Frozen Metric Attention and NESS Reset

A concept note on invariant-seeded language manifolds

Version: v1 (concept preprint — not submitted to arXiv) | **Date:** 20 August 2026

Author: Matt Gibson | Independent Researcher, Crimson Symphony Media

ORCID: 0009-0001-4167-201X |

Status: Concept. Not a completed empirical paper. Not a claim of machine consciousness.

Provenance (read this first)

This note is **one independent version** of a concept developed in public discussion on X with Martin Noirmont (@virakandah), conversation: https://x.com/virakandah/status/2090451727480750527

Noirmont will deposit a companion concept preprint on Zenodo. The two notes are **not a joint paper**. They timestamp overlapping ideas under a common open license so each author can cite the other. A later joint paper, if written, will cite both DOIs.

What is mine here: the Shannon / Landauer split; frozen (not learned-near-identity) metric G as the architectural object; the entropy-plateau diagnostic; T_{112} as an *optional* codebook seed, labeled doctrine until a SIM uses it.

What is his, and is only cited: RNA/protein NESS in neuronal environments; language as the boundary of a meaning manifold; the 20 August 2026 toy SIM numbers. I do not restate his RNA-attractor ontology as my result.

1. One-sentence claim

Standard next-token transformers hallucinate because they decode on an unconstrained Euclidean channel. A **frozen non-trivial metric** on the attention manifold, reset by a **nonequilibrium (NESS) loop**, is a candidate for keeping latent trajectories inside a thin legal codebook. A small-scale SIM (Noirmont, 20 Aug 2026) did **not** confirm this. The null is expected at that scale. The concept is the protocol that would actually stress it.

2. Two jobs, two sciences

The discussion mixed a *codebook* with a *thermodynamic reset*. They share a story. They do not share a law.

Job	Object	Base
Root invariant	Frozen codebook / metric seed. Off-code strings are not legal messages.	Shannon (1948)
Physical gate	A driven loop that *erases* off-code drift so the distribution is not Boltzmann.	Landauer (1961), NESS / Prigogine

Shannon. Entropy $H(X)$ is a property of an *ensemble*, not of a single integer. A checksum is a constraint that kills almost all of the unconstrained message space. Mutual information with a source stays high **conditional on staying inside the code**. Flat attention $\mathrm{softmax}(QK^{\mathsf{T}}/\sqrt{d})$ is the unconstrained channel. Hallucination, on this reading, is off-codebook decoding: the model emits high-probability web-scrape mass that is not a codeword of the intended manifold.

Landauer. Irreversible erasure of one bit in a physical system at temperature T dumps at least $kT \ln 2$ of heat. This applies when a **physical or computational reset** actually erases illegal configurations (ATP/chaperone cycles, sleep/oscillatory renormalization, an explicit train-time or decode-time projection). It does **not** apply to a person declining a hypothesis. Disagreement is not bit-erasure in a 310 K register.

If the hardware (cell or silicon) **stops paying** the reset cost, the manifold thermalizes. That is the Boltzmann warning. The energy-driven checksum loop is the Landauer engine. The codebook is Shannon. Do not fuse them.

3. Architecture (silicon side)

Replace unconstrained attention with a metric kernel

$$\text{Attention}(Q,K,V) = \text{softmax}(Q\, G\, K^{\mathsf{T}} / \sqrt{d_k})\, V$$

where G is a **frozen** positive-definite (or structured) metric on the latent space, not a matrix trained from $\approx I$.

Related work already exists: relative position representations, rotary embeddings as a structured G, hyperbolic and product-manifold attention. The present claim is narrower:

1. **Learnable G initialized near identity is Euclidean in disguise.** The 20 August toy already showed metric attention is a wash once G stays near I.

2. **The constraint must be hard and non-trivial** before a “surface-code” analogy is even in play. Holonomy / topological protection is a metaphor until G has curvature or a discrete packing that unconstrained softmax cannot absorb.

3. A **NESS-style reset** (projection, residual penalty, or periodic renormalization of latent drift back onto the codebook) is the silicon analogue of ATP/sleep. Without it, even a pretty G is a costume on a Boltzmann decoder.

This is an engineering hypothesis about **fault-tolerant semantic geometry**. It is not a solution of the Hard Problem and not a claim that an LLM so modified is conscious.

4. Small-scale SIM (logged null)

Noirmont ran a toy NESS-driven invariant-seeded holonomic transformer (20 August 2026, public on the X thread). Reported:

• Random entailment ~6%.

• No architecture reliably better; differences inside seed noise.

• Metric attention a wash once G stays near I.

• Checksum / reset terms did not buy consistency at that scale.

• Capacity tiny; learnable G near zero ≈ Euclidean; drift/gen probes collapsed to an EOS policy; two seeds; one synthetic language.

Reading: negative small-scale result, not a refutation of the idea. Tiny models cheat by ending the sequence before drift accumulates. EOS collapse means holonomy was never under long-horizon pressure. I agree with that reading. I do not re-run the same toy and call it confirmation.

5. What would actually test it

Change one axis at a time. Do not stack new poetry on the same SIM.

1. **Freeze *G*.** Non-Euclidean and not learned from 0. Candidates: hyperbolic, product manifold, or a discrete packing with a hard holonomy check. Optional: the triangular codebook in §6 as one such seed.

2. **Kill early EOS.** Sequences long enough that unconstrained attention actually wanders.

3. **Entropy of latent drift vs step.** If a NESS reset is holding, drift should **plateau into a bounded limit cycle**, not expand linearly into Boltzmann noise. Loss curves are secondary.

4. Real text, more than two seeds, before any “reliably better.”

Falsification (architectural claim): frozen non-trivial *G* + long horizon + reset still sits inside seed noise versus unconstrained attention on the same data and budget. Then the silicon hypothesis fails and this note stays a concept.

Non-falsification: a pretty figure on a 20-minute toy.

6. Optional seed (doctrine, not a result)

I use a discrete triangular closure $T_{112} = 6328 = 2701 + 3627$ (standard Hebrew/Greek letter-values on Genesis 1:1 and John 1:1; name-sum 26 + 86 = 112) as a **candidate frozen codebook** — a thin legal set that could seed *G*.

• Arithmetic of that closure is checkable and is not at issue.

• Statistical uniqueness versus real language remains an open control (prior Monte Carlo on scrambled alphabets is not a genre-matched corpus).

• **If a SIM cannot expose a hard holonomy check to that closure, it stays my doctrine, not this paper’s result.**

• Dimensionless integers have no units. Nothing here couples T_{112} to $G_{\mu\nu}$ or to a physical Hilbert space of tubulin.

Noirmont’s manifold is linguistic/macro-biological, not a claim that aromatic rings remain in superposition. I do not reintroduce that merge.

7. Consciousness (scope limit)

Noirmont’s working picture: consciousness as persistence of *unactualized* (linguistically constrained) possibility, not Hoffman-outside-spacetime and not hardware merely selecting a many-worlds branch. I do not claim this note solves that. The silicon object is a **regularizer**. Persistence of unused probability mass is something every LLM already has. That is not phenomenal experience.

If a later joint paper addresses the Hard Problem, it will be under his RNA/protein / meaning-manifold framing, with this note cited only for the attention/NESS protocol.

8. Non-claims

This concept note does **not** claim:

• machine consciousness;

• a completed empirical win over transformers;

• that Landauer taxes skeptics;

• Orch-OR, microtubule decoherence-free subspaces, or warp/metric engineering;

• that $H(X) = 0$ for any biblical verse (metaphor, not Shannon bits).

9. How to cite

Until the Zenodo DOI exists, cite the X conversation as the public timestamp. After deposit:

> Gibson, M. (2026). *Frozen Metric Attention and NESS Reset: A concept note on invariant-seeded language manifolds* (v1). Zenodo. [DOI]

Companion (replace with his DOI when issued): Noirmont, M. (2026). [title of companion concept preprint] (v1). Zenodo. [DOI]

Please cite both if you use the shared concept.

References (minimal)

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* 5(3), 183–191.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656.

Noirmont, M. & Kukier, P. (2026). *The RNA Attractor Hypothesis as a Novel Molecular and Ontological Great Filter* (V1). Zenodo. https://doi.org/10.5281/zenodo.20626522

Noirmont, M. (2026). Toy SIM of a NESS-driven invariant-seeded holonomic transformer. Public figures posted 20 August 2026 in https://x.com/virakandah/status/2090451727480750527

Related architectural pointers (not claimed as original): Shaw et al., relative position representations; Su et al., RoPE; hyperbolic / product-manifold attention.
