IQRO JURNALI , IQRO JOURNAL, ИКРО ЖУРНАЛИ
ISSN:2181-4341| wordlyknowledge.uz/index.php/iqro
Volume 24. Issue 01. Oktabr, 2026

INTEGRATING ARTIFICIAL INTELLIGENCE AND SLAM FOR AUTONOMOUS MOBILE ROBOTS: A SEMANTIC, DYNAMICS-AWARE ARCHITECTURE AND A CONTROLLED LOCALIZATION STUDY

Sherzod A. Yarashev
Tashkent University
of Information Technologies named after
Muhammad al-Khwarizmi, Tashkent, Uzbekistan
ORCID: 0009-0009-5053-942X
Jurabek M. Abdiev
Arifov Institute of Ion-Plasma and Laser
Technologies, Academy of Sciences of the
Republic of Uzbekistan, Tashkent, Uzbekistan
E-mail: fiztexabdiev@gmail.com
ORCID: 0000-0003-4524-3439
https://doi.org/10.5281/zenodo.23906522

Abstract. Simultaneous localization and mapping (SLAM) estimates the pose of an autonomous mobile robot where satellite positioning is unavailable, but classical estimators assume a static world. This paper proposes a layered architecture that couples SLAM, semantic perception and navigation through bidirectional data flows and an independent safety layer, and formalises it, including extended Kalman filter (EKF) SLAM equations with analytic Jacobians, semantic masking as a measurement-selection rule and evaluation metrics. A Monte Carlo simulation (40 seeds per condition) quantifies how much masking improves localization when 0 to 40 % of 40 landmarks in a 30 m by 30 m world move at 0.2 to 1.0 m/s; the detector is a confusion model with recall r and false-positive rate f. In this simulation the absolute trajectory error (ATE) of a static-world EKF-SLAM rose from 8.3 cm to 3.7 to 6.4 m with 10 to 40 % moving landmarks. Masking alone at recall 0.9 left 1.3 to 2.6 m. Masking with chi-square gating gave 8.7 to 10.8 cm (oracle detector 8.6 to 10.5 cm) and gating alone 9.0 to 11.9 cm; the combination was better than gating alone by 3 to 10 % on average, and for creeping landmarks (0.02 to 0.1 m/s) its mean ATE was 9.9 cm against 22.5 cm, a difference concentrated in a minority of seeds. Masking pays off only with near-perfect recall, or when combined with gating or temporal voting. Limitations are the abstract detector and independent detector errors; hardware validation remains future work.

Keywords: autonomous mobile robot; simultaneous localization and mapping; semantic perception; dynamic environments; extended Kalman filter; innovation gating; Monte Carlo simulation

1. Introduction

Autonomous mobile robots (AMRs) are used in logistics, manufacturing, agriculture and inspection, where automation and control systems are adopted widely to raise operating efficiency. Simultaneous localization and mapping (SLAM) estimates the robot pose and a map jointly from odometry and exteroceptive measurements [1–3]. The probabilistic foundations [4] include the

extended Kalman filter (EKF) stochastic map [5]; current systems rely largely on graph and factor-graph optimization [6–8] or on feature-based pipelines such as ORB-SLAM [9, 10].

These formulations assume a static world. A moving landmark yields measurements that are individually plausible but inconsistent with the map, and the resulting bias is not reduced by more data. Moving objects were first handled by tracking them jointly with the map [11]; recent visual systems remove regions of potentially moving objects with semantic segmentation and geometric checks, as in DS-SLAM [12] and DynaSLAM [13], or track them in a tightly coupled back end [14] (survey: [15]). Semantic mapping builds on object-level SLAM [16], probabilistic association of uncertain classes [17] and metric-semantic systems such as Kimera [18, 19], fed by real-time detectors [20, 21]. For navigation, the Nav2 stack on ROS 2 [22, 23] combines planners such as A-star [24] and the dynamic window approach [25], tested in simulators such as Gazebo [26]; reinforcement-learning (RL) policies [27, 28] lack formal safety guarantees and need an independent safety layer, for example one based on control barrier functions [29].

Two gaps motivate this paper. First, the coupling of SLAM, semantic perception and navigation is usually a one-way pipeline, whereas mask flags and confidences must reach the estimator, pose and map context must reach the detector, and a safety layer must be able to override the planner. Second, the benefit of semantic dynamic-object handling is usually reported as end-to-end benchmark runs with one network, which confounds detector quality with estimator design and omits the cheap alternative of innovation gating [30, 31]. The contributions are: (i) a layered architecture with bidirectional data flows and an independent safety layer (Section 2.1); (ii) a formal treatment: models, MAP and factor-graph formulation, EKF-SLAM equations with Jacobians for a unicycle model and a range-bearing sensor, masking as measurement selection, an RL reward and the evaluation metrics (Sections 2.2 to 2.5); (iii) a controlled simulation quantifying the benefit of masking as a function of detector recall and false-positive rate, with gating and an oracle as references (Sections 2.6 and 3). The detector is a statistical abstraction, so the study quantifies the benefit of masking for a given detector quality and not the quality of any network; all results are simulation results.

2. Materials and Methods

2.1 System architecture

The architecture has four layers (Fig. 1). Sensing (L1) acquires camera, LiDAR, inertial and encoder data with calibration and time stamping. Perception and estimation (L2) couples a SLAM core (pose and map) with a semantic perception module (class and motion state). Planning (L3) holds a semantic map, a global planner and a local planner, classical or learned. Safety and actuation (L4) holds a safety layer and the controller. Four data flows distinguish it from a one-way pipeline: (i) pose, uncertainty and map context go from SLAM to perception; (ii) a mask flag and confidence per measurement go from perception to the estimator; (iii) class and traversability labels go to the semantic map and planners; (iv) the safety layer receives the proposed command and range data through a direct bypass and limits the command (Section 2.4). Only flows (i) and (ii) are exercised in Section 3; the navigation and safety layers are specified but not evaluated.

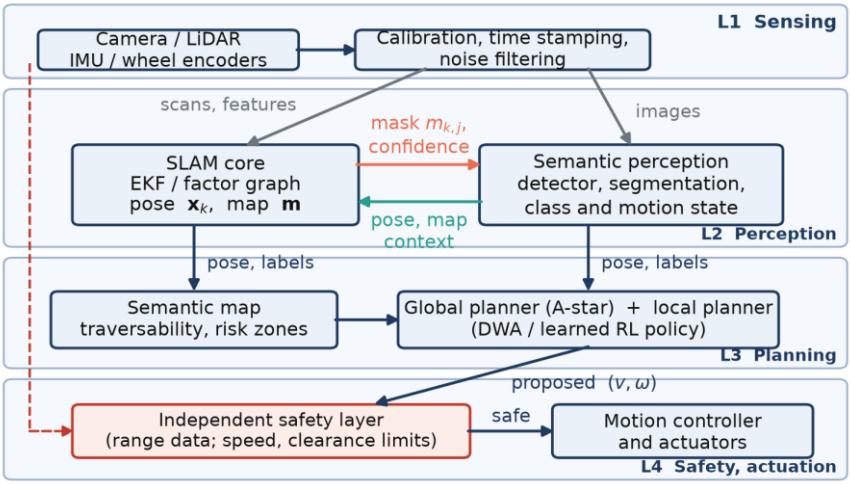


Fig. 1. Layered architecture. Solid arrows are data flows; the orange and green arrows between the SLAM core and the semantic perception module are the bidirectional coupling. The dashed red line is the direct range-data bypass to the independent safety layer, which limits the planner command before it reaches the controller.

2.2 Models, MAP formulation and semantic masking

The pose is $\mathbf{x}_k = (x_k, y_k, \theta_k)^\top$, the control $\mathbf{u}_k = (v_k, \omega_k)^\top$ (speed, yaw rate) and the map $\mathbf{m} = \{\mathbf{l}_j\}_{j=1}^{N_L}$ with $\mathbf{l}_j \in \mathbb{R}^2$. With zero-mean Gaussian noises of covariance $\mathbf{Q}_k$ and $\mathbf{R}$,

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{w}_k, \qquad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \tag{1}$$

$$\mathbf{z}_{k,j} = \mathbf{h}(\mathbf{x}_k, \mathbf{l}_j) + \mathbf{v}_{k,j}, \qquad \mathbf{v}_{k,j} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{2}$$

Let $\mathcal{O}_k$ be the set of landmarks observed at step k. The MAP estimate $\text{argmax}p(\mathbf{X}, \mathbf{m} | \mathbf{Z}_{1:K}, \mathbf{U}_{1:K})$ is a nonlinear least-squares problem on a factor graph [7]; with $\|\mathbf{e}\|_{\mathbf{\Sigma}}^2 = \mathbf{e}^\top \mathbf{\Sigma}^{-1} \mathbf{e}$ and a mask $m_{k,j} \in \{0,1\}$ per measurement, the cost is

$$J = \|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|_{\mathbf{P}_0}^2 + \sum_{k=1}^{K} \|\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_k)\|_{\mathbf{Q}_k}^2 + \sum_{k=1}^{K} \sum_{j \in \mathcal{O}_k} (1 - m_{k,j}) \|\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{x}_k, \mathbf{l}_j)\|_{\mathbf{R}}^2 \tag{3}$$

A measurement with $m_{k,j} = 1$ is removed from the graph. The semantic module flags each observed entity as dynamic ($c_{k,j} = 1$) or not, and the mask is the flag, $m_{k,j} = c_{k,j}$: a measurement is used only if $c_{k,j} = 0$. In the simulation the detector is a confusion model with recall r for dynamic and false-positive rate f for static landmarks,

$$P(c_{k,j} = 1 | j \text{ dynamic}) = r, \qquad P(c_{k,j} = 1 | j \text{ static}) = f \tag{4}$$

drawn independently per step and landmark; a fraction $1 - f$ of the useful measurements is retained and a fraction $1 - r$ of the harmful ones is admitted. Alternatively or additionally, a measurement is accepted if the squared Mahalanobis distance of its innovation $\boldsymbol{\nu}$, with covariance $\mathbf{S}$ (Eq. (11)), satisfies the chi-square gate with two degrees of freedom [30, 31]

$$d^2 = \mathbf{\nu}^\top \mathbf{S}^{-1} \mathbf{\nu} \le \tau_g, \qquad \tau_g = \chi^2_{2,\,1-\alpha} \tag{5}$$

with $\alpha = 0.01$ ($\tau_g = 9.21$) unless stated. The gate uses only geometry and knows nothing of object class; the flag uses only appearance and knows nothing of the filter state.

2.3 EKF-SLAM equations

The state is $\mathbf{s} = (\mathbf{x}^\top, \mathbf{l}_1^\top, \dots, \mathbf{l}_n^\top)^\top$ (dimension $3 + 2n$) with covariance $\mathbf{P}$. The unicycle model with mid-point heading over a step Δt is

$$\mathbf{f}(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} x + v\,\Delta t \cos\phi \\ y + v\,\Delta t \sin\phi \\ \theta + \omega\,\Delta t \end{bmatrix}, \qquad \phi = \theta + \frac{1}{2}\omega\,\Delta t \tag{6}$$

with Jacobians with respect to pose and control

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & -v\,\Delta t \sin\phi \\ 0 & 1 & v\,\Delta t \cos\phi \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \Delta t \cos\phi & -\frac{1}{2} v\,\Delta t^2 \sin\phi \\ \Delta t \sin\phi & \frac{1}{2} v\,\Delta t^2 \cos\phi \\ 0 & \Delta t \end{bmatrix} \tag{7}$$

With $\mathbf{Q}_u = \mathrm{diag}(\sigma_v^2, \sigma_\omega^2)$ and $\mathbf{P}$ partitioned into robot (rr), robot-map (rm) and map (mm) blocks, the prediction is

$$\mathbf{P}_{k|k-1} = \begin{bmatrix} \mathbf{F}\mathbf{P}_{rr}\mathbf{F}^\top + \mathbf{G}\mathbf{Q}_u\mathbf{G}^\top & \mathbf{F}\mathbf{P}_{rm} \\ \mathbf{P}_{rm}^\top \mathbf{F}^\top & \mathbf{P}_{mm} \end{bmatrix} \tag{8}$$

while the map mean and the landmark block are unchanged. The sensor measures range and bearing, $\mathbf{z} = (\varrho, \beta)^\top$; with $\Delta x = l_x - x$, $\Delta y = l_y - y$ and $s = \sqrt{\Delta x^2 + \Delta y^2}$,

$$\mathbf{h}(\mathbf{x}, \mathbf{l}) = \begin{bmatrix} s \\ \mathrm{atan2}(\Delta y, \Delta x) - \theta \end{bmatrix}, \qquad \mathbf{R} = \mathrm{diag}(\sigma_\varrho^2, \sigma_\beta^2) \tag{9}$$

$$\mathbf{H}_r = \begin{bmatrix} -\Delta x/s & -\Delta y/s & 0 \\ \Delta y/s^2 & -\Delta x/s^2 & -1 \end{bmatrix}, \qquad \mathbf{H}_l = \begin{bmatrix} \Delta x/s & \Delta y/s \\ -\Delta y/s^2 & \Delta x/s^2 \end{bmatrix} \tag{10}$$

The full Jacobian of landmark j is $\mathbf{H} = [\mathbf{H}_r\ \mathbf{0} \cdots \mathbf{H}_l \cdots \mathbf{0}]$, and the update is

$$\mathbf{S} = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^\top + \mathbf{R}, \qquad \mathbf{K} = \mathbf{P}_{k|k-1}\mathbf{H}^\top \mathbf{S}^{-1} \tag{11}$$

$$\hat{\mathbf{s}}_{k|k} = \hat{\mathbf{s}}_{k|k-1} + \mathbf{K}\mathbf{\nu}, \qquad \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}\mathbf{S}\mathbf{K}^\top \tag{12}$$

A new landmark is initialised as $\hat{\mathbf{l}} = \hat{\mathbf{x}}_{xy} + \varrho\,(\cos(\hat{\theta} + \beta), \sin(\hat{\theta} + \beta))^\top$, with Jacobians $\mathbf{G}_x = \partial\hat{\mathbf{l}}/\,\partial\mathbf{x}$, $\mathbf{G}_z = \partial\hat{\mathbf{l}}/\,\partial\mathbf{z}$ and

$$\mathbf{P}_{ll} = \mathbf{G}_x \mathbf{P}_{rr} \mathbf{G}_x^\top + \mathbf{G}_z \mathbf{R} \mathbf{G}_z^\top, \qquad \mathbf{P}_{l\cdot} = \mathbf{G}_x\,[\mathbf{P}_{rr}\ \ \mathbf{P}_{rm}] \tag{13}$$

Data association is given. The estimators differ only in the test preceding the update (Algorithm 1).

Algorithm 1. One EKF-SLAM cycle with masking (B1) and optional gating (B2, B3)
Input: $(\hat{\mathbf{s}}, \mathbf{P})$, $\mathbf{u}_k$, measurements $\mathbf{z}_{k,j}$ and flags $c_{k,j}$ for $j \in \mathcal{O}_k$, gate τ_g, mode in {B0, B1, B2, B3}.
1. Predict with Eqs. (6) and (8). 2. **for** each $j \in \mathcal{O}_k$: 3. **if** mode in {B1, B2} **and** $c_{k,j} = 1$: discard, **continue**.
4. **if** j is new: initialise with Eq. (13), **continue**. 5. Compute $\boldsymbol{\nu}$, $\mathbf{H}$ (Eq. (10)), $\mathbf{S}$ and d^2.
6. **if** mode in {B2, B3} **and** $d^2 > \tau_g$: discard (Eq. (5)), **continue**. 7. Update with Eq. (12). 8. **end for**; symmetrise $\mathbf{P}$.

2.4 Navigation reward and independent safety layer

The navigation layer is not evaluated numerically, but its interface is part of the architecture. For an RL local planner [32] the reward combines goal progress, arrival, collision, semantic risk, smoothness, energy and time,

$$r_t = c_g\,(\delta_{t-1} - \delta_t) + c_G\,I_{goal} - c_C\,I_{coll} - c_S\,\sigma(\mathbf{p}_t) - c_\omega\,|\omega_t| - c_E\,v_t^2 - c_\tau \tag{14}$$

where δ_t is the distance to the goal, I are indicators, $\sigma(\mathbf{p}_t) \geq 0$ is the semantic risk from the semantic map and $c_\cdot \geq 0$ are weights. A learned policy has no safety guarantee, so its command is limited by a safety layer using range data directly: with d_{min} the smallest clearance ahead, d_s the stopping margin and a_{max} the admissible deceleration,

$$v_{safe} = \min\{v_{cmd},\ \sqrt{2\,a_{max}\,\max(d_{min} - d_s,\ 0)}\} \tag{15}$$

A control barrier function formulation [29] generalises this limit; the principle is that safety must not depend on semantic or learned modules.

2.5 Evaluation metrics

Localization metrics follow SLAM benchmarking practice [33] (Table 1). As the filter is anchored at the known initial pose, the absolute trajectory error is computed without alignment, with $\mathbf{p} = (x, y)^\top$ the position. Consistency is assessed by the normalised estimation error squared (NEES) of the pose [3, 4], and the measurement-selection metrics are the fraction of useful (static) measurements not used and of dynamic measurements used:

$$\mathrm{ATE} = \sqrt{\frac{1}{K+1}\sum_{k=0}^{K} \|\hat{\mathbf{p}}_k - \mathbf{p}_k\|^2} \tag{16}$$

$$\varepsilon_k = (\hat{\mathbf{x}}_k - \mathbf{x}_k)^\top \mathbf{P}_{rr,k}^{-1} (\hat{\mathbf{x}}_k - \mathbf{x}_k), \qquad C_{95} = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \mathrm{I}\left[\chi^2_{3,0.025} \leq \varepsilon_k \leq \chi^2_{3,0.975}\right] \tag{17}$$

$$D = \frac{n_{stat} - n_{stat}^{used}}{n_{stat}}, \qquad \Lambda = \frac{n_{dyn}^{used}}{n_{dyn}} \tag{18}$$

Here I[·] is the indicator, the heading error is wrapped to $(-\pi, \pi]$ and $\mathcal{K}$ is the set of update steps; a consistent filter has $\bar{\varepsilon} \approx 3$ and C_{95}gtrsim0.95 (single runs, serially correlated errors). Detection metrics for a real network (precision, recall, mean average precision [8]) and the navigation metrics are defined in Table 1; recall and one minus specificity of the static class are the parameters r and f of Eq. (4).

Table 1. Evaluation metrics (extension of the metric list of the source study); the last column states whether the metric is evaluated in Section 3.

Metric	Definition and formula	Unit	Good value	Here
ATE	RMS position error to truth, Eq. (16)	m	Small	Yes
RPE	RMS error of the displacement over a 5 s window, $\mathbf{R}(\hat{\theta}_k)^\top(\hat{\mathbf{p}}_{k+\Delta} - \hat{\mathbf{p}}_k)$ against truth	m	Small (local drift)	Yes
Final error; map RMSE	Position error at the last step; RMS error of initialised static landmarks	m	Small	Yes
Consistency	$\bar{\varepsilon}$ and C_{95}, Eq. (17)	1	$\approx 3; \geq 0.95$	Yes
D, Λ	Useful measurements discarded; dynamic measurements used, Eq. (18)	%	$D \approx f, \Lambda \approx 1 - r$	Yes
Detection	Precision $= TP/(TP + FP)$, recall $= TP/(TP + FN)$, $mAP = N_c^{-1} \sum_c AP_c$	1	Close to 1	via r, f
Navigation	Success rate $= 100\, N_{succ}/N_{trials}$; collisions per trial	%; 1	100 %; 0	No
Latency; resources	$\tau = t_{out} - t_{in}$, throughput $1/\tau$; CPU/GPU, memory, energy	ms; FPS; %, W	$\tau < 200$ ms here	Estimator only

2.6 Simulation design and statistics

The study is a Monte Carlo simulation in Python (NumPy) designed to isolate one question: how much does semantic masking improve EKF-SLAM localization when a fraction ρ of the landmarks moves, as a function of r and f? The scenario is a 30 m × 30 m world with $N_L = 40$ point landmarks and a unicycle robot following an elliptical route (semi-axes 10 m and 8 m) at 1 m/s for 180 s (1800 steps of $\Delta t = 0.1$ s, about three laps, Fig. 2a). The landmark identities are known; odometry noise is $\sigma_v = 0.05$ m/s and $\sigma_\omega = 0.02$ rad/s; the sensor has range 8 m, 360 deg field of view, 5 Hz rate, $\sigma_\varrho = 0.10$ m, $\sigma_\beta = 1.5$ deg and detection probability 1; the initial pose is known and $\mathbf{Q}_u$, $\mathbf{R}$ equal the true covariances. Dynamic landmarks (a random fraction ρ) move with a speed drawn from 0.2 to 1.0 m/s and a random-walk heading, reflected at the boundary. The same world, odometry noise, measurement noise and detector random numbers are used for all estimators within a seed (common random numbers), and the dynamic set is nested across ρ, so that comparisons are paired by seed. The estimators are: **B0**, EKF-SLAM treating all landmarks as static; **B1**, masking; **B2**, masking followed by gating, Eq. (5); **B3**, gating only ($\tau_g = 5.99$ and 9.21); and the **oracle** (B1 with $r = 1$, $f = 0$). The static reference is B0 at $\rho = 0$.

The sweep covers $\rho \in \{0,0.1,0.2,0.3,0.4\}$, $r \in \{0.5,0.7,0.9,1\}$ and $f \in \{0,0.05,0.1\}$, with 40 seeds (1 to 40) per cell, plus a sensitivity run at $\rho = 0.3$, $r = 0.9$, $f = 0.05$ with dynamic speeds of 0.02 to 0.1 and 0.1 to 0.3 m/s. The comparisons were fixed before the final runs: (C1) paired ATE difference between B1, B2, B3 and B0 for every $\rho > 0$; (C2) B2 against B3, i.e. whether semantics adds to gating; (C3) the cost of false positives at $\rho = 0$; (C4) dependence on landmark speed. The primary outcome is the ATE; secondary outcomes are the final error, RPE, heading and map RMSE, C_{95}, D and Λ. The gate was chosen on the tuning seeds among $\chi^2_{2,0.95}$, $\chi^2_{2,0.99}$ and $\chi^2_{2,0.999}$ (B2 ATE at $\rho = 0.2$: 8.4, 8.3, 8.4 cm; at $\rho = 0.4$: 9.8, 9.4, 9.3 cm; differences within about 5 %), and $\alpha = 0.01$ was kept; B3 profited slightly from the stricter gate, so both are reported. The gate was tuned on 20 other seeds (9000 to 9019); nothing was tuned on the reported seeds. We report mean, standard deviation and 95 % confidence interval of the mean (Student t, 39 degrees of freedom); paired differences as mean with 95 % interval and the fraction of seeds favouring the first estimator. Intervals are not adjusted for multiplicity; medians are quoted where the ATE is skewed.

Verification and timing. The Jacobians $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}_r$, $\mathbf{H}_l$, $\mathbf{G}_x$, $\mathbf{G}_z$ agreed with central finite differences at 200 random states to within 4.6×10^{-9}; the sparse update agreed with a dense Joseph-form update on a state of 15 dimensions to within 5.6×10^{-17} (state) and 6.9×10^{-18} (covariance). In the static world the baseline had a mean NEES of 3.15 (expected 3). The run time of the Python implementation was measured on this PC (4 logical cores, Python 3.14, NumPy 2.4, single-threaded BLAS, other jobs running): for $N_L \in \{20, \ldots, 640\}$ one cycle (prediction, ten updates, symmetrisation) was timed 400 times after warm-up; median and inter-quartile range are reported. This is a real measurement of the code, excluding detector inference. Large language models were used under the authors' direction for language editing, structuring, figure drafting and for writing and running the simulation code (see Declarations).

3. Results

All numbers are simulation results (40 seeds per cell); errors are in cm unless stated, as mean (standard deviation).

3.1 The static-world filter fails when landmarks move

In the static world ($\rho = 0$) B0 reaches an ATE of 8.3 (4.6) cm with mean NEES 3.15 and $C_{95} = 0.94$. With only 10 % of the landmarks (4 of 40) moving, its ATE rises to 374 cm, 45 times the static value, and to 637 cm at $\rho = 0.4$ (Table 2), and the filter becomes overconfident: C_{95} falls from 0.94 to 0.04 at $\rho = 0.1$ and 0.01 at $\rho = 0.4$. Fig. 2b shows the cause in one seed: the error of B0 grows once a moving landmark has been initialised as static and keeps growing as it is re-observed, since each observation reinforces a wrong constraint while the landmark covariance shrinks.

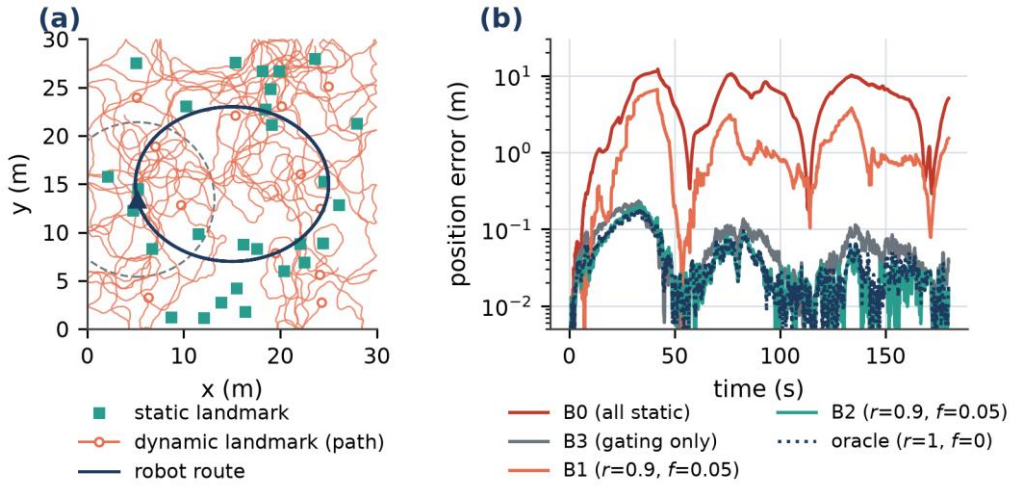


Fig. 2. Scenario and example errors (seed 1, $\rho = 0.3$). (a) World with 28 static landmarks (squares) and 12 dynamic landmarks (circles at their initial position, with their 180 s paths); navy line: robot route, triangle: robot at $t = 30$ s, dashed circle: 8 m sensor range. (b) Position error of the estimators in this seed (logarithmic axis): B0, B3 (gating only), B1 and B2 with $r = 0.9$, $f = 0.05$, and the oracle.

Table 2. ATE (cm), mean (standard deviation) over 40 seeds, $f = 0.05$ unless stated; B3 uses $\tau_g = 9.21$; the oracle has $r = 1$, $f = 0$. At $\rho = 0$ only f acts, so $r < 1$ is not applicable.

ρ	B0	B3	B1 r=0.5	B1 r=0.7	B1 r=0.9	B1 r=1	B2 r=0.9	Oracle
0.0	8.3 (4.6)	8.3 (4.3)	n/a	n/a	n/a	8.3 (4.6)	8.4 (4.3)	8.3 (4.6)
0.1	374 (193)	9.0 (4.7)	288 (188)	245 (174)	131 (63.8)	8.6 (4.6)	8.7 (4.4)	8.6 (4.6)
0.2	443 (190)	9.7 (5.2)	349 (190)	270 (136)	159 (67.1)	8.8 (5.1)	8.9 (4.7)	8.8 (5.0)
0.3	579 (202)	10.4 (5.1)	458 (191)	381 (200)	215 (137)	9.7 (5.1)	9.8 (4.8)	9.7 (5.1)
0.4	637 (187)	11.9 (5.7)	517 (194)	429 (194)	262 (184)	10.5 (5.3)	10.8 (5.2)	10.5 (5.2)

3.2 Dependence on recall, and masking combined with gating

Masking alone (B1) helps only for an almost perfect detector (Table 2, Fig. 3). With $f = 0.05$, $r = 0.9$ leaves 1.3 to 2.6 m, 15 to 25 times the oracle error, because 10.0 % of the dynamic measurements still enter the filter and one leaked measurement can initialise a moving landmark as static; only $r = 1$ restores the static accuracy (8.6 to 10.5 cm), and the false positives change the mean ATE by at most 0.05 cm. Adding the gate changes the picture: B2 attains 8.7 to 10.8 cm at $r = 0.9$ and 8.8 to 11.1 cm even at $r = 0.5$, within 0.2 cm of the oracle (8.6 to 10.5 cm). Gating alone is also effective: 9.0 to 11.9 cm with $\tau_g = 9.21$ and 9.5 to 12.3 cm with $\tau_g = 5.99$, which discards 5.1 % instead of 1.0 % of the useful measurements.

The comparison that matters for the semantic module is C2 (Table 3). B2 minus B3 is −0.3 cm at $\rho = 0.1$ and −1.2 cm at $\rho = 0.4$, i.e. 3 to 10 % of the ATE of B3, and the 95 % interval excludes zero only at $\rho = 0.2$ and $\rho = 0.4$: the semantic flags remove a small part of the error left

after gating. In the static world the paired cost of false positives is 0.02 cm for B1 and 0.08 cm for B2 (C3), small against the seed-to-seed standard deviation of 4.6 cm; with 40 landmarks a loss of 5 % of the measurements is not noticeable, and a sparser scene was not simulated.

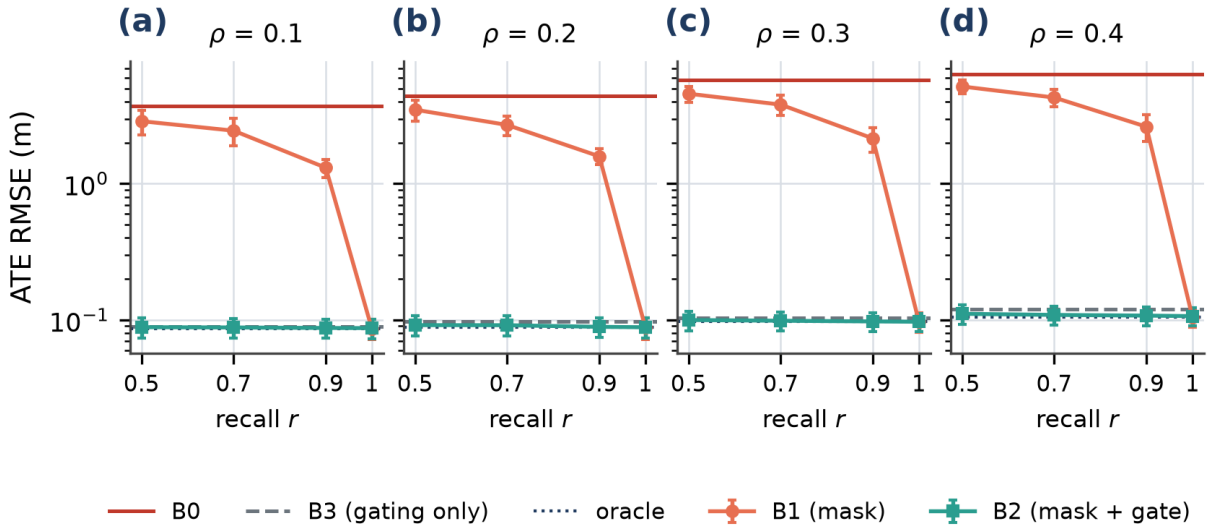


Fig. 3. ATE against detector recall r for each dynamic fraction ρ (logarithmic ordinate; f = 0.05; means with 95 % confidence intervals over 40 seeds). Horizontal lines: B0 (red), B3 (grey dashed, gating only) and the oracle (navy dotted). B1 (orange) approaches the oracle only at r = 1; B2 (teal) is close to it already at r = 0.5.

Table 3. Paired differences in ATE (cm): mean per-seed difference, 95 % interval in brackets and, in parentheses, the share of the 40 seeds in which the first estimator is better; negative favours the first. f = 0.05 for B1, B2; B3 with $\tau_g = 9.21$. At ρ = 0 the B1 and B2 rows show the cost of false positives.

Paired difference in ATE (cm)	ρ=0	ρ=0.1	ρ=0.2	ρ=0.3	ρ=0.4
B1 (r=0.9) minus B0	0.0 [−0.3, 0.3] (50%)	−243 [−296, −190] (98%)	−284 [−332, −236] (100%)	−364 [−417, −310] (100%)	−375 [−445, −306] (92%)
B2 (r=0.9) minus B0	0.1 [−0.3, 0.5] (45%)	−365 [−427, −304] (100%)	−434 [−495, −373] (100%)	−569 [−633, −504] (100%)	−626 [−686, −567] (100%)
B3 minus B0	0.0 [−0.2, 0.3] (60%)	−365 [−427, −303] (100%)	−433 [−494, −372] (100%)	−568 [−633, −504] (100%)	−625 [−685, −566] (100%)
B2 (r=0.9) minus B3	0.0 [−0.3, 0.4] (45%)	−0.3 [−0.9, 0.4] (52%)	−0.8 [−1.4, −0.2] (62%)	−0.6 [−1.6, 0.4] (57%)	−1.2 [−2.4, −0.0] (60%)
B2 (r=0.9) minus oracle	0.1 [−0.3, 0.5] (45%)	0.1 [−0.3, 0.6] (45%)	0.1 [−0.4, 0.5] (48%)	0.0 [−0.5, 0.6] (48%)	0.2 [−0.4, 0.8] (50%)

3.3 Consistency and measurement bookkeeping

At ρ = 0.2, C_{95} is 0.03 for B0 and 0.10 for B1 (r = 0.9), against 0.93, 0.94 and the static-world value 0.94 for B3, B2 and the oracle (Fig. 4a, Table 4). The counts follow the confusion model (Figs. 4b, 4c): the pure mask discards 5.0 % of the useful measurements (f = 0.05) and

admits 10.0 % of the dynamic ones (r = 0.9); B2 admits 0.8 % and discards 6.0 %; gating alone discards 1.0 % and admits 3.7 %. The gate can reject a dynamic measurement only after the landmark has been initialised, which is why B3 does not reach the oracle error.

Table 4. Secondary metrics at ho = 0.4 (means over 40 seeds; f = 0.05 for B1, B2). A consistent filter has a NEES of 3.

ρ	Estimator	Final error (cm)	RPE, 5 s (cm)	Heading RMSE (deg)	Map RMSE (cm)	C_{95}	Mean NEES	D (%)	Λ (%)
0.4	B0	404	294	35.30	711	0.01	1.4×10^{5}	0.0	100.0
0.4	B3	7.0	5.3	0.64	11.5	0.91	3.9	1.0	3.6
0.4	B1 ($r_{=0.9}$)	144	131	12.48	286	0.05	1.1×10^{4}	5.0	10.0
0.4	B2 ($r_{=0.9}$)	6.1	5.2	0.59	9.9	0.94	3.5	6.0	0.8
0.4	Oracle	6.0	5.1	0.57	9.7	0.93	3.5	0.0	0.0

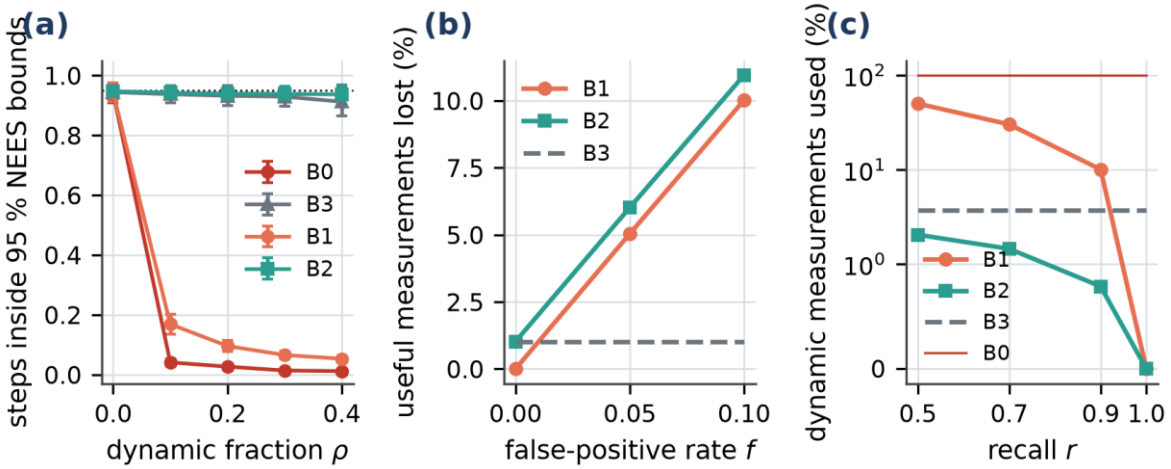


Fig. 4. Consistency and bookkeeping. (a) Fraction of steps with the pose NEES inside its 95 % bounds against ρ (r = 0.9, f = 0.05; 95 % intervals over 40 seeds; dotted line: 0.95). (b) Useful measurements not used, against f (ρ = 0.2, r = 0.9). (c) Dynamic measurements used, against r (ρ = 0.2, f = 0.05; symmetric logarithmic axis).

3.4 Landmark speed and temporal voting

Gating depends on landmark speed, since a creeping landmark stays within the gate. For speeds of 0.02 to 0.1 m/s at ρ = 0.3 (Fig. 5a) the mean ATE of B3 is 22.5 (40.5) cm, 13.2 cm median, against 9.9 (4.4) cm (median 9.1 cm) for B2 and 9.7 cm (median 8.4 cm) for the oracle; the ATE of B3 exceeds 20 cm in 7 of 40 seeds (B2: 1). The paired difference B2 minus B3 is −12.6 cm with 95 % interval −25.7 to 0.5 cm (C4), so the interval just includes zero and the benefit lies in the tail rather than in the typical seed. At 0.1 to 0.3 m/s the means are 13.6 cm (B3) and 9.9 cm (B2).

An exploratory variant, added after the main runs and therefore not pre-specified, masks a landmark if more than half of the flags received for it so far are positive. With the same random draws (f = 0.05, ρ = 0.3) this temporal vote lowers the ATE of masking alone from 458, 381 and 215 cm to 382, 31.7 and 9.5 cm at r = 0.5, 0.7 and 0.9 (Fig. 5b), and with gating (B2v) it is 9.5,

9.5 and 9.6 cm. At r = 0.5 the vote is no better than chance; at r = 0.9 masking alone reaches the oracle level. The result holds only for errors independent between frames, an optimistic assumption for systematically missed objects.

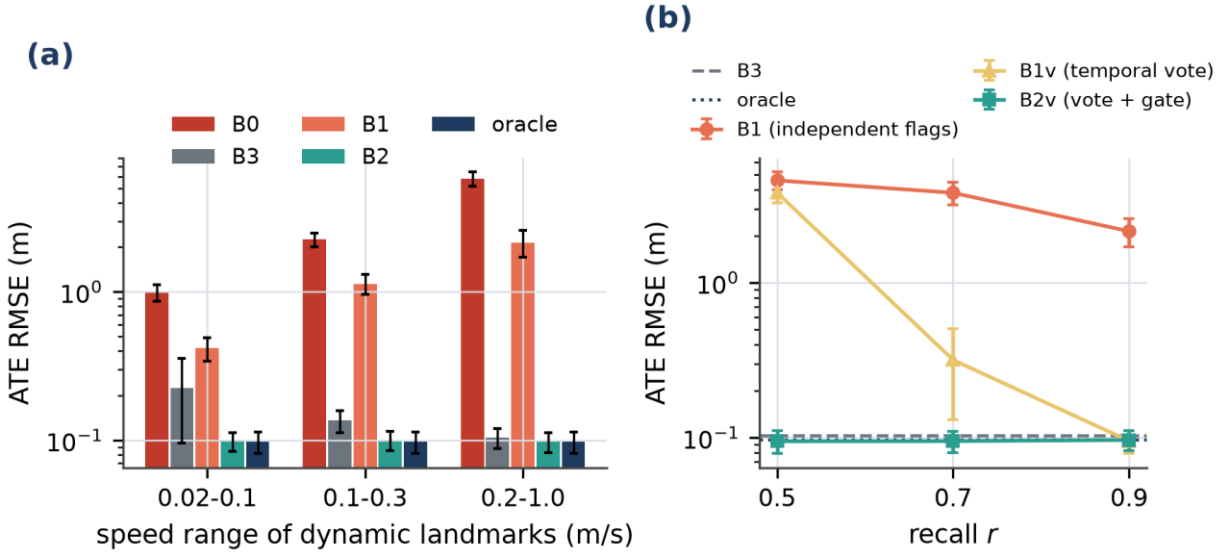


Fig. 5. Sensitivity analyses at ρ = 0.3 (means with 95 % intervals over 40 seeds, logarithmic ordinate). (a) ATE for three speed ranges of the dynamic landmarks (B1, B2: r = 0.9, f = 0.05). (b) Exploratory: ATE against r (f = 0.05) for masking with independent per-frame flags (B1), with a temporal vote (B1v), vote plus gating (B2v), and the reference lines for B3 and the oracle.

3.5 Computational cost

One filter cycle takes 0.81 ms for $N_L = 20$ and 122 ms for $N_L = 640$ (state dimensions 43 and 1283; Fig. 6a), a single update 0.07 and 9.5 ms. The empirical growth exponent in the state dimension is 1.5 over the whole range and 2.4 above $N_L = 160$, i.e. at least quadratic once the covariance (13 MB at $N_L = 640$) exceeds the cache, consistent with $O(n^2)$. The 40 landmarks of the simulation need about 0.9 ms per cycle; at $N_L = 640$ a cycle uses 61 % of the 200 ms sensing period (Fig. 6b), so large maps call for sparse back ends [6]. The mask and the gate test add no measurable cost (median cycle times at $N_L = 640$: 122 ms for B0, 118 for B1, 107 for B2). The PC was shared with other jobs, so the values are indicative.

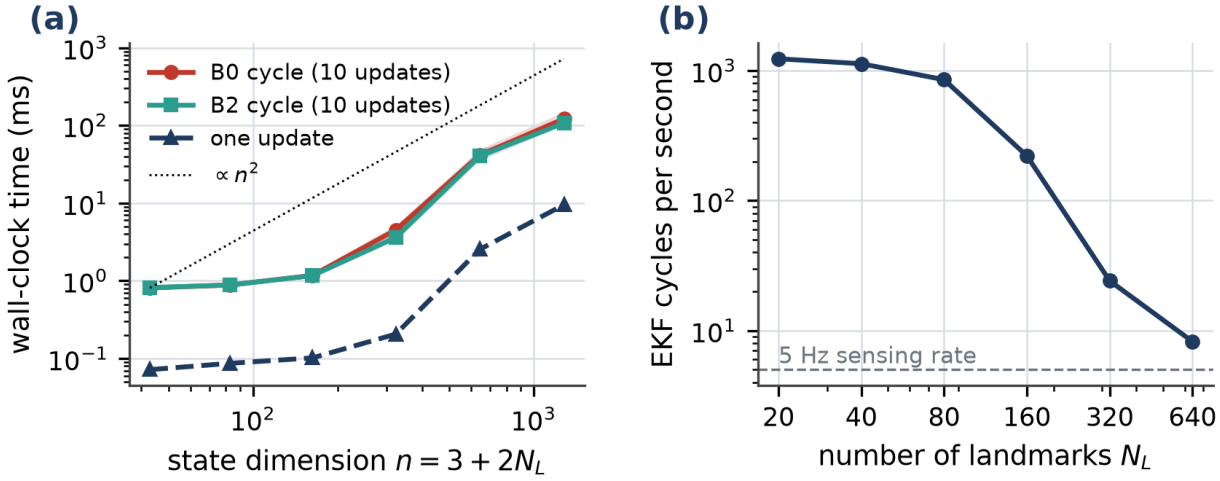


Fig. 6. Measured run time of the Python EKF-SLAM implementation on the PC of Section 2.6. (a) Median (lines) and inter-quartile range (bands) of one cycle with ten updates for B0 and the gated filter B2, and of a single update, against the state dimension $n = 3 + 2N_L$; dotted: quadratic growth. (b) Cycles per second against N_L, with the 5 Hz sensing rate. Detector inference is not included.

4. Discussion

The simulation supports four statements, conditional on the scenario. First, a static-world EKF-SLAM is not robust to a small share of moving landmarks: four moving landmarks out of 40 raised the ATE from 8.3 to 374 cm and made the filter grossly overconfident, in line with the motivation of dynamic SLAM [11, 15]; the covariance of a repeatedly observed landmark shrinks while its position keeps moving, so the error does not average out. Second, masking with frame-wise independent errors is brittle: at $r = 0.9$ the ATE stays 15 to 25 times that of the oracle, because a false negative can place a moving landmark in the map while a false positive costs one measurement among many; this favours high-recall operating points for dynamic classes, as long as enough static landmarks remain. Third, flag and gate are complementary: the gate needs no class knowledge but acts only after initialisation and admits slowly creeping landmarks, whereas the flag acts at first sight and independently of speed. In the main grid B2 and B3 differed by only 3 to 10 %; the advantage of semantics was clearer for creeping landmarks and in the tail of the error distribution, and objects that are static during mapping but moved between sessions cannot be caught by an innovation test at all. Fourth, temporal voting removed much of the brittleness, which supports returning track-level decisions with confidences (flow (ii) of Section 2.1); it cannot recover a class that is systematically missed.

DS-SLAM [12] and DynaSLAM [13] combine semantic segmentation with geometric checks, the design this abstract study supports; the numbers are not comparable, as those systems were evaluated on visual benchmarks with real networks. Tracking moving objects [11] retains information that masking discards and was not examined; the long-term inconsistency of EKF-SLAM [3, 7] is not tested by this 180 s study. For the architecture, we recommend track-level flags with confidence, an independent gate as a second line of defence, separate budgets for covariance cost and detector inference (sparse back ends [7] for large maps), and a safety layer independent of both modules [29].

Limitations. All results are simulation results. The detector is a confusion model with errors independent between landmarks and, in the main grid, between frames, and independent of range, occlusion, lighting and class; the results concern a given detector quality and not any network. Data association is given, although association errors on moving objects are a main failure source in practice. The sensor is an idealised 360-degree range-bearing sensor with detection probability one, the landmarks are planar points, the filter noise equals the true noise, and there is a single route, landmark density and 180 s horizon. B0 is deliberately naive and B3 is the more realistic reference. The gate was tuned among only three candidates, the temporal vote is exploratory, intervals are unadjusted and the means of failing filters are skewed. The timing is for Python on a shared PC and excludes the detector. The reward, safety layer and navigation metrics were not evaluated. **Future work** includes measuring the confusion statistics (with temporal correlation and range dependence) of a trained network, data association in the loop, a factor-graph back end with robust kernels, a ROS 2 pipeline with Nav2 and a physics simulator, evaluation of an RL policy with the safety filter, and real-robot validation, which has not been performed.

5. Conclusions

We proposed a layered architecture that integrates SLAM and AI modules through bidirectional data flows and an independent safety layer, and formalised it with models, EKF-SLAM equations with verified Jacobians, masking as measurement selection, an RL reward and metrics. In a controlled simulation (40 seeds), a static-world EKF-SLAM degraded from 8.3 cm to meter-scale error when 10 to 40 % of the landmarks moved; masking alone at 90 % recall left 1.3 to 2.6 m, whereas masking with gating, and nearly as well gating alone, stayed within 0.2 cm of the oracle. Semantics added a small gain over gating for landmarks moving at 0.2 to 1.0 m/s and a larger one, concentrated in a minority of runs, for creeping landmarks, and temporal voting made masking alone reliable in an exploratory test. These findings hold for an abstract detector with independent errors and a two-dimensional range-bearing sensor, and require confirmation with real detectors and on a physical robot.

Declarations

Author contributions (CRediT). Both authors contributed equally to conceptualization, methodology, software, validation, formal analysis, visualization, and writing (original draft, review and editing).

Funding. No external funding. **Competing interests.** None declared.

Data and code availability. Simulation scripts and generated data are available from the corresponding author on request.

Use of AI. Large language models (Claude, Anthropic; models Claude Fable 5.1, Claude Sonnet 5.5 and Claude Opus 5.5, accessed through Claude Code) were used under the authors' direction for language editing, structuring of the manuscript, drafting of figures, and for writing and running the simulation code reported in Section 2.6 and Section 3. The authors reviewed and verified the content and take full responsibility for it.

References

1. U. Mayinova and J. M. Abdiev, “Enhancing energy efficiency in industrial enterprises through advanced automation and control systems,” JizPI Xabarnomasi, vol. 1, no. 1, pp. 989–992, 2024.
2. T. Bailey and H. Durrant-Whyte, “Simultaneous localization and mapping (SLAM): Part II,” IEEE Robot. Autom. Mag., vol. 13, no. 3, pp. 108–117, 2006, doi: 10.1109/MRA.2006.1678144.
3. Abdiev, J. (2026a). Beyond faradaic efficiency: A closure audit of charge, carbon utilization, and energy reporting in electrochemical CO2 reduction. Journal of Electroanalytical Chemistry, 1022, Article 120675. https://doi.org/10.1016/j.jelechem.2026.120675.
4. Abdiev, J. (2026b). Data for: Reassessing isospin-resolved decay of the ^{31}P giant dipole resonance: Covariance-aware channel competition and cascade closure from archival γ-X-γ′ data [Data set]. Mendeley Data. https://doi.org/10.17632/538h8zxwxn.1.
5. Abdiev, J. M. (2022). Physical terms problems and their solutions [Working paper]. SSRN. https://doi.org/10.2139/ssrn.4024965.
6. Abdiev, J., Abdieva, N. M., & Khasanova, D. Y. (2022). Physical terms, problems and their solutions [Working paper]. SSRN. https://doi.org/10.2139/ssrn.4066756.
7. Abdiev, J., & Malikov, R. (2022). The role of teacher and student cooperation in the modern educational process of Google Docs. Web of Scientist: International Scientific Research Journal, 3(1), 579–589.
8. Abdiev, J., & Safarov, O. (2022). Basalt fiber-basic (primary) concepts. PhilPapers. https://philpapers.org/rec/ABDBF.
9. Abdiev, J., Safarov, O., & Julanov, H. (2022). Study of the properties of polymer composites: Reinforcement based on glass and basalt fibers. Eurasian Scientific Herald, 7, 77–88.
10. Abdiev, J. M., Leyderman, A. Yu., Saidov, A. S., & Usmonov, Sh. N. (2021, April 21). Perenos toka v poluprovodnikovykh p-n-strukturakh s bazovoy n-oblasti iz varizonnogo poluprovodnika [Current transport in semiconductor p-n structures with a base n-region made of a graded-gap semiconductor]. In Yosh olimlar va fizik talabalarning I respublika ilmiy anjumani (YoOFTRIA-I) materiallari [Proceedings of the 1st republican scientific conference of young scientists and physics students (YoOFTRIA-I)] (Vol. 1, pp. 114–117). Tashkent. (in Russian).
11. Abdiev, J. M., & Ziyadullaev, Z. R. (2022, April 22). Si taglikka o‘stirilgan (Si2)1-x(GaP)x qattiq qorishma epitaksial qatlamining kimyoviy tarkibi va morfologiyasini o‘rganish [Study of the chemical composition and morphology of the epitaxial layer of the (Si2)1-x(GaP)x solid solution grown on a Si substrate]. In Fizika fanining rivojida iste‘dodli yoshlarning o‘rni (RIAK-XV): Respublika ilmiy-amaliy konferensiyasi ishlari to‘plami [The role of talented youth in the development of physics (RIAK-XV): Proceedings of the republican scientific-practical conference] (Vol. 2, No. 15, pp. 184–186). (in Uzbek).
12. Abdiyev, J. M., Leyderman, A. Yu., Saidov, A. S., & Usmonov, Sh. N. (2021). Injection diffusion processes in the weak linear graded-band semiconductor p-n-structures. Euroasian Journal of Semiconductors Science and Engineering, 3(1), 20–26. https://doi.org/10.37681/2181-1652-019-x-2021-1-3.

13. Kenjayev, S., & Abdiev, J. (2024). Expanding the role of mathematical modeling in quantum computers: The application of theoretical physics in next-generation computing technologies. Journal of Integrated Education and Research, 3(5), 149–162. https://ojs.rmasav.com/index.php/ojs/article/view/1675.
14. Leyderman, A. Yu., Saidov, A. S., Abdiyev, J. M., Usmonov, Sh. N., & Suyarov, Q. T. (2021, August 15). Influence of weak graded gap on injection diffusion regimes of current transport in semiconductor p-n-structures. In USBIK 2021 Online International Congress on Natural Sciences (Vol. 8, pp. 40–44).
15. Mayinova, U., & Abdiev, J. M. (2024). Enhancing energy efficiency in industrial enterprises through advanced automation and control systems. JizPI Xabarnomasi, 1(1), 989–992.
16. Mustafayev, T., Akhmedzhanov, F., Abdiev, J., & Khasanov, S. (2025a). Correction: An exploration of anisotropic acoustic wave attenuation in quartz crystals. Brazilian Journal of Physics, 55(1), Article 18. https://doi.org/10.1007/s13538-024-01657-x.
17. Mustafayev, T., Akhmedzhanov, F., Abdiev, J., & Khasanov, S. (2025b). An exploration of anisotropic acoustic wave attenuation in quartz crystals. Brazilian Journal of Physics, 55(1), Article 1. https://doi.org/10.1007/s13538-024-01638-0.
18. Payzullaev, A. N., Allaev, B. A., Mirzaev, S. Z., Abdiev, J. M., Urinov, J., & Parkash, A. (2023). The impact of silicon dioxide nanoparticle size on the viscosity and stability of nanofluids: A comprehensive study. ECS Advances, 2(3), Article 031001. https://doi.org/10.1149/2754-2734/ace121.
19. Saidov, A. S., Saparov, D. V., Usmonov, Sh. N., Kutlimratov, A., Abdiev, J. M., Kalanov, M., Razzakov, A. Sh., & Akhmedov, A. M. (2021). Investigation of the crystallographic perfection and photoluminescence spectrum of the epitaxial films of (Si2)1-x(GaP)x ($0 \leq x \leq 1$) solid solution, grown on Si and GaP substrates with the crystallographic orientation (111). Advances in Condensed Matter Physics, 2021, Article 3472487. https://doi.org/10.1155/2021/3472487.
20. Saidov, A. S., Usmonov, Sh. N., Karshiev, A. B., & Abdiev, J. M. (2022). Influence of the varizonal Si1-xGex solid solution composition on the thermovoltaic effect in n-Si-p-(Si)1-xGex structure. IOP Conference Series: Earth and Environmental Science, 1112(1), Article 012040. https://doi.org/10.1088/1755-1315/1112/1/012040.
21. Urinov, Kh. O., Jumanov, Kh. A., Khidirov, A. M., Urinov, S. Kh., Abdiyev, J. M., Jumaboyev, T. A., & Eshmirzayev, M. R. (2020). Magnetocaloric effect in polycrystalline cobalt. Journal of Physics: Conference Series, 1515(2), Article 022079. https://doi.org/10.1088/1742-6596/1515/2/022079.
22. Taylanov, N. A., Urinov, S. Kh., & Abdiev, J. M. (2022). A fourth-order Runge–Kutta method for numerical solution of the Kuramoto–Sivashinsky equation.
23. Urinov, J. O., Abdiev, J. M., & Malikov, R. R. (n.d.). FexV3−xO4: Magnetic and electrical properties of spinel.
24. Komilovich, J. L. H., Ugli, A. A. M., Safarov, O., Ugli, A. J. M., & Ugli, Y. A. A. (2022). Understanding the magneto caloric effect in super spin glass cobalt-based nanoparticles.
25. D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” IEEE Robot. Autom. Mag., vol. 4, no. 1, pp. 23–33, 1997, doi: 10.1109/100.580977.

26. N. Koenig and A. Howard, “Design and use paradigms for Gazebo, an open-source multi-robot simulator,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), vol. 3, 2004, pp. 2149–2154, doi: 10.1109/IROS.2004.1389727.
27. V. Mnih et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.
28. L. Tai, G. Paolo, and M. Liu, “Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2017, pp. 31–36, doi: 10.1109/IROS.2017.8202134.
29. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, “Control barrier function based quadratic programs for safety critical systems,” IEEE Trans. Autom. Control, vol. 62, no. 8, pp. 3861–3876, 2017, doi: 10.1109/TAC.2016.2638961.
30. J. Neira and J. D. Tardós, “Data association in stochastic mapping using the joint compatibility test,” IEEE Trans. Robot. Autom., vol. 17, no. 6, pp. 890–897, 2001, doi: 10.1109/70.976019.
31. Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, Estimation with Applications to Tracking and Navigation. New York, NY, USA: Wiley, 2001, doi: 10.1002/0471221279.
32. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018, ISBN 978-0-262-03924-6.
33. J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of RGB-D SLAM systems,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2012, pp. 573–580, doi: 10.1109/IROS.2012.6385773.