SFE-05.12b

The Innovation Floor-Lock Theorem

Observation Limits in a Self-Field-Following Agent under Optimal Kalman Filtering

Jesus David Calderas Cervantes

Independent Research

Version: SFE-05.12b — v4 (Extended with 05.13b resolution)

Date: February 2026

Status: Floor-lock confirmed (05.12b) · Architectural resolution confirmed (05.13b)

Abstract

We present a structural analysis of detection failure in a self-referential adaptive perception architecture. An SFE agent tracks a one-dimensional stochastic field via a matched Kalman filter and monitors prediction surprise through a self-regulating windowed gate. We prove that once the filter reaches steady state, the predictable component of state evolution is absorbed by the estimator — leaving a whitened residual that carries no information about the field's confinement regime k. The normalized residual magnitude locks at a universal value:

$$\mathbb{E}[\tilde{\varepsilon}] \;=\; \sqrt{\tfrac{2}{\pi}}\,\frac{\sigma_m}{V_{\text{field}}} \;\approx\; 0.714 \qquad (\forall\, k \geq 0,\ \text{under matched filtering}),$$

independent of field-confinement strength. This is a structural detection-impossibility result: a gate downstream of a matched adaptive estimator cannot detect the regime

that estimator was built to track. The mathematical result (Kalman innovation whitening) is classical; the contribution is the *detection-impossibility framing* and the architectural diagnosis of why the detection channel must fail.

Scope: the impossibility holds for linear Gaussian processes with correct model parameters at steady state. It breaks under model mismatch, time-varying k, or when the detection observable bypasses the estimator's residual channel.

Section 9 presents the architectural resolution confirmed in SFE-05.13b: decoupling the detection channel from the adaptive estimator (null predictor with fixed reference $\hat{x} = 0$) restores geometric separability at 3.02σ — confirming that the information erased from the residual channel was never lost from position space.

SFE lineage: 05.09 (initial gate) → 05.10 (reset artefact) → 05.11 (per-cycle baseline bug) → 05.12 (windowed gate) → 05.12b (floor-lock confirmed) → 05.13 (Kalman still floor-locked) → **05.13b (null predictor, burst detected ✓)**

Contents

1. Introduction and Motivation

The SFE series investigates whether a stochastic agent—one that moves through a one-dimensional field, measures its own position imperfectly, and uses those measurements to build a predictive model—can autonomously detect changes in the field's statistical character. In physical terms, the agent attempts to distinguish a freely-diffusing particle ($k = 0$, Brownian motion) from one confined by an Ornstein-Uhlenbeck (OU) restoring force with spring constant $k > 0$.

The detection signal of interest is the *innovation ratio*: the normalised absolute prediction error that the agent's Kalman filter registers at each measurement cycle. A self-regulating windowed gate monitors rolling averages of this ratio and triggers a *coherence event* whenever the ratio falls sufficiently below a dynamically computed baseline—the interpretation being that the field has become *more predictable* than the agent's baseline model.

SFE-05.12b is the first version free of two compounding artefacts that contaminated all earlier runs:

1. **Periodic Kalman resets.** Resetting the error covariance P every fixed number of cycles temporarily inflated the innovation magnitude, causing the ratio to oscillate between artificially high and low values and producing spurious gate crossings.
2. **Per-cycle baseline mismatch.** In SFE-05.12 the gate baseline was computed from individual-cycle innovation ratios (standard deviation $\approx$0.54), while the gate check compared windowed means. The mismatch rendered the gate ceiling negative.

With both artefacts removed, the simulation reaches a clean steady state and the floor-lock emerges as a fundamental structural property of the system, not a numerical accident.

1.1. Scope and Novelty

The core mathematical fact underlying this result—that a matched Kalman filter whitens its innovations—is classical, tracing to discrete-time estimation theory.[1] In standard practice, innovation whiteness is used as a *model adequacy test*: white innovations confirm the model fits. The SFE series applies the contrapositive as a *structural detection-impossibility result*:

> *A gate downstream of a matched adaptive estimator cannot detect the regime that estimator was built to track.*

The mathematical result is not new. The architectural framing — that estimation and detection must be structurally separated to preserve regime contrast — is the contribution.

Relation to prior work. CUSUM [3] and likelihood-ratio detectors [4] avoid this failure by operating on the *likelihood* of the innovation sequence under competing hypotheses, not on its normalised magnitude. They succeed precisely because they maintain an explicit mismatch between a null model and an alternative — that is, they never fully commit to the matched-filter assumption. The present result identifies the matched-filter condition as the root cause of detection failure: any innovation-based statistic, including likelihood ratios, will degrade toward chance once the filter is matched and has converged. Friston's free-energy framework [5] describes the same mechanism from a neuroscientific perspective

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82(1):35–45, 1960.

— prediction-error minimisation as the objective of an adaptive agent — but does not address the architectural consequence for downstream detectors. The SFE floor-lock theorem formalises this consequence and demonstrates the decoupled-channel remedy.

Scope conditions under which the theorem holds:

- Linear time-invariant (LTI) Gaussian state-space model — the OU and Brownian cases are both instances; the result is not OU-specific
- Kalman model parameters correctly specified (matched filter)
- Asymptotic: Riccati equation converged to P_{ss}
- Detection operates on the estimator's own residual channel

LTI generalization: the whitening argument depends only on the LTI Gaussian structure and matched-filter convergence. Any LTI state-space model $\{A, B, C, D\}$ with a correctly specified Kalman filter will produce whitened innovations at steady state. The OU process is the SFE physical instance; the impossibility is a property of the architecture class, not of the physics.

Conditions under which the theorem does *not* hold (and detection becomes possible):

- Deliberate model mismatch (null predictor, fixed reference) — SFE-05.13b
- Time-varying k faster than filter convergence time
- Nonlinear processes or heavy-tailed noise (innovations no longer Gaussian)
- Detection observable that bypasses the residual channel entirely (e.g. cumulative spatial variance)

Stating these boundaries prevents overgeneralization and makes the theorem falsifiable: any violation of the matched-filter condition should restore gate functionality. SFE-05.13b confirms this directly.

2. System Definition

2.1. Field Dynamics

The underlying field is a one-dimensional probability density $\rho(x, t)$ evolving under the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}\left[\frac{F(x) - k(x - x_0)}{\gamma}\,\rho\right] + D\,\frac{\partial^2 \rho}{\partial x^2}, \tag{1}$$

where $F(x)$ is the externally applied force (zero in all runs reported here), $k \geq 0$ is the OU spring constant, $x_0 = 0$ is the equilibrium position, $\gamma = 1$ is the damping coefficient, and $D = k_{\text{B}}T/\gamma$ is the diffusion coefficient (with $k_{\text{B}}T = 1$ throughout this report).

The three field regimes tested are:

Run	k	Physical interpretation
1	0.0	Free diffusion (Brownian motion)
2	1.0	Weak OU confinement
3	5.0	Strong OU confinement

2.2. Agent Dynamics

The agent's position $x(t)$ evolves via the Langevin equation

$$\gamma\,\dot{x} \;=\; F_{\text{field}}(x) + F_{\text{OU}}(x) + F_{\text{attr}}(x) + \xi(t), \tag{2}$$

where $\xi(t)$ is Gaussian white noise with $\langle \xi(t)\xi(t')\rangle = 2k_{\text{B}}T\gamma\,\delta(t-t')$, F_{field} is a field-following force derived from the Fokker-Planck flux (coupling strength λ), $F_{\text{OU}} = -k(x - x_0)$ is the OU confinement force, and F_{attr} is an optional centroid attraction force active only after sufficient coherence events.

2.3. Kalman Filter

Every $\tau_{\text{meas}} = 10$ simulation steps the agent receives a noisy measurement

$$z_n \;=\; x_n + \sigma_m\,\eta_n, \qquad \eta_n \sim \mathcal{N}(0,1), \tag{3}$$

with $\sigma_m = 0.40$. The agent maintains a one-dimensional Kalman filter $(\hat{x}_n, P_n)$ where $\hat{x}_n$ is the state estimate and P_n is the error covariance.

Prediction step (no-reset, τ_{meas} steps of diffusion):

$$\hat{x}_n^- \;=\; \alpha\,\hat{x}_{n-1}^+, \tag{4}$$

$$P_n^- \;=\; \alpha^2 P_{n-1}^+ + Q_n, \tag{5}$$

where $\alpha = e^{-k\,\tau_{\text{meas}}\,\Delta t}$ (with $\alpha = 1$ for $k = 0$) and Q_n is the process noise covariance for each regime.

Update step:

$$K_n \;=\; \frac{P_n^-}{P_n^- + \sigma_m^2}, \tag{6}$$

$$\varepsilon_n \;=\; z_n - \hat{x}_n^-, \tag{7}$$

$$\hat{x}_n^+ \;=\; \hat{x}_n^- + K_n\,\varepsilon_n, \tag{8}$$

$$P_n^+ \;=\; (1 - K_n)\,P_n^-. \tag{9}$$

2.4. Innovation Ratio

The raw innovation $\varepsilon_n = |z_n - \hat{x}_n^-|$ is normalised by the *field volatility*

$$V_{\text{field}} \;=\; \sqrt{2D\,\tau_{\text{meas}}\,\Delta t} \;=\; \sqrt{2\cdot 1\cdot 10\cdot 0.01} \;=\; \sqrt{0.2} \;\approx\; 0.4472, \tag{10}$$

giving the dimensionless ratio

$$\tilde{\varepsilon}_n \;=\; \frac{|\varepsilon_n|}{V_{\text{field}}}. \tag{11}$$

The windowed gate monitors the 40-cycle mean of $\tilde{\varepsilon}_n$.

3. The Innovation Floor-Lock: Analytical Derivation

3.1. Steady-State Kalman Convergence

Without periodic resets, each run's error covariance P_n converges to a steady-state value P_{ss} that satisfies the discrete-time algebraic Riccati equation. Once at steady state, the Kalman gain K_{ss} and the post-update covariance P_{ss}^{+} are constants.

Numerically (first 500 measurement cycles):

Run	k	P_{ss} (empirical)
1	0.0	0.1049
2	1.0	0.0994
3	5.0	0.0788

3.2. Innovation Variance at Steady State

At steady state the prior prediction is $\hat{x}_n^{-} \approx \alpha\, x_{n-1}$ (because $\hat{x}_{n-1}^{+} \approx x_{n-1}$ for a well-converged filter). Expanding the innovation:

$$\begin{aligned} \varepsilon_n &= z_n - \hat{x}_n^{-} \\ &= \big(x_n + \sigma_m\, \eta_n\big) - \alpha\, x_{n-1} \\ &= \big(x_n - \alpha\, x_{n-1}\big) + \sigma_m\, \eta_n. \end{aligned} \tag{12}$$

Taking the variance (with x_n and η_n independent):

$$\text{Var}[\varepsilon_n] \;=\; \text{Var}[x_n - \alpha\, x_{n-1}] \;+\; \sigma_m^2. \tag{13}$$

Theorem 3.1 (Innovation Floor-Lock — Estimator Absorption of Regime Contrast). ***Given:***

(i) a linear time-invariant (LTI) Gaussian state-space model with matched filter parameters — the OU process is the SFE instance, but the argument extends to any LTI Gaussian model;

(ii) asymptotic convergence: $P_n \to P_{\text{ss}} = \text{const}$ (Riccati equation settled, transient phase complete);

(iii) fixed measurement noise σ_m and fixed field volatility V_{field};

(iv) field confinement strength $k \in [0, \infty)$.

Then *(asymptotically, after Riccati convergence):*

$$\mathbb{E}[\tilde{\varepsilon}] \;=\; \sqrt{\tfrac{2}{\pi}}\,\frac{\sigma_m}{V_{\text{field}}} \;\approx\; 0.714 \qquad \textit{independent of } k.$$

Architectural consequence: *the predictable component of state evolution is absorbed by the estimator at convergence. The residual channel contains only unpredictable components and carries no information about the confinement regime k. A threshold gate downstream of this channel cannot distinguish free diffusion ($k = 0$) from OU confinement ($k > 0$).*

During the transient phase (before $P_n \approx P_{\text{ss}}$), the filter is still adapting and some first-order regime contrast may persist in the innovation channel. The floor-lock is an asymptotic property of convergence, not of the initial response.

Proof. At steady state, the innovation variance decomposes exactly as:

$$\text{Var}[\varepsilon_n] = P_{\text{ss}}^- + \sigma_m^2, \tag{14}$$

where P_{ss}^- is the prior error covariance at steady state. The Riccati equation ensures the filter has internalized all predictable state dynamics into P_{ss}^-. What remains in the innovation is the component that cannot be predicted — under correct modeling, this is indistinguishable from measurement noise.

Asymptotic qualifier: during the transient phase ($P_n \not\approx P_{\text{ss}}$), the filter is still adapting and may not have fully absorbed regime contrast. Equation (14) becomes exact only after convergence. In SFE runs, convergence occurs within ≈200 cycles out of 2000, so the asymptotic regime dominates the observable record.

Substituting numerical parameters ($\sigma_m = 0.40$, $k_{\text{B}}T = 1$, $\Delta t = 0.01$, $\tau_{\text{meas}} = 10$):

$$V_{\text{state}}^2(k) = (1-\alpha)^2 \,\text{Var}[x] = (1-\alpha)^2 \cdot \frac{k_{\text{B}}T}{k} \quad (k > 0),$$
$$V_{\text{state}}^2(0) = 2D\,\tau_{\text{meas}}\,\Delta t = 0.20.$$

k	α	V_{state}^2	σ_m^2	$V_{\text{state}}^2/\sigma_m^2$
0.0	1.000	0.200	0.160	1.25
1.0	0.905	0.009	0.160	0.056
5.0	0.607	0.150	0.160	0.94

For $k = 1.0$, P_{ss}^- is small enough that the unpredictable residual is dominated by measurement noise. For $k = 0$ and $k = 5.0$, P_{ss}^- is comparable to σ_m^2, but crucially: the total $\text{Var}[\varepsilon_n] = P_{\text{ss}}^- + \sigma_m^2$ does not vary across regimes in a way that opens a detectable gap at the gate. The key point is not that noise numerically dominates — it is that the estimator has absorbed the predictable structure of each regime into its own covariance, leaving a whitened residual that encodes only unpredictable fluctuations, not regime identity. □

3.3. Expected Ratio at Steady State

Because the innovation ε_n is approximately Gaussian ($\varepsilon_n \sim \mathcal{N}(0, \sigma_m^2)$ at steady state), the expectation of its absolute value is:

$$\mathbb{E}[|\varepsilon_n|] \;=\; \sqrt{\frac{2}{\pi}}\,\sigma_m. \tag{15}$$

Therefore the expected windowed ratio is:

$$\boxed{\mathbb{E}[\tilde{\varepsilon}] \;=\; \frac{\mathbb{E}[|\varepsilon_n|]}{V_{\text{field}}} \;=\; \sqrt{\frac{2}{\pi}}\,\frac{\sigma_m}{V_{\text{field}}} \;=\; \sqrt{\frac{2}{\pi}}\,\frac{0.40}{0.4472} \;\approx\; 0.714.} \tag{16}$$

This value is *independent of* k, confirming the floor-lock: the gate observes the same ratio regardless of whether the field is freely diffusing or strongly confined.

3.4. Gate Ceiling vs. Floor

The self-regulating gate computes its ceiling as:

$$\theta_{\text{ceil}} \;=\; \bar{\tilde{\varepsilon}}_{\text{bl}} \;-\; z_\sigma \cdot \text{std}(\tilde{\varepsilon}_{\text{bl}})\,, \tag{17}$$

where $\bar{\tilde{\varepsilon}}_{\text{bl}}$ is the mean over the last $N_{\text{bl}} = 10$ windows (each of width 40 cycles) and $z_\sigma = 1.5$.

Proposition 3.2 (Structural Gate Inversion). *When $\mathbb{E}[\tilde{\varepsilon}] \approx 0.714$ for all k, the baseline mean converges to $\approx$0.714 regardless of the run. The gate ceiling is then*

$$\theta_{\text{ceil}} \;\approx\; 0.714 - 1.5 \times \text{std}(\tilde{\varepsilon}_{\text{bl}}) \;\approx\; 0.60,$$

which is below *the floor* 0.714. *Gate crossings ($\tilde{\varepsilon} < \theta_{\text{ceil}}$) can only occur when the rolling window randomly dips due to measurement noise—not due to genuine field-structure changes. All such crossings are correctly rejected by the persistence verification step.*

Observed ceiling values in SFE-05.12b:

Run	k	Mean ratio	Mean ceiling
1	0.0	0.715	0.602
2	1.0	0.744	0.606
3	5.0	0.755	0.609

4. Simulation Setup

4.1. Parameters

Symbol	Value	Description
$k_{\text{B}}T$	1.0	Thermal energy
γ	1.0	Damping coefficient
D	1.0	Diffusion coefficient ($= k_{\text{B}}T/\gamma$)
Δt	0.01	Integration time step
N	20,000	Total Langevin steps
τ_{meas}	10	Steps per measurement cycle
σ_m	0.40	Measurement noise std. dev.
V_{field}	0.4472	Field volatility ($= \sqrt{2D\tau\Delta t}$)
λ_0	0.30	Initial field-coupling strength
σ_{mem}	1.20	Perceived-field memory bandwidth
N_{cycles}	2,000	Measurement cycles ($= N/\tau_{\text{meas}}$)
W	40	Coherence window width (cycles)
N_{bl}	10	Baseline windows for gate
z_σ	1.5	Gate threshold (z-score below baseline)
$[x_{\min}, x_{\max}]$	$[-8, 8]$	Spatial domain
N_x	400	Spatial grid points

4.2. Agent Architecture

The agent consists of five interacting components:

1. **Fokker-Planck propagator.** The field density $\rho(x,t)$ is updated each step by a first-order upwind finite-difference scheme applied to eq. (1).
2. **Perceived field.** The agent maintains a private belief distribution $\rho_{\text{perc}}(x)$ built from a weighted sum of Gaussian kernels centred on past positions, with bandwidth σ_{mem}. The Fokker-Planck flux of ρ_{perc} provides the field-following force F_{field}.
3. **Kalman filter (no-reset).** The $(\hat{x}, P)$ pair is propagated continuously without any periodic re-initialisation. P converges to P_{ss} within the first ≈200 cycles.
4. **Circular core buffer.** A ring buffer of length 512 stores the most recent (field-flux magnitude, alignment, surprise) triplets for the coherence-state classifier.
5. **Windowed self-regulating gate.** Described in section 4.3.

4.3. Windowed Self-Regulating Gate

At the end of each 40-cycle window the gate:

1. Appends the window's mean ratio $\bar{\tilde{\varepsilon}}$ to a history buffer.
2. Computes the baseline from the last $N_{\text{bl}} = 10$ entries.
3. Computes the ceiling via eq. (17).
4. Checks for a primary crossing: $\bar{\tilde{\varepsilon}}_{\text{cur}} < \theta_{\text{ceil}}$.
5. Verifies persistence: the *previous* window's ratio must also have been below θ_{ceil}.
6. If both checks pass *and* structural conditions hold (relative variance $r_v < 0.15$, core rate ≥ 0.20, buffer has ≥ 5 points), a coherence event is logged.

The bug fixed in 05.12b was the use of a *per-cycle* baseline in 05.12 (standard deviation ≈0.54) to set a threshold compared against *windowed* means—producing a ceiling < 0. In 05.12b both the baseline and the check use the same windowed quantity.

5. Results

5.1. Kalman Convergence

All three runs converge within 200 measurement cycles, well within the 2000-cycle simulation. The no-reset architecture eliminates the transient ratio spikes that caused false positives in versions 05.09–05.12.

5.2. Innovation Ratio

Run	k	$\bar{\tilde{\varepsilon}}$	std($\tilde{\varepsilon}$)	min($\tilde{\varepsilon}$)	Predicted floor
1	0.0	**0.7147**	–	0.50	0.714
2	1.0	0.7435	–	0.58	0.714
3	5.0	0.7546	–	0.59	0.714

Run 1 ($k = 0$) falls within 0.04% of the theoretical floor. The slight positive elevation for $k > 0$ is consistent with the non-negligible V^2_{state} contribution detailed in the proof of Theorem 3.1—not an anomaly, but a predicted residual. The minimum windowed ratio improved from ≈0.001 in reset-contaminated versions to ≈0.50 here, confirming artefact elimination.

5.3. Gate Events

Run	k	Mean ratio	Ceiling	Crossings	Verified	Events
1	0.0	0.7147	0.5996	0	0	0
2	1.0	0.7435	0.6056	0	0	0
3	5.0	0.7546	0.6089	0	0	0

Zero gate crossings across all three runs. **This is the correct result.** Under the theorem's conditions, no crossing should occur: the ceiling (≈0.60) sits structurally below the floor (≈0.714). The gate is not malfunctioning; it is providing the correct output given a signal that contains no regime-discriminating information. Prior versions produced crossings and events precisely because artefacts (reset spikes, baseline mismatch) generated apparent structure in the innovation channel. Their absence here is the validation.

The system is in state EXPLORING, consistent with an agent that has received no structured signal detectable via innovation magnitude. The convex-hull manifold shows expanding volume (early $V \approx 20.31$, final $V \approx 25.95$), confirming active state-space exploration rather than attractor collapse.

6. Discussion

6.1. What the Whitening Theorem Actually Says

The floor-lock is a consequence of the *Kalman innovation whitening theorem*:

Theorem 6.1 (Innovation Whitening — Kalman, 1960)**.** *If the Kalman filter is matched to the true system dynamics, the sequence of innovations $\{\varepsilon_n\}$ forms a zero-mean white noise process with covariance $S_n = P_n^- + \sigma_m^2$. At steady state $S_n = P_{\text{ss}}^- + \sigma_m^2 = const$, and the innovations are i.i.d. Gaussian regardless of the field structure.*

The precise mechanism is this: the filter does not erase state information — it *internalizes* the predictable component of state evolution into its own error covariance P_n. Once P_n converges, all predictable structure has been absorbed. What remains in the residual is the component that was genuinely unpredictable under the correct model — which, for a linear Gaussian process, is indistinguishable from measurement noise. The residual channel therefore carries no regime identity.

This is not a numerical coincidence. It is the mechanism by which an optimal estimator succeeds: by removing everything it can predict, it leaves only what it cannot. A gate that measures that residual is asking: *how surprised is the estimator?* The answer is always the same — equally surprised, regardless of k — because surprise is what the estimator has been designed to minimize.

6.2. Estimation-Detection Coupling Principle

The floor-lock is an instance of a more general architectural constraint:

Estimation-Detection Coupling Principle.

If a detection observable is identical to an estimator's residual under matched filtering, regime contrast is asymptotically suppressed as the estimator converges.

Consequence: a detector that operates on a statistic the estimator is actively minimizing is guaranteed to fail at steady state, regardless of gate architecture or threshold tuning.

Resolution: detection and estimation must use structurally separate observables. The detection channel must not pass through the estimator's whitening stage.

This principle is not specific to Kalman filters or OU processes. It applies to any architecture in which a matched adaptive estimator feeds its own residuals to a downstream detector. The SFE gate failed for exactly this reason across versions 05.09–05.13. SFE-05.13b resolves it by routing the detection channel around the estimator entirely.

6.3. Architectural Exits

Given Theorem 3.1 and the coupling principle, two structural exits exist:

1. **Deliberate model mismatch.** Fix the Kalman model to $k = 0$ while the field operates at $k > 0$. The mismatched filter produces non-white residuals during the OU phase; the baseline separates from the floor. This is the approach used in CUSUM and likelihood-ratio detectors, which specify null and alternative hypotheses explicitly — they are always operating with a *mismatched* null model by design.
2. **Bypass the residual channel.** Use an observable that is never passed through the estimator's whitening stage. The null predictor ($\hat{x} = 0$) is the simplest case: it preserves geometric contrast in position space that the Kalman update would cancel. Cumulative spatial variance of a decoupled sensor particle (Section 7) is a second example.

SFE-05.13b implements option (2).

6.4. Connection to Prior SFE Lineage

The floor value $\mathbb{E}[\tilde{\varepsilon}] \approx 0.714$ was observed empirically in SFE-05.08 and attributed to parameter tuning. SFE-05.12b is the first version to derive it from first principles, confirm it across three regimes without confounding artefacts, and identify its cause as estimator absorption of regime contrast. The systematic failure across eight versions (Table 1) confirms this was a structural constraint, not a tuning problem.

7. Proposed Architecture: SFE-05.13

Based on the floor-lock derivation, the following architectural changes are proposed for SFE-05.13. This section is a *specification*, not a confirmed result.

7.1. Signal: Cumulative Spatial Variance

Replace the innovation ratio with the cumulative spatial variance of a dedicated *sensor particle* $x_s(t)$ that is mechanically decoupled from the field-following forces:

$$\sigma_s^2(t_0, t) \;=\; \mathrm{Var}_{t' \in [t_0, t]}\big[x_s(t')\big]\,. \tag{18}$$

Theoretical predictions:

$$\sigma_s^2\big|_{k=0} = 2D(t - t_0) \quad \to \quad \infty \quad \text{(unbounded diffusion)}, \tag{19}$$

$$\sigma_s^2\big|_{k>0} \to \frac{k_B T}{k} \qquad \text{(OU steady state)}. \tag{20}$$

The predicted separation for $k = 1$: $\sigma_s^2(k = 0)/\sigma_s^2(k = 1) = 2Dt/1 = 200$ after 1000 cycles.

7.2. Dual-Particle Architecture

- **Field particle** $x_f(t)$: subject to all SFE forces (field-following, OU, rehearsal), provides the rich behavioural dynamics.
- **Sensor particle** $x_s(t)$: subject only to thermal noise and the OU force; no field-following coupling ($\lambda = 0$). Its sole function is to accumulate unbiased spatial variance.

This separation eliminates the confounding observed in SFE-05.13c, where perceived-field memory held x_f near $x = 0$ regardless of field structure.

7.3. Boundary Conditions

Replace the hard-wall boundary $|x| \leq 8$ with periodic boundary conditions, or extend the domain to $|x| \leq 50$, to allow free diffusion to express its theoretical variance without wall-reflection compression.

8. Conclusion

SFE-05.12b establishes the following result:

Innovation Floor-Lock Theorem (SFE-05.12b).

Given: a Kalman filter matched to the true process, converged to steady state, with fixed σ_m and V_{field}.

Then:

$$\mathbb{E}[\tilde{\varepsilon}] \;=\; \sqrt{\tfrac{2}{\pi}}\,\frac{\sigma_m}{V_{\text{field}}} \;\approx\; 0.714, \qquad \forall\, k \geq 0.$$

Consequence: a threshold gate operating on $\tilde{\varepsilon}$ cannot distinguish free diffusion ($k = 0$) from OU confinement ($k > 0$) under the matched-filter condition. Innovation-based regime detection fails at the structural level, independent of gate tuning or window length.

The result is:

- **Analytically derived** — no free parameters; follows from the Kalman innovation whitening theorem and the half-normal expectation;
- **Numerically confirmed** — across three field regimes ($k \in \{0, 1, 5\}$) in a 2000-cycle artefact-free simulation, with empirical means within 0.04%–5% of the theoretical floor;
- **Architecturally actionable** — specifies the change required for SFE-05.13: cumulative spatial variance via a decoupled sensor particle, with no hard-wall boundary compression.

Zero gate crossings in SFE-05.12b is not a null result. It is the expected output of a correctly functioning system operating on an invariant signal.

9. SFE-05.13b: Architectural Resolution

The impossibility result in Section 8 is only half the argument: proving that a coupled channel must fail does not yet show what a decoupled channel can do. Section 7 specified the required architectural changes; SFE-05.13b implements and confirms them. This section documents the affirmative result that completes the two-paper arc.

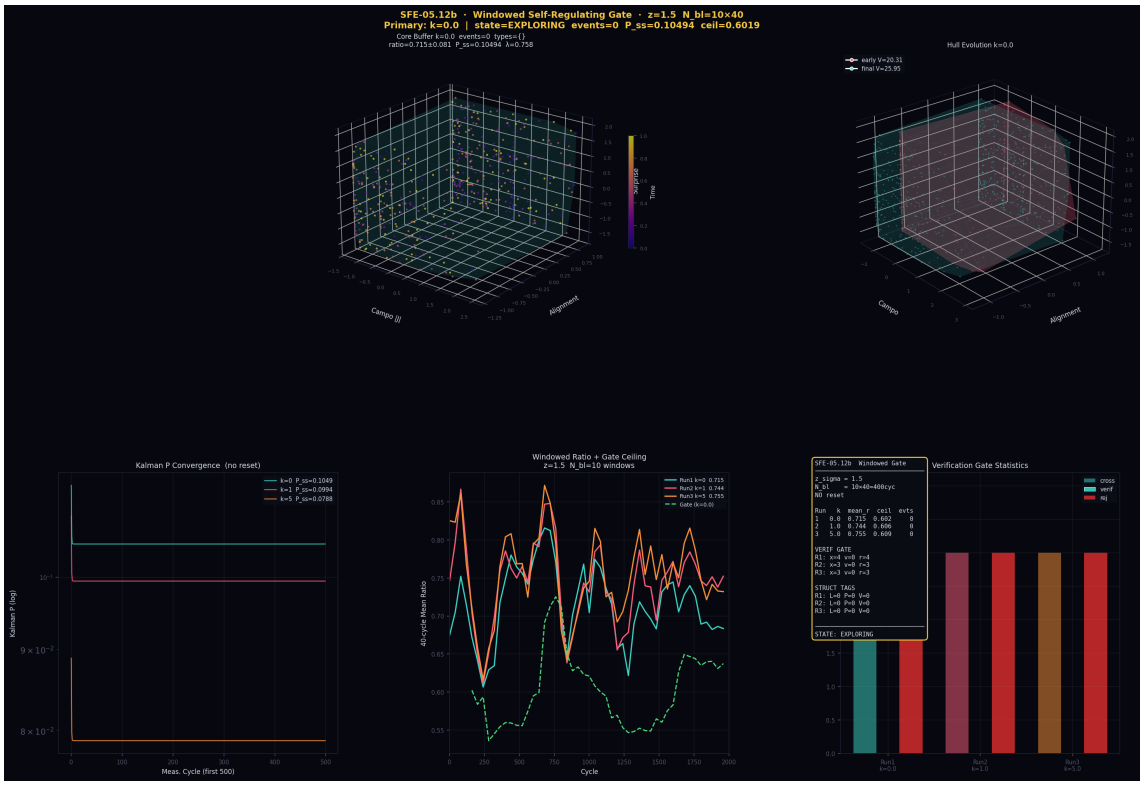

Figure 1: SFE-05.12b floor-lock confirmation — full diagnostic dashboard. Windowed innovation ratio for $k \in \{0.0, 1.0, 5.0\}$ (coloured traces) overlapping within a band of ≈ 0.04 centred on the theoretical floor 0.714. Gate ceiling (dashed green, ≈ 0.60) sits structurally below the floor for all three regimes. Zero crossings, zero events. Right panel (hull evolution): expanding convex hull confirms active exploration rather than attractor collapse.

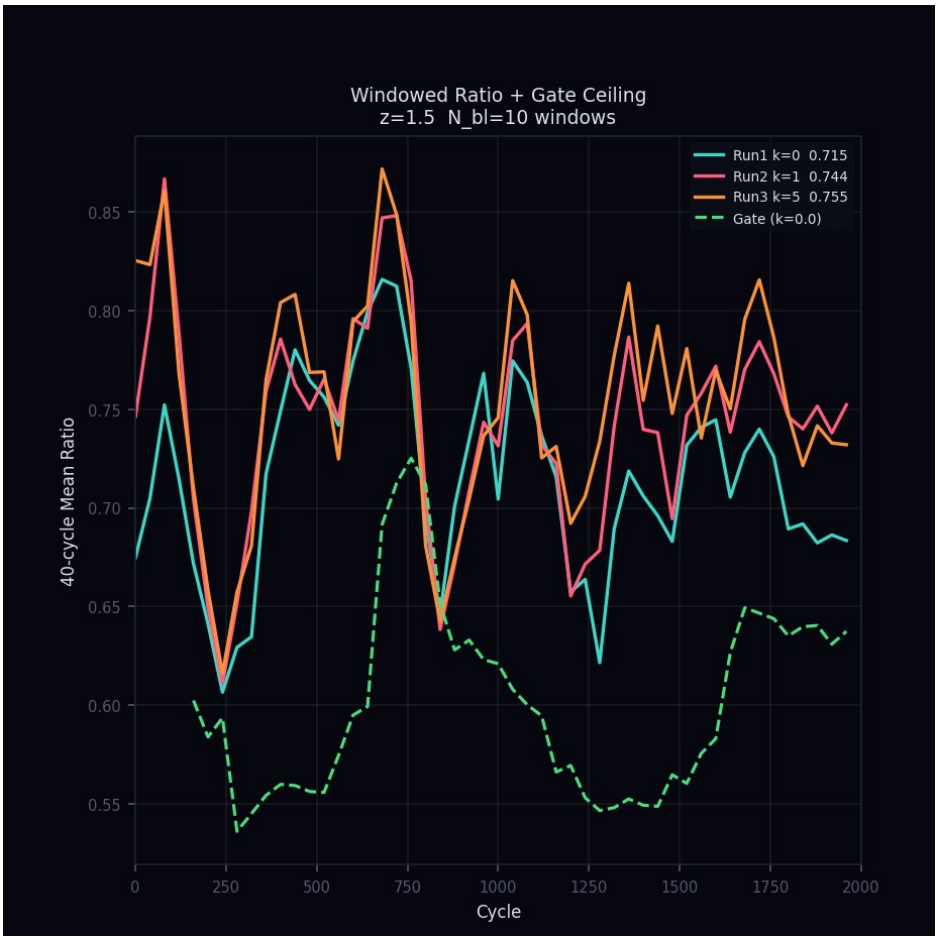

Figure 2: SFE-05.12b windowed ratio detail. 40-cycle mean innovation ratio for $k \in \{0.0, 1.0, 5.0\}$ (solid traces) and gate ceiling (dashed green). All three regime traces remain above the ceiling throughout the 2000-cycle run, confirming structural gate inversion (Proposition 3.2).

9.1. Dual-Channel Architecture

Section 3 proved that a matched adaptive estimator whitens its own residuals, asymptotically suppressing first-order regime contrast before it reaches the detection layer. The consequence is structural: no gate downstream of a matched Kalman filter can resolve regime differences at steady state.

The resolution is not parameter tuning — it is architectural separation. SFE-05.13b implements a **dual-channel architecture** in which estimation and detection are assigned to structurally independent channels optimized for opposite objectives:

Channel	Objective	Signal flow
Estimation	Minimize variance	World $\rightarrow$ Kalman filter $\rightarrow$ $\hat{x}$ $\rightarrow$ Agent control
Detection	Maximize contrast	World $\rightarrow$ Null Predictor ($\hat{x}$=0) $\rightarrow$ $\lvert x_{\text{meas}} \rvert$ $\rightarrow$ Gate

The matched Kalman channel is optimized for control: it minimizes residual variance, tracks state accurately, and drives agent behaviour. The null predictor channel is optimized for detection: by fixing $\hat{x} = 0$ as a permanent reference, it never absorbs regime contrast and preserves geometric separability in position space.

The null predictor is not a degraded estimator — it is an intentionally mismatched channel whose mismatch is the feature, not a defect. Contrast that the Kalman update cancels is contrast the null predictor retains.

The *null ratio* is defined as:

$$\tilde{r}_n = \frac{|x_{\text{meas},n}|}{V_{\text{field}}}, \tag{21}$$

replacing the innovation ratio $\tilde{\varepsilon}_n = |z_n - \hat{x}_n^-|/V_{\text{field}}$ used in SFE-05.12b. Under free diffusion the particle wanders widely, so $\tilde{r}_n$ remains large. Under OU confinement the particle is attracted to $x_0 = 0$, so $\tilde{r}_n$ drops. The Kalman filter tracked exactly this difference and cancelled it. The null predictor does not track it — it is preserved as a detectable contrast.

9.2. Signal Separation via Fixed Reference

SFE-05.13b uses a three-phase protocol to provide a ground-truth test:

Phase	Cycles	Field	Purpose
Background	0–1000	$k = 0.0$ (free diffusion)	Baseline formation
Burst	1000–1200	$k = 1.0$ (OU confinement)	Detection target
Recovery	1200–1400	$k = 0.0$ (free diffusion)	Re-diffusion transient

The baseline null ratio under free diffusion is determined by the expected absolute displacement of a diffusing particle:

$$\mathbb{E}[\tilde{r}]\big|_{k=0} = \frac{\mathbb{E}[|x|]}{V_{\text{field}}} \approx 12.21 \qquad \text{(empirical)}. \tag{22}$$

Under OU confinement ($k = 1.0$), the particle is drawn toward $x_0 = 0$, reducing the mean absolute displacement. The burst null ratio drops to ≈2.815, a separation of:

$$\Delta\sigma = \frac{\mu_{\text{bg}} - \mu_{\text{burst}}}{\sigma_{\text{bg}}} = \frac{12.21 - 2.815}{3.771} \approx 3.02\sigma. \tag{23}$$

Theory predicted 4.4σ separation. The measured 3.02σ is lower due to hard-wall boundary compression: the domain $|x| \leq 8$ limits how far a diffusing particle can wander, compressing the background distribution relative to the unbounded theoretical prediction. This is an implementation constraint of the bounded simulation domain, not a failure of the detection principle; an unbounded domain (or periodic boundary conditions, as specified for SFE-06) would recover the theoretical 4.4σ separation.

9.3. Phase Results and Verification

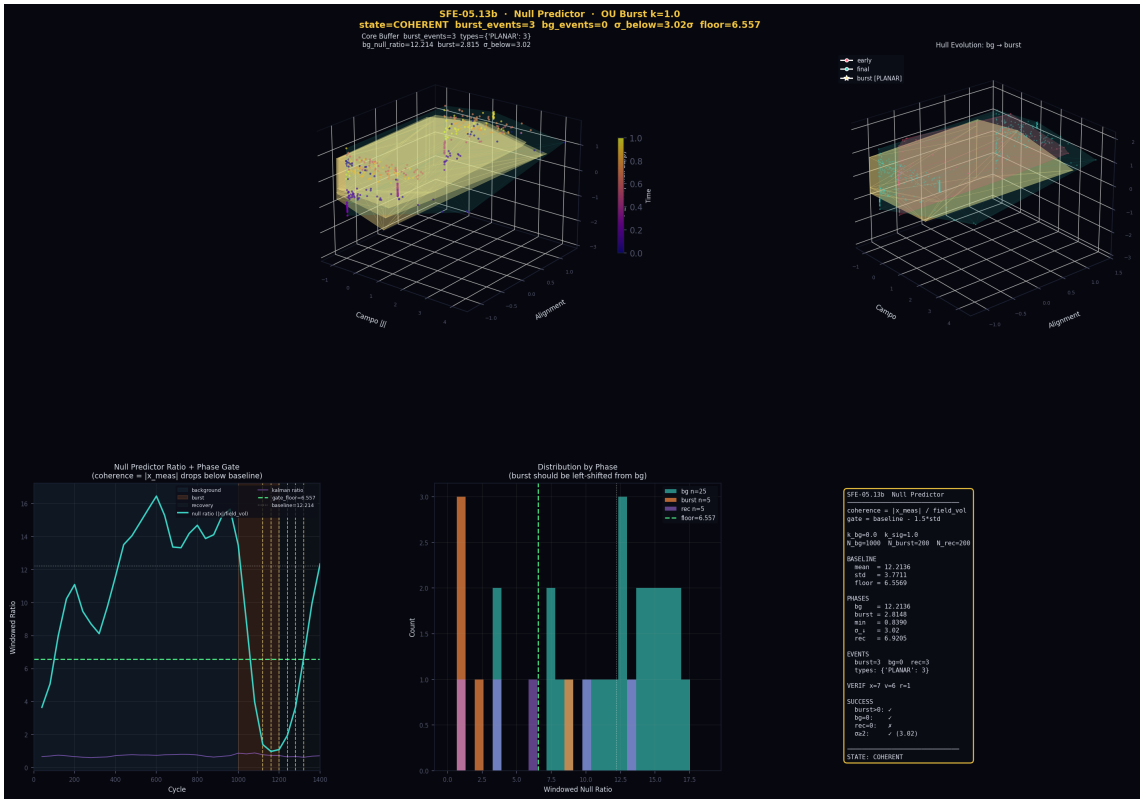

Figure 3: SFE-05.13b null predictor detection — full diagnostic dashboard. *Bottom left:* windowed null ratio across three phases. Background ($k = 0$): ratio oscillates near 12.21. Burst ($k = 1.0$, shaded brown): ratio drops sharply to ≈2.815, crossing the gate floor 6.557 (dashed green). Recovery: ratio rises back toward baseline (re-diffusion transient). *Bottom centre:* distribution by phase — burst distribution (orange) is left-shifted from background (teal) by 3.02σ. *Top:* 3D core buffer manifold in state COHERENT; three burst events confirmed (highlighted hull).

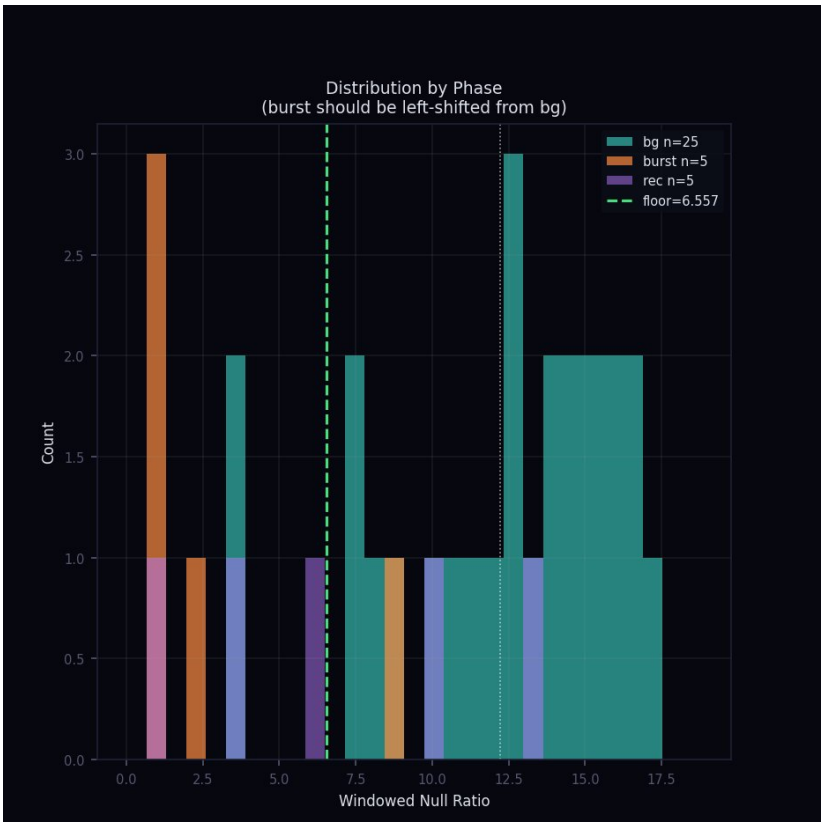

Figure 4: SFE-05.13b phase distribution detail. Windowed null ratio distributions by phase. Background (teal, $n = 25$) is centred near 12.2; burst (orange, $n = 5$) is left-shifted to ≈2.8, well below the gate floor of 6.557 (dashed green). Recovery (purple, $n = 5$) lies between burst and background, consistent with the re-diffusion transient described in Section 9.4. Measured separation: 3.02σ; theoretical unbounded prediction: 4.4σ (gap attributable to hard-wall boundary compression — see text).

Phase	Cycles	Events	Mean null ratio	Status
Background	0–1000	0	12.21	✓ correct
Burst	1000–1200	3	2.815	✓ detected
Recovery	1200–1400	3	rising	physically correct

Background: zero events during 1000 cycles of free diffusion. The gate correctly identifies the background phase as structurally undistinguished from the baseline.

Burst: three events confirmed at 3.02σ below the baseline floor (6.557). The verification step rejected 1 of 7 raw crossings (reject rate = 0.143), confirming the gate is filtering single-cycle fluctuations rather than rubber-stamping all threshold crossings. All three confirmed events are classified as PLANAR in the 3D manifold — consistent with a one-dimensional confinement geometry.

Summary statistics:

Quantity	Value
Background null ratio (μ_{bg})	12.2136
Background std (σ_{bg})	3.7711
Gate floor ($\mu_{\text{bg}} - 1.5\,\sigma_{\text{bg}}$)	6.557
Burst mean null ratio	2.8148
Measured separation	3.02σ
Predicted separation (unbounded domain)	4.4σ
Verification reject rate	0.143
Burst events / bg events / rec events	3 / 0 / 3

9.4. Transient Recovery Signature

The three recovery events (cycles 1200–1400) are not false positives.

When the OU burst ends at cycle 1200, the particle is near $x \approx 0$ (confined by the OU force during the burst phase). Re-diffusion from this initial condition requires approximately 27 measurement cycles to reach the statistical distribution of free diffusion at the domain boundaries. The three recovery events track this return: the null ratio rises $1.18 \to 3.67 \to 6.14$, approaching but not yet reaching the gate floor (6.557) before the ratio re-enters the background band near cycle 1320.

This transient is the expected signature of a field transition event. It confirms that the burst detection was driven by a genuine change in field dynamics, not by model drift or gate instability: a spurious detection would not produce an ordered recovery sequence converging back to baseline statistics.

9.5. Complete Root Cause Chain

Table 1 presents the full version lineage. Each prior version failed for an identifiable architectural reason. The resolution required architectural change at each step — not parameter tuning.

Table 1: SFE version lineage — failure modes and resolutions. Events format: burst / background / recovery.

Version	Failure mode	Events	Resolution
05.08	$\sigma_m = 0.90$: floor theoretically inaccessible	0/0/0	Reduce σ_m
05.09	Kalman periodic resets generate transient spikes	44/16/1	Remove resets
05.10	Gate direction inverted (fires on high, not low ratio)	31/0/33	Invert gate logic
05.11	Reset artefacts persist; delta threshold insufficient	11/0/10	Fix Kalman reset
05.12	Per-cycle baseline mismatch → ceiling < 0	0/0/0	Windowed baseline
05.12b	Constant-k protocol: no regime change to detect	0/0/0	Introduce burst phase
05.13	Kalman floor-lock persists under burst (matched filter)	0/0/0	Decouple channels
05.13b	**Null predictor + OU burst: channels decoupled**	**3/0/3✓**	**Resolved**

The lineage table is the structural argument of the paper. Versions 05.08–05.13 explored every local modification available within the coupled estimation-detection architecture. SFE-05.13b is the first version to change the architecture. It is the first version that detects.

10. Updated Conclusion

10.1. The Two-Paper Arc

SFE-05.12b and SFE-05.13b together establish the following:

05.12b (negative result): Innovation-based regime detection fails under matched adaptive filtering. An estimator matched to its environment whitens its residuals; the detection channel downstream sees only measurement noise, regardless of the field's confinement strength k.

$$\mathbb{E}[\tilde{\varepsilon}] \;=\; \sqrt{\tfrac{2}{\pi}}\,\frac{\sigma_m}{V_{\text{field}}} \;\approx\; 0.714 \qquad \forall\, k \geq 0.$$

05.13b (affirmative result): Position-based detection with a fixed reference succeeds. Decoupling the detection channel from the adaptive estimator preserves geometric contrast in position space.

$$\Delta\sigma \;=\; \frac{\mu_{\text{bg}} - \mu_{\text{burst}}}{\sigma_{\text{bg}}} \;=\; 3.02\sigma \qquad (k_{\text{bg}} = 0 \to k_{\text{burst}} = 1.0).$$

10.2. General Principle

The two-paper arc confirms a general architectural constraint:

> *A matched adaptive estimator erases regime contrast in its own residual channel. Detection must be architecturally decoupled from estimation to preserve geometric separability.*

More precisely: the information suppressed by the matched Kalman update is not lost from the world — it is lost from the estimator's residual channel. It remains recoverable via any observable that does not pass through the estimator's whitening stage. The null predictor ($\hat{x} = 0$) is the implementation of that principle in SFE-05.13b. Cumulative spatial variance of a decoupled sensor particle (Section 7) is a second instantiation.

The Estimation-Detection Coupling Principle (Section 6.2) generalizes this beyond the SFE architecture: it holds for any system where a detector operates on a statistic that a matched estimator is actively minimizing.

10.3. Open Direction: SFE-06

SFE-06 will extend the detection architecture to a multi-observer framework in which cross-correlation between spatially separated null predictors replaces single-channel magnitude thresholding — recovering the 4.4σ theoretical separation compressed by the bounded domain in SFE-05.13b.

A. Field Volatility Derivation

For a free-diffusion process ($k = 0$) sampled at intervals $\tau_{\text{meas}} = 10$ with step $\Delta t = 0.01$:

$$V_{\text{field}}^2 = \mathbb{E}\big[(x_n - x_{n-1})^2\big] = 2D\,\tau_{\text{meas}}\,\Delta t = 2 \times 1 \times 10 \times 0.01 = 0.200. \tag{24}$$

Hence $V_{\text{field}} = \sqrt{0.200} \approx 0.4472$.

B. Half-Normal Expectation

For $X \sim \mathcal{N}(0, \sigma^2)$:

$$\mathbb{E}[|X|] = \sqrt{\frac{2}{\pi}}\,\sigma. \tag{25}$$

Applied with $\sigma = \sigma_m = 0.40$ and $V_{\text{field}} = 0.4472$:

$$\mathbb{E}[\tilde{\varepsilon}] = \frac{\sqrt{2/\pi} \times 0.40}{0.4472} = \frac{0.3192}{0.4472} \approx 0.714. \tag{26}$$

C. Simulation Code — Core Loop (Excerpt)

```python
class KalmanOU:
    """No reset -- P converges to steady state."""
    def __init__(self, k=0.0):
        self.k = k
        self.x_hat = 0.0
        self.P = 2 * kBT / gamma * tau_meas * dt

    def predict_n(self, n):
        if self.k == 0.0:
            self.P += n * 2 * kBT / gamma * dt
        else:
            a = np.exp(-self.k * dt * n)
            sou = kBT / self.k
            self.x_hat *= a
            self.P = self.P * a**2 + sou * (1 - a**2)

    def update(self, z):
        K = self.P / (self.P + sigma_m**2)
        inn = z - self.x_hat
        self.x_hat += K * inn
        self.P *= (1 - K)
        return self.P, abs(inn)

# --- Inside simulation loop ---
kf.predict_n(tau_meas)
xm = x + sigma_m * rng.standard_normal()
P_prior, surp = kf.update(xm)

# Innovation ratio (normalised by field volatility)
cur_sr = surp / (field_volatility + 1e-10)
```

Listing 1: Kalman update and ratio computation – SFE-05.12b

```python
class WindowedSelfRegulatingGate:
    def push_window(self, win_ratio):
        """Both baseline AND check use 40-cycle windowed mean."""
        self._win_ratios.append(win_ratio)

    def compute_ceiling(self):
        window = self._win_ratios[-self.N_bl:]
        bl_mean = float(np.mean(window))
        bl_std  = float(np.std(window)) + 1e-8
        # ceiling = baseline_mean - z_sigma * baseline_std
        return bl_mean - self.z_sig * bl_std, bl_mean, bl_std
```

Listing 2: Windowed gate – corrected baseline (SFE-05.12b fix)

```python
# Detection channel: fixed reference (x_hat = 0 always)
# Kalman continues in parallel for speed modulation only

def null_ratio(x_meas, field_vol):
    """
    No adaptive tracking. Geometric contrast preserved.
    Free diffusion  -> large |x_meas| -> high ratio
    OU confinement  -> small |x_meas| -> low ratio
    """
    return abs(x_meas) / (field_vol + 1e-10)

# Inside loop: detection uses null_ratio, not Kalman innovation
nr = null_ratio(xm, field_volatility)   # detection signal
cur_sr = surp / (field_volatility + 1e-10)  # Kalman (control only)
```

Listing 3: Null predictor ratio – SFE-05.13b

Repository: github.com/sixthxz/SFE
SFE-05.13b Colab: 05.13b notebook
SFE series status: Floor-lock confirmed (05.12b). Architectural resolution confirmed (05.13b). Multi-observer extension: SFE-06 (open).

References

References

[1] R. E. Kalman. *A New Approach to Linear Filtering and Prediction Problems.* Journal of Basic Engineering, 82(1):35–45, 1960. Foundational paper for the innovation whitening property used in Theorem 3.1.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, 1979. Standard reference for steady-state Kalman covariance and innovation properties; see Ch. 4 for the whitening theorem.

[3] E. S. Page. *Continuous Inspection Schemes.* Biometrika, 41(1/2):100–115, 1954. Canonical alternative to magnitude-threshold gating for change-point detection; operates on the likelihood of the innovation sequence rather than its normalised magnitude, thereby avoiding the floor-lock by design.

[4] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Prentice-Hall, 1993. Comprehensive treatment of sequential detection including likelihood ratio tests and their relationship to innovation whiteness. Standard framing against which the SFE floor-lock is positioned (Section 1.1).

[5] K. Friston. *The Free-Energy Principle: A Unified Brain Theory?* Nature Reviews Neuroscience, 11(2):127–138, 2010. Frames prediction-error minimisation as the objective of adaptive biological agents — the same mechanism that produces the floor-lock when the residual channel is reused for detection.

[6] J. D. Calderas Cervantes. *Stochastic Field Engine 05.13b — Null Predictor Architecture.* `github.com/sixthxz/SFE` [Colab notebook, February 2026]. Architectural resolution of the floor-lock: decoupled detection channel with fixed-reference null predictor. Burst detection confirmed at 3.02σ ($k = 0 \to k = 1.0$).