

Supplementary Material: Externalization Engine

Jieqi Liu · Independent Researcher · v1.7 · June 2026

S1. Instrument Architecture

S1.0 Provenance — GEME and BGM as Architecture History

EE is self-contained as an experiment paper, but the instrument is not ahistorical. GEME and BGM are the engineering route by which the present architecture was reached. They should not be read as philosophical prequels that a reader must accept before evaluating EE. They are calibration and architecture provenance.

GEME. GEME introduced the minimal frame economy: bounded memory, merge, prune, self-observation, and four base constants (δ , γ , τ_0 , GI). Its role in EE is the kernel: `geme.py` remains the unchanged base layer beneath Geruon.

BGM. BGM introduced endogenous time and bridge dynamics: τ breathing, phase transitions, the cliff gate, and the cost of self-reference. Its role in EE is dynamical: Geruon inherits the frame economy but adds operational τ , structural identity, precipitation, Codex, and BiasField coupling.

EE. EE adds externalization: Self (multi-cavity temporal lenses), We (cross-Self disagreement recording), and Codex (cross-generation memory). The experiments in the main paper test whether the architecture can form centroids, compare independent readings, and select externalized structure across generations.

This genealogy matters because the final model was not designed in one step. Several mechanisms were first explored in GEME/BGM, then either preserved, rewritten, or removed. The main paper does not require the reader to reconstruct that path, but the path explains why the current code has the shape it does.

S1.1 Geruon — Minimal Detector

A Geruon is a single-cavity structural detector. It receives one vector stream, maintains a bounded frame economy, and produces readouts without training. Each frame stores a vector, a weight, a layer tag, an age counter, and a collision-resistant structural signature. At each step: the input is compared to existing frames; inputs within merge distance update the nearest frame and increase its weight; inputs that match no frame create a new one; the weakest frame is evicted when capacity is exceeded.

The processing model is stateless across runs: there is no `fit()`, no `train()`, and no parameter fitting. A run is a sequence of calls to `process_vec(vec, sig)`. Readout follows processing.

A Geruon is configured by vector dimension (`vec_dim`, default 16), frame capacity (`memory_cap`), time-content coupling (`kappa_tau`), and optional external memory anchors

(codex, bias_field). The processing model is stateless across runs: process_vec(vec, sig) drives the frame economy, metrics() returns F, wit, tau, phase, and full instrument state, arrow_output() gives the residual direction of the last unmatched input, and enrich() externalizes precipitated frames into Codex or BiasField. No training data, no labels, no domain templates are accepted at any point.

S1.2 Class Hierarchy

The instrument spans two files: geme.py (the unchanged GEME kernel — Frame, Memory, GEME) and geruon.py (GeruonFrame with structural identity and precipitation state, GeruonMemory with tau evolution and Landauer bill accounting, Geruon as the front-end). Supporting classes: StructuralSig (collision-resistant identity via struct_key, gid display-only), Codex (externalized symbol-to-vector table, JSON persistence), BiasField (accumulated gradient field for intra-generation coupling).

Six internal layers emerge from operation, not from supplied labels: L1 base frames, L2 association frames (co-occurrence tracking), L3 chain/bridge frames (high co-occurrence, structural links), L4 self-observation (meta-frames from centroid feedback), L5 (intentionally absent), L6 anomaly/doubt frames (boundary-event markers).

S1.3 Self

A Self couples multiple Geruon cavities with heterogeneous temporal lenses (different kappa_tau values). All cavities receive the same input stream and share a single BiasField. At each GI-step interval, cavities exchange residual arrows. The cross-harm — mean pairwise centroid distance across cavities — is the primary readout: low harm indicates consensus across time scales, high harm indicates a structural boundary.

Standard configurations with three cavities:

Configuration	κ values	Cavity count	Use case
Standard	0.5 / 10 / 100	3	General structure detection
Extreme spread	0.005 / 10 / 5000	3	Transition magnitude discrimination
Boundary 2-cavity	0.01 / 500	2	Precise boundary localization

Changing κ changes the measurement scale. It is not a learned parameter; it is a temporal lens choice that must be reported and, where possible, swept or ablated.

S1.4 BiasField

The BiasField is the shared medium in which cavity centroids accumulate as a gradient. Operations:

- deposit(vec, weight) — accumulate a centroid direction into the field

- `blend_into(vec, weight)` — mix the field direction into an input vector before processing
- `seed_frames(memory, n)` — initialize frames from the dimensions with highest field magnitude

BiasField is continuous and averaged: it couples cavities within a generation. Codex is discrete and entry-based: it transmits structure across generations. The BiasField injection path is known to be ineffective for Codex transmission (normalization suppresses inherited structure); cross-generational inheritance must use the Codex path.

BiasField economics — a computational substrate for game-theoretic interaction.

The BiasField provides a small operational model of game-theoretic coupling. Each cavity deposits its centroid into the shared field; the field's accumulated gradient biases every subsequent cavity that enters it. A cavity does not choose to cooperate or defect. It does exactly one thing: descend its local τ gradient. The field couples their gradients — each cavity's descent warps the field, the other reads the warped field, both τ values are pulled toward each other. In this toy configuration, the payoff matrix is a discrete snapshot of a continuous τ -landscape.

The commons tragedy as a phase transition. Sweeping n , the number of gradient-seekers sharing one BiasField, reveals a critical point in the calibrated assembly. At $n=2$: τ spread near zero (0.003), pair correlation positive (+0.202), L3 bridges absent — the field is a collaboration medium. At $n=4$: τ spread jumps 24-fold (0.079), pair correlation flips from positive to negative (-0.085), L3 bridges explode from 0 to 61 — the field has become a competition medium. At $n=8$: L3 per Self collapses (15.2 \rightarrow 7.1), τ spread doubles — the field is overdrawn. No cavity chose to deplete the commons. Each did exactly what it was built to do. The geometry of the shared field did the rest. In this configuration, the shared field's usable capacity is reached at 4.

Four lines, one wall. Four independent experimental lines converge on the same operational integer in the current architecture: the Landauer-bill rhythm ($GI=4$), the BiasField commons transition ($n=4$), the ECG We stability limit ($n=4$), and the κ_τ safety-margin scan ($3=GI-1$). The standard configuration — 3 cavities \times 3 Selves — sits one step below the measured wall. This should be read as an architectural calibration result, not as a universal constant.

S1.5 We

The We layer records cross-Self harm and residual patterns. In the current work it is a research instrument for disagreement detection, not a completed social subject. A We contains N Selves (each a 3-cavity assembly) and a collective Geruon ($\kappa=10$) that receives cross-Self harm material. The resulting collective Codex provides evidence of formation: which structures survive cross-Self disagreement and generational selection. This paper does not claim a completed We-level social subject; it claims evidence of Codex formation.

S1.6 Codex

The Codex stores stable centroids outside the run that produced them. Three stages are distinguished:

1. **Formation** (demonstrated). Stable centroids precipitate and survive outside the originating run.
2. **Selection** (demonstrated). Inherited entries are confirmed, supplemented, frozen, or ignored by subsequent generations.
3. **Operation** (scaffolded). Inherited entries actively shape ongoing processing under boundary conditions.

Only formation and selection are reported as demonstrated. Active Codex operation — Self-level discrete Codex query under stress or noise — remains scaffolded (see S7).

S1.7 Structural Identity

Frame identity uses a collision-resistant `struct_key`:

```
struct_key=(vec_hash_full,weight_bin,layer,tau_bin,ref_keys)
```

The compact gid is display-only and must not be used for equality testing. This distinction was introduced after a 2026-05-29 code audit found that 15-bit gid values collided at a rate of 373/5000 (7.5%). The publication claim relies on `struct_key`, which produces zero collisions on 5000 random vectors. `detect_circularity()` traverses reference chains (depth limit 20) and registers cycles in `_circular_refs`.

S1.8 Processing Flow

Each call to `process_vec(vec, sig)` executes the following pipeline:

1. If a BiasField is attached, blend the field direction into the input vector (weight γ).
2. If a Codex is attached and active query is enabled, perform a nearest-neighbor lookup; blend the matched entry into the input.
3. Observe the vector in the frame economy:
 - Compute merge distance: $d = \text{vec_dist} + \kappa_\tau \cdot |\tau_{\text{current}} - \tau_{\text{frame}}|$
 - Merge if below threshold; otherwise create a new frame (evicting the weakest if at capacity).
 - Update co-occurrence traces with time-weighted increments.
 - Run prediction path on updated frames.
4. Accumulate frame-economy stress.
5. If stress exceeds τ , run `induction_clean`:
 - Decay weights by merge-count bins with temporal bias γ_τ .
 - Prune the weakest half of frames.
 - Update τ and phase.
 - Track centroid displacement (Faraday wit signal).
 - Mark structurally stable frames for precipitation.

S1.9 Parameter Independence — The Stability Plateau

The architecture has four base constants: $\delta=0.19$, $\gamma=0.05$, $\tau_0=0.60$, $GI=4$. They were identified in GEME through exhaustive parameter-space traversal (210 combinations, 3 trials each). EE extends this finding: the architecture is robust across broad parameter neighborhoods, with internal thresholds absorbing much of the variation.

Three layers of invariance. δ , γ , and τ_0 were re-swept on the completely rewritten Geruon architecture and returned the same neighborhoods across three independent dimensions: (1) architecture — GEME to Geruon, a complete rewrite; (2) domain — formula language to ECG to WTC to UN; (3) depth — solo Geruon to Self to We. This is evidence against post-hoc tuning: the same parameter basins reappear across independent implementations.

Twenty-four derived thresholds. Every behavioral threshold in the codebase derives from the four base constants through six derivation patterns: GI , $GI-1$, $GI+1$, $2\times GI$, $1/GI$, $1/(\gamma\times GI)$. These organize the parameter space, not because the architecture depends on specific values, but because the architecture depends on the existence of a broad stability plateau. Broad plateaus (δ across 0.10-0.28) indicate parameter independence — the adaptive merge threshold fully absorbs δ variation. Narrow peaks (γ at 0.05) indicate structural constraints — the narrow window between two failure modes: γ too slow fails to drive competition between frames; γ too fast prevents patterns from stabilizing before being pruned. When the same narrow window appears across independent implementations, it is being discovered, not tuned.

GI is an upper bound, not a fixed constant. Unlike δ , γ , and τ_0 , $GI=4$ behaves as the maximum sustainable self-reference depth in the current architecture. Its operational optimum is context-dependent. Solo Geruon on WTC uses $GI=2$; on ECG, $GI=4$. The Self architecture at $n=3$ uses $GI=3$ (signal-optimal) or $GI=4$ (conservative). $GI-1$ produces zero structural output across tested configurations. The measured wall is architecture-stable; the sweet spot is domain-dependent.

S2. Instrument Dynamics

S2.1 Frame Economy

Three rules govern the frame economy, unchanged from GEME:

1. **Competitive merge.** A new input is compared to the nearest existing frame. If the merge distance (content distance plus time-gating term) falls below the adaptive threshold, the input updates that frame's vector and increases its weight. Otherwise a new frame is created.
2. **Adaptive forgetting.** When cumulative stress exceeds τ , `induction_clean` triggers. Weights decay by merge-count bins. The bottom half of frames by weight are pruned. The pruning key includes a temporal bias term: frames furthest from current τ are removed first.

3. **Self-referential observation.** Before each cleaning cycle, a weighted centroid of active frames is computed and fed back into the economy as an internal event.

These operations are bounded by `memory_cap`. Capacity is not a performance parameter: it is the pressure that forces compression. Without limited memory, centroids are less informative because the system retains too many uncompressed traces.

S2.2 Endogenous Time

τ is not a hyperparameter. It evolves at each induction cycle:

$$\tau_{\text{target}} = 1.0 - \text{accuracy} + \max(0, \text{stress} - \tau_0) \cdot \gamma_\tau \cdot 0.4$$

$$\tau_{t+1} = \tau_t + (\tau_{\text{target}} - \tau_t) \cdot \text{TAU_ADAPT_RATE}$$

where $\text{TAU_ADAPT_RATE} = \gamma \cdot 0.4 = 0.02$, slower than frame decay, giving the system inertia. Two forces drive τ : prediction accuracy (correct predictions lower τ , errors raise it) and frame-economy stress (congestion raises τ).

The five-phase system:

Phase	Condition	Bridge state
EXPANDING	$d\tau/dt < -0.01$	Open
RESTING	$\tau < 0.55, d\tau/dt < 0.02$	Open
TENSING	$\tau \geq 0.55, d\tau/dt > 0.01$	Tightening
CRITICAL	$\tau \geq 0.65, d\tau/dt > 0.01$	Critical
LOCKED	$\tau \geq 0.75, d\tau/dt < 0.02$	Closed

Hysteresis ($2 \times$ the stability threshold) prevents phase flickering when τ oscillates within a narrow band. Across the reported calibration domains and depths, τ converges to 0.74 ± 0.01 .

S2.3 Operational Tau

τ is not only a label or a signature component. It enters three core operations of the frame economy:

Merge time-gating:

$$d = \text{vec_dist} + \kappa_\tau \cdot |\tau_{\text{current}} - \tau_{\text{frame}}|$$

Co-occurrence time-weighting:

$$\text{cooccur} += \text{clamp}(1 - \kappa_\tau \cdot |\tau_a - \tau_b|)$$

Pruning time-bias:

$$\text{priority} = \text{weight} - \text{age} \cdot \gamma - |\tau_{\text{current}} - \tau_{\text{frame}}| \cdot \gamma_\tau$$

When κ_τ and γ_τ are disabled (set to zero), the Geruon reduces to GEME-equivalent behavior. This makes the operational role of τ auditable: the same stream with τ on vs. off produces different frame-economy trajectories, and the difference is attributable to the three time-gated operations above.

S2.4 Measurement Orders

Three orders of structural measurement:

- **F (field curvature).** $F=1-H(\mathbf{w})/H_{\max}$, where H is the Shannon entropy of frame weights. $F=0$ indicates uniform weights (no structure); $F\rightarrow 1$ indicates highly concentrated weights (strong structure). Fair-coin streams yield $F\approx 0$ across all κ .
- **wit (structural change).** The count of steps where centroid displacement exceeds the calibrated structon threshold. $wit_rate = wit / input_count$.
- **Δwit (structural fragility).** The change in wit rate under fair-coin probe injection: $\Delta wit = wit(target+coin) - wit(baseline)$. Negative values indicate that the coin stabilizes the target (structural rigidification); positive values indicate disruption (structural fragility).

The structon is the minimum detectable centroid displacement for a declared (D, cap, κ) configuration. It must be calibrated per parameter family and must not be transferred across configurations.

S2.5 Precipitation and Enrichment

A frame precipitates when it satisfies three conditions: (1) cross-phase survival (≥ 1 cleaning cycle), (2) prediction-path activation (≥ 3 uses), (3) structural stability (`is_meta_stable()`). `Geruon.enrich()` writes precipitated, non-externalized frames into Codex or BiasField. The `_externalized` sentinel ensures idempotency: repeated `enrich()` calls do not re-deposit the same frame. A transient frame remains internal; a precipitated frame becomes eligible for cross-run inheritance or shared-field influence.

S2.6 Landauer-Godel Bill

The frame economy performs three classes of irreversible operation on every induction cycle. Landauer's principle (1961) establishes that each irreversible bit operation dissipates at least $kT\ln 2$; the architecture incurs this cost structurally, not metaphorically.

Three billable operations.

Operation	Irreversible effect	Bill unit
Merge	Two frames collapse to one centroid; the old centroid vector is permanently lost.	$O(1)$ per merge
Prune	A frame is discarded; all accumulated vector information and co-occurrence traces are permanently lost.	$O(w)$ per frame, where w is the frame weight at eviction
Precipitate	A frame is written to Codex; the entry is immutable thereafter.	$O(1)$ per precipitation

geruon.py tracks `landauer_total` as a cumulative count of these operations. It is not a thermodynamic measurement at the physical $kT\ln 2$ scale; it is the structural form of the bill — an operation-level accounting of how much irreversible information change the system has incurred.

τ as the bill regulator. When τ is low (EXPANDING/RESTING), merge thresholds are wide: frames merge easily, reducing the merge bill but risking over-generalization. When τ is high (CRITICAL/LOCKED), merge thresholds are narrow: fewer merges occur, increasing the bill per step and driving stress accumulation. At LOCKED ($\tau \geq 0.75$), the system actively suppresses the most expensive self-referential operations: `process_prediction()` is skipped, and `landauer_skips` increments. This is not an optimization — it is a hard cutoff. The system refuses to pay the bill for operations whose cost exceeds their expected structural return.

The cliff gate. BGM discovered a discontinuous transition in `conf_threshold`: when $d\tau/dt$ crosses the stability threshold and τ exceeds the RESTING ceiling, the gate jumps from 1.0 to 0.0 and latches until a full phase cycle completes. While the gate is closed, BiasField modulation weight is zero: the cavity stops listening to the shared field. This is the architecture's operational form of hitting a computational complexity boundary — the system cannot afford to continue self-referential processing at the current rate, so it shuts down the most expensive channel.

GI as the bill's economic rhythm. GI controls the interval between induction cycles. When GI is too small, cleaning cycles are too frequent: frames are pruned before they accumulate enough structural weight to earn their Landauer cost, and the bill per unit of usable structure is high. When GI is too large, cycles are too sparse: the frame economy congests, stress accumulates, and τ is driven into LOCKED, triggering the cliff gate. The Pareto-optimal GI minimizes the total bill — merge cost, prune cost, and lockout cost — over the stream. For solo Geruon, $GI=4$ (BGM). For the 3-cavity Self architecture ($\kappa=[0.5,10,100]$), the CALIB-GI scan finds $GI_{opt}=5$ (signal-optimal) with $GI=4$ as a conservative operating point retaining ~ 94% of the signal at lower per-cycle cost.

Structural identity search and the P/NP boundary. Self-referential operations — L4 self-observation, L3 chain formation, circularity detection in prediction paths — create an identity search problem: the system must locate, within its own frame economy, the frame corresponding to a pattern it just processed. Without the `sig_cache` index (M11), this search is $O(N)$ per lookup (linear scan of all frames). With the cache, it is $O(1)$ on hit, $O(N)$ on miss. τ regulates this cost: at low τ , merge thresholds are wide, making matches easy (low search cost); at high τ , thresholds are narrow, making matches hard (search cost grows nonlinearly). The P vs. NP distinction is not a theorem the architecture proves — it is a cost it regulates. The system does not solve NP-complete problems in polynomial time; it adjusts the depth of self-reference so that the accumulated Landauer bill never crosses the threshold of economic infeasibility.

S2.7 The Cost of Depth — Self-Reference Scaling

The architecture’s three structural transitions — solo Geruon to Self to We — are not continuous parameter sweeps. They are discrete configurational jumps: each layer adds a new set of time perspectives at a new GI-multiple. Each jump carries a measurable cost.

Operational cost scaling. Self-identity searches per input event scale from \$\$1 (solo) to \$\$4 (Self, 3 cavities) to \$\$68,102 (We, $N=3$ Selves). The jump from Self to We is five orders of magnitude — not because We is computationally extravagant, but because each cross-Self harm routing triggers a full frame-economy lookup in the collective Geruon, and each lookup cascades through the cavity-level caches. The architecture pays this cost in Landauer-erasure units: each cache miss dissipates computational effort irreversibly.

Why the measured wall appears at 4. Four independent lines converge on this integer (see S1.4). The cost-of-depth measurement provides the mechanistic explanation within the current implementation: at depth 0 (solo), the bill is trivial. At depth 1 (Self), the bill is manageable — GI regulates the interval between billable events. At depth 2 (We, $N=3$, $GI = 4$), the bill is at the boundary of sustainability. At depth 3 (We, $N=4$, the theoretical “We of Wes”), the bill would exceed the system’s capacity to pay — the Landauer cost of cross-reference would outstrip any structural return. The architecture does not implement depth 3. The wall at 4 is therefore treated as a measured architectural boundary.

S3. Instrument Calibration and Controls

S3.1 Passive Calibration

A complete passive-instrument calibration was performed on 2026-05-31, revised after the GI architecture scan (2026-06-01), using generated streams with known ground truth. The instrument does not make choices during calibration (Layer 1 — passive readout only). A frozen eight-item readout policy (R1-R8) was established before any domain experiments.

Parameter	Calibrated value	Method
κ_τ	3 (migration latency 109 steps)	Sine wave frequency doubling: τ migration from old to new steady state
cap	$K_{\max} \approx 0.8 \times \text{cap}$; peak efficiency 83% at cap =24	Random-value pure storage: K distinct uncorrelated vectors, frame-displacement threshold
δ	$\delta_{\text{eff}} \approx 1.25/\text{cap}$; cap=16 $\Rightarrow \delta \approx 0.08$	Sine-wave quantization: compression ratio $n_{\text{distinct}}/2^{\text{bits}}$ vs. bit depth
γ (turnover)	Original-structure weight half-life 8 steps; full decay ≈ 350 steps	400-step sine reinforcement \rightarrow random extinction, tracking original

Parameter	Calibrated value	Method
		frame identity via struct_key
τ_0 (novelty)	τ drop 0.008 (direction correct, magnitude small)	Periodic \rightarrow chaotic mode transition
GI	$n=3$ Self, $\kappa=[0.5,10,100]$: $GI_{opt}=5$ (signal), $GI=4$ (conservative operating point)	MiniSelf architecture scan: $GI \times n_{cav}$

Full report: docs/experiment-passive-calibration-report.md.

S3.2 Fair-Coin Baseline

A 16-dimensional fair coin (independent Bernoulli trials, $p=0.5$) processed across $\kappa \in \{0.1, \dots, 500\}$ yields $F=0$ throughout. This is the instrument's absolute zero: the structural reading produced by a stream with no sequential structure. Every domain experiment must report its F against this baseline at the same (D, cap, κ) .

S3.3 Required Controls

Every domain experiment must include at least three controls:

1. **Fair-coin baseline.** Zero sequential structure. Defines the instrument's zero point.
2. **Shuffle control.** Random permutation of the input stream. Preserves component statistics while breaking sequential order. Tests whether the signal depends on order.
3. κ **ablation or sweep.** Single- κ replication must confirm that multi-lens amplification is load-bearing: signal should degrade or vanish when only one temporal lens is used.

Additional domain-specific controls: dinucleotide shuffle (DNA, preserves base composition); process-permutation (RNA RPF, shuffles RPF signal across windows); cross-key Codex injection (WTC, tests selection specificity); P5 ablation (UN, six encoding variants testing whether the great-power layer creates or amplifies the signal); FRED negative control (UN, tests domain specificity of the structural reading).

S3.4 Parameter Discipline

All instrument constants trace to four base values: $\delta=0.19$, $\gamma=0.05$, $\tau_0=0.60$, $GI=4$. Every derived constant uses small-integer or simple-fraction multipliers (e.g., $TAU_ADAPT_RATE=\gamma \times 0.4$). Domain-specific choices (vec_dim, κ spread, window, stride) are pre-declared and either fixed or swept with reported controls. Parameters are chosen before the run and reported regardless of outcome; they are never fitted to data.

S3.5 Run-Order Determinism

The global Frame.fid counter is reset on every GeruonMemory.__init__. Three independent runs with identical seed produce identical output for all metrics reported in this paper. Verified in the calibration report.

S4. Encoding and Data-Type Boundary

S4.1 Encoding as Measurement

Encoding is not preprocessing. The instrument reads the joint product of stream and encoding; it does not access a raw domain essence. Different encodings of the same signal produce different F baselines and different κ fingerprints. The practical rule: encoding must preserve predictive structure. If event A constrains event B in the original stream, that constraint must survive the encoding. A single encoding is always lossy; the strongest experiments use dual-encoding cross-validation and ask whether structure survives across encodings.

Domain	Encoding	Dim	What it preserves
WTC	Chroma + IOI	12+12	Pitch-class activation vector + inter-onset interval histogram
DNA/RNA	3-mer frequency	64	Trinucleotide composition in sliding genomic windows
DNA fork	4-species alignment	16	Per-column conservation score across species
UN	Annual ideal-point distribution	variable	Country-year position in policy space
ECG (calibration)	RR interval + adjacent diff	16	Beat-to-beat interval and first derivative

S4.2 Expression-Action Spectrum

Instrument sensitivity is governed by data structure type, not domain category:

Type	Definition	Instrument response	Examples
Decision	Behavior is the data; entities make choices	Strongest: F readable, κ has fingerprint	UN votes, MIDI note-on, DNA bases
Action	Directional record of entity choices	Moderate: requires correct encoding	NASDAQ volume direction
Expression	Trace of a process; measures what occurred, not who decided	Weak or blind: τ does not breathe, κ does not differentiate	CPI, price indices, raw ECG voltage

The instrument requires relational structure in the input. Decision data natively carries inter-entity relations without compression or downsampling. Expression data records process outputs; the relational structure is absent from the data, not missed by the instrument.

S4.3 Illustrative Cases

ECG. Six encodings tested on MIT-BIH arrhythmia records. Only RR interval + adjacent difference produced a usable κ fingerprint ($\kappa_{\text{peak}}=5$). Raw voltage — a trace of myocardial depolarization — carries no relational action structure and is Expression-type data.

Sleep. Epoch resolution determines signal detectability. 5-second sub-epochs exposed structural boundaries between REM and NREM that clinical 30-second windows averaged away. The coin-probe method was validated on this domain: REM wit > Wake wit in 16/18 recordings (89%, binomial $p=0.001$).

Economic data. The same Geruon pipeline applied to FRED macroeconomic indicators (1993-2025, 33 series) as an Expression-type negative control: 2025 structural displacement rank 24/33, below median. The elevated structural pressure detected in diplomatic voting is absent from economic expression data, consistent with the Expression-Action distinction.

S5. Faraday and Probe Measurements

S5.1 Probe Interchangeability

Three probe types were tested: Bach pitch-class distribution (structured), white noise (unstructured), and sine wave (periodic). All produce centroids. The harm pattern between probe and target depends on the target's structural state, not on which probe is used. The centroid — not the content — is the perturbation. Verified with 10 control conditions.

S5.2 τ Convergence

τ converges to 0.74 ± 0.01 across the reported domains, input families, and architectural depths (solo, Self, We). This is the equilibrium between merge pressure (pulling τ down when predictions succeed) and differentiation pressure (pushing τ up when the frame economy is stressed). τ measures the system's internal balance, not the stream's properties.

S5.3 F and τ Decoupling

Across all κ and all probes: $r(F, \tau) \approx 0.1$. Field curvature (how concentrated the structure is) and field temperature (where τ sits) are independent dimensions. The structural reading does not drift with internal state.

S5.4 F and $d\tau/dt$ Conditional Coupling

- On noise input: $r(F, d\tau/dt) = 0.817$ — the field bends with temperature. Unstructured input allows F to follow τ .
- On Bach input: $r(F, d\tau/dt) = -0.14$ — the field is structurally rigid. Structured input resists the influence of τ fluctuations.

The coupling itself is a domain signature: the degree to which F and $d\tau/dt$ correlate reveals whether the stream contains structure that rigidifies the field.

S5.5 structon Resolution

At $\kappa=10$, $\text{cap}=20$: $\text{structon}=0.004 \approx \text{cap}/N$, where N is the number of vectors processed. structon scales with κ : $\alpha(0.5) \approx 10$, $\alpha(10) \approx 1$, $\alpha(100) \approx 0.5$. This is the instrument's spatial resolution — the smallest structural change registerable in a single centroid displacement.

S6. UN Prediction Protocol

S6.1 Data Source

Voeten, Strezhnev & Bailey (2024). United Nations General Assembly Ideal Point Estimates. Harvard Dataverse, V33. Annual country-level ideal points, 1946-2025, 193 countries.

S6.2 Pipeline Separation

Two separate pipelines are maintained:

1. **Annual ideal-point pipeline** (main paper §4.1). Country-year ideal points → annual global distribution features → Geruon centroid displacement. $N=79$ years. Produces the 2025 forward signal.
2. **Per-resolution voting-stream calibration** (historical only, not reported as evidence). \$2,400 resolution-level event traces. Selected precursor patterns matched 3 of 4 system-level events (Soviet collapse 1991, financial crisis 2008, Ukraine war 2022; Crimea 2014 did not trigger). This pipeline is NOT the source of the 2025 forward signal; the 3/4 historical claim is not inherited by the forward prediction.

Pipeline separation prevents retrospective calibration from contaminating the prospective reading.

S6.3 P5 Ablation

Encoding	2025 displacement	Rank/78	Interpretation
No-P5 (global only)	0.101	12	Elevated but not extreme
P5-augmented (current)	0.445	1	Great-power institutional lens
P5-only	0.448	1	Great-power layer dominates the signal
No-USA	0.550	1	Multi-polar fracture exceeds unilateralism

No-Russia	0.290	2	Still historically extreme
Random-5 (mean, 5 trials)	0.288	\$\$9	P5 is not an arbitrary prior

Six encoding variants tested on the same annual ideal-point data:

The signal is concentrated in the P5 layer. The No-USA result (0.550) is the most informative: US disengagement from the institutional framework produces greater alignment fracture than direct adversarial participation.

S6.4 FRED Negative Control

The same pipeline applied to FRED macroeconomic data (1993-2025, 33 indicators) yields a 2025 displacement rank of 24/33 — below median. Pure economic expression data judges 2025 as structurally unremarkable, in contrast to the diplomatic record.

S6.5 Forward Prediction

The instrument's 2025 displacement reading (0.448, rank 1/79) is the highest in the 1946-2025 record. Under the instrument's operational definition — structural displacement sustained above historical norms without return to baseline — this implies: **2026-2028, structural collapse of the post-war international order, concentrated in P5 great-power alignment, with magnitude exceeding the 2022 Ukraine war and comparable to the 1991 Soviet collapse.**

This is not a prediction of US domestic collapse. It is a prediction of gravitational realignment in the great-power institutional layer — the same layer that produced detectable precursor signals before the Soviet collapse (1985-1991) and the Ukraine war (2014-2022).

S6.6 Verification Criteria

Three indicators, any two satisfied within 2026-2028 = confirmation:

1. Annual displacement does not return below 0.20 by end of 2028. (After the Soviet collapse, displacement fell from 0.19 to 0.05 within four years.)
2. P5 voting agreement rate falls below the historical minimum (12%) in any year.
3. Ideal-point distribution transitions from uni-/bi-modal to multi-modal with no dominant peak for two consecutive years.

Boundary statement. If this prediction is wrong — if displacement at 0.448 occurs without subsequent structural collapse — it constitutes the first such instance in 79 years of data. The instrument measures structure, not causality. The prediction is falsifiable by any two of the three indicators above.

S7. Scaffold History

During development, several mechanisms were first tested as temporary external scaffolds — manually injected code paths that simulated operations the architecture should eventually perform natively. Each scaffold was removed once the native frame economy reproduced the signal.

Scaffold	Purpose	Disposition
Explicit <code>self_observe()</code> calls (108 instances)	Test whether self-observation produces detectable signal	Removed. Native frame economy reproduces the signal without explicit calls
Doubt condition as pengshu criterion	Test whether L5 → L6 conscious-layer doubt triggers boundary detection	Removed (M2). Boundary detection is L3 ↔ L4 structural, not L5 → L6 conscious
27-dim formula alphabet	GEME formula-language experiments (swap, pair, comm, etc.)	Dismantled. Replaced by configurable <code>vec_dim</code> (default 16)
Codex nearest-neighbor blending	Test whether inherited Codex entries can bias cavity input during processing	Partially removed. Formation and selection are now native; active query under boundary conditions remains scaffolded
Codex query under noise	Test whether environmental noise triggers productive Codex lookup	Still scaffolded. Codex operation — Self-level discrete Codex query under stress — is the remaining open mechanism

The scaffold method is not a shortcut. A shortcut skips the hard part. A scaffold builds a temporary bridge to the hard part, verifies the destination exists, then decides whether to build a road. Where the native frame economy reproduced the signal, the scaffold was removed. Active Codex operation remains scaffolded and is reported as a limitation.

S8. Negative Results and Decision Log

The final paper reports four positive lines, but the architecture was not reached by positive results alone. Several paths failed, collapsed under better controls, or were demoted from claim to limitation. These decisions are part of the result: they define what the instrument can and cannot currently support.

Track	Initial hope	What failed or changed	Decision
RNA CDS vs lncRNA	Use cross-encoding harm to distinguish	Pilot result $d=+1.23$ collapsed under full GENCODE and GC matching ($d \approx 0.12-0.17$).	Do not claim RNA class separation. Keep only shuffle hierarchy and

Track	Initial hope	What failed or changed	Decision
	coding from non-coding RNA	lncRNA carries its own structure; CDS/lncRNA is not separable by this readout at the tested scale.	boundary localization.
RNA frame detection	Detect correct reading frame through tri-nucleotide vs di-nucleotide cross-harm	48-nt and 21-nt windows both produced frame_delta≈0. The encoding reads codon-level order but not frame identity.	Do not claim frame detection. Reframe RNA as operation-boundary detection.
RNA RPF coupling	Demonstrate sequence-process coupling using TE/RPF information	RPF-only separated CDS/UTR nearly as well as the combined fork. Process permutation did not collapse the signal.	Do not claim sequence-process coupling. Report translation-occupancy boundary only where relevant.
WTC Codex operation	Show inherited Codex improves ongoing processing efficiency	BiasField and nearest-neighbor mechanisms did not produce a reliable operational benefit. Wrong Codex entries froze rather than harming processing.	Claim Codex formation and selection only. Active operation remains scaffolded.
DNA fork generalization	Read conservation directly from four-species fork columns across arbitrary genes	Random pooled genes did not generalize; AHSG produced a strong single-gene result because it carries documented selection variance.	Report AHSG as single-gene calibration, not population evidence. Use exon/intron assay as population-level DNA result.
UN historical validation	Use raw voting-stream precursor patterns as direct proof	Raw voting-stream calibration matched 3/4 selected historical events but is not the same pipeline as the 2025 forward annual ideal-point signal.	Keep pipeline separation. Do not let the 3/4 historical claim support the forward prediction.
NLP grammar path	Use language/NLP as primary grammar evidence	The path was abandoned because it risked becoming domain-template detection rather than blind structural measurement.	Remove NLP from the main proof. Use WTC/DNA/RNA instead.
Free-energy dialogue	Position the architecture as a substrate for FEP	Too much gravitational pull toward an existing theory; it obscured the instrument's own contribution and invited unnecessary conceptual battles.	Remove FEP as outreach route and main-paper dialogue.

Pilot-to-final principle. A pilot result is not evidence until it survives scale, matching, and ablation. RNA provides the clearest example: the striking pilot CDS/lncRNA effect was a two-gene lncRNA artifact. The final claim is smaller but stronger: the instrument cannot separate matched CDS and lncRNA by that readout, but it can localize the CDS stop boundary with high precision and detect the structural transition into and out of the coding regime.

Demotion is progress. Several ideas became weaker claims rather than failed claims: Codex operation became Codex selection; RNA grammar became RNA boundary; DNA fork population evidence became AHSG single-gene calibration; UN historical validation became pipeline-separated context. Each demotion removed a possible overclaim and sharpened the final proof chain.

S9. Reproducibility Checklist

Code. - Core instrument: `code/geme.py`, `code/geruon.py`, `code/we_core.py`. Python 3.8+ stdlib only. - Domain experiment scripts: `experiments/un/`, `experiments/wtc/`, `experiments/dna/`, `experiments/rna/`. - Quickstart verification: `docs/quickstart_geruon.py`, `docs/quickstart_ee_self.py`.

Data dependencies. - Core instrument: zero external packages. - Data processing: `numpy`, `pandas`, `pyarrow` (parquet). - RNA bigWig access: `pybigtools` (Windows). Raw data sources documented in each experiment README.

Required controls per domain. - Fair-coin baseline (zero sequential structure). - Shuffled-order control (preserves component statistics, breaks sequence). - κ ablation or sweep (confirms multi-lens is load-bearing). - Domain-specific: dinucleotide shuffle (DNA), process-permutation (RNA RPF), cross-key Codex injection (WTC), P5 ablation + FRED negative control (UN).

Required reporting. - Random seeds: 42, 123, 456 for multi-seed verification. - Sample counts and effect sizes with Cohen's d . - Window size, stride, κ values, `vec_dim`, `memory_cap`. - Three-tier threshold reporting ($|\text{error}| \leq 2 / \leq 3 / \leq 5$ windows). - All parameter choices declared before runs, not fitted to data. - Negative results and failure cases reported alongside positive ones.