FRUSTRATION CERTIFICATES FOR INDEPENDENTLY CALIBRATED SENSOR SYSTEMS

JUAN CARLOS PAREDES

Abstract. Multi-sensor systems are calibrated pairwise, in sessions, at different times — and a calibration that drifts between sessions produces a fault that no pairwise monitor can see, because every sensor pair remains internally self-consistent. We prove a dichotomy that organizes when such faults are even detectable: if all pairwise relations are estimated from one joint sample, frustration of the signed relation graph is geometrically capped (a frustrated cycle of length L forces some edge confidence below $\cos(\pi/L)$, for any joint distribution), whereas independently calibrated relations admit unbounded frustration — so observed frustration above the cap is a distribution-free certificate that no single consistent calibration exists, and a leave-one-out filtration names the suspect sensor. On real fusion-diagnostic data from two machines of different confinement physics and data provenance — MAST magnetics and LHD bolometry — the certificate detects planted polarity drift at AUC 0.976–1.000 and localizes it at 0.93–1.00, under a single untuned pipeline, while the fault remains statistically invisible to pairwise inconsistency (0.218 faulted vs 0.223 null). A third, deliberately adversarial study characterizes the operating envelope and yields the method's preconditions: a shared latent and relation stationarity across sessions. We close with an honest composition- budgeting rule: a pre-registered test falsifies the product-form capacity rule we ourselves expected, and the operational law is two-regime, $\max(1.405\sigma\sqrt{k},\ 0.6B/\sqrt{n})$.

1. Introduction

Deployed sensor networks are not calibrated all at once. Pairs of instruments are cross-calibrated in sessions — at installation, after maintenance, on different days, against different references — and the system's working model of "consistent" is the union of those pairwise sessions. This paper concerns a fault class native to that reality: **inter-calibration inconsistency**, in which the relation between sensors changes between sessions (a swapped lead, a sign convention applied inconsistently, a re-mounted probe) while every individual sensor, and every individual calibrated pair, continues to function correctly.

This fault class has a structural property that makes it dangerous: it is invisible to pairwise monitoring. Each calibration session is internally consistent; each pair's residuals are clean. In our primary case study, the mean pairwise inconsistency of a faulted network is statistically indistinguishable from the fault-free null (0.218 versus 0.223), while the fault is plainly present and consequential. Whatever detects it must read structure that no pair contains.

The structure that works is *frustration*: the failure of the pairwise relation signs to admit any global assignment, accumulated around cycles. But frustration is only evidence if it could not have arisen innocently, and our organizing result is a dichotomy that makes this precise. If all relations are estimated from **one joint sample**, the signed relation graph cannot be strongly frustrated: for any joint distribution whatsoever, a frustrated cycle of length L forces at least one edge confidence below $\cos(\pi/L)$ — a frustrated triangle can never have all three $|r| > 1/2$. If instead the relations are estimated in **independent sessions**, no such constraint exists, and a single polarity change between sessions produces fully frustrated cycles at full confidence.

Date: July 5, 2026.

Frustration is therefore uninformative exactly when calibration is joint, and a *certificate about the calibration structure itself* exactly when calibration is independent — which is the deployed case.

Our contributions: (1) the dichotomy theorem with a self-contained proof; (2) a distribution-free certificate — a frustrated cycle whose Fisher-z lower confidence bounds all clear $\cos(\pi/L)$ rejects, at stated confidence, the hypothesis that any single joint law generated the edges; (3) localization by a leave-one-out frustration filtration; (4) evidence on two physically and institutionally distinct machines under one untuned pipeline; (5) an operating-envelope study that converts the method's failure mode into its preconditions; and (6) a composition-budgeting analysis whose pre-registered headline is the falsification of the budgeting rule we expected to validate.

We are explicit about what is not claimed. The cycle inequality behind Theorem A is elliptope geometry [Laurent–Poljak]; the frustration index is classical balance theory [Harary]; sublinear balance testing is published [Adriaens–Apers]; the ingredients of the budget law are textbook uncertainty propagation and detection theory [GUM; detection]. The claims are the dichotomy framing, the certificate semantics and threshold, the localization mechanism, and the empirical record.

2. The Dichotomy

Setting. Nodes $V = \{1,\dots,n\}$ carry real observables. A *signed relation graph* assigns each edge $e = (i,j)$ a sign $s_e \in \{\pm 1\}$ and a confidence $|r_e| \in [0,1]$, in practice the sign and magnitude of an estimated correlation. A cycle C is *frustrated* if $\prod_{e\in C} s_e = -1$. The *frustration radius* of the graph is the minimum confidence-weighted fraction of edges whose deletion (equivalently, sign flip) renders the graph balanced; it is the natural H^0 obstruction to a global sign assignment and coincides with the correlation-clustering distance-from-balance.

Theorem A (joint sampling caps frustration). Let all edges be estimated from a single joint sample: $r_{ij} = \mathrm{corr}(X_i, X_j)$ under one joint distribution. Then every frustrated cycle C of length L satisfies

$\Sigma_{e\in C} \arccos|r_e| \geq \pi$, and hence $\min_{e\in C} |r_e| \leq \cos(\pi/L)$.

Proof. Standardize and embed $X_i \mapsto u_i$ as unit vectors with $r_{ij} = \langle u_i, u_j\rangle$, and let $\varphi_e = \arccos|r_e| \in [0, \pi/2]$ be the projective angle. Walk the cycle $1\to 2\to\dots\to L\to 1$ choosing gauges $\zeta_1 = +1$, $\zeta_{k+1} = \zeta_k s_{k,k+1}$, so each traversed edge is seen at its acute angle φ. By the spherical triangle inequality along the chain, $\angle(\zeta_1 u_1, \zeta_L u_L) \leq \Sigma_{k<L} \varphi_k$. Frustration forces the closing edge to be seen at its obtuse representative: $\zeta_L \zeta_1 \cdot \mathrm{sign}(r_{L,1}) = -1$, so $\angle(\zeta_L u_L, \zeta_1 u_1) = \pi - \varphi_{L,1}$. Combining gives $\pi - \varphi_{L,1} \leq \Sigma_{k<L} \varphi_k$. ∎

Two special cases anchor intuition. For $L = 3$ the cap is $1/2$: an antiferromagnetic triangle with all $|r| > 1/2$ is impossible under any joint law — the equicorrelation bound on the elliptope, where the inequality is classical [Laurent–Poljak]; we keep the proof for self-containment. In the noiseless limit $|r_e| = 1$, signs are forced to a coboundary $s_{ij} = z_i z_j$: exact balance. With N joint samples the cap inherits an $O(1/\sqrt{N})$ Fisher-z margin.

Observation B (independent calibration removes the cap). If each edge is estimated from its own session — disjoint data: different time windows, different operating periods, separate calibration runs — the collection $\{(s_e, |r_e|)\}$ satisfies no positive-semidefiniteness constraint, because no single joint law is ever instantiated. Any signed graph is reachable, including fully frustrated cycles with every $|\hat{r}_e| \to 1$. The generic mechanism is a polarity change at one node between sessions: edges calibrated before and after disagree in sign, and any cycle mixing pre- and post-change edges is frustrated at the full confidence of its constituent sessions.

The dichotomy as a lens. Under joint estimation, strong frustration is impossible — the world seen through one joint distribution is always realizable up to the cap; empty "high-frustration, high-confidence" quadrants in jointly-sampled studies are theorems, not empirical luck. Under independent calibration, frustration above the cap cannot be explained by *any* plant behavior, because Theorem A quantifies over all joint laws; it can only be a property of the measurement system. That inversion — the certificate indicts the calibration structure, not the world — is what the rest of the paper operationalizes.

3. The Certificate

Radius and gauge. With gauge variables z $\in$ {$\pm$1}^n, the frustration radius is min_z Σ_{e:s_e $\neq$ z_i z_j} c_e / Σ_e c_e, with c_e = $|\hat{r}$_e|. Exact minimization is 2^{n$-$1}; Section 4 gives the estimator used at scale.

The threshold test. For an edge estimated from T samples, the one-sided Fisher-z lower confidence bound is r_lo = tanh(atanh$|\hat{r}|$ $-$ z_{α/L}/$\sqrt{}$(T$-$3)), Bonferroni-corrected over the cycle's L edges. The certificate condition for a frustrated cycle C is

min_{e$\in$C} r_lo,e > cos(π/L).

When it holds, Theorem A excludes — at simultaneous confidence 1$-\alpha$, with no parametric assumption on the observables, since correlation matrices of any joint law are PSD — every joint distribution as the source of C's edges. The conclusion is categorical: the calibration sessions are mutually inconsistent.

Localization. The filtration score of node v is the drop in frustration radius upon deleting v. The faulted sensor maximizes this score because the frustrated cycles concentrate through it; ranking by score names the suspect sensor and, via the session assignments of its incident edges, the suspect sessions. In all studies below the top-ranked node is the planted fault in 93–100% of trials.

Protocol. Every experiment reports against a matched null (identical pipeline and session randomization, no fault) with pre-registered gates — detection AUC $\geq$ 0.90, localization hit-rate $\geq$ 0.80 — and all outcomes are reported, including the two that came out negative (Sections 5.3 and 6).

4. Locality and Scale

Three facts make the certificate practical at deployment size. First, deciding balanced-versus-ε-far in the dense model needs only $\tilde{O}(1/\varepsilon)$ queries [Adriaens–Apers, via Zaslavsky's reduction to bipartiteness testing]; our own measurements (constant query size $\approx$ 5 across n = 20. . . 320) merely confirm their result in implementation and we claim nothing here. Second, exact radius computation can be replaced by multi-restart greedy flip descent: across 40 trials at n = 12 the greedy estimate matched brute-force exactly (maximum deviation 0.0000), replacing the 2^{n$-$1}-assignment exact search with O(restarts · n · |E|) local descent. Third, and least obvious: the certificate *improves* with scale. Repeating the full Section-5 protocol at n = 24 and n = 29 sensors yields detection AUC and localization both 1.000 — each added sensor contributes independent cycles through any faulted node, so evidence accumulates with network size rather than diluting.

5. Case Studies

Common design. Nodes are sensors; each edge (i,j) is calibrated in its *own* session, drawn independently from a time-ordered archive, with the sign and confidence of the edge computed only from that session's data — genuine estimation independence, the Observation-B precondition, met honestly rather than simulated. The planted fault is a polarity inversion (lead-swap class) at one randomly chosen node, effective from a randomly chosen mid-archive changepoint onward, so that the faulted node's pre- and post-changepoint edges disagree. The matched null runs the identical pipeline with no fault. Trials: 60 fault + 60 null per configuration. The planted flip is exact rather than approximate physics: $\mathrm{corr}(-\mathrm{x}, \mathrm{y}) = -\mathrm{corr}(\mathrm{x}, \mathrm{y})$ identically, and a raw-data replay (flipping the probe series and recomputing) reproduces every number.

5.1 MAST magnetics (tokamak). Twelve poloidal-field probes (of 29 live across all 39 archived discharges), 66 edges, sessions = discharges. Results over four seeds: detection AUC 0.987–1.000, localization 0.95–1.00. The null radius is 0.001–0.002 — these diagnostics genuinely cohere when calibration is consistent, which is precisely what makes the certificate's floor clean [companion gluing studies]. The faulted radius is 0.057–0.061, fifteen standard deviations above the floor. The lead result is the quadrant plot (Fig. 1): **mean pairwise inconsistency is statistically identical between fault and null (0.218 vs 0.223)** while the frustration radius separates the populations essentially completely. The fault is invisible where deployed monitors look, and obvious where the cycle structure looks. The Theorem-A test fires accordingly: 15.7 certificate triangles per faulted trial versus 0.0 per null trial (trial-level rates 0.95 vs 0.00) at working threshold $0.55 = \cos(\pi/3) + \text{margin}$.

5.2 LHD bolometry (stellarator). The identical pipeline — same code path, same gates, no per-machine tuning — applied to 12 of 60 bolometer channels across 40 discharges of a different machine, different confinement physics, and different data provenance (NIFS Labcom archive versus MAST Zarr). Results over three seeds: AUC 0.976–0.998, localization 0.93–1.00; null radius 0.001; certificate rates 0.85 vs 0.00. Baseline edge confidences are high (median $|\hat{r}|$ = 0.97), as the channels co-view a single plasma. The two studies jointly support the claim that the certificate is a method, not a dataset artifact.

5.3 The envelope: where the certificate correctly hesitates. A deliberately adversarial third study uses seven MAST global scalars of *different physics* (current, densities, radiated power, solenoid current, gas flux), where edge confidences are moderate (median 0.62; only 56% above the working threshold) and — critically — the sign of a pair's relation can change with operating regime. The certificate degrades exactly as Theorem A predicts: AUC 0.796, localization 0.533, and a nonzero null radius (0.083) with the threshold test firing in 45% of *fault-free* trials. We emphasize the correct reading: those null firings are not false positives of the test. Edges calibrated in different discharges genuinely do not arise from one joint law when the underlying relations are regime-dependent; the certificate truthfully reports cross-regime inconsistency — which, for the calibration-fault application, is a confound. The envelope study thereby yields the method's preconditions: **(i) a shared latent** (high edge confidence), and **(ii) relation stationarity across sessions** (edge signs stable in the absence of faults). The fusion diagnostics of 5.1–5.2 satisfy both; discordant global scalars violate (ii) — a phenomenon first visible as a shot-specific correlation in earlier multi-discharge work [relation-deflation record], systematized here.

6. Composition Budgeting

When the certificate's inputs traverse composed acquisition stages (quantization, transmission, re-quantization), its sampling margins must be budgeted. We expected to validate a product-form provisioning rule, $D^* = \frac{1}{2}(B/(2\sigma\sqrt{n}))^2$, which prescribes per-stage bandwidth B_min for a target end-to-end capacity. The pre-registered test (gate: predicted-to-measured capacity within $\times 1.30$ per cell) **failed, at $\times$2.15 worst-cell** — and the failure is the section's content. Measured end-to-end discriminability (tie- corrected midrank AUC, threshold AUC $\geq$ 0.84) is bandwidth-independent across 0.7–1.4 $\times$ B_min in the noise-dominated regime, matching $\delta^* = 1.405\sigma\sqrt{k}$ to $\leq 3\%$; in the quantization- dominated regime it is depth-independent, matching $\delta^* \approx 0.6{\cdot}B/\sqrt{n}$; and at coarse cells composition is *regenerative* — re-quantization cleans accumulated noise, the classical digital-repeater effect, so deeper chains can outperform the noise law. The operational budgeting rule is therefore two-regime:

δ^*_operational $\approx \max(\ 1.405{\cdot}\sigma{\cdot}\sqrt{k},\ 0.6{\cdot}B/\sqrt{n}\ )$,

two separate classical constraints — RSS noise accumulation [GUM] and the quantizer resolution floor — of which exactly one binds in any deployment. Guidance: identify the regime, check both constraints, and grow the certificate's Fisher-z margins with $\sqrt{k}$ on the noise side. We report the product rule's failure rather than quietly replacing it because the rule's ingredients are widely assumed to compose multiplicatively; our pre-registered negative is, to our knowledge, the first direct operational test.

7. Related Work

Cycle consistency in vision. Structure-from-motion pipelines reject outlier relative poses by loop-closure tests [Zach et al.; sequential cycle inference]. There, a globally consistent configuration is presumed to exist and bad measurements are discarded to reach it. Our certificate tests the opposite hypothesis — whether any consistent global calibration exists — with the dichotomy supplying the null and the threshold distribution-free.

Group synchronization. Angular and group synchronization recover global states from noisy pairwise relations, with sharp thresholds [Singer; Bandeira]. Recovery presumes existence; we certify existence (or its failure) and localize, without reconstruction.

Closure phase / interferometric self-calibration. In radio interferometry, per-antenna calibration errors cancel around a loop, so the **closure phase** is the gauge-invariant of which our frustration is the Z/2 (sign) reduction, and self-calibration is the quotient by the node gauge [Jennison 1958; self-cal literature]. A six-decade- validated instance of the certificate's invariant in the U(1) setting; our contribution is the provenance/dichotomy semantics — the same loop quantity is vacuous under joint estimation and evidentiary under independent calibration — and the sign-valued, distribution-free certificate. (The non-invertible/lossy extension is the companion papers' territory.)

Balance theory. The frustration index of signed graphs is classical [Harary]. The contribution here is provenance semantics: the same index is vacuous or evidentiary depending on how the signs were estimated.

Elliptope geometry. Theorem A's inequality is of the cut-polytope/elliptope family [Laurent–Poljak]; we contribute the dichotomy framing and the certificate use.

Property testing. Sublinear balance testing is due to [Adriaens–Apers]; we cite and confirm, and claim nothing.

Industrial anomaly detection. ICS/sensor benchmarks target plant faults and attacks [SWaT; WADI]; calibration-structure faults — invisible pairwise, by Section 2 — are a complementary class that standard residual monitors do not address.

8. Limitations and Outlook

The certificate's scope is set by its preconditions: a shared latent and relation stationarity across sessions; Section 5.3 shows the diagnostic signature when (ii) fails. The fault class demonstrated is polarity-type; gain drift with preserved sign is invisible to a sign-valued graph and requires transport-valued edges (rotations or partial isometries), whose certificate theory raises a stability question — when is an almost-consistent family of lossy transports near an exactly consistent one — that we develop separately on the (matrix-norm $\times$ defect-functional) grid suggested by recent work on stability of group representations [BLSW; GLMR]. All faults here are planted; the natural real-fault tiers are industrial datasets with root-cause labels [SWaT; WADI] and meteorological station networks, whose documented instrument changes provide ground-truth calibration discontinuities [GHCN homogenization]. Both are external-validation work, not method work: the method's claim — that calibration inconsistency is a provable blind spot of pairwise monitoring, and that cycle structure converts it into a certified, localized detection — stands on the evidence presented.

References

Elliptope / cut polytope (§1, §2, §7). - **Laurent–Poljak:** M. Laurent, S. Poljak, *On a positive semidefinite relaxation of the cut polytope*, Linear Algebra Appl. **223/224** (1995), 439–461. (The elliptope; the cycle inequalities on its correlation-matrix points.)

Balance theory (§1, §7). - **Harary:** F. Harary, *On the notion of balance of a signed graph*, Michigan Math. J. **2** (1953), 143–146. (Frustration/balance; the frustration index.) - T. Zaslavsky, *Signed graphs*, Discrete Appl. Math. **4** (1982), 47–74. (The balance–bipartiteness reduction used by the testing literature.)

Property testing (§1, §4, §7). - **Adriaens–Apers:** F. Adriaens, S. Apers, *Testing properties of signed graphs*, arXiv:2102.07587 (2021). ($\tilde{O}(1/\varepsilon)$-query balance testing in the dense model; our constant-query measurements are implementation confirmation only.)

Detection theory / uncertainty budgeting (§1, §6). - **GUM:** JCGM 100:2008, *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*, BIPM (2008). (RSS accumulation.) - **detection:** S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory*, Prentice Hall (1998). (d′ and threshold detection; the classical ingredients of both regimes.)

Cycle consistency in vision (§7). - **Zach et al.:** C. Zach, M. Klopschitz, M. Pollefeys, *Disambiguating visual relations using loop constraints*, CVPR 2010. (Loop-closure outlier rejection.) - **sequential cycle inference:** T. Xiao, X. Wang, F. Deng, et al., *Sequential cycle consistency inference for eliminating incorrect relative orientations in structure from motion*, PFG – J. Photogramm. Remote Sens. Geoinf. Sci. **89** (2021). doi:10.1007/s41064-021-00152-1.

Group synchronization (§7). - **Singer:** A. Singer, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. Comput. Harmon. Anal. **30** (2011), 20–36. - **Bandeira:** A. S. Bandeira, N. Boumal, A. Singer, *Tightness of the maximum likelihood semidefinite relaxation for angular synchronization*, Math. Program. **163** (2017), 145–167.

Closure phase / self-calibration (§7). - **Jennison 1958:** R. C. Jennison, *A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent*, MNRAS **118** (1958), 276–284. (Closure phase.) - **self-cal literature:** T. J. Pearson, A. C. S. Readhead, *Image formation by self-calibration in radio astronomy*, Annu. Rev. Astron. Astrophys. **22** (1984), 97–130.

Stability of representations (§8). - **BLSW:** *Stability and instability of lattices in semisimple groups*, J. Anal. Math. **151** (2023), 1–23. arXiv:2303.08943. - **GLMR:** L. Glebsky, A. Lubotzky, N. Monod, B. Rangarajan, *Asymptotic cohomology and uniform stability for lattices in semisimple groups*, arXiv:2301.00476 (2023).

Industrial / external-validation datasets (§7, §8). - **SWaT:** J. Goh, S. Adepu, K. N. Junejo, A. Mathur, *A dataset to support research in the design of secure water treatment systems*, CRITIS 2016. - **WADI:** C. M. Ahmed, V. R. Palleti, A. P. Mathur, *WADI: a water distribution testbed for research in the design of secure cyber physical systems*, CySWATER 2017. - **GHCN homogenization:** M. J. Menne, C. N. Williams Jr., *Homogenization of temperature series via pairwise comparisons*, J. Climate **22** (2009), 1700–1717. (Documented instrument changes as ground-truth calibration discontinuities.)

Internal (program record; available on request). - **companion gluing studies:** the fusion-diagnostics gluing record — MAST magnetics cohere across an extensive battery of attempted obstructions. - **relation-deflation record:** first observation of relation non-stationarity (a shot-specific I_sol–gas_out correlation, −0.96 in one discharge, absent when pooled).

Figures

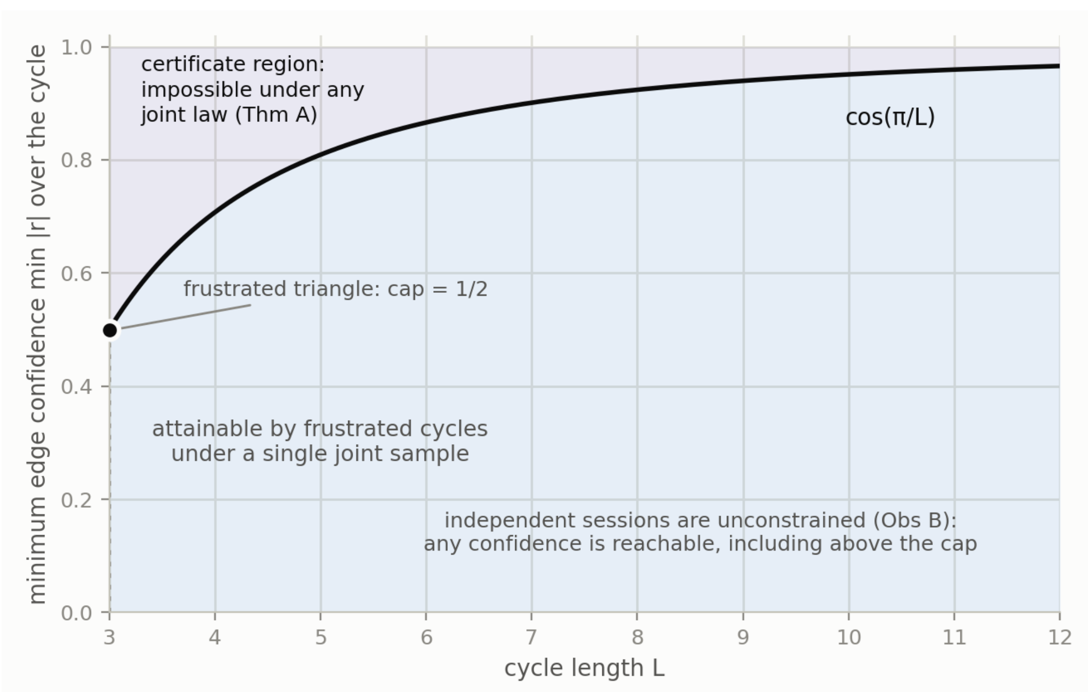

Figure 1. Theorem A caps the minimum confidence of a frustrated L-cycle at $\cos(\pi/L)$ under any joint law; independent calibration is unconstrained, so observations in the upper region certify calibration inconsistency.

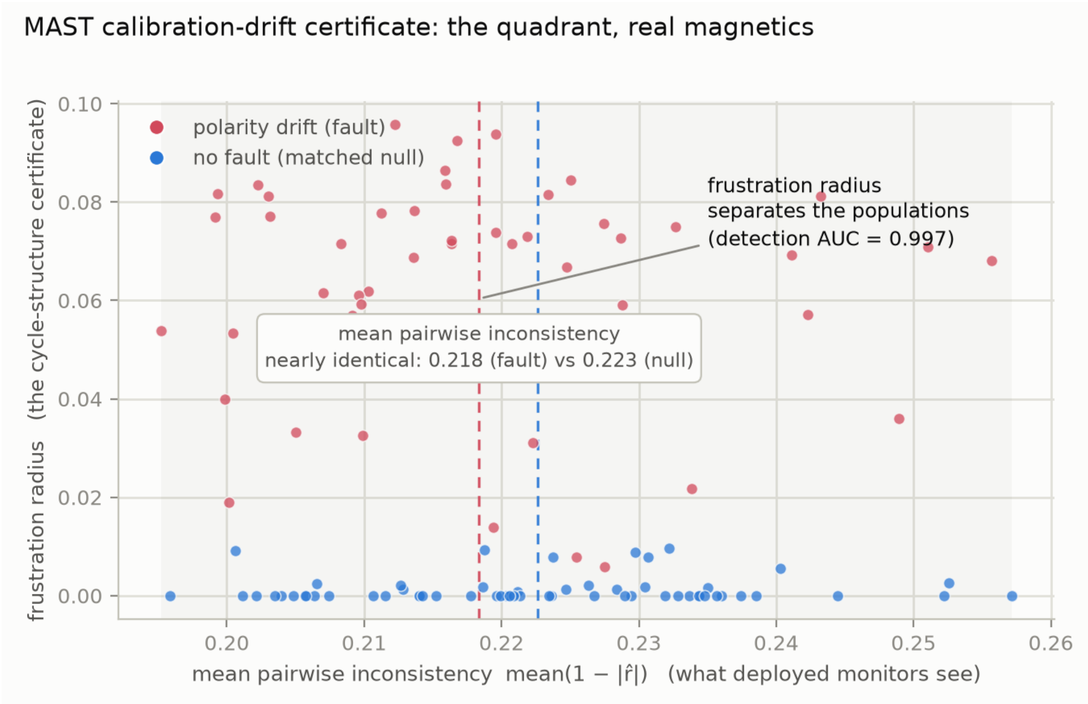

Figure 2. Quadrant plot (Sec. 5.1): mean pairwise inconsistency is statistically identical between fault and null while the frustration radius separates the populations.

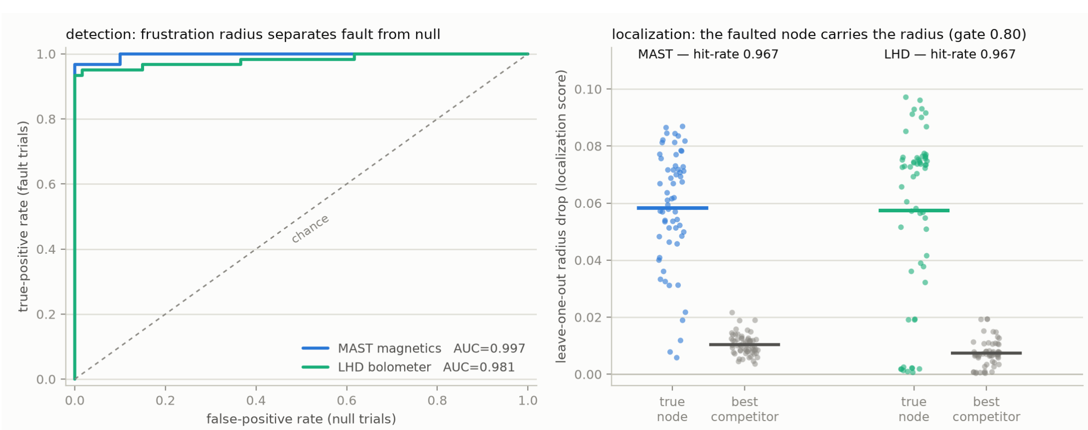

Figure 3. Detection and localization panels (Secs. 5.1–5.2).

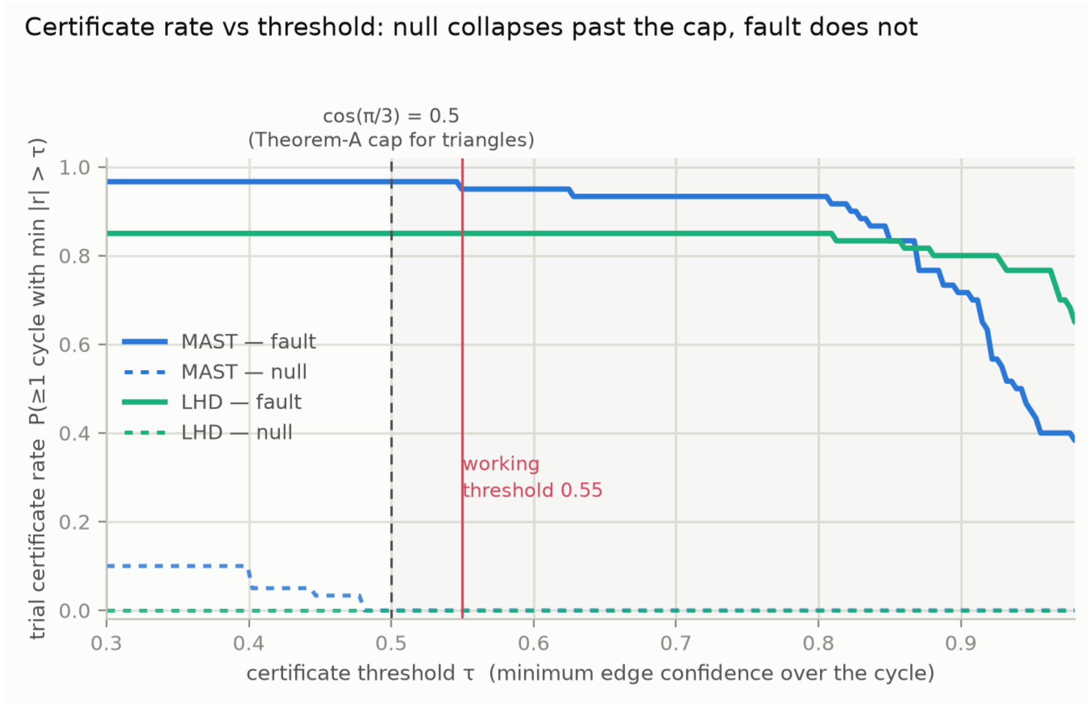

Figure 4. Certificate firing rates vs threshold.

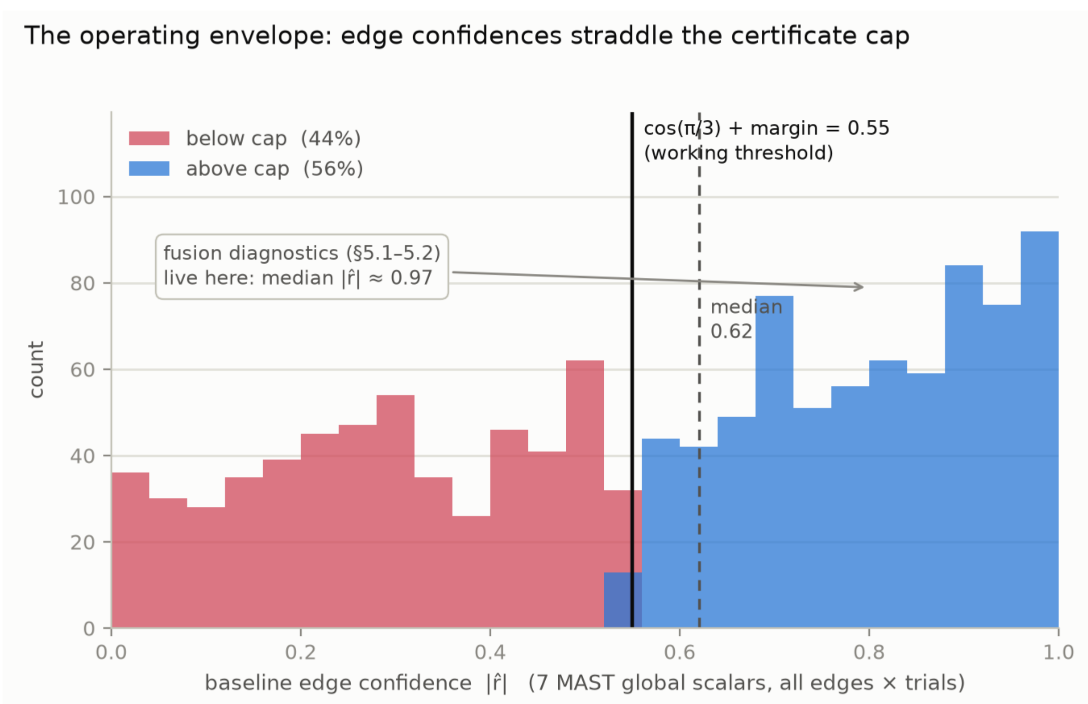

Figure 5. Envelope study edge-confidence histogram (Sec. 5.3).

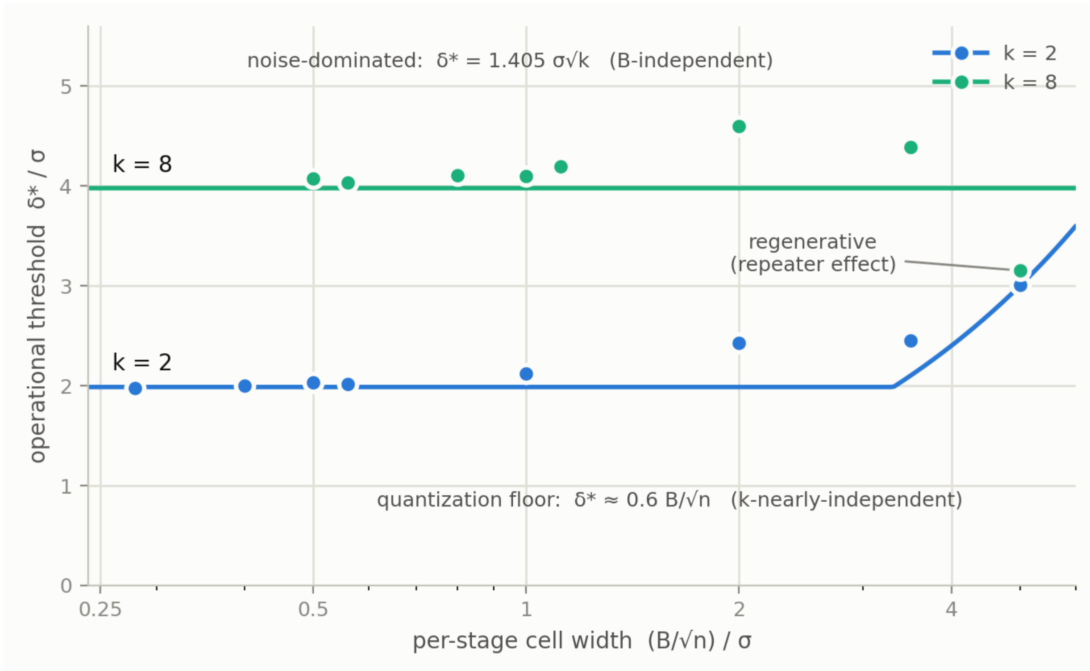

Figure 6. The operational two-regime law $\delta^* = \max(1.405\,\sigma\sqrt{k},\ 0.6\,B/\sqrt{n})$ (lines) against measured δ^* (dots); at coarse cells deep chains are regenerative.

Authorship and AI-assistance disclosure

This research program — its conception, its original ideas, its questions, and its direction — is the work of the author, Juan Carlos Paredes. The strategic pivots, the falsification discipline, the decisions about what to pursue, what to test, what to claim, and what not to claim are his throughout. In the course of the program he directed large language models (Anthropic Claude — Opus and Fable model families) as research assistants: implementing and running numerical experiments he commissioned, formalizing proofs in Lean 4, drafting and revising prose, and retrieving literature, all under his direction and review. Formal results are machine-checked (Lean 4) where stated; numerical claims are reproducible from the released scripts. Juan Carlos Paredes is the author of this work; the assistance is disclosed in the spirit of the full transparency this venue requires.

Independent researcher
Email address: cpt66778811@gmail.com