

Supplementary Material

This supplement documents the frozen study configuration underlying the main paper. Its role is documentary and reproducibility-oriented: it records exact parameter grids, numerical settings, detector settings, oracle thresholds, robustness case definitions, case-level validation summaries, and archived integrity artifacts. It does not modify the detector, alter the structural score after inspection, or enlarge the central claim.

S1. Exact sweep grids and numerical settings

The frozen paper suite used explicit parameter lists rather than generated uniform grids, so every reported structural boundary is tied to a sampled parameter value and every reported qualitative boundary is tied to a fixed interpolation rule on a frozen support. Table S1 records the exact sweep grids, fixed parameters, and simulation settings for all four sweeps.

For FitzHugh–Nagumo, the main sweep used

$$\epsilon \in \{0.02, 0.0268, 0.0359, 0.0481, 0.0645, 0.0864, 0.1, 0.1158, 0.13, 0.145, 0.1551, 0.2078, 0.2785, 0.3732, 0.44, 0.47, 0.5\}.$$

with fixed parameters $a = 0.7$, $b = 0.8$, $I = 0.5$. The nuisance sweep fixed $\epsilon = 0.08$ and used

$$I \in \{0.495, 0.49625, 0.4975, 0.49875, 0.5, 0.50125, 0.5025, 0.50375, 0.505\}.$$

Both sweeps used Radau with output spacing $\Delta t = 0.03$, relative tolerance 10^{-7} , absolute tolerance 10^{-9} , and time reset after transient removal. The main sweep used transient length 180 and post-transient window length 260; the nuisance sweep used transient length 120 and post-transient window length 220.

For the autonomous forced van der Pol benchmark, the main sweep used

$$A \in \{0.2, 0.35, 0.5, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0, 1.025, 1.05, 1.075, 1.1, 1.15, 1.2, 1.25, 1.4, 1.6\},$$

with fixed parameters $\mu = 8.0$ and $\omega = 0.9$. The nuisance sweep fixed $A = 0.8$ and $\mu = 8.0$ and used

$$\omega \in \{0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94\}.$$

Both sweeps used RK45 with output spacing $\Delta t = 0.05$, maximum internal step 0.05, relative tolerance 10^{-6} , absolute tolerance 10^{-8} , and time reset after transient removal. Both used transient length 80 and post-transient window length 160.

Structural and oracle sweeps shared the same parameter grid within each benchmark. This matched-grid design is important for interpretation because it ensures that structural and qualitative boundaries are compared on the same ordered support.

Benchmark	FitzHugh-Nagumo	FitzHugh-Nagumo	Autonomous forced van der Pol	Autonomous forced van der Pol
Sweep role	main	nuisance	main	nuisance

Control parameter	epsilon	I	A	omega
Parameter values	0.02, 0.0268, 0.0359, 0.0481, 0.0645, 0.0864, 0.1, 0.1158, 0.13, 0.145, 0.1551, 0.2078, 0.2785, 0.3732, 0.44, 0.47, 0.5	0.495, 0.49625, 0.4975, 0.49875, 0.5, 0.50125, 0.5025, 0.50375, 0.505	0.2, 0.35, 0.5, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0, 1.025, 1.05, 1.075, 1.1, 1.15, 1.2, 1.25, 1.4, 1.6	0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94
Fixed parameters	a=0.7, b=0.8, I=0.5	epsilon=0.08, a=0.7, b=0.8	mu=8.0, omega=0.9	A=0.8, mu=8.0
Solver	Radau	Radau	RK45	RK45
dt	0.03	0.03	0.05	0.05
max_step			0.05	0.05
transient	180	120	80	80
post_transient_window	260	220	160	160
rtol	1.00E-07	1.00E-07	1.00E-06	1.00E-06
atol	1.00E-09	1.00E-09	1.00E-08	1.00E-08
Observation to detector	(v,w) full state	(v,w) full state	(x1,x2) full state; phi excluded	(x1,x2) full state; phi excluded
Oracle / qualitative rule	Tail amplitude threshold 0.2 on v	Tail amplitude threshold 0.2 on v	Stroboscopic cluster-count threshold 4	Stroboscopic cluster-count threshold 4

TABLE S1: Exact sweep grids and numerical settings

S2. Detector, boundary, and oracle settings

The structural sweep saved two trajectory-only indicator families: an occupancy/secant stack and a delay-space divergence diagnostic. The full saved atlas therefore contains

$\{time_pr, arc_pr, occupancy_gap, speed_cv, speed_log_spread, delay_div_rate, delay_div_quality\}$,

together with the delay-divergence regression correlation. The reported structural boundary, however, is driven only by the contract-fixed subset

$\{time_pr, occupancy_gap\}$,

with equal weights and fixed orientation

$time_pr \mapsto increasing, \quad occupancy_gap \mapsto decreasing.$

For each boundary-driving metric, the frozen pipeline applied oriented robust standardization based on the median and median absolute deviation, with standard-deviation fallback when the median absolute deviation degenerated. The standardized curves were then smoothed by a centered moving average of width 3, and the structural score was

defined as the equal-weight average of the two smoothed component curves. Table S2 records the exact score-construction settings.

Structural boundaries were extracted by the two-segment change-point rule described in the main text. The reported main-run structural boundary is the last sampled parameter in the left segment. The main-run algorithmic sensitivity interval was obtained by rescanning the detector over smoothing windows $\{1, 3, 5\}$ and admissible minimum segment sizes drawn from $\{1, 2, 3, \lfloor N/6 \rfloor\}$, filtered to preserve two admissible segments.

The oracle settings were also frozen. For FitzHugh–Nagumo, the primary oracle used the final quarter of the post-transient v -signal and declared oscillation whenever the tail amplitude was at least 0.2. The primary qualitative boundary was the first threshold crossing of that amplitude through 0.2, linearly interpolated between adjacent sweep samples. An auxiliary Hopf-style reference was also recorded for the FitzHugh–Nagumo benchmark but did not enter any structural computation.

For the autonomous forced van der Pol benchmark, the oracle used a stroboscopic section over the final 70% of the post-transient trajectory, sampling one (x_1, x_2) point per 2π -wrap of the lifted phase. The frozen oracle computed stroboscopic cluster count, normalized spread, and tail peak-period coefficient of variation. The saved oracle label `complex_response` was declared whenever the cluster count exceeded 4, the period coefficient of variation exceeded 0.03, or the normalized spread exceeded 0.75. The primary qualitative boundary was the first threshold crossing of the cluster-count curve through 4. The cluster-connectivity tolerance was 0.05 times the tail phase-plane scale, clipped below by 10^{-3} .

Finally, the structural contract explicitly forbade oracle tail amplitude, Lyapunov-QR quantities, and equilibrium-stability quantities from entering the core detector. Those quantities could be saved only in the oracle or auxiliary layer.

Section	Setting	Value
Structural atlas	Saved trajectory-only metrics	time_pr, arc_pr, occupancy_gap, speed_cv, speed_log_spread, delay_div_rate, delay_div_quality, delay-divergence correlation
Structural score	Boundary-driving subset	time_pr, occupancy_gap
Structural score	Weights	time_pr=1.0, occupancy_gap=1.0
Structural score	Orientation	time_pr increasing; occupancy_gap decreasing
Structural score	Standardization	Median/MAD robust z-score; SD fallback if MAD degenerates
Structural score	Smoothing	Centered moving average, width=3
Boundary extraction	Detector	Single change point via two-segment linear regression
Boundary extraction	Main-run sensitivity scan	Smoothing windows $\{1,3,5\}$; minimum segment sizes $\{1,2,3,\text{floor}(N/6)\}$
FHN oracle	Primary rule	Tail amplitude of v on final quarter of

		post-transient trajectory
FHN oracle	Threshold	Amplitude ≥ 0.2 defines oscillation; qualitative boundary = first crossing through 0.2
FHN oracle	Auxiliary reference	Trace-zero Hopf-style epsilon_H reference (saved separately; not used in core detector)
VdP oracle	Primary rule	Stroboscopic cluster-count boundary on lifted phase section
VdP oracle	Threshold	First crossing of cluster count through 4
VdP oracle	Additional label rules	complex_response if cluster count > 4 or period CV > 0.03 or normalized spread > 0.75
VdP oracle	Cluster tolerance	0.05 x tail phase-plane scale, clipped below by 1e-3
Contract	Forbidden core quantities	oracle_tail_amplitude, lyapunov_qr, equilibrium_stability

TABLE S2: Detector, boundary, and oracle settings

S3. Extended robustness case definitions

Robustness summaries in this study were computed relative to a replicate-based robustness baseline rather than relative to the single deterministic main-run structural boundary. For case c , the reported robustness boundary shift is therefore

$$\delta\lambda_c = \lambda_c - \lambda_{baseline,rob}$$

where $\lambda_{baseline,rob}$ denotes the boundary extracted from the robustness baseline case for that benchmark. This distinction should be kept explicit when comparing the robustness tables with the main-text single-run boundaries.

Table S3 lists the full robustness design. In both benchmarks, the core tier contained `baseline`, `full_state_noisy`, `coarse_sampling`, `shorter_window`, and `solver_crosscheck`; the supplemental tier contained `delay_scalar_clean` and `delay_scalar_noisy`; and the stress tier contained `full_state_noisy_stress`. The case definitions were benchmark-specific only in numerical detail.

For FitzHugh–Nagumo, the full-state noisy core case used relative observation noise $\sigma = 0.005$, the coarse-sampling case used $\Delta t = 0.06$, the shorter-window case used transient 126 and post-transient window 182, and the solver cross-check used BDF. The scalar-delay cases used embedding dimension 3 with delay 3; the noisy scalar-delay case used relative observation noise $\sigma = 2 * 10^{-4}$. Replicate counts were 3 for the baseline, 5 for full-state noisy, 3 for coarse sampling, 3 for shorter window, 3 for solver cross-check, 4 for clean scalar delay, 14 for noisy scalar delay, and 4 for the stress case.

For the autonomous forced van der Pol benchmark, the full-state noisy core case used relative observation noise $\sigma = 0.001$, the coarse-sampling case used $\Delta t = 0.055$, the shorter-window case used transient 80 and

post-transient window 148, and the solver cross-check used Radau. The scalar-delay cases used embedding dimension 3 with delay 24; the noisy scalar-delay case used relative observation noise $\sigma = 3 * 10^{-4}$. Replicate counts were 3 for the baseline, 6 for full-state noisy, 4 for coarse sampling, 4 for shorter window, 4 for solver cross-check, 4 for clean scalar delay, 12 for noisy scalar delay, and 6 for the stress case.

Acceptance thresholds followed the main-text definitions: core thresholds of 0.08, 0.75, and 0.25 for boundary shift, aligned Spearman correlation, and interval width; supplemental thresholds of 0.12, 0.60, and 0.25; and stress thresholds of 0.20, 0.35, and 0.45.

Benchmark	Case	Tier	Description	Representation	Replicates	dt	Transient	Post-transient window	Method	Noise sigma	Relative noise	Embedding dim	Delay	Notes
FitzHugh-Nagumo	baseline	core	full-state baseline with replicate ICs	full_state	3					0.00E+00	FALSE	3	10	
FitzHugh-Nagumo	full_state_noisy	core	full-state trajectory with relative observation noise	full_state	5					5.00E-03	TRUE	3	10	
FitzHugh-Nagumo	coarse_sampling	core	coarser temporal sampling	full_state	3	0.06				0.00E+00	FALSE	3	10	
FitzHugh-Nagumo	shorter_window	core	shorter post-transient window	full_state	3		126	182		0.00E+00	FALSE	3	10	
FitzHugh-Nagumo	solver_crosscheck	core	solver cross-check using BDF	full_state	3				BDF	0	FALSE	3	10	BDF
FitzHugh-Nagumo	delay_scalar_clean	supplemental	scalar observation with delay embedding	delay_scalar	4					0	FALSE	3	3	

FitzHugh-Nagumo	delay_scalar_noise	supplemental	scalar delay embedding with relative noise	delay_scalar	14					0.0002	TRUE	3	3	
FitzHugh-Nagumo	full_state_noisy_stress	stress	full-state relative-noise stress test	full_state	4					0.005	TRUE	3	10	
Autonomous forced van der Pol	baseline	core	full-state baseline with replicate ICs	full_state	3					0	FALSE	3	10	
Autonomous forced van der Pol	full_state_noisy	core	full-state trajectory with relative observation noise	full_state	6					0.001	TRUE	3	10	
Autonomous forced van der Pol	coarse_sampling	core	coarser temporal sampling	full_state	4	0.055				0	FALSE	3	10	
Autonomous forced van der Pol	shorter_window	core	shorter post-transient window	full_state	4		80	148		0	FALSE	3	10	
Autonomous forced van der Pol	solver_crosscheck	core	solver cross-check using Radau	full_state	4				Radau	0	FALSE	3	10	Radau
Autonomous forced van der Pol	delay_scalar_clean	supplemental	scalar observation with delay embedding	delay_scalar	4					0	FALSE	3	24	

Autonomous forced van der Pol	delay_scalar_noise	supplemental	scalar delay embedding with relative noise	delay_scalar	12					0.0003	TRUE	3	24
Autonomous forced van der Pol	full_state_noisy_stress	stress	full-state relative-noise stress test	full_state	6					0.002	TRUE	3	10

TABLE S3: Extended robustness case definitions

S4. Extended robustness results

Table S4 reports case-level robustness outcomes, including the extracted boundary, boundary interval, interval width, aligned rank correlations for the boundary-driving metrics and structural score, and any sign alignment used during comparison.

In FitzHugh–Nagumo, the robustness baseline produced boundary 0.1551 with interval [0.1158,0.1551]. The noisy full-state and stress cases both produced boundary 0.13, while the coarse-sampling, shorter-window, and solver-crosscheck cases each produced 0.1551. The clean scalar-delay case produced 0.1158, and the noisy scalar-delay case produced 0.1551. The minimum aligned Spearman correlation of the structural score across core cases was 0.9559. In the supplemental scalar-delay cases, the *time_prchannel* required the permitted sign reversal relative to the full-state baseline, whereas *occupancy_gap* remained aligned without reversal.

In the autonomous forced van der Pol benchmark, the robustness baseline produced boundary 1.05 with interval [1.025,1.05]. The noisy full-state case produced 1.025, the coarse-sampling case 1.05, the shorter-window case 1.0, the solver-crosscheck case 1.05, and both supplemental scalar-delay cases 1.025. The stress case produced 1.075 with interval [0.9,1.1]. In this benchmark, the aligned Spearman correlation of the structural score was 1.0 in every core and supplemental case, and all component alignments remained positive.

The supplement therefore makes explicit a point only summarized in the main text: the robustness module is conservative enough to permit nontrivial movement of the replicate-based boundary object while still preserving the ordering that matters for the paper’s claim.

Benchmark	Case	Tier	Replicas	Boundary	CI low	CI high	CI width	Boundary shift	Aligned Spearman (score)	Aligned Spearman (time_pr)	Aligned Spearman (occupancy_gap)	Alignment sign (score)	Alignment sign (time_pr)	Alignment sign (occupancy_gap)
FitzHugh-Nagumo	baseline	core	3	0.1551	0.1158	0.1551	0.0393	0	1	1.00E+00	1.00E+00	1	1	1

FitzHug h-Nagu mo	full_stat e_noisy	core	5	0.13	0.13	0.13	0	-0.0251	0.99509 80392	9.93E-0 1	9.88E-0 1	1	1	1
FitzHug h-Nagu mo	coarse_s ampling	core	3	0.1551	0.1551	0.1551	0	0	0.95588 23529	9.85E-0 1	9.63E-0 1	1	1	1
FitzHug h-Nagu mo	shorter_ window	core	3	0.1551	0.1551	0.1551	0	0	0.99754 90196	9.98E-0 1	1.00E+0 0	1	1	1
FitzHug h-Nagu mo	solver_c rosschec k	core	3	0.1551	0.1551	0.1551	0	0	0.99754 90196	0.99264 70588	1	1	1	1
FitzHug h-Nagu mo	delay_sc alar_cle an	supplem ental	4	0.1158	0.1158	0.1158	0	-0.0393	0.49754 90196	0.94117 64706	0.99019 60784	1	-1	1
FitzHug h-Nagu mo	delay_sc alar_noi sy	supplem ental	14	0.1551	0.1551	0.1551	0	0	0.82107 84314	0.70098 03922	0.99019 60784	1	-1	1
FitzHug h-Nagu mo	full_stat e_noisy _stress	stress	4	0.13	0.13	0.13	0	-0.0251	0.99509 80392	0.99264 70588	0.98774 5098	1	1	1
Autono mous forced van der Pol	baseline	core	3	1.05	1.025	1.05	0.025	0	1	1	1	1	1	1
Autono mous forced van der Pol	full_stat e_noisy	core	6.00E+0 0	1.03E+0 0	1	1.05	0.05	-0.025	1	0.99793 60165	0.99793 60165	1	1	1
Autono mous forced van der Pol	coarse_s ampling	core	4.00E+0 0	1.05E+0 0	1.025	1.05	0.025	0	1	1	1	1	1	1
Autono mous forced van der Pol	shorter_ window	core	4	1	1	1	0	-0.05	1	1	1	1	1	1
Autono mous forced van der Pol	solver_c rosschec k	core	4	1.05	1	1.05	0.05	0	1	1	1	1	1	1

Autonomous forced van der Pol	delay_sclar_clean	supplemental	4	1.025	1.025	1.025	0	-0.025	1	0.9979360165	1	1	1	1
Autonomous forced van der Pol	delay_sclar_noisy	supplemental	12	1.025	1	1.025	0.025	-0.025	1	1	1	1	1	1
Autonomous forced van der Pol	full_state_noisy_stress	stress	6	1.075	0.9	1.1	0.2	0.025	1	0.9938080495	0.991744066	1	1	1

TABLE S4: Extended robustness results

S5. Specificity and ablation detail

Table S5 records the nuisance/main raw-span ratios and the ablation-variant summaries.

For FitzHugh–Nagumo, the nuisance/main span ratios were 0.00453 for `time_pr` and 0.00688 for `occupancy_gap`, with oracle label constancy throughout the nuisance sweep. The ablation suite returned: `atlas_full`, boundary 0.1158 with interval [0.1158,0.145] and lead distance 0.38139; `single_occupancy_gap`, boundary 0.1 with interval [0.1,0.13] and lead distance 0.39719; and `single_time_pr`, boundary 0.1551 with interval [0.1551,0.1551] and lead distance 0.34209.

For the autonomous forced van der Pol benchmark, the nuisance/main span ratios were 0.07049 for `time_pr` and 0.08713 for `occupancy_gap`, again with oracle label constancy throughout the nuisance sweep. The ablation suite returned the same structural boundary for all three variants: 1.025, with lead distance 0.225.

These expanded results are useful mainly as criticism-hardening detail. They do not change the interpretation given in the main text.

Benchmark	Autonomous forced van der Pol	FitzHugh-Nagumo
Main control parameter	A	epsilon
Nuisance control parameter	omega	I
Oracle label unique count (nuisance)	1	1
Nuisance qualitative boundary		
time_pr ratio	0.070488	0.004527
occupancy_gap ratio	0.087126	0.006879
Max ratio	0.087126	0.006879
Tolerance	0.35	0.35
Passed	TRUE	TRUE

Benchmark	Variant	Structural boundary	CI low	CI high	Lead distance	Metric names
FitzHugh-Nagumo	atlas_full	0.1158	0.1158	0.145	0.3813902598	time_pr, occupancy_gap
FitzHugh-Nagumo	single_occupancy_gap	0.1	0.1	0.13	0.3971902598	occupancy_gap
FitzHugh-Nagumo	single_time_pr	0.1551	0.1551	0.1551	0.3420902598	time_pr
Autonomous forced van der Pol	atlas_full	1.025	1.025	1.025	0.225	time_pr, occupancy_gap
Autonomous forced van der Pol	single_occupancy_gap	1.025	1.025	1.025	0.225	occupancy_gap
Autonomous forced van der Pol	single_time_pr	1.025	1.025	1.025	0.225	time_pr

TABLE S5: Specificity and ablation detail

S6. Archived artifacts, integrity checks, and reproduction

Each saved sweep bundle wrote `results.json`, `config.json`, `provenance.json`, and `manifest.json`, together with selected trajectory arrays where requested. Main benchmark runs additionally wrote `summary.json`, `claim_compliance.json`, a `robustness/robustness_summary.json` bundle, and figure files for the structural score, boundary overlay, boundary-driving metrics, and robustness summary. Nuisance runs wrote their own summary and figure bundles.

At study level, the frozen pipeline wrote `study_index.json`, `study_summary.csv`, `study_manifest.json`, `specificity_report.json`, `ablation_report.json`, `integrity_report.json`, and `contract_report.json`. Manifest files recorded relative paths, byte counts, and SHA-256 hashes. Integrity validation checked both per-run manifests and the study-level manifest. Contract validation checked the presence of the required benchmark families and sweep roles, trajectory-only separation of the core detector, presence of qualitative boundaries in the main runs, positive lead distance in the main runs, passed robustness, passed specificity, passed ablation, and passed integrity.

A clean reproduction therefore consists of four stages: environment creation and editable installation, execution of the unit-test suite, execution of the frozen paper study, and validation of the saved manifests and study-level reports. The purpose of documenting this sequence is not to treat reproducibility as a substitute for scientific content, but to make the frozen evidence bundle auditable.

Artifact	Present
ablation_report.json	TRUE
contract_report.json	TRUE

TABLE S6: Archived artifacts, integrity checks, and reproduction