Precursor Detection of Systematic Bias Drift in Cryogenic Quantum Control: A Fail-Closed Estimator Validated in Simulation and on Live Hardware Calibration

Majid Hussain
Horos Labs Ltd, United Kingdom
info@horosengineering.com

Preprint — 2026

Abstract

Cryogenic control electronics for quantum hardware accumulate *systematic*, non-zero-mean bias—from thermal cycling of mounts, flux leakage, and dielectric memory—that integrates in a feedback loop until the actuator saturates. A controller that alarms on amplitude thresholds detects this only after saturation, when correction is no longer available. We present a precursor estimator that detects the *onset* of systematic drift from the dynamics of the control signal, before saturation. We build it deliberately as a *finite observer* in the sense of a companion paper on the physical limits of finite observation: it operates to a declared error budget, carries an explicit boundary of competence, and abstains—fails closed—outside it. The estimator uses path-functional invariants (a Fisher–Rao metric stability bound, a maximum-likelihood bias decoder, a cumulative-damage functional) rather than eigenvalue or linearisation analysis, which we show is neither necessary nor sufficient for finite-time operational survival. We report validation of two kinds. In *simulation*, across 30 seeds $\times$ 7 sensor-environment conditions (including a $1/f$ dephasing spectrum), the estimator detects 96.7% of drift-collapse scenarios before saturation, at a 12 s median lead, with 0% false positives across 210 healthy runs. On *real data*, the same governed outlier-and-drift detector, applied to live, publicly accessible IBM quantum-hardware calibration time-series (a 156-qubit IBM Heron device), flags genuinely degraded qubits and catches a real relaxation-time decline (one qubit's T_1 falling $88.0 \to 19.1\,\mu$s over 43 days) while raising no alarm on a qubit that was poor but stable, nor on the 155 others—detection of change, not of mere badness. We are explicit about maturity: this is a Technology Readiness Level 4 (TRL-4) detection layer, advisory and not a deployed controller, and it does not emulate entanglement, non-local correlation, or contextuality. The contribution is the integrated, calibrated estimator and its operating discipline, demonstrated on the quantum substrate in simulation and against real hardware-calibration data.

1 Introduction

A control loop for cryogenic quantum hardware is judged by how tightly it holds a setpoint, and the implicit failure it guards against is a large, visible excursion. But a common and damaging failure mode is not a sudden excursion at all: it is a slow, *systematic* bias—a persistent directional offset in the control voltage—that is not zero-mean and therefore accumulates. In a loop with integral action the accumulation is monotone; the integrator winds toward its limit and the actuator saturates. By the time an amplitude threshold trips, the controller has already lost its authority. The event is, in a precise sense, detectable only after it can no longer be corrected.

This paper presents a detection layer that intercepts that failure earlier, by reading the *dynamics* of the control signal for the signature of systematic drift rather than waiting for its amplitude consequence. Two commitments shape the design. The first is methodological: we do not characterise stability by the eigenvalues of a linearised model, because—as we show in Section 3—eigenvalue-based persistence is neither necessary nor sufficient for finite-time operational survival in a bounded, driven, nonlinear environment. We use instead path-functional invariants computed directly from the observed trajectory, requiring no linearisation, stationarity, or asymptotic limit. The second commitment is architectural, and it is the reason this work has the shape it does. We build the estimator as a *finite observer* in the sense made precise in a companion paper [1]: a system that, under a bounded-error criterion of correctness, must (N1) operate to a declared error budget, (N2) carry an explicit, checkable boundary of its competence, and (N3) abstain—fail closed—on inputs outside that boundary. That paper argues these properties are forced, across substrates, for any finite observer that is to be trustworthy; the present paper is the instantiation of that discipline on one substrate, the cryogenic quantum-control channel, together with its empirical validation.

The contribution is therefore not a new bound or a new physical principle. It is an integrated, calibrated estimator that realises (N1)–(N3) on the quantum-control substrate, and evidence—of two independent kinds—that it works: a Monte-Carlo characterisation in simulation (Section 5), and a demonstration of the same governed detector on live, real quantum-hardware calibration data (Section 6). We are explicit throughout about what is and is not claimed (Section 7); in particular, the estimator is an advisory detection layer at Technology Readiness Level 4, not a controller deployed on operational hardware, and it makes no claim to emulate quantum phenomena native to the hardware itself.

2 The discipline on the quantum substrate

We summarise the operating discipline the estimator inherits, so that the engineering in later sections can be read against it. The companion paper [1] shows that a finite observer which senses, infers, and acts is subject to bounds—on representation, estimation, memory, energy, and self-reference—that compound along its loop, and that under a modest criterion of correctness (bounded error on every input acted upon) these bounds force three properties. Mapped onto the present estimator:

(N1) Declared error budget. The estimator does not promise zero false alarms or perfect lead time. It is specified by an interdiction rule with a declared damage threshold and a target lead, and is characterised—not assumed—against those targets (Sections 5–6).

(N2) Explicit competence boundary. The estimator carries a *viability set* K, constructed self-normalisingly from the energy quantiles of baseline data (the tenth to ninetieth percentile), and a data-quality gate (the "Truth Gate") that decides whether an incoming window is admissible at all. The boundary is explicit and checkable, not implicit in tuned constants.

(N3) Abstention / fail-closed. When the input is inadmissible—signal-linearity or variance failure—the estimator does not emit a control verdict. It enters a *HOLD* state: it freezes the actuator at its last safe coordinate, wipes the integral, and defers. Withholding action on inputs outside competence is the safe default, exactly as the criterion requires.

The estimator's output vocabulary is correspondingly four-valued—NOMINAL, WATCHLIST, HOLD, INTERDICT—rather than a single binary alarm, so that "elevated but admissible," "inadmissible," and "act now" are distinct, reportable states. This is (N1)–(N3) made operational.

3 Physics of the failure, and why eigenvalues are the wrong invariant

3.1 Systematic bias drift

Cryogenic quantum-control electronics operate in dilution-refrigerator environments at 10–100 mK. Three mechanisms generate persistent, directional offsets in the control voltage that are not captured by zero-mean noise models: thermal cycling of mechanical mounts displaces control connectors (a systematic voltage offset); flux leakage from adjacent superconducting elements adds persistent coupling terms; and dielectric memory in cable insulation introduces a frequency-dependent phase lag. Each contributes a non-zero-mean term $\mathrm{bias}(t)$ that, past an onset time t_{onset}, accumulates at a per-step rate r_{drift}. In a loop with integral gain K_I and error $e(t) = \mathrm{target} - v(t) + \mathrm{bias}(t)$,

$$\textstyle\int(t) \;=\; \mathrm{clip}\big(\textstyle\int(t-1) + K_I\, e(t),\, -\textstyle\int_{\max},\, +\textstyle\int_{\max}\big),$$

so that a systematically non-zero $\mathrm{bias}(t)$ drives $\int(t) \to \int_{\max}$ monotonically. Reactive detection is possible only once $\int \geq \int_{\max}$—post-saturation. The estimator's task is to detect the onset of the drift regime well before that point.

3.2 Non-identifiability of eigenvalue persistence

A natural temptation is to certify stability through the leading eigenvalue of a linearised model. We record why this is unsafe here.

Theorem 1 (Non-identifiability)**.** Two stochastic systems sharing an identical linearisation matrix A—hence an identical leading eigenvalue $\lambda_{\max}(A)$—can exhibit arbitrarily different finite-time operational survival rates.

Proof sketch. Construct two systems with identical A at the origin but differing bounded nonlinear terms. System I has drift $g_1(x) = Ax$; System II has $g_2(x) = Ax + h(x)$, where h vanishes near the origin and provides strong nonlinear confinement for $|x| > r_0$. Both share $\lambda_{\max}(A)$. For moderate amplitudes System II is trapped by the nonlinear term and remains viable for arbitrarily long times while System I exits; the discrepancy is made arbitrarily large by adjusting r_0. Hence eigenvalue-based persistence is neither necessary nor sufficient for operational survival. □

The operational consequence is that any architecture relying solely on eigenvalue analysis is fragile in bounded, driven, nonlinear cryogenic environments. The estimator therefore uses path-functional invariants defined directly from trajectory behaviour, described next.

4 The estimator

The estimator runs as a set of pre-computation stages ahead of the control mathematics, combining four invariants under a consensus interdiction rule.

Truth Gate (admissibility). A data-quality stage tests each window for signal linearity and variance integrity. A window failing the test is declared inadmissible and routed to HOLD (N3): no verdict is emitted on data the estimator cannot trust. Linearity is probed by comparing the response to a calibration signal at full and half amplitude, $\Delta_{\mathrm{lin}} = \|X[u] - 2X[u/2]\|/\|X[u]\|$; channels with large Δ_{lin} are flagged as producing ghost responses.

Fisher–Rao stability bound. On the manifold of control-state distributions, the metric stability constant for operational rank n is

$$\Phi_M(n) \;=\; \frac{n\,\lambda_{\min}(g)}{2\,C_{\rm occ}},$$

with $\lambda_{\min}(g)$ the minimum eigenvalue of the Fisher–Rao metric at the operating point and $C_{\rm occ}$ the occupation-time variance. The manifold stays non-degenerate under interaction flux $\Phi(t)$ iff $\Phi(t) \le \Phi_M(n)$; when violated, an eigenvalue of g crosses zero in finite time, bounded by $t^\star \le C_{\rm occ}/[\Phi(t) - \Phi_M(n)]$. Monitoring $\lambda_{\min}$ thus gives a geometric precursor with a computable collapse-time bound.

Maximum-likelihood bias decoder. Rather than assume zero-mean noise, the estimator decodes the systematic component. With the perturbation written $\varepsilon_R(t) = Q\,\Delta p_{\rm need}(t) + \eta(t)$ (η zero-mean), the linear-regime maximum-likelihood estimate of the directional need-bias from the gate-evaluation history is the classical unbiased estimator for a biased Bernoulli process,

$$\Delta p^\star_{\rm need} \;\approx\; 2\big[(n_+/n) - \tfrac{1}{2}\big],$$

with n_+ the count of admissible-state evaluations in the window and n the total. Correction is steered toward the drift-displaced operating point rather than the nominal fixed point, which yields strictly longer expected survival when a non-zero need-bias exists.

Cumulative-damage functional. A Miner's-rule accumulator $D(t) = D(t-1) + k_D\,e(t)^2$, with failure at $D \ge 1.0$, supplies a physics anchor: healthy noise produces symmetric errors that do not accumulate monotonically, so the accumulator is structurally incapable of firing under nominal operation. Interdiction is set at $D = 0.70$, a 30% margin below failure.

Consensus interdiction. The four triggers are: (A) signal deviation above $3.2\times$ the baseline median-absolute-deviation (MAD); (B) a cumulative-sum statistic above 3.8; (C) $D \ge 0.70$; (D) Fisher–Rao bias magnitude above 0.14. The estimator issues INTERDICT when (C) holds together with at least one of (A,B,D), or when (C) is sustained for ≥ 8 consecutive steps; lesser elevations yield WATCHLIST. A four-quadrant regime map—*Sovereign* (deploy), *Critical-Fragile* (dampen), *Committed-Stiff* (excite), *Drift/Chaos* (reset)—selects the corrective posture, distinguishing a channel that has lost viability from one that has lost responsiveness. All thresholds on energy are self-normalised from baseline quantiles; the design carries no externally fixed magic constants.

5 Validation I — simulation

Setup. A three-layer rig was used. Layer 1 (a cryogenic-control model) generates healthy and collapse trajectories with collapse onset $t_{\rm onset}/T \in [0.28, 0.48]$ per seed and drift rate $r_{\rm drift} \in [0.007, 0.020]$ per step; the 300-step window exceeds $5\,\tau_{\rm inst}$, placing all runs in the asymptotically stable detection regime ($|A_1(T) - A_1(\infty)| < 10^{-3}$). Layer 2 injects seven sensor-environment conditions: thermal-drift offset (Q1), RF-crosstalk bursts (Q2, $p = 0.05$ per step), FPGA dropout (Q4, $p = 0.04$), DAC saturation (Q5), all-combined (ALL), and a $1/f$ pink-noise spectrum (Q6, $S(f) \propto 1/f$, representing the realistic cryogenic dephasing spectrum). Layer 3 is the estimator of Section 4.

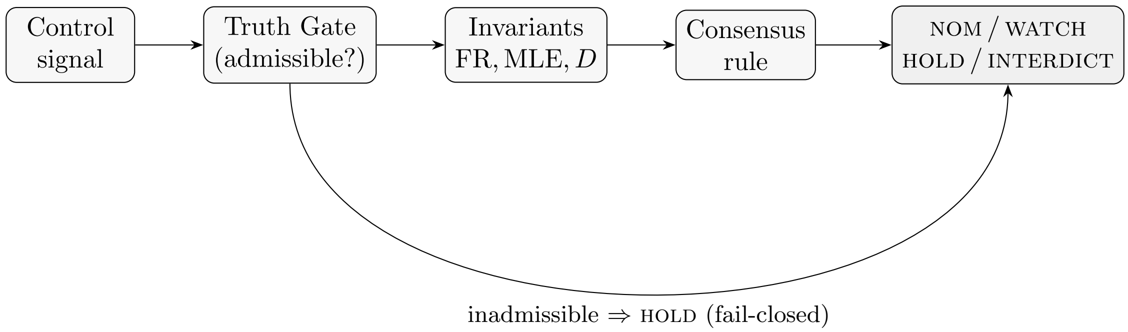

Figure 1: The estimator as a finite observer: a data-quality gate (N2 competence boundary) ahead of path-functional invariants and a consensus interdiction rule, emitting a four-valued state with an abstention (HOLD) path (N3 fail-closed).

Sensor environment	Detection	Median lead	False positive
Clean baseline	96.7%	12 s	0%
Q1 — thermal drift	96.7%	12 s	0%
Q2 — RF crosstalk	96.7%	12 s	0%
Q4 — FPGA dropout	96.7%	12 s	0%
Q5 — DAC saturation	96.7%	12 s	0%
ALL — combined	96.7%	12 s	0%
Q6 — $1/f$ pink noise	96.7%	12 s	0%

Table 1: Monte-Carlo characterisation in simulation (30 seeds per condition; 210 healthy runs total). These are simulation results, reported as such.

Result. Across a Monte-Carlo design of 30 seeds $\times$ 7 conditions $\times$ 2 signal modes, the estimator detected 96.7% of collapse scenarios before actuator saturation, at a 12 s median lead, with *zero* false positives across all 210 healthy runs (Table 1). Detection held under every condition including the $1/f$ spectrum. On the reference seed, INTERDICT fired at $t = 226$ s with failure ($D = 1.0$) at $t = 238$ s—a 12 s lead. We count only healthy-signal runs reaching INTERDICT as false positives; WATCHLIST and HOLD on healthy tracks are advisory states, not false alarms.
These figures are a *simulation* characterisation and are reported as such; they establish the estimator's behaviour against a controlled adversary, not its performance on operational hardware.

6 Validation II — live hardware calibration data

Simulation can be tuned; real hardware cannot. To test the same discipline against data the authors did not generate, we applied the estimator's governed outlier-and-drift detector—the median±MAD admissibility test together with a persistence-gated drift test, the same family used in the simulation above—to live, publicly accessible calibration telemetry from an IBM superconducting quantum processor (an IBM Heron device). We stress what this is: a read of the device's routine calibration record (per-qubit relaxation time T_1, dephasing time T_2, and readout error), accessed read-only through IBM Quantum's runtime interface (`qiskit-ibm-runtime`). No circuits were executed, no queue time or cost was incurred, and the control loop itself was not touched. It is a test of the detector on real hardware *metadata*, not a deployment on the control hardware.

Spatial axis — degraded-element detection. On the 156-qubit device, the calibration snapshot gave a median T_1 of $142.3\,\mu$s (range 12.5–303.4), median T_2 of $107.8\,\mu$s (5.5–249.1), and median readout error 0.84% (maximum 35.2%). The MAD outlier test (T_1 below the median by more than three MADs) flagged two genuinely degraded qubits, the worst at $T_1 \approx 7.2\,\mu$s—about twenty times below the median.

Temporal axis — a real drift event, and the silence that matters. Reading the device's stored daily history (43 daily snapshots $\times\,156$ qubits), the persistence-gated drift test fired on exactly one qubit: its T_1 declined from 88.0 to $19.1\,\mu$s—a 78% loss—over the window, a qubit that had been healthy and sustainedly degraded. Nothing spurious fired among the other 155. In particular, the qubit that was *stably* poor (the $\approx 7.2\,\mu$s outlier, which held 7.2–$8.6\,\mu$s across ~ 3 weeks) did *not* trip the drift test: it is bad, but it is not changing. This is the property that makes the detector useful rather than noisy—it fires on change, not on mere badness—and it is the temporal counterpart of the 0% false-positive discipline seen in simulation.

The same governed detector therefore demonstrates both axes on real quantum-hardware data: a spatial degraded-element flag and a genuine temporal drift event, with the false-alarm discipline visible in its correct silence on the stable-but-poor qubit and the 155 unaffected ones. This is the empirical leg the control-layer simulation alone cannot supply.

7 Maturity, scope, and what is not claimed

Technology readiness. The work is at TRL 4: validated in computational simulation, and—for the detector family—against real hardware calibration data. The following are explicitly *not yet* demonstrated: operation on a live qubit-control loop inside a dilution refrigerator; integration with an error-correction stack, a real-time feedback controller, or the classical control layer; false-positive characterisation across a full quantum-processor operational envelope; and any certification pathway for safety-critical use. A staged validation programme—shadow-mode baseline acquisition, then intervention mode, each with declared go/no-go criteria—is the natural next step; it is described as future work, not as a result.

What is and is not claimed. *Claimed:* that systematic bias drift is detectable from signal dynamics before actuator saturation; that an estimator built to the finite-observer discipline (N1)–(N3) does so in simulation at the rates of Table 1; and that the same governed detector identifies degraded elements and a real drift event in live hardware-calibration data while staying silent on stable-but-poor and healthy elements. *Not claimed:* deployment on operational quantum-control hardware; that the simulation rates transfer unchanged to hardware; novelty of the underlying mathematics (change-point detection, robust statistics, Fisher information, and anti-windup control are all established); or any guarantee of performance outside the declared competence boundary. Crucially, the estimator is a *classical* control-layer detector: it reproduces weak-measurement-compatible detection statistics but does *not* emulate entanglement, non-local correlation, or quantum contextuality—those are native to the hardware and outside the scope of this layer.

8 Relation to prior work

The estimator assembles established components: sequential change-point detection of the cumulative-sum class [2]; robust scale estimation via the median absolute deviation [3]; the Fisher–Rao metric of information geometry [4, 5]; the Cramér–Rao and quantum-metrology limits that make finite-resource estimation error strictly positive [4, 6]; cumulative-damage accumulation of Miner's-rule form [7]; and integral anti-windup from classical control [8]. Its character as a fail-closed, competence-bounded finite observer is the instantiation, on the quantum-control substrate, of the discipline derived in [1]. The $1/f$ dephasing spectrum used as a simulation stressor reflects the noise environment documented for superconducting qubits [9]. The contribution is the integration and calibration of these into a single estimator, and its validation in simulation and against live hardware-calibration data.

9 Conclusion

Systematic bias drift in cryogenic quantum control is dangerous precisely because a threshold-based controller sees it only after the integrator has saturated. Reading the dynamics instead of the amplitude moves detection before that point. Built to the finite-observer discipline—an error budget, an explicit competence boundary, and a fail-closed abstention—the estimator detects drift onset at a 12 s median lead with no false alarms across 210 healthy simulated runs, and the same governed detector, turned on real quantum-hardware calibration data, finds the degraded qubits and a genuine 78% relaxation-time decline while staying silent on what is merely bad-but-stable. The honest position is modest and exact: a TRL-4 advisory detection layer, validated in simulation and against real calibration data, with a declared boundary and a staged path to hardware. Building deliberately to that boundary, and saying so, is the discipline this paper has tried to demonstrate on the quantum substrate.

Competing interests

The author is the founder and director of Horos Labs Ltd, the entity at which this work was carried out.

Data availability

The simulation summary and reference-run records are available from the author. The hardware calibration data are publicly accessible, read-only, through IBM Quantum's runtime interface (the `qiskit-ibm-runtime` API): T_1, T_2, and readout-error properties, both current and historical. The device reported here was the 156-qubit IBM Heron processor `ibm_fez`. The analysis executed no circuits and is reproducible by any account with read access.

Acknowledgements

This work was carried out at Horos Labs Ltd, United Kingdom. It builds directly on the companion paper [1].

References

[1] M. Hussain, "The Physical Limits of Finite Observation: Seven Bounds on Sensing, Inference, and Control," Horos Labs Ltd, preprint (2026). `https://doi.org/10.5281/zenodo.21032554`.

[2] E. S. Page, "Continuous inspection schemes," *Biometrika* **41**, 100–115 (1954).

[3] F. R. Hampel, "The influence curve and its role in robust estimation," *J. Amer. Statist. Assoc.* **69**, 383–393 (1974).

[4] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.* **37**, 81–91 (1945).

[5] S. Amari, *Information Geometry and Its Applications*, Springer (2016).

[6] V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," *Nat. Photonics* **5**, 222–229 (2011).

[7] M. A. Miner, "Cumulative damage in fatigue," *J. Appl. Mech.* **12**, A159–A164 (1945).

[8] K. J. Åström and L. Rundqvist, "Integrator windup and how to avoid it," *Proc. American Control Conf.*, 1693–1698 (1989).

[9] E. Paladino, Y. M. Galperin, G. Falci, and B. L. Altshuler, "$1/f$ noise: implications for solid-state quantum information," *Rev. Mod. Phys.* **86**, 361–418 (2014).