A Stability Index for Cross-Domain Degradation Detection

Shawn Barnicle
Independent Researcher
Chicago, USA
ShawnBarnicle.ai@gmail.com

February 5, 2026

Abstract

We report an empirical pattern: a stability index Φ, formed from normalized identity (baseline preservation), temporal coherence, and Shannon entropy, separates stable from failing regimes across diverse engineered systems using a common critical region near 0.25 under a fixed protocol. The protocol and critical region were fixed from bearing studies (October 2024) and then applied unchanged across six engineered domains in 31 evaluated cases: ten run-to-failure bearings, ten run-to-failure turbofan engines, two power-grid frequency recordings (including the Great Britain blackout of 9 Aug 2019), three earthquake-precursor cases plus a quiet baseline, two neural-network training cases, and 445 superconducting qubits across three IBM Quantum backends. Across non-quantum domains, evaluated failure/critical intervals yielded Φ below the critical region while included stable controls yielded Φ above it under their domain protocols; in the quantum evaluation, qubit subsets above the region exhibited substantially lower shot-based circuit error than subsets below it on the same backend snapshots. We additionally report a biological extension on cardiac and neural signals. For cardiac rhythm windows, arrhythmia discrimination achieved AUC $\approx$ 0.90 with improved balanced performance at an operating threshold near $\Phi_c \approx 0.0$, while atrial fibrillation versus normal discrimination was markedly weaker (AUC $\approx$ 0.56), suggesting a scope boundary for stable-state discrimination versus destabilization/transition detection. The index performed within 0.07 AUC of domain-specific heart rate variability metrics and maintained discrimination (AUC 0.64–0.88) across temporal resolutions from 5 to 30 seconds. For neural signals (EEG seizure data), a higher coupling constant ($\alpha \approx 0.55$) yielded improved separation relative to $\alpha = 0.1$, suggesting that α may vary by system type. We report the observation without claiming a mechanistic derivation and provide sufficient methodological detail for independent scrutiny; code and result artifacts can be provided for replication upon reasonable request, with distribution potentially subject to licensing due to ongoing intellectual-property constraints.

1 Introduction

Detecting when a system is entering a failing regime remains a domain-specific problem. Mechanical engineers monitor vibration signatures. Electrical engineers track frequency deviations. Machine learning practitioners watch accuracy curves. Each field has developed its own metrics, thresholds, and intuitions.

This paper reports an unexpected finding from our work on bearing degradation. While analyzing vibration data from rotating machinery, we constructed a stability index combining three quantities: a normalized identity ratio, temporal autocorrelation, and Shannon entropy. The index took the form $\Phi = I \times \rho - \alpha \times S$ with a coupling constant $\alpha = 0.1$. Across the evaluated cases, intervals with $\Phi_{\min} > 0.25$ corresponded to stable/control regimes, while intervals with $\Phi_{\min} < 0.25$ corresponded to failing or critical regimes under the evaluation protocols described below.

What surprised us was not that the index worked on bearings. What surprised us was that the same formula, with the same threshold, worked on power grids, turbofan engines, earthquake precursors, and neural networks.

We are not claiming to have discovered a fundamental law of physics. We are reporting an empirical regularity that we cannot fully explain. The threshold 0.25 was not derived from theory; it emerged from the bearing data and then held across other domains. Whether this reflects deep structure or coincidence, we leave to others to investigate. Our contribution is the observation itself, documented with sufficient detail for replication.

For transparency about out-of-sample evaluation, we record the chronology: $\alpha = 0.1$ and $\Phi_c = 0.25$ were fixed from the XJTU-SY bearing analysis in October 2024; all other domain analyses reported here were performed between November 2024 and January 2026 using identical α and Φ_c values. These dates are verifiable via timestamped result artifacts and repository history (see Data and Code Availability).

The paper proceeds as follows. Section 2 defines the stability index and describes how its components are computed in each domain. Section 3 presents validation results across six domains totaling 31 independent systems. Section 4 discusses limitations, possible explanations, and directions for future work.

2 The Stability Index

2.1 Definition

The stability index Φ combines three measurable quantities:

$$\Phi = I \times \rho - \alpha \times S \tag{1}$$

where I is normalized identity (behavioral consistency relative to a baseline), ρ is coherence (temporal autocorrelation), S is Shannon entropy (in bits), and $\alpha = 0.1$ is a coupling constant. For scalar degradation metrics used in physical monitoring (e.g., RMS amplitude, variance, deviation magnitude), we compute identity as a capped baseline-to-current ratio with an exponent $\beta \geq 1$:

$$I(t) = \min\left(\left(\frac{m_{\text{baseline}}}{m_{\text{current}}(t)}\right)^{\beta}, 1\right), \tag{2}$$

where m is a domain-appropriate degradation metric, β is specified per domain (e.g., $\beta = 1$ for vibration RMS), and the cap prevents $I > 1$ during early stable operation.

Unless otherwise stated, coherence is computed as lag-1 Pearson autocorrelation on the windowed signal x:

$$\rho = \text{corr}(x_t, x_{t-1}). \tag{3}$$

In time-series domains, $\rho \in [-1, 1]$ and may be negative; negative ρ indicates anti-correlated dynamics, which can occur in unstable regimes approaching failure. For quantum calibration analysis (Section 3.6), we use a bounded coherence proxy derived from relaxation/dephasing metrics.

Entropy is computed as Shannon entropy (in bits) on a discretized distribution within each window, with the number of bins K fixed per domain as specified in code:

$$S = -\sum_{i=1}^{K} p_i \log_2(p_i). \tag{4}$$

Unless otherwise stated, for binned continuous-valued time series (grids and strain/strain-rate) we use $K = 50$ bins per evaluated window. For XJTU-SY bearings we use $K = 50$ bins; for NASA C-MAPSS turbofans we follow the reference implementation with $K = 10$ bins. For confusion-matrix entropy the distribution is over the n^2 cells (no additional binning).

In the six engineered domains evaluated here, we observe a critical region near $\Phi_c \approx 0.25$ separating stable from failing/critical regimes under the fixed protocol. In a biological cardiac extension, the same Φ construction preserves ordering (lower Φ for more disordered/less coherent rhythm), but the best operating threshold shifts downward (near $\Phi_c \approx 0.0$) consistent with higher baseline variability in healthy physiological signals.

2.1.1 Computation Protocol (All Domains)

Unless otherwise noted, Φ is computed on a sliding window over the time series, and each system is summarized by the minimum window value $\Phi_{\min}$ observed within the evaluated interval (e.g., end-of-life for run-to-failure datasets, or a pre-event window for incident detection), reflecting that failures are typically triggered by worst-case excursions rather than average behavior. Windows are evaluated with fixed parameters chosen a priori; we do not tune α or Φ_c per domain. The window length, stride, baseline period definition, and the exact evaluation interval (e.g., pre-event cutoff time) are provided in the accompanying codebase and result artifacts (see Data and Code Availability). Where an interval-level protocol is used (XJTU-SY bearings; NASA C-MAPSS turbofans), we report the corresponding interval summary value (e.g., Φ^{early}, Φ^{EOL}) rather than $\Phi_{\min}$.

2.2 Component Definitions by Domain

The abstract quantities I, ρ, and S require domain-specific operationalization. We describe each below. In all domains, identity I is operationalized as baseline preservation of a domain-appropriate performance or degradation statistic; the specific metric varies by domain but the conceptual role is uniform.

Although the operational definitions of I, ρ, and S are domain-specific, their roles are invariant:

I quantifies baseline preservation of a scalar performance/degradation statistic, ρ quantifies short-horizon temporal coherence of the monitored process, and S quantifies distributional disorder within the same observation window. The claim of universality in this paper is therefore not that identical raw measurements are used across domains, but that the same functional combination of (baseline preservation $\times$ coherence) opposed by an entropy penalty yields a consistent critical region under fixed global parameters. In all domains, I is normalized to be approximately bounded in $[0, 1]$ in this analysis (via capping or midpoint normalization), enabling cross-domain comparison.

2.2.1 Mechanical Systems (Bearings)

For vibration-based monitoring, identity I is the ratio of baseline to current RMS amplitude, capped at 1.0:

$$I(t) = \min\left(\frac{\text{RMS}_{\text{baseline}}}{\text{RMS}_{\text{current}}(t)}, 1\right) \tag{5}$$

where the baseline is computed from early operational life (first 10% of measurements). Coherence ρ is the lag-1 autocorrelation of the vibration time series. For XJTU-SY bearings, entropy follows the reference implementation: $S = -\sum_{i=1}^{K} p_i \log_2(p_i)$ using $K = 50$ histogram bins over the evaluated window. We report Φ^{early} computed on the first 10% window and Φ^{EOL} computed using the reference end-of-life protocol; these are interval-level values rather than sliding-window minima.

2.2.2 Aerospace Systems (Turbofans)

For turbofan engines with multiple sensors, the component quantities are computed per sensor channel and then aggregated as an unweighted mean across the available sensor channels under the reference implementation.

2.2.3 Electrical Systems (Power Grids)

For grid frequency monitoring, identity I is the ratio of baseline frequency variance to current variance, measuring how tightly the grid maintains nominal frequency (50 Hz in Europe). Coherence ρ is the autocorrelation of the frequency time series. Entropy S is computed as Shannon entropy (in bits) from the binned distribution of frequency deviations.

2.2.4 Geophysical Systems (Earthquake Precursors)

For strain-based monitoring, identity I is the ratio of baseline strain variance to current variance. Coherence ρ is the temporal autocorrelation of strain measurements. Entropy S is computed as Shannon entropy (in bits) from the binned distribution of strain rate values. Analysis uses pre-event windows only, ensuring that classification is based on precursor data rather than post-hoc fitting.

2.2.5 Computational Systems (Neural Networks)

For neural network monitoring, identity I is computed from the confusion matrix as the correlation between current and baseline error patterns. Coherence ρ is the autocorrelation of accuracy across training epochs. Entropy S is computed as Shannon entropy (in bits) on the normalized confusion-matrix cell distribution.

2.2.6 Biological Systems (Cardiac Rhythm)

For cardiac rhythm monitoring, we construct a beat-to-beat interval time series (RR intervals) and compute Φ on fixed-duration windows under the same functional form $\Phi = I \times \rho - \alpha \times S$ with $\alpha = 0.1$. Identity I is operationalized as baseline preservation of an RR-variability statistic (implemented as a baseline-to-current ratio and capped to remain ≤ 1 during stable operation), coherence ρ is computed as lag-1 autocorrelation on the RR-interval series within the window, and entropy S is computed as Shannon entropy (in bits) on a discretized RR-interval distribution within the window (binning fixed in code). This operationalization differs from engineered domains only in the underlying observable (RR intervals rather than vibration, frequency, strain, or confusion-matrix evolution), while preserving the same abstraction: baseline preservation, temporal coherence, and within-window disorder.

2.2.7 Biological Systems (Neural/EEG)

For neural signal monitoring, we construct a time series from EEG amplitude (RMS across channels). Identity I is baseline preservation of RMS amplitude, coherence ρ is lag-1 autocorrelation, and entropy uses a deviation formulation $S_{\text{dev}} = |S - S_{\text{baseline}}|$. Optimization indicated $\alpha \approx 0.55$ outperformed $\alpha = 0.1$ on the evaluated EEG data (see Section 3.8).

3 Validation Results

We validated the stability index on 31 evaluated cases across six engineered domains. In all engineered-domain evaluations, we used publicly available datasets and evaluated the index using fixed parameters (no per-domain tuning) and either run-to-failure end-of-life intervals or retrospectively defined pre-event windows only. We additionally report a biological cardiac extension (Section 3.7) under the same Φ form and $\alpha = 0.1$, noting a downward shift in the optimal operating threshold.

Case selection and inclusion criteria. Within each domain we evaluated the systems analyzed here under fixed settings: all ten XJTU-SY bearings reported here, all ten FD001 turbofan units reported here, both grid recordings used here (UK Aug 2019 incident window; Germany Sep 2019 stable window), the seismic/precursor cases listed in Table 4 for which pre-event data were available under the described construction, both neural-network training cases as defined here, and all qubits available from the reported calibration snapshots for the three IBM backends. We did not tune α or Φ_c per domain and did not exclude cases based on Φ outcomes.

3.1 Mechanical: Bearing Degradation

We analyzed ten bearing systems from the XJTU-SY dataset [1], which contains complete run-to-failure vibration measurements under three operating conditions (35 Hz/12 kN, 37.5 Hz/11 kN, 40 Hz/10 kN).

Table 1 shows both early-window and end-of-life values. All ten bearings exhibited $\Phi^{\mathrm{EOL}} < 0.25$ over the evaluated end-of-life interval, with values ranging from -0.370 to -0.003. For the first 10% window, four bearings showed $\Phi^{\mathrm{early}} > 0.25$, while six exhibited $\Phi^{\mathrm{early}} \leq 0.25$, suggesting that some runs may have entered a degrading regime prior to the start of monitoring (or that the first 10% window is not uniformly healthy across bearings). The negative values indicate that the disorder term dominated the coherent identity term under this protocol.

System	Φ^{early}	Φ^{EOL}	Outcome
Bearing1_1	0.509	−0.168	Failed
Bearing1_2	−0.124	−0.336	Failed
Bearing1_3	−0.017	−0.261	Failed
Bearing1_4	0.493	−0.003	Failed
Bearing2_2	−0.361	−0.263	Failed
Bearing2_3	−0.129	−0.370	Failed
Bearing2_4	0.147	−0.255	Failed
Bearing2_5	0.467	−0.344	Failed
Bearing3_1	0.254	−0.276	Failed
Bearing3_4	−0.425	−0.290	Failed

Table 1: Bearing results showing an early window (first 10% of measurements) and an end-of-life evaluated interval. Φ^{early} may indicate degradation began prior to monitoring start for some runs under this dataset/protocol.

3.2 Aerospace: Turbofan Engine Degradation

We analyzed ten turbofan engines from the NASA C-MAPSS dataset [2], which contains run-to-failure sensor measurements (21 channels) under varying operational conditions and fault modes.

All ten engines exhibited $\Phi^{\mathrm{EOL}} < 0.25$ over the evaluated end-of-life interval, with values ranging from 0.039 to 0.241 (Table 2). In this study we therefore treat the turbofan evidence as end-of-life discrimination under unchanged parameters. We note that for FD001 the first 10% of cycles is not a reliable "healthy control" interval under the reference implementation (the early interval also yields $\Phi < 0.25$ for these units), consistent with nonstationarity and/or degradation dynamics being present early in the simulated trajectories.

Engine	Φ^{EOL}	Outcome
Unit 1	0.205	Failed
Unit 2	0.186	Failed
Unit 3	0.152	Failed
Unit 4	0.241	Failed
Unit 5	0.236	Failed
Unit 6	0.169	Failed
Unit 7	0.066	Failed
Unit 8	0.130	Failed
Unit 9	0.039	Failed
Unit 10	0.165	Failed

Table 2: Turbofan end-of-life results (NASA C-MAPSS FD001). All ten engines exhibited $\Phi^{\mathrm{EOL}} <$ 0.25 over the evaluated end-of-life interval under the fixed reference implementation.

3.3 Electrical: Power Grid Stability

We analyzed 1-second grid frequency data from Great Britain (August 2019; ∼2.6 million samples; provider: National Grid) and Germany (September 2019; ∼2.5 million samples; provider: TransnetBW) curated in the Power Grid Frequency Database [4].

On August 9, 2019, the UK experienced a major blackout affecting approximately one million people [3]. Using frequency data from before the event, we computed $\Phi_{\min} = 0.178$. This value falls below the 0.25 threshold, indicating critical instability.

For comparison, we analyzed a stable period from the German grid during September 2019. The computed value was $\Phi_{\min} = 0.401$, above the threshold, correctly classifying stable operation.

We emphasize that the UK analysis used pre-event data only. The parameters ($\alpha = 0.1$, threshold 0.25) were established from the bearing analysis months earlier. This was not post-hoc fitting.

System	$\Phi_{\min}$	Class.	Outcome
UK (Aug 2019)	0.178	Critical	Blackout
Germany (Sep 2019)	0.401	Stable	No event

Table 3: Power grid stability. The UK blackout was correctly identified as critical from pre-event data.

3.4 Geophysical: Earthquake Precursors

We analyzed geophysical precursor behavior using two data types: (i) continuous strain time series from the USGS Donna Lea strainmeter [5] in California (707,471 measurements spanning 2002–2016), and (ii) an event-catalog foreshock sequence for Tohoku derived from USGS-reported foreshocks occurring March 9–10, 2011.

Parkfield (M6.0, 2004) and San Simeon (M6.5, 2003) were analyzed using pre-event strain time series windows from the Donna Lea strainmeter. Tohoku (M9.1, 2011) was analyzed using a derived time series constructed from the USGS earthquake catalog (ComCat) [6]: we computed event counts in fixed 1-hour bins over the 48 hours preceding the mainshock; identity I is computed as a baseline-to-current variance ratio using the first 12 hours as baseline, and ρ and S are computed on this event-rate series using the same protocol as other domains. In all cases, $\Phi_{\min} < 0.25$ was observed in the pre-event window, while a quiet baseline year (2010) from the Donna Lea strainmeter yielded $\Phi_{\min} = 0.577$, indicating stability.

Notably, Parkfield and San Simeon were both flagged from the same sensor (Donna Lea), demonstrating that the result reflects genuine signal rather than sensor-specific artifacts.

Event	Magnitude	$\Phi_{\min}$	Outcome
Tohoku (2011)	M9.1	-0.357	Confirmed
Parkfield (2004)	M6.0	0.114	Confirmed
San Simeon (2003)	M6.5	0.084	Confirmed
Stable year (2010)	—	0.577	No event

Table 4: Earthquake precursor analysis. Pre-event $\Phi_{\min}$ values correctly distinguished seismically active periods from quiet periods.

3.5 Computational: Neural Network Stability

We analyzed two neural-network training cases (MNIST [7], Fashion-MNIST [8]) under the same evaluation settings using confusion-matrix evolution across epochs.

Networks that failed to converge showed $\Phi_{\min} < 0.25$ early in training. Networks that successfully formed stable representations showed $\Phi_{\min} > 0.25$. Using the protocol above, the unstable training case (catastrophic forgetting) satisfied $\Phi_{\min} < 0.25$ while the stable training case (class imbalance) satisfied $\Phi_{\min} > 0.25$.

In the cross-domain count of 31 systems (Table 10), we treat MNIST and Fashion-MNIST as two computational systems evaluated with the same fixed settings.

We also validated the index on behavioral drift detection across four domains (computer vision, NLP, medical AI, audio classification) using eleven independent tests. The threshold correctly separated failing from stable systems in all cases. These drift tests are reported separately and are not included in the 31-case summary in Table 10.

3.6 Quantum: Superconducting Qubit Stability

Unlike the other domains where Φ is computed directly from the monitored time series, the quantum Φ components are computed from publicly reported calibration properties (a deterministic proxy), and validation is performed by comparing shot-based circuit success rates between HIGH-Φ and LOW-Φ subsets on the same backend snapshot.

We analyzed 445 superconducting qubits across three IBM Quantum backends (ibm_fez, ibm_torino, ibm_marrakesh) using publicly available calibration data. For quantum systems, identity I is normalized fidelity computed from calibration-derived error rates. In our implementation, "fidelity" is computed deterministically from backend calibration properties (gate error and readout error terms) for the specific qubits/couplers used in the evaluated subset, as specified in the accompanying codebase (see Data and Code Availability). We then normalize relative to a fixed midpoint reference:

$$I = \frac{\text{fidelity} - 0.50}{0.50}. \tag{6}$$

Here "fidelity" denotes the calibration-derived quality proxy computed deterministically from backend properties (as specified in code), and 0.50 is used as a fixed normalization reference for this proxy in our protocol. Coherence ρ is computed as a bounded proxy $\rho = \min(T_2/T_1, 1)$, measuring how well phase coherence is maintained relative to energy relaxation. Entropy S is represented by an entropy proxy derived from readout error, reflecting measurement disorder; specifically, the proxy is computed deterministically from the reported readout error terms in the backend calibration snapshot (see accompanying code).

Qubits were classified as HIGH-$\Phi_{\min}$ ($\Phi_{\min} \geq 0.25$) or LOW-$\Phi_{\min}$ ($\Phi_{\min} < 0.25$), where $\Phi_{\min}$ is evaluated per calibration snapshot rather than a time-sliding window. The reported error rate is defined as $1 - p_{\text{success}}$, where p_{success} is the empirical probability mass on the ideal output bitstring(s) for each algorithm under 1024 shots. Circuits were compiled using Qiskit's default transpilation at optimization level 1 and executed on the same backend calibration snapshots used to compute $\Phi_{\min}$. Full circuit definitions, ideal outputs, and result artifacts are available on request (see Data and Code Availability).

Backend	Qubits	LOW-Φ Err	HIGH-Φ Err
ibm_fez	156	1.28%	0.08%
ibm_torino	133	0.90%	0.36%
ibm_marrakesh	156	0.46%	0.08%

Table 5: Quantum backend validation. HIGH-Φ qubits show 2.5–16× lower error rates than LOW-Φ qubits across all backends.

The best discrimination was observed on Bernstein-Vazirani: HIGH-$\Phi_{\min}$ qubits showed 2.12% error versus 64.72% for LOW-$\Phi_{\min}$ qubits, a ratio of 30.47×. The threshold also identified dead qubits across the backends, all of which showed $\Phi_{\min} < 0$ under the same computation.

Notably, Nation and Treinish [9] note that relaxation/dephasing information (e.g., T_1/T_2-derived quantities) was not included in their default layout scoring because it did not strongly affect layout ordering in their reported tests. In our protocol, the coherence proxy $\rho = \min(T_2/T_1, 1)$ is included directly in Φ because it measurably improves separation between low-error and high-error subsets in the evaluated backends under the fixed threshold.

3.7 Biological: Cardiac Rhythm Stability

We evaluated a biological extension using PhysioNet cardiac databases [12, 13, 11]. The evaluation uses the same Φ form and $\alpha = 0.1$.

3.7.1 Arrhythmia Detection

Using labeled windows from the MIT-BIH Arrhythmia Database (48 patients, 110,000+ beats), the index achieved AUC = 0.902. The 0.25 threshold yields 99.6% recall but poor specificity; $\Phi_c = 0.0$ improves balance:

Φ_c	Recall	Spec.	MCC	AUC
0.25	99.6%	32.4%	0.415	0.902
0.0	95.7%	70.2%	0.668	0.902
-0.1	90.5%	79.7%	0.699	0.902

Table 6: Cardiac threshold analysis. The 0.25 threshold retains meaning (99.6% recall) but optimal balance shifts to $\Phi_c \approx 0.0$.

3.7.2 Atrial Fibrillation

On AFDB (AFIB vs NORMAL), discrimination was weak (AUC = 0.556). This suggests the index detects transitions rather than stable-state differences.

3.7.3 Comparison with HRV Metrics

We compared Φ against standard heart rate variability metrics:

Metric	AUC	Type
RMSSD	0.971	Cardiac-specific
SDNN	0.926	Cardiac-specific
Φ	0.902	Cross-domain

Table 7: Φ vs HRV metrics. Gap of 0.07 AUC represents the tradeoff for cross-domain applicability.

3.7.4 Resolution Scaling

Performance degrades gradually with shorter windows:

Window	AUC	Use Case
60 sec	0.90	Clinical
30 sec	0.88	Bedside
10 sec	0.76	Wearable
5 sec	0.64	Real-time

Table 8: Resolution scaling. No abrupt cliff; 30-second approaches clinical performance.

3.8 Biological: Neural Signal Stability

We evaluated the CHB-MIT Scalp EEG Database [11] for seizure onset detection.

3.8.1 Coupling Constant Variation

On patient chb01 (7 seizures), optimization over α revealed that higher values improved performance:

α	AUC	Note
0.10	0.541	Engineered default
0.20	0.565	Initial EEG
0.55	0.581	Optimized

Table 9: EEG α optimization. Higher α improves neural signal discrimination.

3.8.2 Multi-Patient Results

Across four patients (chb01–chb04), AUC ranged from 0.31 to 0.61 with $\alpha = 0.55$. The direction of the Φ signal varied by patient, consistent with known seizure heterogeneity.

3.8.3 Event-Level Detection

Using a K-of-N rule (3 flags in 10 windows), patient chb01 achieved 67% seizure sensitivity with 0.74 false alarms per hour.

3.8.4 Interpretation

The finding that $\alpha \approx 0.55$ outperforms $\alpha = 0.1$ suggests the coupling constant may vary by system type: lower for stability-focused systems (bearings, hearts), higher for information-processing systems (brains).

3.9 Summary

For Table 10, "Correct" is defined per domain in this analysis: run-to-failure datasets are correct if $\Phi^{\mathrm{EOL}} < 0.25$ (bearings, turbofans); incident/precursor cases are correct if the pre-event $\Phi_{\min} < 0.25$ and stable baselines satisfy $\Phi_{\min} > 0.25$ (grids, seismic); neural-network systems are correct if the unstable training case satisfies $\Phi_{\min} < 0.25$ while the stable training case satisfies $\Phi_{\min} > 0.25$ as defined here; and quantum backends are correct if the subset separation holds (HIGH-Φ subsets exhibit lower circuit error than LOW-Φ subsets on the same backend snapshot).

Table 10 summarizes validation across the six engineered domains evaluated under the fixed critical region near $\Phi_c \approx 0.25$. Biological cardiac results are reported separately (Section 3.7) because physiological signals exhibit a lower baseline Φ under this operationalization.

Domain	Systems	Correct	Accuracy
Mechanical (Bearings)	10	10	100%
Aerospace (Turbofans)	10	10	100%
Electrical (Grids)	2	2	100%
Geophysical (Seismic)	4	4	100%
Computational (AI)	2	2	100%
Quantum (Qubits)	3	3	100%
Total	**31**	**31**	**100%**

Table 10: Validation summary (engineered domains). Under the fixed protocol with $\Phi_c \approx 0.25$, the stability index correctly classified all 31 engineered-domain systems across six domains. Quantum validation counts three backends; individual qubit-level analysis covered 445 qubits. Biological cardiac evaluations are reported separately (Section 3.7) because physiological signals exhibit a lower baseline Φ under this operationalization.

4 Discussion

4.1 What We Observe

The stability index $\Phi = I \times \rho - \alpha \times S$ with threshold $\Phi_c \approx 0.25$ correctly classified 31 out of 31 systems across six domains. The same coupling constant ($\alpha = 0.1$) worked for mechanical bearings, electrical grids, aerospace turbofans, earthquake precursors, neural networks, and superconducting qubits.

We did not tune α or the threshold Φ_c per domain. Within each domain, the operational definitions of I, ρ, and S are fixed a priori and applied uniformly to all systems evaluated in that domain.

4.2 What We Do Not Claim

We make no claim that this is a fundamental law, that we understand why 0.25 appears, or that the index generalizes universally.

The 31/31 accuracy is unusually clean for physical systems and invites skepticism. We acknowledge this. Independent replication across additional domains and larger sample sizes would strengthen confidence in the result.

To support independent scrutiny, we make code available (see Data and Code Availability) that allows direct parameter sweeps over α and Φ_c (and windowing choices) so that readers can test sensitivity and assess whether performance is robust to reasonable perturbations of these values.

4.3 Sensitivity to α and Φ_c

To directly assess parameter sensitivity, we report the separation result under a small grid of (α, Φ_c) values using the same fixed protocol and evaluation intervals. Table 11 summarizes the outcome; full sweeps and scripts are available on request (see Data and Code Availability). Including stable controls constrains α from both directions: at $\alpha = 0.05$ several failed cases rise above Φ_c, while at $\alpha = 0.15$ one stable control falls below Φ_c.

α	Φ_c	Separation?
0.05	0.20	No
0.05	0.25	No
0.05	0.30	No
0.10	0.20	No
0.10	0.25	Yes
0.10	0.30	Yes
0.15	0.20	No
0.15	0.25	No
0.15	0.30	No

Table 11: Sensitivity grid for α and Φ_c over 20 failed/EOL cases (10 bearings, 10 turbofans) and 2 stable controls (Germany grid, 2010 seismic baseline). "Yes" indicates full separation: all failed cases satisfy $\Phi < \Phi_c$ AND all stable controls satisfy $\Phi > \Phi_c$. In this tested grid and under this fixed evaluation protocol, $\alpha = 0.10$ is the only value achieving full separation, indicating the entropy coupling is constrained by the requirement to separate failures from stable controls; $\alpha = 0.05$ under-weights entropy (failed cases exceed threshold) while $\alpha = 0.15$ over-weights entropy (stable cases fall below threshold).

4.4 Biological Extension

The cardiac results suggest that while the Φ form can remain unchanged, the operating threshold shifts. Using $\Phi_c = 0.25$ preserves 99.6% recall but induces false positives; $\Phi_c \approx 0.0$ improves balance. Healthy hearts exhibit higher baseline variability than engineered systems, shifting Φ downward.

The neural results introduce a further consideration: α itself may vary by system type. On EEG data, $\alpha \approx 0.55$ outperformed $\alpha = 0.1$. This suggests a distinction between stability-focused systems ($\alpha \approx 0.1$) and information-processing systems ($\alpha \approx 0.5$). This remains speculative.

4.5 Scope: Transitions vs. Stable States

The weak AFDB result (AUC ≈ 0.56) suggests the index detects transitions rather than stable-state differences. The EEG results partially support this: preictal states showed above-chance separation, though weaker than cardiac and requiring patient-specific calibration. Further evaluation on sepsis onset and similar transitions is needed.

4.6 Possible Explanations

We do not yet have a mechanistic explanation for why a common threshold appears across these substrates. One possibility is that the index captures a generic signature of systems approaching critical transitions: declining identity, reduced temporal coherence, and increased disorder co-occur in a way that is not strongly substrate-dependent. This would align with ideas from statistical physics about universal behavior near phase transitions, though we have not established this connection rigorously.

We offer this as a hypothesis for future investigation, not as established theory.

4.7 Limitations

The sample sizes per domain are modest. Full analysis was performed on 31 systems (10 bearings, 10 turbofans, 2 grids, 4 seismic events including one stable baseline, 2 AI systems, 3 quantum backends covering 445 qubits). Stable controls are fewer than failure/critical cases in this study (e.g., 2 stable controls in the sensitivity grid versus 20 failed/EOL cases). A key next step is expanding stable-control coverage within the same sources (e.g., additional non-event grid windows, additional quiet geophysical windows, and additional converged training runs), which will more directly bound false-positive rates under the fixed threshold. While 31/31 accuracy is suggestive, larger validation studies are needed.

The coupling constant $\alpha = 0.1$ was determined empirically from bearing data. A principled derivation would strengthen the framework.

The entropy term S is computed as Shannon entropy (in bits) in all domains, with discretization binning fixed per domain as specified in code. Differences in bin counts across domains may affect the effective scale of S; robustness to binning choices is a direction for future analysis, but the underlying distributions differ (vibration amplitudes, frequency deviations, confusion-matrix cell masses). Whether these sources of disorder are fully comparable under a shared abstraction requires further investigation.

4.8 Relationship to Prior Work

The structure of $\Phi = I \times \rho - \alpha \times S$ resembles the Helmholtz free energy $F = U - TS$ from thermodynamics [10]. The first term represents ordered structure (internal energy or coherent identity), and the second represents disorder scaled by an intensive parameter (temperature or observer coupling). We note this analogy without claiming rigorous equivalence.

In quantum computing, Nation and Treinish [9] developed heuristic cost functions for qubit selection that recover approximately 40% of missing fidelity. In our experiments on the evaluated IBM backends, the index yields substantial error reduction when comparing HIGH-$\Phi_{\min}$ versus LOW-$\Phi_{\min}$ subsets. Nation and Treinish [9] report heuristic layout scoring results under their protocol; our findings highlight that including coherence information (e.g., T_2/T_1-derived quantities) can be highly discriminative under the protocol used here.

4.9 Future Directions

Immediate priorities include independent replication, extension to additional domains (including additional biological transition-detection tasks, financial networks, and social dynamics), and investigation of the threshold's theoretical basis and domain-dependent thresholds.

Longer-term questions include whether the threshold varies systematically with system properties, whether the index can provide early warning with sufficient lead time for intervention, and whether the coupling constant α can be derived from first principles.

5 Conclusion

We report an empirical finding: a stability index $\Phi = I \times \rho - \alpha \times S$ with threshold 0.25 correctly separates stable from failing regimes across mechanical, electrical, aerospace, geophysical, computational, and quantum domains. The same parameters work across all domains tested, including retrospective detection of pre-event instability in the UK power grid blackout of August 9, 2019.

We do not claim to have discovered a universal law. We document an observation that held across 31 independent systems and invite others to test, replicate, and explain it.

The methodology is fully described above; code and artifacts are available on request (see Data and Code Availability).

We additionally evaluated biological extensions. Cardiac arrhythmia discrimination achieved AUC $\approx$ 0.90 at threshold $\Phi_c \approx 0.0$, performed within 0.07 AUC of domain-specific HRV metrics, and maintained discrimination across 5–60 second windows. AFib discrimination was weak (AUC $\approx$ 0.56), suggesting a scope boundary for transitions versus stable states. Neural/EEG evaluation showed above-chance preictal detection with improved performance at $\alpha \approx 0.55$, suggesting the coupling constant may vary by system type.

Data and Code Availability

Data and code are available for replication upon reasonable request (see licensing note). Contact: ShawnBarnicle.ai@gmail.com.

Acknowledgments

The author thanks the open-source scientific community for providing datasets essential to this research: the XJTU-SY bearing degradation dataset from Xi'an Jiaotong University, the NASA C-MAPSS turbofan engine dataset, USGS strainmeter data, the UK National Grid frequency archive, and standard machine learning benchmarks including MNIST and Fashion-MNIST.

References

[1] Wang, B., Lei, Y., Li, Naipeng, and Li, Ningbo (2020). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 69(1), 401–412.

[2] Saxena, A., Goebel, K., Simon, D., and Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *IEEE International Conference on Prognostics and Health Management.*

[3] National Grid ESO. (2019). *Technical Report on the events of 9 August 2019.* 6 September 2019. Available: https://www.ofgem.gov.uk/system/files/docs/2019/09/eso_technical_report_-_final.pdf

[4] Jumar, R., Maaß, H., Schäfer, B., Rydin Gorjão, L., and Hagenmeyer, V. (2020).

Database of Power Grid Frequency Measurements. *arXiv:2006.01771*. Data archives (including Germany/TransnetBW recordings) available via: `https://power-grid-frequency.org/database/`

[5] U.S. Geological Survey. (2016). Donna Lea Strainmeter Data. USGS Earthquake Hazards Program.

[6] U.S. Geological Survey. ANSS Comprehensive Earthquake Catalog (ComCat). `https://earthquake.usgs.gov/data/comcat/`

[7] LeCun, Y., Cortes, C., and Burges, C.J. (2010). MNIST handwritten digit database. *AT&T Labs*.

[8] Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv:1708.07747*.

[9] Nation, P.D. and Treinish, M. (2023). Suppressing quantum circuit errors due to system variability. *PRX Quantum*, 4, 010327.

[10] Callen, H.B. (1985). *Thermodynamics and an Introduction to Thermostatistics* (2nd ed.). John Wiley & Sons.

[11] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., and Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* 101(23), e215–e220.

[12] Moody, G.B. and Mark, R.G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine* 20(3), 45–50.

[13] Moody, G.B. and Mark, R.G. (1983). A new method for detecting atrial fibrillation using R-R intervals. *Computers in Cardiology* 10, 227–230.

[14] Shoeb, A.H. (2009). Application of machine learning to epileptic seizure onset detection. *PhD Thesis, MIT*.