Table 1 (Sec. 5.2). They share the rooms, frozen belief (seed 42), 1 m SENSE snap, 0.3 m motion clearance and budgets (64 acquisitions, 512 decisions) with the policy.

Candidates are the integer-metre points within the planner's radius that a reachability search over the decoded belief can reach from the agent's integer-metre position, excluding acquired and refused positions; they are sorted by coverage and capped at 32 per decision.

Utility. Raycast information gain marches 72 rays out to 8 m from the candidate pose and sums the belief entropy the rays would resolve (the lattice study's implementation, re-expressed in vectorised form and checked against it to 10^{-9}). The imagined-observation planners write an imagined observation into a copy of the belief, decode it at the belief's calibrated global completion temperature and score the clipped predictive entropy removed inside the candidate's window. The imagined observation is either the nearest acquired observation warped rigidly into the candidate frame (*copy*) or the knowledge head read at the candidate pose with a zero echo raster and the calibrated knowledge temperature (*knowledge read*), the observation sources of Appendix D.10. The settings transferred from the lattice study also carry $\tau = 2$ (Appendix D.9), the quiet-decision stop threshold rather than a temperature; τ has no effect here because the planners never choose STOP.

The score subtracts route length, remembered collisions along the route and revisits from utility, weighted by $\lambda_{\text{distance}}$, $\lambda_{\text{collision}}$ and λ_{revisit}, respectively.

Walk-and-sense loop. The planner makes one decision per acquisition. A route follower executes the returned route as the environment's unit cardinal moves, so it never leaves the lattice and uses the same clearance rule. A move either lands on the waypoint or collides under the environment's own rule. A collision is stamped into the belief by the environment, blocks that edge, is remembered and triggers a replan; so does a route step that the legal mask forbids. At the goal the route follower SENSEs when the nearest node is unacquired and within the snap radius. Otherwise it refuses the goal, which then joins the acquired positions so that it is never proposed again, and replans. With no candidates it SENSEs under the same nearest-node check or takes a seeded random cardinal step avoiding remembered collisions; it replans after every such decision for raycast information gain and after every fourth for the imagined-observation planners. STOP is never chosen; episodes end at the

acquisition cap or the decision limit. Under this contract the planners spend 62.1–63.0 of their 64 acquisitions while the capped policies take 35.5–36.8.

The lattice protocol of Appendix D.10 differs. It follows a route for at most eight hops and acquires an observation at *every* route node, so every arrival is an acquisition and the planner replans at least every eight hops. Here, the route follower walks the whole route without sensing, senses at the goal only and replans on a SENSE, a refused goal or a collision.

Settings of raycast information gain and the copy planner were reselected inside the environment rather than transferred from the lattice study. The grid crosses radius $\in \{4, 8\}$ m with cost weight $\in \{0.3, 1, 3\}$ (with $\lambda_{\text{collision}} = 3\times$ the cost weight and, for the imagined-observation planners, $\lambda_{\text{revisit}} =$ the cost weight). One further point keeps the lattice-study radius and cost weight ($r = 8$ m, $c = 1$) but uses the lattice platform clearance of 1.15 m instead of the environment's 0.3 m. The seven points per planner were run on the 69 usable selection rooms with one start each and scored by the policy selection objective (mean failure-penalised acquisitions-to-90, then success, then final F_1). Raycast information gain and the copy planner both selected radius 8 m, $\lambda_{\text{distance}} = 0.3$, $\lambda_{\text{collision}} = 0.9$, and $\lambda_{\text{revisit}} = 0.3$ for the imagined-observation planners (raycast information gain keeps the lattice study's 3.0).

On the 69 usable selection rooms the selected point scores 32.1 acquisitions-to-90 with 81.2 % success for raycast information gain against 39.8 and 60.9 % at the lattice-study point, and 29.8 with 79.7 % for the copy planner against 34.3 and 72.5 %. The 1.15 m clearance point was worse for both (42.7 and 34.5; Table 23, reference-environment columns). The knowledge-read planner was not searched separately; it inherits the copy planner's selected settings.

On the test split the planners spend 0.106 (raycast information gain), 0.268 (copy) and 0.237 s (read) per decision with 31.0–31.6 candidates scored, 2.5, 0.9 and 0.9 % of decisions without a candidate and 2.4, 1.9 and 2.1 % fallback decisions. These latencies were measured on shared GPUs and characterise the supplied implementations, not an inherent cost.

The planner rows of Table 1 run the same three planners, with the same walk contract and the settings selected here, in the main method's environment (fine-tuned belief with support-gated fusion, window-centre anchor, collision memory). They cover all 355 rooms at 128 and at 64 acquisitions. The grid was re-run in that environment and both planners reselected the same settings (Appendix E.7; Table 23). The raycast planner was also run with the policy's move set (Appendix E.7).

E.5 Reference-recipe arms: memory-free belief and fixed sensing schedule

Two diagnostic arms ask whether the reference recipe's policy needs the recurrent context memory and whether it needs to choose when to sense. Both retrain the reference recipe's capped configuration unchanged (15 M steps, the same reward, networks and selection rule) and are evaluated with the same protocol (Table 22). Each is a single run at seed 42; the capped policy's own seed spread is 30.2 against 30.9 failure-penalised acquisitions-to-90.

Memory-free belief. PPO is trained on the geometric-only belief of Sec. 5.4, the same belief trained without the recurrent context memory H. It uses legacy fusion, so Stage 2 predictions still reach its decoders through G. The environment reproduces this belief's decoder by allocating a zero H canvas, skipping its writes and feeding zeros in place of H to the global read. The configuration differs from the capped policy only in the belief and its calibrated temperature. Relative to the capped policy (memory-free minus capped), there is no resolved difference in failure-penalised acquisitions-to-90 (-0.5 [-1.1, 0.2]) or success ($+0.8$ [-1.6, 3.4] points), but final F_1 and interior IoU show resolved losses of -2.0 [-2.8, -1.3] and -2.8 [-3.7, -1.9] points (Table 22). This near-null policy effect contrasts with the 22.8-point success deficit of the main method's without-memory-pathway arm (Appendix E.12). That arm also differs in recipe, budget and fusion rule: under legacy fusion, Stage 2 predictions stay in G when H is removed, whereas under that arm's support-gated fusion H is the route for retaining past local predictions beyond the support-gated evidence.

Fixed sensing schedule. SENSE is removed from the action set and an acquisition is instead taken automatically on every N-th MOVE decision, with $N = 13$ the capped seed-42 policy's own rate on this test split; the arm is therefore a *retrospective* control. The acquisition does not consume a decision, so 512 decisions allow at most 40 acquisitions, and the arm takes 38.8 sensing actions on average. Against the capped policy (fixed schedule minus capped) it shows no resolved difference in failure-penalised acquisitions-to-90 ($+0.2$ [-0.5, 0.9]) and a success change of -5.3 [-7.6, -3.1] points (Table 22). Because the sensing period was selected retrospectively from test behaviour and automatic acquisitions used different decision accounting, this experiment does not isolate the effect

Table 22: Paired room bootstraps on the held-out test split (246 usable rooms × 3 starts, 64-acquisition cap, STOP disabled for the capped policies): the capped policy (seed 42) against the two strongest planners run inside the policy environment (Appendix E.4), and the two retrained arms of Appendix E.5 against the capped policy. Episodes are averaged within a room before resampling; entries are the mean paired difference, first minus second, with its 95 % percentile interval, in failure-penalised acquisitions-to-90 for the first column and in percentage points for the others.

Paired room bootstrap (10,000 resamples), first − second	Acq. to 90% [95% CI]	Success (points)	Final F1 (points)	Interior IoU (points)	Rooms
Capped policy (seed 42) − imagined planner, read	−6.5 [−7.6, −5.3]	+9.6 [7.2, 12.3]	+6.4 [5.1, 7.8]	+6.7 [5.3, 8.2]	246
Capped policy (seed 42) − raycast IG planner	−8.8 [−10.4, −7.2]	+13.6 [10.0, 17.3]	+4.9 [3.8, 6.0]	+6.2 [4.9, 7.6]	246
Memory-free capped policy − capped policy (seed 42)	−0.5 [−1.1, 0.2]	+0.8 [−1.6, 3.4]	−2.0 [−2.8, −1.3]	−2.8 [−3.7, −1.9]	246
Fixed-schedule capped policy − capped policy (seed 42)	+0.2 [−0.5, 0.9]	−5.3 [−7.6, −3.1]	−3.7 [−4.7, −2.8]	−3.4 [−4.3, −2.5]	246

of adaptive sensing timing, and we do not use it to establish a timing-specific advantage for the main method.

E.6 STOP floor and budget controls of the reference recipe

This subsection explains why the reference recipe uses a STOP floor of 24 and how the floor and the budget change its results. An initial run with STOP legal from 4 acquisitions ended nearly every episode at four acquisitions (training-rollout stop fraction 1.0); training was restarted with the floor raised to 24. With this reward, sensing concentrates near the floor and the policy tends to stop as soon as STOP becomes legal.

When the selected checkpoint is re-run on the 246 usable test rooms, success rises with the floor (55.8, 63.7, 66.8, 68.4 and 70.6 % for floors 16, 24, 32, 48 and 64). The failure-penalised acquisitions-to-90 score stays almost flat above floor 24 (32.1–33.8 for floors 24–64; 36.7 at floor 16), so raising the floor above 24 chiefly changes when the policy stops. With the acquisition cap raised to 128 and a 1,024-decision budget, the trained-STOP policy (floor 24) stops as before (33.2 senses, 63.8 % success). Re-run with STOP disabled (floor 128), the same checkpoint instead reaches 82.8 % success with 94.2 % final recall but 56.0 collisions per episode. The actor's acquisitions/64 scalar is rescaled to the new cap, so these runs are references outside the training distribution. The floor-128 re-run is the row *Reference checkpoint, floor 128* of Table 27.

That the policy stops as soon as its floor allows is consistent with this reward design: under the main method's recipe, whose terminal block is task-aligned (Sec. 4.4), the policy learns to stop within the same action space (Appendix E.7, *Where it stops*). That recipe also changes the belief, the actor and the STOP-floor schedule, so this contrast does not isolate the reward.

E.7 Main method: training, freeze and comparisons

This subsection expands the budgeted comparisons of Sec. 5.2. It reports the main method (seed 42, STOP floor 4, 128 acquisitions; Sec. 4.4) and the *64-trained runs*: two runs of the main method's recipe trained at 64 acquisitions on the reference recipe's 3,177 usable training rooms (seeds 42 and 43). It covers training and freeze, evaluation protocols, planner settings, the information contract, the planner and stopping comparisons, test populations and tiers, and the 246-room comparison with the reference recipe.

Training and freeze. The main method, its seed-43 replicate and the ablation arms of Appendix E.12 use the same implementation and training schedule, with the intervention specified for each arm: 15 M transitions on all 3,447 training rooms at 128 acquisitions and 1,024 decisions. The canvas is anchored at the room-window centre (node spans of about 116 m admitted), and rooms are sampled in four node-count bands. The main method uses the fine-tuned belief of Appendix D.6, frozen throughout PPO. Its auxiliary losses are collision and current-completion binary cross-entropies, each weighted 0.1; the reference-recipe arm removes them.

Checkpoints were selected on the 75 selection rooms of the validation split, never on test rooms, by success, then wall F_1, then acquisitions. The selected updates (with training time) are 3,600 for the main method (seed 42; 14.75 M of its 15 M transitions, 35.3 h on one A40), 2,600 for seed 43 (35.6 h) and 3,600 for seed 44 (31.7 h, resumed once from its update-300 checkpoint). For the arms they are 2,600 for the reference-recipe arm (25.9 h), 3,000 for no predictor (32.5 h), 2,200 for no collision memory (32.7 h), 2,400 and 1,200 for the legacy-fusion arm at seeds 42 and 43 (35.2

and 40.3 h) and 3,400 and 3,600 for the without-memory-pathway arm at seeds 42 and 43 (34.1 and 32.7 h). For the controls they are 400 for the no-acoustic control (29.4 h) and 2,200 for the start-anchored control (32.7 h), which was selected on the 69 usable selection rooms. Each checkpoint, belief, recipe and evaluation protocol was frozen after training and before its test evaluation, and no checkpoint was re-selected after a test read. A provisional checkpoint of the seed-42 without-memory-pathway arm (update 2,400), taken while training was still running, was evaluated on the validation selection rooms only, not on the test rooms; after training finished, the same rule selected update 3,400 on those validation rooms, and the arm was evaluated on the test rooms only after that freeze. Planner checkpoints and configurations were likewise frozen before evaluation. Checkpoints, splits and tables are identified by hash in the release provenance.

The 64-trained runs used the 3,177 rooms within 52 m, 64 acquisitions and 512 decisions (seeds 42 and 43; 19.0 and 18.7 h). Updates 3,400 and 3,600 were selected on the 69 usable selection rooms and frozen before their test read.

Test-read history. The room-disjoint test split was read four times: (i) 242 original rooms at the reference recipe's freeze; (ii) all 355 after adding 113 large rooms without changing the recipe; (iii) an initial seed-42 run of the main method's recipe and the 64-trained runs; and (iv) the reported evaluation of the main method, its seed replicates, the ablation arms and the input controls, each read once after its own freeze.

Protocols and rooms. Every row uses the paper's evaluation contract (3 starts per room, seed 2000, the same (room, start) pairs as the reference recipe's test evaluations; Appendix F) under one of three protocols, named as in the tables: *128 acquisitions* (STOP floor 4, 128-acquisition cap, 1,024 decisions; the main method's own), *64 acquisitions* (STOP floor 4, 64-acquisition cap, 512 decisions) and *STOP disabled* (floor 64, 64-acquisition cap, 512 decisions; the reference recipe's capped protocol). In the two STOP-on protocols the reference-recipe arm and the walkers keep floor 24; the planners never choose STOP.

We evaluate the main method on all 355 rooms: the 246 supported by the reference recipe (104 sealed-holdout rooms, 107 validation rooms and 35 added rooms, all with node spans of at most 52 m), and the remaining 109 rooms (78 added and 31 original) beyond this limit. For rooms beyond 52 m, comparisons are limited to planners and walkers run under the main method's configuration.

Paired differences are room bootstraps (10,000 resamples): on all 355 rooms against the walkers, the planners and the ablation arms (Tables 1 and 2), on the 246 against the reference recipe with the main method restricted to them (Table 27). The main method's 355-room means by population and tier are in Tables 28 and 29.

Rows under the 64-acquisition cap run the checkpoint trained at 128 acquisitions, with the decision limit and the collision-memory capacity lowered from 1,024 to 512. Their recipe hash therefore differs from the checkpoint's, and that one hash check is overridden. Interior IoU and coverage use hole-aware truth (Appendix F). Tables 1 and 27 therefore report interior IoU for every row evaluated in the main method's environment (the restricted rows over their 246 rooms) and a dash for the reference recipe's rows, whose interior IoU under the frozen pipeline's hole-blind truth is not compared with the hole-aware values.

Planner settings under the main method's environment. The raycast information-gain and copy planners of Table 1 carry the settings selected in the reference recipe's environment (Appendix E.4), the *transferred setting* below. To check that these settings still suit the new environment, the same grid was re-run in the main method's environment (the fine-tuned belief with support-gated fusion, window-centre anchor, all-rooms configuration, 64 acquisitions and 512 decisions) on its 75 selection rooms with one start each (seed 2000). It covers radius $\in \{4, 8\}$ m $\times$ cost weight $\in \{0.3, 1, 3\}$ plus the lattice-clearance point ($r = 8$ m, $c = 1$, clearance 1.15 m), scored by the same objective as the earlier grid (mean failure-penalised acquisitions-to-90, censored at 65, then success, then final F_1; Table 23).

Both planners reselect $r = 8$ m, $c = 0.3$, the corner of the grid (its largest radius and smallest cost weight), as in every radius and cost grid of this appendix; larger radii and smaller cost weights are untested, so the planners may be under-tuned in that direction. Raycast information gain scores 37.3 acquisitions-to-90 against 41.4 for the runner-up ($r = 4$ m, $c = 0.3$), a margin of 4.2 acquisitions on the 75 rooms and 4.5 on the 69 rooms the earlier grid used. The copy planner scores 32.2 against 36.0 for its runner-up ($r = 8$ m, $c = 1$), margins of 3.8 and 4.1.

Table 23: Planner setting selection grids (Appendices E.4 and E.7). Top two blocks: 64 acquisitions, one start per room, seed 2000, in the reference recipe's environment (*reference env.*, 69 usable selection rooms) and in the main method's environment (*main env.*, 75 selection rooms; last column restricted to the 69 shared rooms). Acq. 90 = failure-penalised acquisitions-to-90 (censored at 65); success and final F_1 in %; objective = mean Acq. 90, then success, then final F_1; bold = selected. The main-environment columns report the sample-matched wall F_1 at 0.30 m. Both planners reselect the transferred setting ($r = 8$ m, $c = 0.3$, selected in the reference recipe's environment). Bottom block: the raycast grid at 128 acquisitions (1,024 decisions, three starts per room, 75 selection rooms; Acq. 90 censored at 129), objective = success, then Acq. 90, then final F_1, with mean senses, collisions and wall-clock time per point; the setting selected at 64 acquisitions is selected again.

Setting	Acq. 90 reference env.	Succ. reference env.	F_1 reference env.	Acq. 90 main env.	Succ. main env.	F_1 main env.	Acq. 90 main env., 69 rooms
Raycast IG planner (the transferred setting is reselected; 75 selection rooms, 69 shared with the reference grid)							
r = 4 m, c = 0.3	37.7	66.7	91.8	41.4	62.7	87.8	39.4
r = 4 m, c = 1.0	42.6	56.5	89.5	44.7	52.0	83.7	42.9
r = 4 m, c = 3.0	47.8	47.8	85.6	50.5	38.7	81.5	49.2
r = 8 m, c = 0.3	**32.1**	**81.2**	**94.4**	**37.3**	**61.3**	**88.5**	**34.8**
r = 8 m, c = 1.0	39.8	60.9	92.1	41.7	56.0	87.0	39.7
r = 8 m, c = 3.0	47.6	46.4	86.8	49.7	45.3	82.9	48.3
r = 8 m, c = 1.0, clearance 1.15 m	42.7	52.2	88.7	44.9	52.0	83.8	43.1
Imagined planner, copy (the transferred setting is reselected; 75 selection rooms, 69 shared with the reference grid)							
r = 4 m, c = 0.3	41.2	62.3	91.4	46.2	56.0	84.6	44.6
r = 4 m, c = 1.0	42.5	55.1	87.8	48.5	45.3	82.1	47.1
r = 4 m, c = 3.0	50.1	37.7	82.5	51.5	33.3	76.8	50.3
r = 8 m, c = 0.3	**29.8**	**79.7**	**94.7**	**32.2**	**73.3**	**89.4**	**29.3**
r = 8 m, c = 1.0	34.3	72.5	93.5	36.0	66.7	88.0	33.5
r = 8 m, c = 3.0	42.5	55.1	89.4	48.9	36.0	81.3	47.5
r = 8 m, c = 1.0, clearance 1.15 m	34.5	69.6	91.9	42.1	54.7	85.2	40.1

Raycast IG planner, 128 acquisitions, 3 starts, success first (the 64-acquisition setting is reselected; 75 selection rooms)

Setting	Succ.	Acq. 90	F_1	Senses	Collisions	Wall clock (h)
r = 4 m, c = 0.3	73.8	61.2	92.0	115.3	20.9	1.0
r = 4 m, c = 1.0	68.9	68.2	91.1	115.2	20.1	1.1
r = 4 m, c = 3.0	64.9	77.3	89.9	115.5	14.9	1.1
r = 8 m, c = 0.3	**76.0**	**54.7**	**93.4**	**115.4**	**23.0**	**2.0**
r = 8 m, c = 1.0	71.6	62.7	92.3	115.4	21.1	1.8
r = 8 m, c = 3.0	69.3	75.1	90.4	115.4	14.5	1.6
r = 8 m, c = 1.0, clearance 1.15 m	72.4	65.8	90.5	112.7	22.7	1.7

On the 69 shared rooms, with identical starts, the copy planner at the selected point differs from its reference-environment result by -0.5 acquisitions-to-90 and 0.0 success points. Raycast information gain needs 2.7 more acquisitions there and loses 14.5 success points. The mean senses are 61.5–62.2 at every point in both environments, so acquisition counts vary little and the performance differences reflect the selected trajectories (the 75-room columns include six added large rooms and are not a like-for-like comparison). Since neither setting changed, the planner rows of Table 1 are the evaluations at the transferred setting.

The raycast grid (the same seven points) was then re-run at the main method's own protocol: 128 acquisitions and 1,024 decisions with three starts per room on the 75 selection rooms (225 episodes and 1.0–2.0 h per point). It was scored on success first, then failure-penalised acquisitions-to-90 (censored at 129), then final F_1 (Table 23). It selects $r = 8$ m, $c = 0.3$ again, with 76.0 % success and 54.7 acquisitions-to-90 against 73.8 % and 61.2 for the runner-up ($r = 4$ m, $c = 0.3$). The 128-acquisition raycast row of Table 1 therefore uses the setting selected at 128 acquisitions, which is the same point; the imagined-observation planners keep the setting selected at 64 acquisitions.

Information contract. Table 24 states the shared environment and available-information contract and the main method's actor encoding. The ablation arms, planners and walkers have their own encodings, motion sets and STOP rules; sharing available information does not give each comparator the main actor's input features. The reference recipe also differs in the actor channels of Appendix E.1, the start-anchored canvas and the absence of collision memory.

Scope of comparisons. No external learned comparator is evaluated, as adaptation requires matched acoustics, motion and acquisition accounting. Without a rate-matched periodic-sensing controller, these comparisons assess the joint exploration strategy without isolating the contribution of adaptive timing.

Table 24: Information contract of the policy environment: what the actor receives at every decision and what is privileged.

Item	Statement
Acoustics	The cached observation (the frozen Stage 1–2 outputs of the node's heading-0 RIR) at every lattice node, written at the node's true pose when SENSE is taken within the 1 m snap radius; there is no line-of-sight test between the platform and the node (audit of the reference recipe's trained-STOP policy and the walkers: Table 20).
Pose	Exact position (float64) and the last move's heading; no pose noise.
Workspace	A room-specific square window supplied before mapping: its origin and side come from the bounding box of the admissible nodes (14 m margin) and the polygon corners (2.25 m margin), rounded up to a multiple of 8 m; the belief canvas is anchored at its centre, the actor's coarse agent stamp is placed relative to that anchor, its support scalars are normalised by the window area, and a move whose target leaves the window fails. A declared spatial prior, not a map.
Contacts	A failed move is reported as its outcome code and the contact point on the swept segment (the intersection with the nearest wall segment, or the closest point on it), snapped to a 0.125 m cell and stamped into the collision channel of G; the collision memory masks that (pose, move) pair for the rest of the episode.
Acquisition legality	SENSE is legal when the nearest admissible node is within 1 m and the acquisition cap is not reached; STOP is legal at or above the STOP floor; the admissible node set and the window are task metadata shared by every method. The actor receives no map of admissible nodes: beyond the window, the node set reaches it only through SENSE legality, the new or repeat outcome of a SENSE and the node poses at which acquisitions are registered; the planners and walkers use acquired positions and nearest-node checks.
Actor	Local (13 channels) and coarse (7 channels) belief windows, 65 scalars and 15 per-move features (Sec. 4.4).
Reward and critic	Wall recall, F_1 and the known reachable fraction against the true polygon during training (Table 21); the privileged critic reads the true class rasters and an unacquired-node stamp; evaluation scores the decoded belief against the same truth.

Planner comparisons and stopping controls. The main method's lead over the planners (Sec. 5.2) holds on success, efficiency and collisions. At 128 acquisitions it succeeds in 70.1 % [66.0, 74.2] of episodes with 5.6 collisions, against 15.1–22.3 collisions for the three planners with cardinal moves (Table 1). The belief-frontier walker (Yamauchi, 1997) reaches 21.2 % success collision-free and never stops; the random walker reaches 7.3 %, stopping once STOP becomes legal at its floor of 24. Paired over all 355 rooms, the main method needs fewer failure-penalised acquisitions-to-90 than each of the three planners with cardinal moves (main method minus planner: -13.2 [-15.8, -10.5] to -17.3 [-20.2, -14.4] at 128 acquisitions and -4.0 [-5.0, -3.0] to -6.1 [-7.4, -4.8] at 64 acquisitions). With STOP disabled, it retains a success advantage of $+11.0$ [8.5, 13.6] to $+15.1$ [12.4, 18.0] points (54.5 % success) while averaging 58.8 senses against 62.6–62.8 (Table 28). At 64 acquisitions it uses 21.6 fewer senses than raycast information gain, with 3.5 against 14.3 collisions (Table 1).

Matched motion primitives. This paragraph supports the motion control of Sec. 5.2. The planner rows above execute their routes as unit cardinal lattice moves (the walk-and-sense loop of Appendix E.4), whereas the policy chooses among 16 headings $\times$ {0.5, 1} m. To separate the move set from the decision rule, the raycast planner was also run with the policy's move set (*policy motion* in Table 1). Its decision (candidates, information score, goal and route) is then unchanged, and the route follower walks the route with the environment's full move table. At each decision it takes, among the legal moves not remembered as failed that shorten the straight distance to the current waypoint, the one that shortens it most. The waypoint cursor only advances, past route points within 0.25 m of the agent and on to the farthest contiguous route point within 2 m, so staircase routes are cut diagonally. A refused move enters collision memory, blocks the route edge and triggers a replan as before, and the fallback random move is drawn from the full table. The raycast planner was re-run this way at its selected setting on all 355 rooms at 128 and at 64 acquisitions (Table 25); the imagined-observation planners keep cardinal moves.

Table 25: The raycast information-gain planner with cardinal lattice moves and with the policy's move set (the same decisions, the routes walked with 16 headings $\times$ {0.5, 1} m) at the selected setting ($r = 8$ m, $c = 0.3$) on all 355 test rooms, 3 starts per room, seed 2000: success in %, mean senses, distance and collisions per episode. Δ rows are paired room bootstraps (10,000 resamples, 95 % CI) of the policy-move-set row minus the cardinal row and of the main method minus the policy-move-set row; an interval containing zero is no resolved difference.

Motion executor, all 355 test rooms	Succ. (%)	Senses	Distance (m)	Collisions
Raycast IG planner, 128 acquisitions, 1,024 decisions (the main method's own protocol)				
Cardinal lattice moves	49.4	124.3	429.6	22.3
Policy move set (16 headings $\times$ {0.5, 1} m)	49.5	124.4	350.7	32.1
Δ policy move set $-$ cardinal [CI]	+0.1 [-1.7, 1.9]	+0.1 [-0.2, 0.3]	-78.9 [-86.4, -71.3]	+9.9 [8.0, 11.9]
Δ EchoBelief (full) $-$ policy move set [CI]	+20.7 [16.2, 25.2]	-65.1 [-69.2, -61.0]	-2.8 [-32.0, 26.9]	-26.5 [-29.8, -23.4]
Raycast IG planner, 64 acquisitions, 512 decisions				
Cardinal lattice moves	39.3	62.6	229.7	14.3
Policy move set (16 headings $\times$ {0.5, 1} m)	38.5	62.9	188.7	17.8
Δ policy move set $-$ cardinal [CI]	-0.8 [-2.5, 0.8]	+0.4 [0.2, 0.6]	-40.9 [-44.5, -37.6]	+3.5 [2.5, 4.5]
Δ EchoBelief (full) $-$ policy move set [CI]	+10.9 [8.0, 13.9]	-22.0 [-24.3, -19.7]	+15.3 [6.3, 24.7]	-14.3 [-16.0, -12.6]

The policy's move set does not change the raycast planner's success resolvably and does not close its gap to the main method. With it, the planner reaches 49.5 % success at 128 acquisitions and 38.5 % at 64, against 49.4 and 39.3 % with cardinal moves ($+0.1$ [-1.7, 1.9] and -0.8 [-2.5, 0.8] points, no resolved difference). Its senses change by $+0.1$ [-0.2, 0.3] and $+0.4$ [0.2, 0.6], the latter a small resolved increase. It travels 350.7 m instead of 429.6 (-78.9 [-86.4, -71.3]) and 188.7 instead of 229.7 (-40.9 [-44.5, -37.6]) while colliding 32.1 instead of 22.3 times ($+9.9$ [8.0, 11.9]) and 17.8 instead of 14.3 ($+3.5$ [2.5, 4.5]). The oblique and shorter moves shorten the walked route, and the greedy waypoint following collides more often. The main method's success margin over the planner with the policy's move set is $+20.7$ [16.2, 25.2] points at 128 acquisitions and $+10.9$ [8.0, 13.9] at 64, so the planners' cardinal move set does not account for the gap. The corresponding differences are -65.1 [-69.2, -61.0] and -22.0 [-24.3, -19.7] senses and -26.5 [-29.8, -23.4] and -14.3 [-16.0, -12.6] collisions.

A stop rule for the planner. The planners never choose STOP, so part of their sensing gap to the policy could be a stopping effect; this paragraph supports the stopping control of Sec. 5.2. The planners' implementation includes a quiet-decision rule: it requests a stop once the best candidate score (utility minus path cost) stays below the threshold τ (Appendix D.9) for a number of consecutive decisions (the patience). With the rule enabled, the planner chooses STOP as soon as the STOP floor of 4 allows; below the floor it walks the decision's goal and asks again at the next decision.

Only $\tau \in \{0.25, 0.5, 1, 2, 4\}$ and patience $\in \{2, 4\}$ were tuned, for raycast information gain with the policy's move set at $r = 8$ m, $c = 0.3$, on the 75 selection rooms (three starts, 128 acquisitions). The selection rule was fixed before the grid ran: the fewest senses among the points within 2 success points of the same planner without the rule, otherwise the most successful point. No point came within 2 points of the no-rule baseline (76.9 %). Success falls with sensing, from 72.4 % at 103.9 senses ($\tau = 0.25$, patience 4, selected) through 52.4 % at 54.6 to 26.7 % at 11.3 (Table 26). On the 355 test rooms the selected rule stops 17.6 % of episodes and changes raycast information gain by -2.7 [-3.8, -1.7] success points and -8.2 [-9.8, -6.6] senses. The main method leads it by $+23.4$ [19.0, 27.8] points with 56.9 [52.8, 61.1] fewer senses and 22.1 [19.4, 24.9] fewer collisions.

Where it stops. This paragraph supports the learned-stopping part of the efficiency claim in Sec. 5.2. Fig. 15a bins the acquisitions and wall recall at STOP of every STOP-ended episode on the 246 usable test rooms. The trained-STOP policy stops in 89.2 % of episodes, from its floor of 24 upwards, with 71 % of those stops at or above 90 % recall. The main method stops in 89.6 % of its episodes at 128 acquisitions and 77.6 % at 64 acquisitions, with 83 % of the stops at or above the threshold under both protocols. The 64-trained runs stop between 5 and 63 acquisitions in 82.9 and 84.4 % of episodes, with 79 and 76 % of their stops at or above the threshold.

On all 355 rooms the main method ends 78.8 % of its episodes at 128 acquisitions by STOP, 10.3 % at the cap and 10.9 % at the decision limit, none for want of a legal action (at 64 acquisitions: 55.3 % by STOP, 39.1 % at the cap).

Table 26: Raycast information gain with its quiet-decision stop rule (128 acquisitions, 1,024 decisions, STOP floor 4, the policy's move set, $r = 8$ m, $c = 0.3$). Top: the selection grid over τ and patience on the 75 selection rooms, sorted by success; the no-rule baseline never stops. Bottom: the selected rule on the 355 test rooms against the same planner without it and the main method (EchoBelief (full)), with paired room-bootstrap differences (percentage points for success, recall and stop). Success bar = no-rule success minus 2 points (the selection rule); Stop = share of episodes ended by STOP; Quiet stops = number of episodes ended by the rule.

Raycast information gain, 128 acquisitions, STOP floor 4	Success (%)	Senses	Recall (%)	Collisions	Stop (%)	Quiet stops
Selection partition (75 rooms × 3 starts): only τ and patience vary; success bar 74.9 %						
No stop rule (baseline)	76.9	115.3	92.4	42.1	0.0	0
$\tau = 0.25$, patience 4 (selected)	72.4	103.9	91.0	32.2	25.8	58
$\tau = 0.25$, patience 2	68.4	84.6	89.5	15.4	70.2	158
$\tau = 0.5$, patience 4	68.4	86.6	88.8	26.2	51.1	115
$\tau = 0.5$, patience 2	61.3	71.0	86.5	14.4	77.8	175
$\tau = 1$, patience 4	52.4	54.6	82.4	19.0	78.7	177
$\tau = 1$, patience 2	44.4	42.1	80.1	12.6	92.4	208
$\tau = 2$, patience 4	38.7	25.6	77.1	12.0	99.1	223
$\tau = 2$, patience 2	34.2	19.0	74.1	10.9	100.0	225
$\tau = 4$, patience 4	31.1	14.7	72.6	10.5	100.0	225
$\tau = 4$, patience 2	26.7	11.3	69.5	9.8	100.0	225
Test split (355 rooms × 3 starts), room means						
Raycast, no stop rule (policy move set)	49.5	124.4	79.1	32.1	0.0	
Raycast with the selected stop rule	46.8	116.2	77.2	27.7	17.6	187
EchoBelief (full, seed 42)	70.1	59.3	90.7	5.6	78.8	
Paired room bootstrap (10,000 resamples) [95% CI]						
Stop rule − no stop rule	−2.7 [−3.8, −1.7]	−8.2 [−9.8, −6.6]	−1.9 [−2.5, −1.4]	−4.4 [−6.1, −2.9]	+17.6 [14.7, 20.6]	
EchoBelief (full) − stop rule	+23.4 [19.0, 27.8]	−56.9 [−61.1, −52.8]	+13.5 [11.6, 15.4]	−22.1 [−24.9, −19.4]	+61.2 [56.8, 65.6]	

Table 30 decomposes the failure-penalised score as $(1 - c)\, T_{\text{hit}} + c\,(\text{cap} + 1)$, where c is the fraction of episodes that never cross 90 % recall and T_{hit} the mean acquisitions to the crossing in the others. For the main method at 128 acquisitions on all 355 rooms, $c = 29.5$ % and $T_{\text{hit}} = 40.6$ acquisitions ($66.6 \approx 0.705 \times 40.6 + 0.295 \times 129$). Over the 292 rooms with a crossing, it takes 2.6 collisions and 220.1 m of travel up to the crossing, and 28 % of a crossing episode's travel comes after the crossing (room means).

Populations and tiers. This paragraph supports the room-size and population results of Sec. 5.2. Table 28 splits every row by test population. At 128 acquisitions the main method's success is highest on the sealed holdout and lowest on the rooms beyond the 52 m span cap, which only its larger window admits. On the 120 sealed-holdout rooms it scores 78.1 % [71.7, 83.9] success at 128 acquisitions, 68.1 % [60.3, 75.3] at 64 acquisitions and 76.9 % with STOP disabled, with 5.2, 3.2 and 4.3 collisions. The reference recipe loads 104 of these rooms. There its trained-STOP policy scores 73.7 % with 6.2 collisions, and its capped seeds 85.9 and 84.0 % with 12.5 and 23.0 collisions. Restricted to the same 104 rooms, the main method scores 77.6 % at 64 acquisitions and 87.8 % with STOP disabled, with 2.3 and 3.7 collisions. On the 122 validation rooms rooms the main method scores 75.1, 68.3 and 73.5 % under the three protocols. On the 35 added rooms within the span cap, where the trained-STOP policy succeeds in 1.9 % of episodes, it scores 65.7 % at 128 acquisitions and 20.0 % at 64. Beyond the cap it scores 52.1 % [42.7, 61.5] on the 78 added and 39.8 % [23.7, 54.8] on the 31 original rooms, at 103.7 and 108.7 senses and 7.3 and 16.7 collisions, so substantial failures remain there.

At 64 acquisitions the 64-trained runs gain over the trained-STOP policy mainly on the 35 added rooms (15.2 and 18.1 against 1.9 % success). On the 104 holdout rooms they are within a point of it (74.7 and 74.0 against 73.7 %) with fewer collisions (2.8 and 2.3 against 6.2), and on the 107 validation rooms rooms they are near it (77.6 and 73.8 against 74.1 %). Their recipe thus shows no resolved advantage on the rooms the trained-STOP policy already maps, but has fewer collisions there and succeeds more often on the large rooms.

Table 29 gives the node-count and difficulty tiers, and Fig. 15b plots success by node tier. At 128 acquisitions the main method succeeds in 97.9, 87.6, 75.8 and 52.9 % [46.5, 59.2] of episodes on the $<$100, 100–400, 400–600 and $\geq$600-node tiers (16, 129, 40 and 170 rooms). The reference checkpoint at floor 128 reaches 50.3 % [41.5, 59.6] on the 61 rooms of the largest tier it can load, and the main method succeeds on none of the 10 extreme rooms. At 64 acquisitions the 64-trained runs' gain over the trained-STOP policy concentrates in the $\geq$600-node tier (16.4 and 14.2 against 4.4 %) and on the 19 very complex rooms (22.8 and 24.6 against 1.8 %). In the $\geq$600-node tier they keep

Table 27: The main method restricted to the 246 test rooms the reference recipe loads (3 starts per room, seed 2000) against the reference checkpoint re-run at STOP floor 128 and, at 64 acquisitions, the trained-STOP policy and the two capped seeds (STOP disabled); the columns of Table 1, with Δ = restricted main method minus row in success points, paired over the 246 rooms (10,000 resamples, 95 % CI; an interval containing zero is no resolved difference). Interior IoU is reported for the main method's rows over the 246 rooms and dashed for the reference recipe's rows, whose interior IoU under the frozen pipeline's hole-blind truth is not compared with the hole-aware values (Appendix F); the main method's 64-acquisition rows run its checkpoint trained at 128 acquisitions with a 512-decision limit.

Policy	Succ.	F_1	IoU	Acq. 90	Senses	Coll.	Dist.	Stop	Δ succ. [CI]
The 246 rooms the reference recipe loads, 128 acquisitions; the main method restricted to them, Δ = restricted main method − row									
EchoBelief (full, seed 42, floor 4), 246 rooms	79.7	93.5	95.3	47.6	39.0	3.7	229.5	89.6	
Reference checkpoint, floor 128	82.8	95.4	–	46.9	126.3	56.0	347.0	0.0	−3.1 [−7.2, 0.8]
The 246 rooms the reference recipe loads, 64 acquisitions (capped rows: Δ against the restricted main method with STOP disabled)									
EchoBelief (full, seed 42, floor 4), 246 rooms	69.1	91.5	93.1	32.9	31.6	2.7	164.0	77.6	
Trained-STOP policy (floor 24)	63.7	88.1	–	33.8	31.3	6.4	112.2	89.2	+5.4 [2.2, 8.8]
EchoBelief (full), STOP disabled, 246 rooms	76.4	94.2	94.4	31.0	57.1	3.8	408.1	0.0	
Capped policy, seed 42	73.0	92.5	–	30.2	36.8	14.0	405.6	0.0	+3.4 [1.2, 5.6]
Capped policy, seed 43	69.6	90.5	–	30.9	35.5	19.5	392.2	0.0	+6.8 [4.2, 9.5]

sensing (seed 42: 57.5 against 50.2 senses) instead of stopping at the floor. Fig. 16 plots recall and F_1 against the acquisition count on all 355 rooms, the 246 usable rooms and the two largest tiers.

The 246-room comparison with the reference recipe. Table 27 pairs the main method, restricted to the 246 usable rooms, with the reference recipe's rows under the same protocol. At 128 acquisitions the main method shows no resolved success difference from the reference checkpoint at floor 128, with far fewer senses and collisions; at 64 acquisitions and with STOP disabled it has resolved success gains over the trained-STOP and capped policies. At 128 acquisitions it stops in 89.6 % of episodes after 39.0 senses with 3.7 collisions and reaches 79.7 % success (-3.1 [-7.2, 0.8] points against the reference checkpoint, which senses 126.3 times and collides 56.0 times). At 64 acquisitions it reaches 69.1 % success, $+5.4$ [2.2, 8.8] points over the trained-STOP policy, and with STOP disabled 76.4 %, $+3.4$ [1.2, 5.6] and $+6.8$ [4.2, 9.5] points over the capped seeds with 3.8 against 14.0 and 19.5 collisions. Paired against the trained-STOP policy, the 64-trained runs gain $+3.8$ [0.5, 7.0] (seed 42) and $+2.3$ [-0.9, 5.7] (seed 43, no resolved difference) success points, and seed 42 has half the collisions (-3.4 [-5.7, -1.5]).

Auxiliary stopping-head diagnostic. This diagnostic asks, on the selection rooms, whether stopping can be retrofitted to a frozen explorer. Stopping heads were fitted on 610,846 labelled decisions of the 64-trained seed-42 run at update 600 with STOP masked. They used either a "recall ≥ 0.9 now" label (the current-quality head) or a continuation label, the discounted task return over the next h acquisitions (the continuation head). Both were run with real termination on the 138 selection-evaluation episodes at floor 4. The current-quality head reaches 54.3 % success and the continuation head 68.1 % at threshold 0.9 and 73.2 % at 0.99, against 71.0 % for the explorer's own STOP; the continuation label improves success by $+13.8$ points over the current-quality label. This selection-set diagnostic did not establish an improvement over end-to-end stopping, which is used by the main method.

Attribution. The main method differs from the reference recipe's policies in belief, fusion, truth raster, anchor, actor, reward and stopping, so the 246-room comparison does not isolate any one change. The arms of Appendix E.12 test specified interventions on the main method's recipe, including the bundled reference-recipe arm, and the legacy-fusion arm (Table 2) is the belief-package control.

E.8 Live acoustic simulation instead of the cached lattice

This subsection asks whether sensing from the cached lattice, rather than simulating each RIR at the true pose, changes the results; Table 31 details the live comparison summarised in Sec. 5.2. Live sensing shows no resolved difference in success, recall or failure-penalised acquisitions-to-90, a small resolved loss of wall F_1, and fewer senses, less travel and fewer collisions. The comparison runs the main method (its frozen seed-42 checkpoint of Appendix E.7, update 3,600, STOP floor 4,

Table 28: Main method and its seed replicates, ablation arms, planners, walkers and reference policies by test population (room means; success and F_1 in %, acquisitions to 90 % failure-penalised, senses, collisions; 3 starts per room, seed 2000; per-group intervals are 1,000-resample room bootstraps and the text quotes them where used). Checkpoints were frozen before the 120-room holdout was read (Appendix E.7). The holdout and validation rooms columns use the full 120 and 122 rooms for all-rooms rows (main method and its replicates, ablation arms, planners and walkers), and the restricted 104 and 107 rooms for reference-recipe and 64-trained rows; the column headers' restricted counts apply only to the latter rows. Seed-replicate and ablation rows are those of Table 2 at 128 acquisitions; planner rows run under the main method's configuration. [c]No earlier policy can load the 109 rooms whose node span exceeds 52 m; the planners and walkers under the all-rooms configuration are the only comparators there.

		Sealed holdout (104 of 120)					Validation rooms (107 of 122)					Added, span ≤ 52 m (35)					Added, span > 52 m[c] (78)					Original, span > 52 m[c] (31)					All 355 rooms				
Policy	Rooms	Succ.	F1	Acq.	Senses	Coll.	Succ.	F1	Acq.	Senses	Coll.	Succ.	F1	Acq.	Senses	Coll.	Succ.	F1	Acq.	Senses	Coll.	Succ.	F1	Acq.	Senses	Coll.	Succ.	F1	Acq.	Senses	Coll.
STOP on, 64-acquisition cap																														
Trained-STOP policy (floor 24)	246	73.7	91.4	29.2	27.7	6.2	74.1	90.6	28.0	27.9	6.6	1.9	70.6	65.0	52.3	6.3	–	–	–	–	–	–	–	–	–	–	–	–	–	–
64-trained run, seed 42, floor 4	246	74.7	91.6	28.4	26.4	2.8	77.6	91.8	26.3	25.9	2.6	15.2	82.1	64.0	58.8	4.8	–	–	–	–	–	–	–	–	–	–	–	–	–	–
64-trained run, seed 43, floor 4	246	74.0	92.0	28.6	24.5	2.3	73.8	91.2	27.5	24.3	2.5	18.1	82.6	63.6	58.2	4.4	–	–	–	–	–	–	–	–	–	–	–	–	–	–
EchoBelief (full), floor 4	355	68.1	88.7	33.0	32.0	3.2	68.3	88.5	31.7	30.5	3.4	20.0	86.1	63.5	60.0	3.9	4.3	71.8	64.8	62.6	4.0	6.5	62.2	62.9	60.8	8.9	49.4	84.6	42.5	40.9	3.5
STOP disabled, 64-acquisition cap																														
Capped policy, seed 42	246	85.9	94.9	24.1	33.0	12.5	81.0	93.3	24.7	32.4	13.8	10.5	82.6	64.7	61.6	18.9	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Capped policy, seed 43	246	84.0	93.8	24.9	31.6	23.0	77.9	92.5	25.6	31.0	15.5	1.9	75.0	64.9	61.2	21.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–
64-trained run, seed 42	246	84.6	95.1	27.0	63.0	3.8	83.2	94.4	25.3	63.0	3.6	10.5	82.0	64.3	63.0	4.5	–	–	–	–	–	–	–	–	–	–	–	–	–	–
64-trained run, seed 43	246	85.3	95.7	26.7	63.0	3.0	83.8	94.6	25.5	63.0	3.2	13.3	83.0	64.1	63.0	3.6	–	–	–	–	–	–	–	–	–	–	–	–	–	–
EchoBelief (full)	355	76.9	91.3	30.6	57.0	4.3	73.5	91.1	30.2	56.9	4.4	22.9	85.9	63.4	63.0	3.8	4.3	70.8	64.8	62.8	3.7	6.5	61.4	62.9	62.3	8.0	54.5	86.2	41.2	58.8	4.2
Belief-frontier walker, all-rooms configuration	355	15.8	63.2	60.0	59.1	0.0	21.6	66.8	58.6	59.9	0.0	0.0	34.3	65.0	59.3	0.0	0.0	25.6	65.0	62.0	0.0	0.0	24.2	65.0	61.5	0.0	12.8	53.3	61.1	60.0	0.0
Random walker, all-rooms configuration	355	4.4	46.0	62.6	15.7	48.1	6.3	49.4	61.7	15.3	47.0	0.0	21.0	65.0	15.2	49.8	0.0	14.7	65.0	15.5	52.8	0.0	14.3	65.0	15.5	61.5	3.7	37.8	63.0	15.5	48.9
Raycast IG planner, all-rooms configuration	355	56.9	84.1	41.9	62.6	14.7	58.2	84.8	40.1	62.6	13.1	1.0	65.8	65.0	62.9	15.1	0.0	52.0	65.0	62.3	15.2	1.1	47.4	64.7	60.7	31.8	39.3	75.5	48.6	62.6	14.3
Imagined planner, copy, all-rooms configuration	355	64.2	83.9	38.9	62.9	9.3	63.4	83.6	37.0	62.8	8.4	0.0	55.6	65.0	62.9	8.1	0.0	43.6	65.0	62.8	9.1	3.2	36.5	64.2	62.7	14.1	43.5	72.1	46.5	62.8	8.8
Imagined planner, read, all-rooms configuration	355	60.8	85.8	39.8	62.5	12.6	64.5	86.6	36.7	62.8	11.4	1.0	68.6	65.0	62.6	11.8	0.0	58.8	65.0	62.5	13.3	4.3	52.0	63.6	60.8	31.8	42.8	78.4	46.7	62.6	12.3
128-acquisition cap, 1,024 decisions																														
Reference checkpoint, STOP floor 128	246	88.5	96.8	35.0	126.4	55.9	86.0	96.0	37.0	126.4	49.7	56.2	89.2	112.2	125.5	75.9	–	–	–	–	–	–	–	–	–	–	–	–	–	–
EchoBelief (full), floor 4	355	78.1	92.2	47.7	43.4	5.2	75.1	92.0	50.8	41.9	5.2	65.7	90.3	92.2	75.7	4.6	52.1	87.2	109.0	103.7	7.3	39.8	79.3	111.3	108.7	16.7	70.1	90.9	66.6	59.3	5.6
EchoBelief (full), seed 43, floor 4	355	81.4	92.9	46.0	48.7	6.5	78.4	92.7	46.8	46.3	6.3	61.9	90.2	96.9	80.7	5.8	51.3	86.3	109.6	105.7	9.7	36.6	78.9	111.8	108.8	22.8	71.8	91.1	65.3	63.5	7.1
EchoBelief (full), seed 44, floor 4	355	77.5	92.3	48.5	41.7	4.4	75.7	92.4	48.2	40.5	4.4	66.7	90.4	90.0	73.2	4.7	53.4	88.0	109.3	103.0	7.2	39.8	79.4	111.0	108.3	16.2	70.5	91.2	65.9	57.9	5.1
Ablation: legacy belief, floor 4	355	73.3	91.3	51.7	40.2	5.6	72.1	91.2	51.7	39.4	6.2	73.3	92.6	92.9	90.8	7.6	49.1	86.4	111.1	108.2	8.6	30.1	76.4	114.8	112.5	15.8	67.6	90.3	68.8	59.8	6.7
Ablation: legacy belief, seed 43, floor 4	355	73.1	91.5	53.3	46.0	5.5	71.6	91.6	52.5	47.9	6.2	58.1	91.2	101.8	91.9	4.9	38.9	84.6	115.5	109.2	6.5	23.7	75.9	118.3	113.6	15.4	63.6	90.0	71.5	65.1	5.9
Ablation: reference recipe, floor 24	355	68.6	89.1	53.2	45.5	22.7	71.9	90.2	50.3	43.8	18.1	44.8	84.4	102.7	73.2	27.0	25.2	78.7	119.3	91.6	18.3	18.3	71.4	120.5	101.2	51.4	57.8	86.7	71.6	57.8	20.6
Ablation: no predictor, floor 4	355	77.8	91.6	49.4	42.5	5.2	75.7	91.7	48.0	41.6	4.9	56.2	89.6	97.3	78.0	5.0	50.0	86.6	111.0	108.3	8.1	43.0	78.0	110.0	112.1	15.5	68.8	90.3	67.2	60.2	5.7
Ablation: no collision memory, floor 4	355	73.9	91.2	54.0	47.4	4.8	71.0	91.0	55.1	43.4	11.4	68.6	89.9	96.1	89.0	5.8	47.4	86.5	113.2	111.9	6.8	37.6	76.6	113.5	112.8	13.4	66.6	89.9	71.5	64.3	7.6
Ablation: without memory pathway (belief + policy retrained), floor 4	355	63.9	85.1	67.3	53.2	5.2	60.9	85.1	68.0	51.8	5.3	30.5	86.5	121.5	111.1	9.0	8.1	75.9	128.0	122.3	9.3	9.7	68.3	125.3	121.9	16.1	47.3	83.2	86.2	73.6	6.5
Ablation: without memory pathway, seed 43, floor 4	355	65.6	84.9	64.8	53.7	5.6	63.7	85.7	63.7	52.4	5.1	23.8	86.3	123.8	108.9	9.0	8.1	75.4	127.8	121.4	10.2	5.4	67.2	125.2	121.0	17.4	48.2	83.2	84.1	73.6	6.8
Belief-frontier walker, all-rooms configuration	355	31.4	73.2	107.7	103.4	0.0	30.9	75.0	106.2	102.0	0.0	0.0	50.4	129.0	116.3	0.0	0.0	39.4	129.0	123.2	0.0	0.0	36.5	129.0	122.3	0.0	21.2	64.1	114.0	108.5	0.0
Random walker, all-rooms configuration	355	10.0	52.3	117.5	23.7	75.9	11.5	55.4	115.9	23.7	72.4	0.0	25.2	129.0	23.9	75.7	0.0	17.5	129.0	23.9	83.2	0.0	16.6	129.0	23.6	97.0	7.3	43.1	120.6	23.7	76.3
Raycast IG planner, all-rooms configuration	355	70.6	90.8	64.8	123.3	24.0	69.7	90.9	62.4	122.9	22.2	11.4	80.3	125.9	127.0	19.0	2.1	64.1	128.5	126.9	21.2	2.2	64.2	127.6	125.2	52.9	49.4	83.9	84.0	124.3	22.3
Imagined planner, copy, all-rooms configuration	355	75.0	89.0	56.7	123.3	16.0	71.0	88.5	57.3	123.1	14.6	1.0	71.9	128.9	126.7	12.6	0.9	59.2	128.9	127.0	15.5	5.4	49.9	125.4	126.5	23.9	50.0	80.6	79.9	124.4	15.1
Imagined planner, read, all-rooms configuration	355	74.2	91.1	59.3	123.4	20.6	72.7	91.0	56.1	123.3	17.9	10.5	80.3	125.8	127.0	14.7	4.3	70.6	127.8	126.8	19.6	5.4	64.3	124.7	126.5	45.8	52.0	85.5	79.8	124.4	18.9

Table 29: Success, wall F_1 (%) and senses by lattice-node tier and by the corpus manifest's difficulty tier (room means; 3 starts per room, seed 2000; the text's per-tier intervals are 1,000-resample room bootstraps). Tier room counts are given for the reference recipe's configuration (246 rooms) and for all 355; the extreme tier lies entirely beyond the 52 m span cap, so only the all-rooms rows populate it. Seed-replicate and ablation rows are those of Table 2 at 128 acquisitions; planner rows run under the main method's configuration.

Node-count tiers (rooms: 246-room configuration / all 355)

		<100 nodes (16 / 16)			100–400 nodes (129 / 129)			400–600 nodes (40 / 40)			≥600 nodes (61 / 170)		
Policy	Rooms	Succ.	F1	Senses	Succ.	F1	Senses	Succ.	F1	Senses	Succ.	F1	Senses
STOP on, 64-acquisition cap													
Trained-STOP policy (floor 24)	246	100.0	98.5	23.1	84.5	95.0	24.8	72.5	93.5	26.6	4.4	67.2	50.2
64-trained run, seed 42, floor 4	246	100.0	96.0	10.0	85.0	94.6	22.3	75.8	92.5	25.7	16.4	78.5	57.5
64-trained run, seed 43, floor 4	246	100.0	96.5	9.6	85.5	94.7	20.6	68.3	91.9	23.8	14.2	78.3	56.1
EchoBelief (full), floor 4	355	97.9	94.6	8.2	86.0	94.9	23.3	76.7	93.7	26.2	10.6	73.8	60.9
STOP disabled, 64-acquisition cap													
Capped policy, seed 42	246	100.0	98.7	14.5	90.4	96.8	28.6	95.0	97.4	36.7	14.8	78.4	60.1
Capped policy, seed 43	246	100.0	98.1	13.6	89.9	96.5	27.2	85.0	96.3	34.5	8.7	72.2	59.6
64-trained run, seed 42	246	100.0	99.6	63.0	93.3	97.7	63.0	88.3	96.1	62.9	14.8	78.9	63.0
64-trained run, seed 43	246	100.0	99.5	63.0	94.6	98.2	63.0	89.2	97.4	63.0	15.3	79.0	62.9
EchoBelief (full)	355	100.0	99.6	59.3	92.5	97.9	54.3	93.3	97.8	57.0	12.2	73.2	62.7
Belief-frontier walker, all-rooms configuration	355	95.8	98.3	53.6	23.3	78.2	59.6	0.0	60.9	59.4	0.0	28.4	61.2
Random walker, all-rooms configuration	355	43.8	84.2	16.6	4.7	57.8	15.5	0.0	44.3	15.0	0.0	16.7	15.5
Raycast IG planner, all-rooms configuration	355	100.0	99.5	63.0	76.2	94.3	63.0	59.2	89.4	63.0	1.0	55.7	62.1
Imagined planner, copy, all-rooms configuration	355	100.0	99.8	63.0	84.0	96.3	63.0	67.5	93.3	62.8	1.8	46.2	62.7
Imagined planner, read, all-rooms configuration	355	100.0	99.7	63.0	81.4	95.4	63.0	65.0	92.2	62.8	2.9	60.4	62.2
128-acquisition cap, 1,024 decisions													
Reference checkpoint, STOP floor 128	246	100.0	98.8	127.0	93.0	98.4	126.8	92.5	98.5	125.7	50.3	85.9	125.2
EchoBelief (full), floor 4	355	97.9	94.6	8.8	87.6	95.2	27.1	75.8	93.3	27.6	52.9	86.6	95.9
EchoBelief (full), seed 43, floor 4	355	100.0	95.7	9.7	91.0	96.0	32.2	88.3	95.3	36.8	50.8	86.1	98.7
EchoBelief (full), seed 44, floor 4	355	97.9	96.0	9.9	86.6	95.2	23.7	80.8	94.3	30.0	53.3	87.0	94.9
Ablation: legacy belief, floor 4	355	81.2	93.2	4.8	82.9	94.5	20.6	78.3	94.0	23.8	52.2	86.1	103.3
Ablation: legacy belief, seed 43, floor 4	355	91.7	94.3	6.1	85.0	95.3	31.8	72.5	93.6	28.0	42.5	84.7	104.5
Ablation: reference recipe, floor 24	355	100.0	98.6	23.0	85.0	93.7	28.3	75.8	92.6	36.5	29.0	78.9	88.4
Ablation: no predictor, floor 4	355	95.8	94.7	6.2	85.5	94.6	23.9	84.2	95.0	32.1	50.0	85.6	99.3
Ablation: no collision memory, floor 4	355	95.8	95.3	8.4	84.5	94.3	29.0	67.5	93.0	30.5	50.0	85.4	104.3
Ablation: without memory pathway (belief + policy retrained), floor 4	355	93.8	91.3	11.9	76.7	88.4	31.8	67.5	87.5	45.4	15.9	77.5	117.9
Ablation: without memory pathway, seed 43, floor 4	355	97.9	91.7	12.0	82.9	89.1	32.8	70.0	87.8	46.8	12.0	76.9	116.6
Belief-frontier walker, all-rooms configuration	355	93.8	98.1	67.8	44.2	85.6	99.1	8.3	72.9	102.8	0.0	42.6	120.9
Random walker, all-rooms configuration	355	70.8	90.8	23.5	11.4	65.9	23.7	0.0	49.2	23.7	0.0	19.8	23.8
Raycast IG planner, all-rooms configuration	355	100.0	99.3	86.3	90.7	97.6	125.2	78.3	95.3	127.0	6.5	69.4	126.6
Imagined planner, copy, all-rooms configuration	355	100.0	99.7	86.2	92.5	98.3	125.1	87.5	97.8	127.0	4.3	61.3	126.8
Imagined planner, read, all-rooms configuration	355	100.0	99.6	86.2	92.5	97.8	125.2	88.3	97.4	127.0	8.2	72.0	126.8

Difficulty tiers of the corpus manifest (rooms: 246-room configuration / all 355)

		complex (227 / 319)			very complex (19 / 26)			extreme (0 / 10)		
Policy	Rooms	Succ.	F1	Senses	Succ.	F1	Senses	Succ.	F1	Senses
STOP on, 64-acquisition cap										
Trained-STOP policy (floor 24)	246	68.9	90.2	30.3	1.8	63.5	42.7	–	–	–
64-trained run, seed 42, floor 4	246	71.2	91.7	29.0	22.8	74.0	51.8	–	–	–
64-trained run, seed 43, floor 4	246	69.5	91.7	27.3	24.6	74.0	51.8	–	–	–
EchoBelief (full), floor 4	355	53.8	87.6	38.9	14.1	66.9	57.6	0.0	36.6	63.0
STOP disabled, 64-acquisition cap										
Capped policy, seed 42	246	78.3	94.3	35.2	10.5	70.8	55.9	–	–	–
Capped policy, seed 43	246	75.0	92.5	34.2	5.3	66.8	51.6	–	–	–
64-trained run, seed 42	246	76.8	94.4	63.0	33.3	75.0	63.0	–	–	–
64-trained run, seed 43	246	78.0	95.0	63.0	31.6	74.2	63.0	–	–	–
EchoBelief (full)	355	59.4	89.4	58.6	15.4	66.2	59.9	0.0	36.0	63.0
Belief-frontier walker, all-rooms configuration	355	14.2	55.8	59.8	0.0	37.6	62.0	0.0	14.5	62.4
Random walker, all-rooms configuration	355	4.1	39.9	15.5	0.0	24.1	15.7	0.0	8.0	15.0
Raycast IG planner, all-rooms configuration	355	42.6	77.7	62.8	14.1	63.9	60.7	0.0	33.8	58.4
Imagined planner, copy, all-rooms configuration	355	47.9	75.4	62.9	6.4	52.6	62.6	0.0	19.4	62.8
Imagined planner, read, all-rooms configuration	355	47.4	81.4	62.7	2.6	59.3	62.1	0.0	35.2	59.5
128-acquisition cap, 1,024 decisions										
Reference checkpoint, STOP floor 128	246	88.0	96.9	126.3	21.1	77.2	125.6	–	–	–
EchoBelief (full), floor 4	355	76.4	92.7	53.8	20.5	80.8	100.8	0.0	58.2	127.0
EchoBelief (full), seed 43, floor 4	355	78.6	93.1	58.5	16.7	79.3	101.4	0.0	58.9	125.6
EchoBelief (full), seed 44, floor 4	355	77.4	93.2	52.8	12.8	79.7	93.0	0.0	57.9	127.0
Ablation: legacy belief, floor 4	355	74.1	92.6	54.6	14.1	76.9	98.0	0.0	52.6	126.6
Ablation: legacy belief, seed 43, floor 4	355	69.6	92.3	59.4	14.1	76.6	110.7	0.0	52.3	126.6
Ablation: reference recipe, floor 24	355	64.3	89.2	53.7	1.3	70.7	82.5	0.0	49.8	123.4
Ablation: no predictor, floor 4	355	75.2	92.5	55.3	16.7	77.3	94.4	0.0	55.0	127.0
Ablation: no collision memory, floor 4	355	72.5	92.3	59.4	19.2	76.0	100.2	0.0	52.8	127.0
Ablation: without memory pathway (belief + policy retrained), floor 4	355	52.5	85.5	70.1	2.6	70.1	96.6	0.0	44.7	127.0
Ablation: without memory pathway, seed 43, floor 4	355	53.3	85.4	69.9	3.8	70.8	97.7	0.0	44.0	126.9
Belief-frontier walker, all-rooms configuration	355	23.3	66.3	107.0	3.8	53.2	121.0	0.0	24.7	126.4
Random walker, all-rooms configuration	355	8.2	45.3	23.7	0.0	28.5	23.8	0.0	9.9	23.6
Raycast IG planner, all-rooms configuration	355	53.2	85.6	124.2	21.8	75.5	126.2	0.0	52.1	123.0
Imagined planner, copy, all-rooms configuration	355	54.1	83.6	124.1	19.2	63.4	126.4	0.0	30.5	127.0
Imagined planner, read, all-rooms configuration	355	56.4	88.0	124.2	17.9	69.7	126.8	0.0	47.8	125.0

Table 30: Resources up to the first crossing of 90 % wall recall: the first-crossing columns average each room's crossing starts, then the rooms with a crossing (95 % room-bootstrap CI). In the failure-penalised decomposition, T_{hit} instead averages all crossing episodes equally and c is the fraction of all episodes never crossing, so the first-crossing acquisition column and T_{hit} use different averaging conventions. The score (failure-penalised acquisitions-to-90) is $(1-c)\,T_{\text{hit}} + c\,(\text{cap}+1)$, for the main method, its seed replicates and ablation arms and the 64-trained runs of its recipe beside the reference recipe's trained-STOP and capped seed-42 policies (246 rooms $\times$ 3 starts; the all-rooms rows over their 355). The 64-trained runs' advantage over the trained-STOP policy is the failure term (c 32.1 and 34.0 against 36.3 %), not the hit time; with STOP disabled they reach the crossing later than the capped policy with half its collisions and less travel up to it.

Policy	Crossed (%) [95% CI]	Rooms with a crossing	Acq. to 90%	Decisions to 90%	Distance to 90% (m)	Collisions to 90%	Success (%)	STOP / cap / decision limit (%)	Censored c (%)	T_{hit}	$(1-c)\,T_{\text{hit}}$	$c\,(\text{cap}+1)$	Score (acq.)
64-acquisition cap, 512 decisions; first-crossing quantities over the episodes that crossed 90% recall, the decomposition over all episodes: score $= (1-c)\,T_{\text{hit}} + c\,(\text{cap}+1)$													
Trained-STOP policy (floor 24)	63.7 [58.1, 69.2]	177	17.6	85.1	66.7	1.6	63.7	89.2 / 10.4 / 0.4	36.3	16.0	10.2	23.6	33.8
Capped policy, seed 42	73.0 [67.8, 78.2]	193	19.4	121.4	99.6	2.7	73.0	0.0 / 21.3 / 78.7	27.0	17.3	12.6	17.5	30.2
64-trained run, seed 42, floor 4	67.9 [62.7, 72.9]	193	19.3	97.4	77.6	1.4	67.5	82.9 / 17.1 / 0.0	32.1	17.2	11.7	20.9	32.6
64-trained run, seed 43, floor 4	66.0 [61.0, 71.1]	193	19.3	102.9	83.1	1.4	66.0	84.4 / 14.4 / 1.2	34.0	16.6	11.0	22.1	33.1
EchoBelief (full), floor 4 (355 rooms)	49.4 [44.6, 54.1]	208	22.6	124.5	101.3	1.4	49.4	55.3 / 39.1 / 5.6	50.6	19.5	9.6	32.9	42.5
64-trained run, seed 42, STOP disabled	73.4 [68.2, 78.6]	193	20.9	99.2	77.7	1.4	73.4	0.0 / 99.7 / 0.3	26.6	19.5	14.3	17.3	31.6
64-trained run, seed 43, STOP disabled	74.4 [69.4, 79.4]	199	22.4	107.1	84.4	1.2	74.4	0.0 / 99.6 / 0.4	25.6	20.0	14.9	16.6	31.5
EchoBelief (full), STOP disabled (355 rooms)	54.5 [49.6, 59.4]	210	23.8	122.1	97.8	1.3	54.5	0.0 / 51.1 / 48.9	45.5	21.3	11.6	29.6	41.2
EchoBelief (full), seed 43, floor 4 (355 rooms)	50.2 [45.3, 55.2]	197	22.0	115.1	92.4	1.5	50.1	50.7 / 44.1 / 5.2	49.8	19.8	9.9	32.3	42.3
EchoBelief (full), seed 44, floor 4 (355 rooms)	51.3 [46.4, 56.2]	210	22.9	121.6	98.3	1.2	51.2	57.4 / 40.8 / 1.9	48.7	20.1	10.3	31.7	42.0
Ablation: legacy belief, floor 4 (355 rooms)	48.0 [43.3, 52.8]	198	19.8	120.9	100.0	1.9	47.7	51.9 / 42.8 / 5.3	52.0	17.9	8.6	33.8	42.4
Ablation: legacy belief, seed 43, floor 4 (355 rooms)	47.4 [42.7, 52.2]	195	20.4	112.3	90.7	2.0	47.0	52.0 / 46.8 / 1.2	52.6	18.3	8.7	34.2	42.8
Ablation: reference recipe, floor 24 (355 rooms)	48.4 [43.6, 53.2]	197	20.5	117.6	94.6	3.5	48.3	61.1 / 38.3 / 0.6	51.6	17.6	8.5	33.6	42.1
Ablation: no predictor, floor 4 (355 rooms)	49.1 [44.4, 53.9]	204	21.6	124.9	102.5	1.7	48.9	55.6 / 39.8 / 4.6	50.9	19.2	9.4	33.1	42.5
Ablation: no collision memory, floor 4 (355 rooms)	47.7 [42.9, 52.6]	195	21.6	110.4	88.1	1.5	47.7	53.7 / 45.5 / 0.8	52.3	19.6	9.3	34.0	43.3
Ablation: without memory pathway (belief + policy retrained), floor 4 (355 rooms)	40.4 [35.8, 45.1]	171	28.2	121.0	92.2	1.5	39.8	51.0 / 48.0 / 1.0	59.6	27.0	10.9	38.8	49.6
Ablation: without memory pathway, seed 43, floor 4 (355 rooms)	42.1 [37.3, 46.9]	175	29.2	129.1	99.1	1.8	41.5	50.5 / 46.7 / 2.8	57.9	27.3	11.5	37.7	49.2
128-acquisition cap, 1,024 decisions; first-crossing quantities over the episodes that crossed 90% recall, the decomposition over all episodes: score $= (1-c)\,T_{\text{hit}} + c\,(\text{cap}+1)$													
EchoBelief (full), floor 4 (355 rooms)	70.5 [66.4, 74.5]	292	44.0	265.9	220.1	2.6	70.1	78.8 / 10.3 / 10.9	29.5	40.6	28.6	38.0	66.6
EchoBelief (full), seed 43, floor 4 (355 rooms)	71.9 [67.9, 75.9]	291	44.7	267.0	219.7	3.2	71.8	73.6 / 8.7 / 17.7	28.1	40.4	29.1	36.2	65.3
EchoBelief (full), seed 44, floor 4 (355 rooms)	71.0 [66.9, 74.9]	292	43.1	244.6	199.8	2.5	70.5	84.0 / 9.0 / 6.9	29.0	40.1	28.4	37.4	65.9
Ablation: legacy belief, floor 4 (355 rooms)	67.8 [63.8, 71.8]	287	43.4	275.2	229.0	3.5	67.6	67.9 / 13.6 / 18.5	32.2	40.2	27.3	41.5	68.8
Ablation: legacy belief, seed 43, floor 4 (355 rooms)	64.0 [59.8, 68.2]	275	43.8	248.5	202.4	3.0	63.6	74.7 / 20.1 / 5.2	36.0	39.2	25.1	46.4	71.5
Ablation: reference recipe, floor 24 (355 rooms)	57.8 [53.3, 62.3]	258	37.2	220.7	179.2	5.1	57.8	87.5 / 11.9 / 0.6	42.2	29.8	17.2	54.4	71.6
Ablation: no predictor, floor 4 (355 rooms)	69.2 [65.2, 73.1]	291	43.3	265.2	219.6	2.9	68.8	77.8 / 11.5 / 10.7	30.8	39.6	27.4	39.7	67.2
Ablation: no collision memory, floor 4 (355 rooms)	66.7 [62.5, 70.8]	282	46.3	247.8	199.4	2.7	66.6	76.2 / 18.1 / 5.6	33.3	42.8	28.5	43.0	71.5
Ablation: without memory pathway (belief + policy retrained), floor 4 (355 rooms)	48.1 [43.7, 52.5]	221	45.9	222.2	174.1	3.0	47.3	64.9 / 29.4 / 5.7	51.9	40.0	19.2	67.0	86.2
Ablation: without memory pathway, seed 43, floor 4 (355 rooms)	48.9 [44.2, 53.6]	212	42.9	200.2	155.3	2.8	48.2	69.1 / 25.4 / 5.5	51.1	37.2	18.2	65.9	84.1

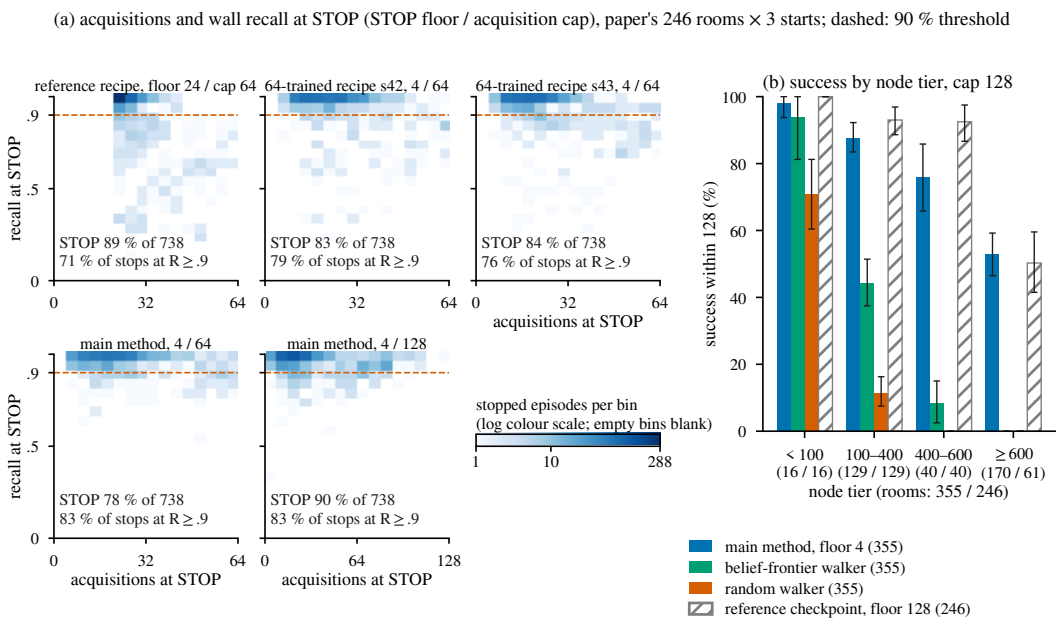

Figure 15: (a) Where the policies stop: acquisitions and wall recall at STOP on the reference recipe's 246 rooms (3 starts, seed 2000; STOP-ended episodes only, log colour scale). The trained-STOP policy stops when its floor of 24 allows; the 64-trained runs at floor 4 stop between 5 and 63 acquisitions (seed 42: 6–63) and the main method at floor 4 across its 64- or 128-acquisition budget, 76–83 % of their stops at or above the 90 % threshold. (b) Success within the 128-acquisition cap by node tier for the main method and the two walkers under its configuration on all 355 rooms, with the reference checkpoint at STOP floor 128 on its 246 (bars, room means with 95 % room-bootstrap CIs, 1,000 resamples).

128 acquisitions) under the paper's cached-lattice sensing contract and under a live protocol, with one start per room (seed 2000) on all 355 test rooms.

Live rendering uses the corpus's acoustic configuration with a 0.45 s ray horizon, compared with 3 s for the cached corpus. The room impulse response is simulated at the robot's true position (heading 0 as in the corpus, 12-microphone ring), encoded by the frozen polar echo encoder (Stage 1) and local geometry mapper (Stage 2) and written into the belief at that position on the 0.25 m memory grid (at most 0.177 m of rounding). SENSE is legal anywhere, and a repeat within 0.5 m of an earlier sensing position counts as a repeat sense. Live mode therefore changes both how observations are generated and the freedom to sense away from the nodes. The paired rows are live minus cached on identical (room, start) episodes with rooms as resampling units (10,000 draws); acquisitions to 90 % recall are failure-penalised (129 for episodes that never cross).

The cached side is a separate one-start pass of the same checkpoint, not the 3-start rows of Table 1. On it the main method succeeds in 70.4 % of the rooms with final F_1 90.8 % and recall 90.6 % after 60.3 senses, 362.5 m and 5.6 collisions per episode (66.4 failure-penalised acquisitions to 90 %). Live, on the same episodes, it succeeds in 69.0 % with final F_1 90.1 % and recall 89.9 % after 57.1 senses, 327.8 m and 5.0 collisions (66.7 acquisitions to 90 %).

Only 5 of the 355 live episodes keep the cached episode's decision and sense counts (4 of them also its outcome; 298 episodes keep the outcome), so the live observations do alter the trajectories. The paired differences in success (-1.4 points $[-5.4, 2.8]$), recall (-0.68 $[-1.43, 0.05]$) and acquisitions to 90 % ($+0.4$ $[-2.7, 3.5]$) all include zero, so there is no resolved difference. Final F_1 shows a resolved change of -0.7 $[-1.2, -0.3]$, and the costs fall: -3.2 $[-4.4, -2.2]$ senses, -34.6 $[-51.0, -19.3]$ m and -0.7 $[-1.1, -0.2]$ collisions per episode.

No node tier resolves a success difference ($+0.00$ $[-18.75, 18.75]$ on the 16 rooms under 100 nodes, -2.3 $[-8.5, 3.9]$ on the 129 rooms with 100–400 nodes, $+5.0$ $[-7.5, 20.0]$ on the 40 rooms with 400–600 nodes and -2.4 $[-8.8, 4.1]$ on the 170 rooms with at least 600 nodes). The saving in

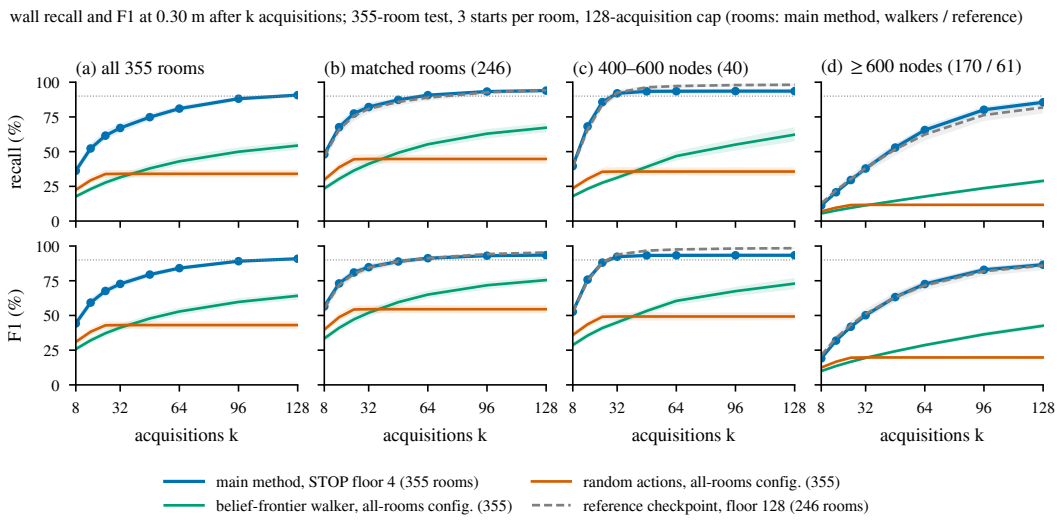

Figure 16: Wall recall (top) and F_1 at 0.30 m (bottom) after k acquisitions under the 128-acquisition cap (3 starts per room, seed 2000; an episode that ends before k carries its final value forward; bands, bootstrap 95 % CIs over room means) for the main method at STOP floor 4 and the two walkers under its configuration, with the reference checkpoint at STOP floor 128 dashed: (a) all 355 rooms, (b) the 246 rooms the reference recipe loads, (c) 400–600-node rooms (40 / 40), (d) ≥600-node rooms (170 / 61; rooms for the main method / the reference). The main method's and walkers' curves are reconstructed from the recorded per-step reward components; the reference's tiers are re-aggregated from its episode curves.

senses and travel comes from the 100–400-node rooms (-4.1 [-6.0, -2.3] senses, -38.0 [-68.5, -11.3] m) and the ≥600-node rooms (-3.2 [-4.9, -1.6] senses, -37.4 [-61.4, -14.8] m).

The SENSE-to-lattice distance summaries are close in both modes: median 0.40 m cached and 0.41 m live, 90th percentile 0.59 m in both, and 25.5 and 25.3 % of SENSE steps beyond 0.5 m. None of the 21,408 cached and 0.1 % of the 20,258 live SENSE steps lie beyond the 1 m eligibility radius, so sensing beyond it is rare and the comparison mainly tests fresh acoustic rendering at the true pose.

Scope of the sensing evaluation. Both cached and live evaluations use the same simulator on unfurnished, extruded floor plans without transmission or diffraction. RIRs are noiseless and use a point source and omnidirectional microphones at a fixed heading. Measured RIRs remain untested (for physical sensing, see Brunetto et al., 2023; Dümbgen et al., 2023).

E.9 Reconstructed maps of single episodes

Figures 4 and 17 show what the main method reconstructs in single episodes with the settings of Table 1 (the frozen seed-42 checkpoint at update 3,600, STOP floor 4, 128 acquisitions, 1,024 decisions), one episode per room. Every map is replayed from the recorded actions through the frozen belief, so the picture and the episode's numbers come from the same computation. A stretch of true wall is drawn in teal when the final completion calls a wall cell within the 0.30 m tolerance of it (the evaluator's rule behind wall recall at 0.30 m), and the weighted teal fraction is the recall printed in the panel's header. Whatever stays grey was not reconstructed. In Fig. 4 the cells the belief calls wall without any echo support are shaded violet: structure the belief completes beyond what the echoes covered. The gallery omits that layer, which is illegible at tile size, and keeps the recovered and unrecovered walls. The trajectory, the acquisition positions, the collisions and the final pose are drawn on top; the walls themselves are context and are never an input to the policy or the belief.

Figure 4 (Sec. 5.2) shows the largest room of four layout families, each in a successful episode that ended by STOP. Figure 17 samples 36 rooms roughly in proportion to the test split's layout-family and difficulty classes, taking the largest rooms of each class and breaking ties by room id, so the selection is fixed by rule rather than by eye. Successes and failures are both included: 13 of the 36 episodes reach 90 % wall recall, and the other 23 carry an orange label. A star marks a STOP ending and an

Table 31: Paired cached and live evaluation of the main method on all 355 test rooms, one identical start per room (seed 2000; 128 acquisitions, STOP floor 4). Live mode uses a 0.45 s ray horizon versus 3 s for the cached corpus, renders at the true pose, writes on the 0.25 m memory grid and permits SENSE anywhere (Appendix E.8). The cached pass is separate from the three-start evaluation. Differences are live minus cached, with 95 % room-bootstrap intervals (10,000 resamples); tiers with at least 10 rooms are shown. Acquisitions to 90 % recall are failure-penalised at 129. Distance summaries measure SENSE-to-nearest-node distance.

Test split (355 rooms), EchoBelief (full): 355 rooms in both passes, one start per room, 128 acquisitions, STOP floor 4; the cached side is a separate one-start pass, not the 3-start rows of Table 1. Columns: cached = cached lattice; live = the same checkpoint with live RIRs at the true pose and SENSE legal anywhere

	All 355 rooms		<100 nodes 16 rooms		100–400 nodes 129 rooms		400–600 nodes 40 rooms		≥600 nodes 170 rooms	
	cached	live	cached	live	cached	live	cached	live	cached	live
Success (%)	70.4	69.0	93.8	93.8	89.1	86.8	75.0	80.0	52.9	50.6
Final F_1 (%)	90.8	90.1	95.7	93.8	95.4	94.4	93.1	92.8	86.3	85.7
Final recall (%)	90.6	89.9	97.2	95.2	96.1	95.2	93.0	93.2	85.2	84.6
Senses	60.3	57.1	8.4	7.8	29.6	25.5	27.9	26.1	96.1	93.0
Distance (m)	362.5	327.8	29.0	26.5	212.3	174.3	152.8	127.7	557.1	519.8
Collisions	5.6	5.0	0.2	0.4	3.2	2.4	1.6	1.4	8.9	8.2
Acq. to 90 %	66.4	66.7	4.8	12.6	30.2	31.3	47.9	45.1	104.0	103.8

	Paired live − cached [95% CI]				
	All 355 rooms	<100 nodes 16 rooms	100–400 nodes 129 rooms	400–600 nodes 40 rooms	≥600 nodes 170 rooms
Success (%)	−1.4 [−5.4, 2.8]	+0.00 [−18.75, 18.75]	−2.3 [−8.5, 3.9]	+5.0 [−7.5, 20.0]	−2.4 [−8.8, 4.1]
Final F_1 (%)	−0.7 [−1.2, −0.3]	−1.9 [−3.6, −0.5]	−1.0 [−1.9, −0.1]	−0.3 [−1.7, 1.4]	−0.59 [−1.27, 0.05]
Final recall (%)	−0.68 [−1.43, 0.05]	−1.97 [−4.34, −0.01]	−0.9 [−2.1, 0.3]	+0.2 [−2.0, 2.8]	−0.6 [−1.7, 0.4]
Senses	−3.2 [−4.4, −2.2]	−0.6 [−1.6, 0.2]	−4.1 [−6.0, −2.3]	−1.8 [−4.7, 0.5]	−3.2 [−4.9, −1.6]
Distance (m)	−34.6 [−51.0, −19.3]	−2.5 [−6.2, 1.2]	−38.0 [−68.5, −11.3]	−25.1 [−62.6, 1.9]	−37.4 [−61.4, −14.8]
Collisions	−0.7 [−1.1, −0.2]	+0.19 [0.00, 0.38]	−0.8 [−1.5, −0.3]	−0.2 [−0.6, 0.2]	−0.7 [−1.6, 0.1]
Acq. to 90 %	+0.4 [−2.7, 3.5]	+7.81 [0.00, 22.94]	+1.1 [−5.0, 7.4]	−2.7 [−16.3, 10.8]	−0.2 [−3.1, 2.6]

Episodes identical in (decisions, senses): 5; identical success: 298; both: 4.

Distance from each SENSE position to the nearest lattice node. Cached lattice: median 0.40 m, 90th percentile 0.59 m, mean 0.40 m, 0.0 % beyond 1 m, 25.5 % beyond 0.5 m (21,408 SENSE steps over 355 episodes, 60.3 per episode); live (the same checkpoint, RIRs at the true pose, SENSE legal anywhere): median 0.41 m, 90th percentile 0.59 m, mean 0.40 m, 0.1 % beyond 1 m, 25.3 % beyond 0.5 m (20,258 SENSE steps over 355 episodes, 57.1 per episode).

Cached mode senses within 1.0 m of a node and observes the node's cached encoding; live mode simulates the RIR at the true pose (heading 0) and writes the belief there on the 0.25 m memory grid (at most 0.177 m Euclidean rounding). Acq. to 90 %: acquisitions to 90 % wall recall, failure-penalised (129 when never crossed).

open circle any other ending, independently of whether the episode succeeded. In both successful and failed episodes, grey stretches are true walls not recovered by the final completion.

E.10 Geometry-disjoint confirmation cohort

Because the 355-room test split was read more than once during development (Sec. 5.1), a cohort of 60 new, geometry-disjoint rooms re-tests the main comparison against raycast information gain (Sec. 5.2). The confirmation cohort was generated under a new seed and frozen before the main method's checkpoint. The main method's read was declared before it ran and is its only read of the cohort; no method or reporting choice for the main method was made after looking at cohort outcomes.

Generation. Sixty rooms were drawn with the generator's production presets, with catalogue types allocated in proportion to the test split's counts by largest remainder: 32 rooms from legacy named types, 10 building, 9 curved, 7 maze and 2 irregular curved rooms; 55 complex, 4 very complex and 1 extreme; 8–3,297 walls and 76–2,226 lattice nodes. The corpus's RIR gates were evaluated on 16 lattice poses per room, and all 60 rooms were accepted at the first draw. Geometry identities (invariant to rigid motion, reflection and uniform scale) were checked against all 4,026 corpus rooms with no collision.

Freeze. A three-acquisition smoke test on one room preceded the cohort freeze. The freeze fixed the room set with its geometry identities, the start node of each room (seed 2000, one start) and the protocol (128 acquisitions, 1,024 decisions, STOP floor 4, RIRs simulated under the live acoustic configuration of Appendix E.8 at the robot's true pose, SENSE legal anywhere). It also fixed the comparators (raycast information gain with the transferred setting of Appendix E.7; the belief-frontier walker), the endpoints and the rule that every generated room is evaluated and failures are reported. An addendum declared the main method's checkpoint and configuration hashes before its read, retaining the cohort and protocol.

Results. In Table 32, the main method succeeds on 66.7 % [55.0, 78.3] of the rooms with final F_1 90.6 % [88.1, 92.9] after 58.2 senses and with 5.4 collisions per episode (68.0 [55.0, 81.3] failure-penalised acquisitions to 90 %). Paired on the same rooms, it is $+15.0$ [1.7, 28.3] success points and $+8.3$ [4.4, 12.5] F_1 points ahead of raycast information gain (51.7 % [38.3, 65.0] success, F_1 82.3 % after 124.3 senses and 20.1 collisions), with -66.1 [-76.8, -55.2] senses and -14.7 [-19.6, -10.2] collisions. It is $+66.7$ [55.0, 78.3] success points ahead of the belief-frontier walker as frozen. As elsewhere, an interval that contains zero is no resolved difference.

The node tiers repeat the test-split pattern. On the 27 rooms with at least 600 nodes the paired success difference from raycast information gain is $+40.7$ [22.2, 59.3] points. On the 22 rooms with 100–400 nodes it is $+4.55$ [0.00, 13.64], no resolved difference in success, with F_1 2.0 [0.8, 3.2] points lower and 100.3 [91.5, 107.5] fewer senses. These are the paper's standard node tiers, used here in place of the evaluator's tiers named in the freeze, so this split is post hoc.

The belief-frontier walker as frozen deadlocked under live sensing: every episode spent its budget on 127 senses of which 119.5 were repeats, travelled 26.8 m and succeeded in 0 % of the rooms, because its lattice bookkeeping never sees repeat acquisitions discarded by the environment's 0.5 m repeat rule. The original row is retained.

A disclosed post-freeze fix of the walker alone (the repeat rule respected; sensing only where the pose clears every wall by the corpus's 0.45 m ring clearance) was evaluated separately. It gives 18.3 % [10.0, 28.3] success and F_1 61.6 % [54.6, 68.6] after 106.8 senses, against the main method's 66.7 % [55.0, 78.3] and 90.6 % [88.1, 92.9] on the same rooms; that re-run is reported beside the frozen rows and not paired. Evaluations of the same policies on the 355-room test split are shown alongside; they are other cohorts under their own observation modes and are not paired with the cohort.

E.11 Fresh rooms beyond the corpus

To accommodate longer exploration paths in large layouts, we extend the acquisition/decision budgets from 128/1,024 to 256/2,048 and 512/4,096, with the same frozen checkpoint, STOP floor 4 and live sensing configuration of Appendix. E.8. The policy maps 78.0 % of the 500 fresh rooms to 90 % recall.

E.12 Ablations and input controls of the main method

This subsection details the ablations of Sec. 5.3 and the input controls of Sec. 5.2. At 128 acquisitions, removing the memory pathway costs the most success (22.8 points) and the bundled reference-recipe

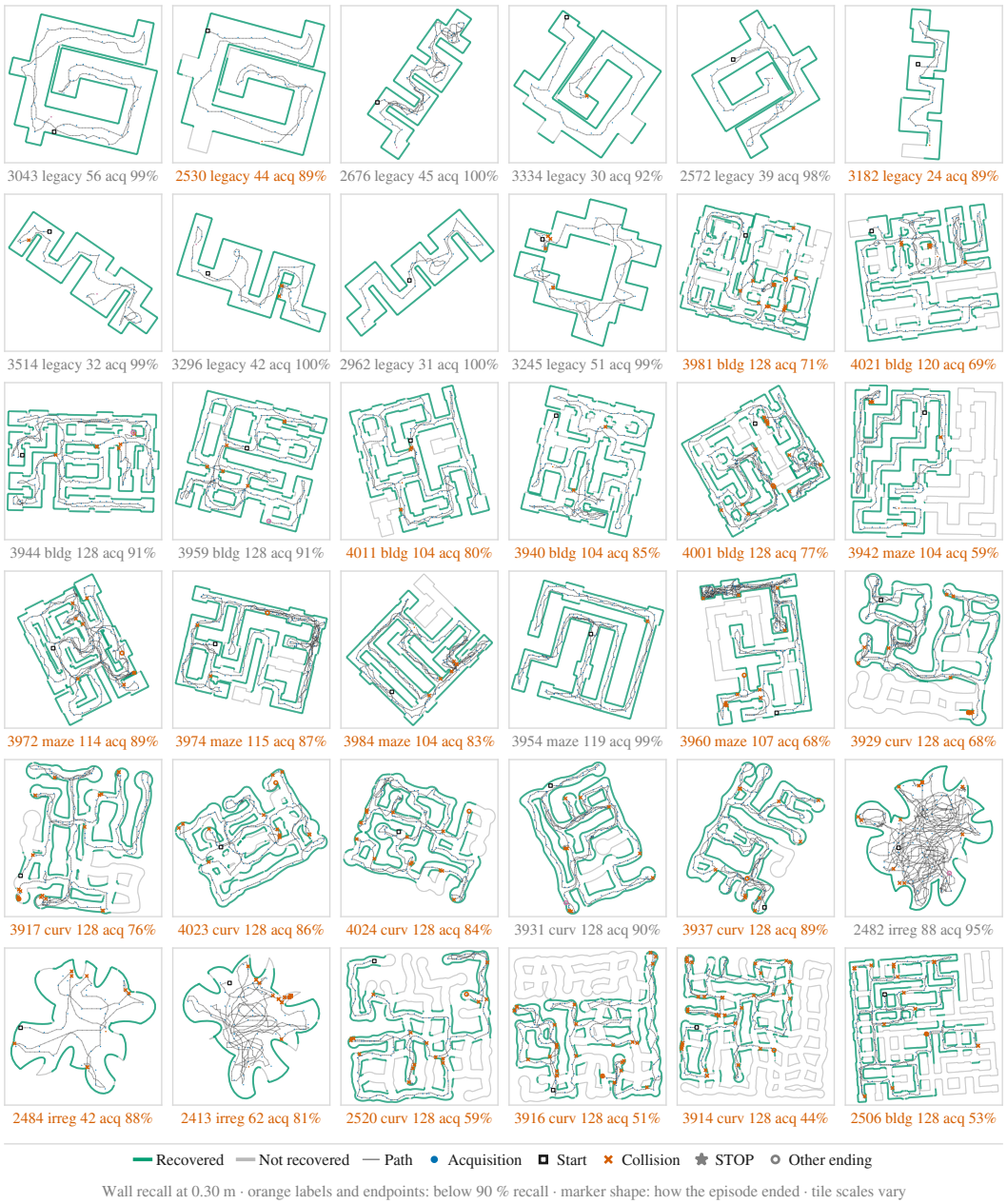

Figure 17: Thirty-six episodes of the main method's test evaluation, one per room, sampled across layout families and difficulty tiers (Appendix E.9). The reconstruction criterion follows Fig. 4: teal, wall recovered within 0.30 m; grey, wall not recovered. The predicted-wall layers are omitted at this size and tiles are scaled independently. Each tile gives the room, its layout family, the acquisitions used and the final wall recall. Thirteen episodes reach 90 % recall; the other 23 carry an orange label. A star marks a STOP ending and an open circle any other ending.

Table 32: Confirmation on 60 new rooms, geometry-disjoint from all 4,026 corpus rooms. The cohort and protocol were frozen before the main method's checkpoint; its sole read was declared in advance (Appendix E.10). One start per room (seed 2000), 128 acquisitions, 1,024 decisions, STOP floor 4 for the main method, and live sensing at the true pose with SENSE legal anywhere. Intervals are 95 % room bootstraps (10,000 resamples); paired differences are main method minus comparator, and tiers with at least 10 rooms are shown. The post-freeze walker correction is a separate, unpaired evaluation. Test-split references use the main method's live pass and the comparators' cached three-start passes and are not paired with this cohort. Acquisitions to 90 % recall are failure-penalised at 129.

Confirmation cohort: 60 newly generated rooms read once after the protocol was frozen, one start per room, 128 acquisitions, RIRs simulated at the true pose, SENSE legal anywhere; geometry identities disjoint from all 4,026 corpus rooms

	EchoBelief (full)	Raycast IG	Walker (frozen read)	Walker (post-freeze fix)	Δ main − raycast IG [95% CI]	Δ main − walker (frozen) [95% CI]
Success (%)	66.7	51.7	0.0	18.3	+15.0 [1.7, 28.3]	+66.7 [55.0, 78.3]
95% CI	[55.0, 78.3]	[38.3, 65.0]	[0.0, 0.0]	[10.0, 28.3]		
Final F_1 (%)	90.6	82.3	25.2	61.6	+8.3 [4.4, 12.5]	+65.4 [61.0, 69.6]
95% CI	[88.1, 92.9]	[77.2, 87.3]	[20.4, 30.4]	[54.6, 68.6]		
Final recall (%)	90.4	77.6	17.4	52.3	+12.8 [7.5, 18.5]	+73.0 [69.0, 76.9]
95% CI	[87.2, 93.3]	[70.9, 84.1]	[13.5, 21.7]	[44.3, 60.5]		
Senses	58.2	124.3	127.0	106.8	−66.1 [−76.8, −55.2]	−68.8 [−79.8, −57.7]
95% CI	[47.2, 69.3]	[121.6, 126.3]	[127.0, 127.0]	[98.8, 114.1]		
Distance (m)	340.5	413.3	26.8	419.7	−72.8 [−159.3, 11.0]	+313.7 [234.7, 396.5]
95% CI	[265.5, 419.9]	[365.6, 466.7]	[16.1, 39.0]	[388.3, 451.3]		
Collisions	5.4	20.1	0.0	0.0	−14.7 [−19.6, −10.2]	+5.4 [3.4, 7.8]
95% CI	[3.4, 7.8]	[14.4, 26.6]	[0.0, 0.0]	[0.0, 0.0]		
Acq. to 90 %	68.0	85.3	129.0	116.3	−17.3 [−27.0, −7.4]	−61.0 [−74.0, −47.7]
95% CI	[55.0, 81.3]	[72.7, 97.6]	[129.0, 129.0]	[107.7, 123.7]		
Rooms with 100–400 nodes (22 rooms)						
Success (%)	100.0	95.5	0.0	36.4	+4.55 [0.00, 13.64]	+100.0 [100.0, 100.0]
Final F_1 (%)	96.7	98.7	43.6	83.8	−2.0 [−3.2, −0.8]	+53.1 [46.1, 60.2]
Final recall (%)	98.0	99.2	32.1	78.0	−1.27 [−2.42, −0.01]	+65.8 [59.2, 72.5]
Senses	24.9	125.1	127.0	91.8	−100.3 [−107.5, −91.5]	−102.1 [−109.6, −92.9]
Distance (m)	192.4	339.0	45.9	511.2	−146.6 [−303.9, 17.2]	+146.5 [26.1, 296.2]
Collisions	2.3	11.1	0.0	0.0	−8.8 [−15.0, −3.3]	+2.3 [0.8, 4.4]
Acq. to 90 %	16.5	42.5	129.0	106.1	−26.1 [−38.0, −14.9]	−112.5 [−116.2, −108.3]
Rooms with ≥600 nodes (27 rooms)						
Success (%)	48.1	7.4	0.0	0.0	+40.7 [22.2, 59.3]	+48.1 [29.6, 66.7]
Final F_1 (%)	85.5	63.6	8.5	35.9	+21.9 [16.6, 27.2]	+76.9 [73.0, 80.4]
Final recall (%)	84.4	53.0	4.7	23.2	+31.4 [24.0, 38.8]	+79.7 [74.1, 84.6]
Senses	100.5	125.9	127.0	119.3	−25.4 [−35.5, −15.7]	−26.5 [−36.0, −17.6]
Distance (m)	558.2	470.8	16.9	339.0	+87.4 [4.7, 171.2]	+541.3 [461.0, 621.1]
Collisions	9.4	27.0	0.0	0.0	−17.6 [−26.1, −10.4]	+9.4 [5.7, 14.2]
Acq. to 90 %	106.7	124.5	129.0	129.0	−17.8 [−27.6, −8.6]	−22.3 [−33.0, −12.3]
Rooms with complex difficulty (55 rooms)						
Success (%)	69.1	54.5	0.0	16.4	+14.5 [1.8, 29.1]	+69.1 [56.4, 80.0]
Final F_1 (%)	92.3	83.8	26.4	62.7	+8.5 [4.3, 13.0]	+65.8 [61.1, 70.4]
Final recall (%)	92.5	79.2	18.3	53.1	+13.3 [7.5, 19.4]	+74.2 [70.3, 78.0]
Senses	53.8	124.6	127.0	104.9	−70.8 [−81.3, −59.8]	−73.2 [−84.2, −61.8]
Distance (m)	305.2	395.4	28.6	422.8	−90.2 [−178.4, −4.2]	+276.6 [199.8, 359.4]
Collisions	3.5	15.5	0.0	0.0	−12.0 [−15.7, −8.4]	+3.5 [2.4, 4.6]
Acq. to 90 %	65.7	83.7	129.0	115.7	−18.0 [−28.4, −7.2]	−63.3 [−77.0, −49.6]
Test-split references: the same policies on the 355-room test split (other cohorts; not paired with the rows above)						
Rooms / starts	355 / 1	355 / 3	355 / 3	–	–	–
Success (%)	69.0	49.4	21.2	–	–	–
Final F_1 (%)	90.1	83.9	64.1	–	–	–
Final recall (%)	89.9	79.5	54.4	–	–	–
Senses	57.1	124.3	108.5	–	–	–
Distance (m)	327.8	429.6	397.1	–	–	–
Collisions	5.0	22.3	0.0	–	–	–
Acq. to 90 %	66.7	84.0	114.0	–	–	–

Walker (frozen read): belief-frontier walker, live (frozen read; repeated acquisitions, see text); Walker (post-freeze fix): belief-frontier walker, live, post-freeze fix (see text). Raycast IG: raycast information gain. Paired columns on the same 60 rooms: Δ main − raycast IG: EchoBelief (full) − raycast IG; Δ main − walker (frozen): EchoBelief (full) − belief-frontier walker (frozen read). Test-split references: EchoBelief (full), live read of the 355-room test split; raycast IG and belief-frontier walker, cached-lattice runs on the 355-room test split (3 starts). Tiers with fewer than 10 rooms are omitted.

Table 33: Main method on 500 new rooms, by layout family and complexity (live protocol, Sec. E.11). Success means final wall recall $\geq 90\,\%$; wall F_1 and recall use a 0.30 m tolerance. Acquisitions to 90 % recall count failures as the budget. Legacy classes are grouped by corner count, other families by generator complexity. Balanced means weight the ten classes equally; pooled means weight rooms equally.

Class	Rooms	Success (%)	Final F1 (%)	Final recall (%)	Interior IoU (%)	Acq. to 90%	Senses	Collisions	Stop (%)
Main method: 500 fresh rooms, STOP floor 4									
legacy / medium	36	100.0	97.1	98.2	97.2	9.2	16.5	0.6	97.2
legacy / complex	102	100.0	96.3	97.6	96.6	22.2	28.6	2.1	96.1
building / standard	46	82.6	89.9	91.7	93.1	135.4	112.2	6.7	91.3
building / hard	46	60.9	89.8	90.5	89.4	291.3	184.8	23.5	82.6
building / extreme	45	8.9	87.3	86.2	83.9	433.1	247.8	52.7	73.3
maze / standard	45	100.0	93.1	94.0	93.6	87.7	90.3	2.9	93.3
maze / hard	45	84.4	92.3	92.0	90.1	184.0	158.3	19.0	71.1
maze / extreme	45	28.9	88.9	87.0	84.2	390.6	240.1	52.8	75.6
curved / standard	45	100.0	92.6	95.9	93.7	102.1	111.8	10.3	82.2
curved / hard	45	91.1	89.9	93.3	90.1	211.3	213.2	42.0	60.0
balanced over the 10 classes	500	75.7	91.7	92.6	91.2	186.7	140.4	21.3	82.3
pooled over the 500 rooms	500	78.0	92.1	93.1	91.7	171.2	129.9	19.4	83.6

Table 34: Summary of the 500 rooms in Table 33, one metric per row. Each cell gives the main method's balanced or pooled mean with a 95 % room-bootstrap interval (10,000 resamples). For balanced means, rooms are resampled within each class and the class means are averaged.

Metric	Balanced (10 classes)	Pooled (500 rooms)
Main method: mean [95% CI], extended live evaluation of 500 rooms		
Success (%)	75.7 [73.0, 78.5]	78.0 [74.4, 81.6]
Final F_1 (%)	91.7 [91.5, 91.9]	92.1 [91.8, 92.5]
Final recall (%)	92.6 [92.4, 92.9]	93.1 [92.7, 93.5]
Interior IoU (%)	91.2 [91.0, 91.4]	91.7 [91.3, 92.1]
Acq. to 90%	186.7 [178.2, 195.1]	171.2 [156.9, 185.9]
Never reached 90% (%)	24.3 [21.5, 27.0]	22.0 [18.4, 25.6]
Senses	140.4 [137.7, 143.0]	129.9 [122.3, 137.4]
Decisions	1283.9 [1220.5, 1348.7]	1181.5 [1083.6, 1280.6]
Distance (m)	1118.6 [1058.5, 1180.1]	1028.8 [939.5, 1118.9]
Collisions	21.3 [20.0, 22.6]	19.4 [17.4, 21.6]
Stop (%)	82.3 [78.8, 85.6]	83.6 [80.4, 86.8]

arm the next most (12.3 points). For the other seed-42 arms, main method minus arm is $+1.3$ to $+3.6$ points, and the main method's seed-43 and seed-44 replicates show no resolved success difference under any protocol. Each ablation arm retrains the main method's recipe for 15 M steps with the intervention specified for it. The arms are evaluated under the three protocols of Appendix E.7 on all 355 rooms, each paired against the main method over rooms (10,000 resamples; Table 2). Their population and tier splits are in Tables 28 and 29, and their costs in Table 30. Each arm has one training run at seed 42, except the legacy-fusion and without-memory-pathway arms, which also have a seed-43 replicate (Table 35, paired with the seed-42 main method); the contrasts below are the seed-42 runs unless stated. Checkpoints and configurations were frozen before evaluation. Each paragraph reports the protocols in the order 128 acquisitions, 64 acquisitions, STOP disabled; Δ is main method minus arm, with paired 95 % intervals.

Seed 43. The replicate trained with seed 43, otherwise unchanged, shows no resolved success difference under any protocol but senses and collides more than the main method. It scores 71.8 % success at 128 acquisitions (Δ -1.7 [-4.7, 1.3]), 50.1 % at 64 acquisitions (Δ -0.8 [-2.8, 1.3]) and 53.1 % with STOP disabled (Δ $+1.4$ [-0.1, 2.9]). The main method senses less than seed 43 (Δ -4.2 [-5.4, -3.1], -3.4 [-4.2, -2.6] and -2.9 [-3.4, -2.3] senses) and collides less (Δ -1.5 [-2.0, -0.9], -0.9 [-1.3, -0.6] and -1.4 [-1.9, -0.9]) under all three protocols. The three-seed mean of Table 1 averages seeds 42, 43 and 44 within each room before resampling.

Seed 44. A third replicate again shows no resolved success difference and collides slightly less than the main method. It scores 70.5 % at 128 acquisitions (Δ -0.4 [-3.2, 2.3]), 51.2 % at 64 acquisitions (Δ -1.8 [-4.0, 0.5]) and 54.6 % with STOP disabled (Δ -0.1 [-1.6, 1.4]), with Δ collisions of

Table 35: Seed replicates of two ablation arms (355 test rooms, three starts each; 15 M training steps; Δ = main method minus arm, paired 95 % CI). The seed-42 rows of these arms are in Table 2.

Policy (all 355 test rooms)	Success (%)	Stop (%)	Main − arm: Δ success (pts)	Main − arm: Δ acq. to 90%	Main − arm: Δ collisions	Main − arm: Δ senses
STOP floor 4, 128-acquisition cap, 1,024 decisions (the main method's own protocol)						
EchoBelief (full, seed 42)	70.1	78.8				
Legacy fusion, seed 43	63.6	74.7	+6.6 [3.1, 10.0]	−4.8 [−7.3, −2.3]	−0.3 [−1.4, 0.6]	−5.7 [−7.6, −4.0]
Without memory pathway, seed 43 (belief and policy retrained)	48.2	69.1	+22.0 [17.7, 26.2]	−17.5 [−20.2, −14.7]	−1.2 [−1.7, −0.7]	−14.3 [−16.0, −12.6]
STOP floor 4, 64-acquisition cap, 512 decisions						
EchoBelief (full, seed 42)	49.4	55.3				
Legacy fusion, seed 43	47.0	52.0	+2.3 [0.2, 4.5]	−0.3 [−1.0, 0.5]	−0.58 [−1.43, 0.02]	−1.1 [−1.8, −0.5]
Without memory pathway, seed 43 (belief and policy retrained)	41.5	50.5	+7.9 [5.3, 10.6]	−6.6 [−7.7, −5.5]	−0.1 [−0.5, 0.2]	−5.8 [−6.7, −5.0]
STOP disabled (floor 64), 64-acquisition cap, 512 decisions						
EchoBelief (full, seed 42)	54.5	0.0				
Legacy fusion, seed 43	51.0	0.0	+3.5 [2.0, 5.1]	−0.7 [−1.1, −0.3]	+0.2 [−0.6, 0.9]	−3.8 [−4.4, −3.3]
Without memory pathway, seed 43 (belief and policy retrained)	48.7	0.0	+5.7 [3.7, 7.9]	−6.6 [−7.4, −5.8]	−0.6 [−0.9, −0.3]	−0.4 [−1.0, 0.2]

+0.6 [0.2, 0.9], +0.28 [0.05, 0.50] and +0.4 [0.1, 0.6]. The three seeds span 70.1–71.8 % success at 128 acquisitions.

Legacy-fusion arm (*Legacy fusion* in tables). This arm tests the belief package as a whole, fusion rule and fine-tuning together. Against the seed-42 arm, the support-gated belief buys fewer collisions under every protocol and a resolved success gain only with STOP disabled. Against the seed-43 replicate, the success gain is resolved under all three protocols, including 128 acquisitions. The arm replaces the fine-tuned belief with support-gated fusion by the reference recipe's seed-42 belief with legacy fusion (a cell is written when $p_{\text{wall}} + p_{\text{free}} \geq 0.5$; temperature 1.3112), keeping the remaining settings of the main method. It scores 67.6 % success at 128 acquisitions (Δ +2.5 [−1.0, 6.1] points, no resolved difference; collisions −1.1 [−1.8, −0.5]; failure-penalised acquisitions-to-90 −2.2 [−4.9, 0.5], no resolved difference). It scores 47.7 % at 64 acquisitions (+1.7 [−0.7, 3.9], no resolved difference; collisions −0.6 [−1.1, −0.2]; senses +1.8 [1.3, 2.4]) and 51.4 % with STOP disabled (+3.1 [1.6, 4.7] points; collisions −1.6 [−2.2, −1.2]; senses +7.4 [6.4, 8.4]). For the seed-43 replicate, Δ success is +6.6 [3.1, 10.0] points at 128 acquisitions, +2.3 [0.2, 4.5] at 64 acquisitions and +3.5 [2.0, 5.1] with STOP disabled (Table 35).

Reference-recipe arm. This bundled arm has the second-largest success deficit at 128 acquisitions and collides far more than the main method. It uses the reference recipe's network and features (single trunk, 12 local channels, 21 scalars, no per-move features), its terminal block, a fixed STOP floor of 24, no predictor, no collision memory and no auxiliary losses. It is trained in the all-rooms configuration (the fine-tuned belief, window-centre anchor, band sampler, 128-acquisition budget). With its STOP-on rows at its own floor of 24, it scores 57.8 % success at 128 acquisitions (Δ +12.3 [8.7, 16.0]), 48.3 % at 64 acquisitions (Δ +1.1 [−1.1, 3.4], no resolved difference) and 51.5 % with STOP disabled by setting the floor to the cap (Δ +3.0 [1.3, 4.8]). It has 15.0 [11.8, 18.5], 10.7 [8.5, 13.1] and 15.6 [12.6, 18.9] more collisions per episode than the main method under the three protocols, and its failure-penalised acquisitions-to-90 is higher by 5.0 [2.5, 7.5] at 128 acquisitions.

No predictor. Removing the one-step predictor features from the actor's residual, while retaining the belief's recurrent context memory, gives no resolved difference in success or failure-penalised acquisitions-to-90 under any protocol. The arm scores 68.8 % success (Δ +1.3 [−1.7, 4.3]), 48.9 % (Δ +0.5 [−1.8, 2.7]) and 53.5 % (Δ +0.9 [−0.7, 2.6]), with Δ acquisitions-to-90 of −0.5 [−2.7, 1.6], +0.04 [−0.80, 0.88] and −0.2 [−0.6, 0.3].

No collision memory. Without the episode-long mask on failed (pose, move) pairs, the arm has a resolved success deficit at 128 acquisitions and with STOP disabled and senses more, but its collision count shows no resolved difference. It scores 66.6 % success (Δ +3.6 [0.4, 6.8]), 47.7 % (Δ +1.7 [−0.7, 3.9], no resolved difference) and 52.3 % (Δ +2.2 [0.5, 3.8]). Its Δ collisions are −2.0 [−7.5, 1.0], −1.1 [−3.7, 0.4] and +0.42 [0.00, 0.82], the last interval containing zero at its lower bound, and its Δ senses −5.0 [−6.6, −3.5], −1.2 [−1.8, −0.6] and −3.4 [−3.9, −2.8]. Its failure-penalised acquisitions-to-90 is higher by 4.9 [2.3, 7.4] at 128 acquisitions.

Without-memory-pathway arm. Removing the memory pathway is the largest ablation: a resolved success deficit under every protocol, reproduced by the seed-43 replicate. This arm applies the recipe of the fine-tuned belief (support-gated fusion and hole-aware completion targets; Appendix D.6) to

Table 36: Controls of the main method under the three protocols (Δ = main method minus control, paired 95 % CI over the rooms the control covers): no acoustic observations on all 355 test rooms, and the start-anchored window on the 246 rooms its 52 m span cap loads, where the main method's row is restricted to the same rooms.

Policy	Rooms	Success (%)	Stop (%)	Main − control: Δ success (pts)	Main − control: Δ acq. to 90%	Main − control: Δ collisions	Main − control: Δ senses
STOP floor 4, 128-acquisition cap, 1,024 decisions (the main method's own protocol)							
EchoBelief (full, seed 42), all 355 test rooms	355	70.1	78.8				
No acoustic observations	355	0.0	10.8	+70.1 [66.0, 74.1]	−62.4 [−67.1, −57.5]	−65.9 [−68.0, −63.8]	+45.8 [41.5, 50.2]
EchoBelief (full, seed 42), the same 246 rooms	246	79.7	89.6				
Start-anchored window	246	80.5	90.5	−0.8 [−4.3, 2.6]	−2.3 [−5.2, 0.7]	+0.6 [−0.1, 1.2]	−11.4 [−13.7, −9.2]
STOP floor 4, 64-acquisition cap, 512 decisions							
EchoBelief (full, seed 42), all 355 test rooms	355	49.4	55.3				
No acoustic observations	355	0.0	0.5	+49.4 [44.6, 54.1]	−22.5 [−25.0, −20.0]	−33.4 [−34.6, −32.1]	+33.1 [30.8, 35.3]
EchoBelief (full, seed 42), the same 246 rooms	246	69.1	77.6				
Start-anchored window	246	66.5	69.2	+2.6 [−0.5, 5.7]	−0.8 [−2.0, 0.3]	+0.42 [−0.02, 0.81]	−5.4 [−6.5, −4.3]
STOP disabled (floor 64), 64-acquisition cap, 512 decisions							
EchoBelief (full, seed 42), all 355 test rooms	355	54.5	0.0				
No acoustic observations	355	0.0	0.0	+54.5 [49.6, 59.4]	−23.8 [−26.3, −21.3]	−32.1 [−33.3, −30.9]	+50.8 [50.2, 51.4]
EchoBelief (full, seed 42), the same 246 rooms	246	76.4	0.0				
Start-anchored window	246	72.1	0.0	+4.3 [1.9, 6.9]	−1.3 [−2.1, −0.6]	+0.8 [0.4, 1.3]	−5.8 [−6.4, −5.2]

the seed-42 geometric-only fit instead of the seed-42 fit with H. It runs the resulting belief without H as in Appendix E.5 and retrains the policy on it with the remaining settings of the main method unchanged. It scores 47.3 % success (Δ +22.8 [18.8, 27.0]), 39.8 % (Δ +9.6 [6.9, 12.3]) and 48.8 % with STOP disabled (Δ +5.6 [3.6, 7.9]). Its final wall F_1 is lower by 7.6 [7.1, 8.3], 11.1 [10.4, 11.8] and 9.4 [8.6, 10.1] points, its failure-penalised acquisitions-to-90 is higher by 19.6 [16.8, 22.3], 7.1 [6.0, 8.2] and 6.4 [5.6, 7.2], and it senses more (Δ -14.3 [-16.1, -12.5], -5.5 [-6.4, -4.7] and -3.9 [-4.4, -3.4]). Collisions differ little: Δ -0.9 [-1.4, -0.4], $+0.1$ [-0.1, 0.4] and $+0.25$ [-0.05, 0.56], so the main method collides slightly less at 128 acquisitions and the last two show no resolved difference. It stops in 64.9 % of episodes at 128 acquisitions against the main method's 78.8 %. For its seed-43 replicate, main method minus arm is $+22.0$ [17.7, 26.2], $+7.9$ [5.3, 10.6] and $+5.7$ [3.7, 7.9] success points under the three protocols (Table 35). Under support-gated fusion, removing H also removes the route for retaining past local predictions beyond the support-gated evidence, so this arm measures the memory pathway as a whole rather than recurrence alone. The legacy-fusion arm changes the fusion rule and fine-tuning but keeps the memory. Together, the two arms place the belief's contribution to the controller mainly in that memory pathway rather than in the fusion rule. At 128 acquisitions the without-memory-pathway deficit is far larger than the legacy-fusion arm's ($+2.5$ [-1.0, 6.1] for seed 42, no resolved difference; a resolved $+6.6$ [3.1, 10.0] for seed 43).

Controls. Two controls test whether the policy relies on acoustics and on the room window. Each retrains the main method's recipe for 15 M steps with one input removed and is frozen and evaluated like the arms (Table 36).

No acoustic observations: every acquisition writes the no-observation values (unknown class, zero echo and polar maps), so only the pose, the swept footprint and the contact stamps enter the belief. Acquisitions still count against the budget and the STOP floor. The selection rule picks update 400, since no checkpoint succeeds on a selection episode. On the 355 test rooms it succeeds on no episode under any protocol (wall recall 1.0 % at 128 acquisitions, 71.5 collisions), and its final checkpoint (update 3,600) also scores 0.0 % (recall 0.8 %, 34.3 collisions).

Start-anchored window: the belief canvas is anchored at the start node and the actor's known-area scalars are normalised by the canvas rather than the room window, so nothing the actor sees derives from the room-specific window. The start anchor admits node spans up to 52 m, so this control is trained on the usable training rooms and evaluated on the 246 usable test rooms, paired with the main method restricted to the same rooms. It shows no resolved success difference at 128 or 64 acquisitions, senses more under every protocol and trails with STOP disabled. At 128 acquisitions it reaches 80.5 % success against the main method's 79.7 % (main method minus control -0.8 [-4.3, 2.6]) with 11.4 [9.2, 13.7] more senses; at 64 acquisitions the difference is $+2.6$ [-0.5, 5.7] and with STOP disabled $+4.3$ [1.9, 6.9] points. This subset includes only 61 of the 170 test rooms with at least 600 nodes. Randomly translated windows remain untested.

Table 37: Component results on EchoRooms (percentages except acquisition counts). Rows: the polar echo encoder (Stage 1); the local geometry mapper and the history refiner (Stage 2; mean ± sd over seeds); the evidence-anchored spatial memory's primary fit (seed 42, legacy fusion, before fine-tuning; Appendix D) after 64 depth-first acquisitions; and the reference recipe's selected seed-42 policy (*Trained-STOP policy*; acquisitions to 90 % wall recall are failure-penalised; 95 % CIs on test). Validation: 75 reporting rooms (policy: 69 selection rooms × 2 starts); test: 355 rooms (policy: 246 rooms × 3 starts). Main-method results: Table 1.

Model	Metric	Validation	Test	Seeds
Polar echo encoder (Stage 1)	near-range echo F1	97.5	97.0	1 (42)
Local geometry mapper (Stage 2)	current-view near-wall F1	85.9 ± 0.1	80.6 ± 0.2	4 (42–45)
History refiner (Stage 2)	prefix-knowledge near-wall F1	94.1 ± 0.1	91.1 ± 0.1	4 (42–45)
Spatial memory (primary)	prefix-knowledge near-wall F1	92.0	87.5	1 (42)
Spatial memory (primary)	whole-room completion F1 (0.30 m) at 64	67.7	45.2	1 (42)
Spatial memory (primary)	interior IoU at 64	68.6	45.1	1 (42)
Trained-STOP policy	acquisitions to 90% recall (censored at 65)	24.0	33.8 [30.9, 36.8]	1 (42)
Trained-STOP policy	success within 64 acquisitions (%)	79.7	63.7 [58.0, 69.2]	1 (42)

E.13 Collision writes and recall

This subsection checks, for the reference recipe, whether collision stamps inflate recall. A collision stamps its contact cell into the evidence map before the belief is rescored, so wall contacts could in principle improve recall. Replaying the recorded actions of the reference recipe's capped seed-42 policy and of the belief-frontier walker with that stamp suppressed (motion, observations and belief model unchanged) removes the 10,342 collision writes of the capped policy's episodes. Suppressing the stamps changes final recall, F_1 and interior IoU by at most 0.01 points ($+0.01$ [0.00, 0.02] recall) and the failure-penalised score by -0.008 [-0.020, 0.001]; the collision-free walker is unchanged by construction. For the reference recipe, therefore, there is no resolved difference in final recall or failure-penalised acquisitions-to-90 when collision stamps are suppressed. The replay was not repeated for the main method, whose support-gated evidence map also receives contact stamps, so this result does not extend to it.

F Evaluation protocol, metrics and statistics

This appendix details the metrics, censoring rules, bootstraps and compute behind Sec. 5.1; Table 37 summarises the component results.

Local metrics. Local metrics score the egocentric maps of Stage 2 and the local heads of the evidence-anchored spatial memory. A cell is *known* if $p_{\text{wall}} + p_{\text{free}} \geq 0.5$, and a *predicted wall* if it is known and $p_{\text{wall}} \geq p_{\text{free}}$. Using a Euclidean distance transform on the 0.125 m grid, precision is the fraction of predicted wall cells within the tolerance (0.125 m primary, 0.30 m secondary) of a target wall cell, and recall the converse. Near-wall F_1 scores the 0–4 m region and full-radius F_1 the full 8 m. We also report false-free-wall (the fraction of true wall cells predicted free), NLL, Brier score and 15-bin ECE. Counts are summed over each room's measurements, and rooms are averaged with equal weight.

Whole-room metrics. Whole-room metrics score the decoded completion and define policy success. Truth wall segments are sampled every 0.125 m with length weights, and the prediction is the set of centres of known wall cells in the completion crop. Precision is the fraction of predicted points within the tolerance of a truth sample, recall the length-weighted fraction of truth samples within the tolerance of a predicted point; the headline wall F_1 uses 0.30 m. Interior IoU compares the cells whose arg-max class is interior with the polygon interior over the whole truth raster (the room plus a 2 m margin). False-free-wall is length-weighted, and resolved area is the area whose entropy is below 0.1 nats.

Active metrics and censoring. Acquisitions-to-90 counts acquisitions until whole-room wall *recall* at 0.30 m first reaches 0.90; for the lattice planners, attempts-to-90 counts motion attempts. An episode that never reaches 0.90 within its budget is censored and enters the mean at the budget plus

one (65 or 129 acquisitions, or 257 attempts), so these means are failure-penalised; Table 30 reports the censored fractions of the policy comparisons.

A policy episode succeeds if its final wall recall at 0.30 m is at least 0.90, under a cap of 64 acquisitions and 512 decisions or 128 acquisitions and 1,024 decisions. Recall can fall back below 0.90 after first crossing it, so success is not simply one minus the censored fraction. The lattice planner studies (Appendices D.9 and D.10) instead define success as wall F_1 at 0.30 m ≥ 0.90 at the last scored point within 128 attempts, while censoring refers to recall within 256 attempts. Policy results by room size use four lattice-node tiers (below 100, 100–400, 400–600 and at least 600 nodes), and the lattice planner study also reports F_1 on the rooms with at least 64 nodes.

Hole-aware truth. The reference recipe's completion truth uses only the outer polygon, so hole interiors are labelled interior (*hole-blind* truth). The main method's *hole-aware* truth labels them exterior (Appendix E.7); 50 of the 242 original test rooms have holes. This affects interior IoU, coverage, the reachable cells in $\text{Known}_{\text{reach}}$ and frontier shaping, but not wall recall, precision or F_1, whose samples already include hole boundaries. Interior IoU is therefore not compared across the two conventions.

Bootstraps and multiplicity. Paired comparisons resample rooms with replacement 10,000 times, keeping starts and systems paired, and report the 2.5th and 97.5th percentiles of the room-mean difference (Efron & Tibshirani, 1993; Patterson et al., 2024); per-row intervals in a policy table use 1,000 resamples where its caption says so. Holm's step-down procedure (Holm, 1979) corrects the contrasts within each Stage 1 family (microphone counts and encoder ingredients); all other contrasts use uncorrected percentile intervals. A difference is *resolved* when its interval excludes zero.

Compute and latency. All fits ran on single NVIDIA A40 GPUs. The Stage 2 fits took about 78.5 ks (local geometry mapper) and 20.8 ks (history refiner), and the EchoRooms fit of the evidence-anchored spatial memory took 18,657 s. The reference recipe's seed-42 policy trained for 15 M steps at a mean logged rate of about 178 environment steps per second (24.6 h including selection rounds); training times of the main method and its arms are in Appendix E.7. Batch-1 latencies of the frozen modules are in Table 13, with the refiner timed on 63 observations.

These timings exclude planning. A full forecast-based FC-EER decision copies the belief for up to 32 candidates and decodes each copy on the covering crop. In the lattice planner study (Appendix D.10; shared A40), its median and 95th-percentile latencies were 138.5 and 353.1 ms over 5,408 decisions with at least one candidate (20.9 candidates on average), against a median of 616.5 ms for raycast information gain. These timings characterise the supplied implementations, not the planners: the lattice raycast spends 90 % of its decision time in a host-side ray loop.

G Evidence repair: measured versus predicted evidence

The reference recipe's legacy fusion treats Stage 2 predictions as measured evidence wherever predicted wall-plus-free mass reaches 0.5, so predicted free space can overwrite wall evidence (*prediction-induced wall erasure*). On depth-first replays of the selection rooms with the reference recipe's frozen seed-42 belief, 64.5 % of the curved-room wall length that was predicted wall at least once reads as free at 64 acquisitions (8 rooms), against 12.2 % on the 58 polygonal rooms with a full 64 acquisitions. Support-gated fusion (Appendix D.6) reduces this erasure on curved rooms to 7.4 % and raises evidence-map wall F_1 by $+18.1$ [9.6, 26.5] points at 64 acquisitions, but fails one of four pre-registered acceptance criteria: measured free recall drops by -15.3 [-18.5, -12.1] points on polygonal and -30.9 [-38.2, -24.1] on curved rooms, and an echo-threshold sweep did not recover it. Four matched fine-tunes of the seed-42 belief (5,000 updates with hole-aware completion targets; legacy or support-gated fusion; the paper's or extended decoder inputs) show no resolved wall-F_1 difference on identical depth-first observations at 64 acquisitions. Arm C (support-gated fusion, the paper's inputs) was adopted as the main method's fine-tuned belief. Alternative completion heads and joint Stage 1–Stage 2 training did not resolve the curved-room completion deficit and were not adopted.