

Excess Entropy Production as a Candidate Universal Cost of Persistence: A Thermodynamic Foundation for the Attractor Framework; Robert Galida (July 2026) [F]

Abstract

Every dissipative system maintains its attractor through continuous reconfiguration. Reconfiguration requires work; work generates entropy. The recovery rate κ – corrective permeability – is the rate at which a system reconfigures to return to its attractor after perturbation. This paper proposes that κ is a measure of excess entropy generation rate.

We develop an abstract persistence cost framework and prove its equivalence to Lyapunov theory. We then identify entropy production as a physical realization of this cost, deriving: $\kappa = \inf_x \delta(x) \int_0^\infty \sigma_{\text{excess}}(\phi^t(x)) dt$ $\kappa = x \inf_x \int_0^\infty \sigma_{\text{excess}}(\phi^t(x)) dt \delta(x)$

where $\sigma_{\text{excess}} = \sigma - \sigma_{ss}$ $\sigma_{\text{excess}} = \sigma - \sigma_{ss}$ is the excess entropy production rate above the system's steady-state baseline. For physical systems, the baseline is zero (equilibrium); for biological, cognitive, and social systems, the baseline is the steady-state dissipation rate of the healthy, well-coordinated attractor.

This unifies physical, biological, cognitive, and social systems. The framework is grounded in the second law of

thermodynamics and non-equilibrium steady-state thermodynamics, not analogy. Empirical predictions are provided for each domain.

Keywords: entropy generation, excess entropy production, corrective permeability, attractor framework, dissipative structures, reconfiguration, Lyapunov theory, free energy principle, allostatic load

1. Introduction

The attractor framework defines persistence as the ability of a system to maintain its attractor under perturbation. Historically, persistence has been measured kinematically – as distance traveled or time spent away from equilibrium. This paper proposes that the true cost of persistence is thermodynamic: it is the excess entropy generated during reconfiguration and recovery.

Every dissipative system maintains its attractor through continuous reconfiguration. A bacterium reconfigures its metabolism to maintain homeostasis. A brain reconfigures its synaptic connections to maintain predictive models. A society reconfigures its institutions to maintain order. Reconfiguration requires work; work generates entropy. The second law of thermodynamics applies at every level of organization.

We develop an abstract persistence cost framework first, establishing its equivalence to Lyapunov theory. We then identify entropy production as a physical realization of this cost, deriving the relationship between corrective permeability and excess entropy generation.

The framework unifies physical, biological, cognitive, and

social systems. It is grounded in the second law of thermodynamics and non-equilibrium steady-state thermodynamics, not analogy.

2. The Persistence Cost Functional

Let X be a state space, $\phi_t(x)$ the flow of a dynamical system, and $A \subseteq X$ an attractor set. Let $\delta(x) = d(x, A)$ be the distance from x to the attractor. For a treatment of state-space constraints in viability theory, see Aubin (1991).

Definition 1 (Persistence Cost Functional): A persistence cost functional $C(x)$ is a scalar function on X satisfying:

1. $C(x) \geq 0$ for all x
2. $C(x) = 0$ if and only if $x \in A$
3. $C(\phi_t(x)) \in L^1([0, \infty))$ for all x in the basin

Definition 2 (Cumulative Persistence Cost): For a finite horizon $T > 0$: $D_T(x) = \int_0^T C(\phi_t(x)) dt$

For trajectories that converge to the attractor: $D_\infty(x) = \int_0^\infty C(\phi_t(x)) dt$

3. Existence and Lyapunov Equivalence

Theorem 1 (Existence of the Persistence Functional): Assume $C(x) \geq 0$, $C = 0$ only on A ,

and $C(\phi_t(x)) \in L^1([0, \infty))$ $C(\phi_{t^*}(x)) \in L^1([0, \infty))$ for all x in the basin. Assume f is locally Lipschitz, the flow is continuously differentiable in the initial condition, and C is continuous and locally bounded. Then:

1. $D^\infty(x) = \int_0^\infty C(\phi_t(x)) dt$ $D^\infty(x^*) = \int_0^\infty C(\phi_{t^*}(x)) dt$ exists and is finite.
2. $D^\infty D^\infty(x^*)$ is continuous.
3. $D^\infty D^\infty(x^*)$ satisfies the transport equation:

$$\nabla D^\infty(x) \cdot f(x) = -C(x) \quad \nabla D^\infty(x^*) \cdot f(x) = -C(x)$$

Proof: The integral exists and is finite by the $L^1 L^1$ assumption. Continuity follows from the dominated convergence theorem under the stated regularity assumptions. To derive the transport equation, compute: $D(\phi_h(x)) = \int_0^\infty C(\phi_{t+h}(x)) dt = D(x) - \int_0^h C(\phi_t(x)) dt$ $D(\phi_{h^*}(x)) = \int_0^\infty C(\phi_{t^*+h}(x)) dt = D(x) - \int_0^h C(\phi_{t^*}(x)) dt$

$$\text{Then: } D(\phi_h(x)) - D(x) = -\int_0^h C(\phi_t(x)) dt \rightarrow -C(x) \quad D(\phi_{h^*}(x)) - D(x) = -\int_0^h C(\phi_{t^*}(x)) dt \rightarrow -C(x)$$

as $h \rightarrow 0$ $h \rightarrow 0$. By the chain rule: $\nabla D(x) \cdot f(x) = -C(x)$ $\nabla D(x^*) \cdot f(x) = -C(x)$ \square

Corollary (Equivalence to Lyapunov Theory): Any Lyapunov function $V(x)$ (with $V \geq 0$, $V = 0$ on the attractor, and $V' \leq 0$) yields a persistence cost $C(x) = -V'(x)$. Conversely, any persistence cost $C(x)$ satisfying $\nabla D \cdot f = -C$ defines a Lyapunov function $D(x)$.

Proof: If V is a Lyapunov function, then $V' = \nabla V \cdot f \leq 0$. Define $C = -V'$. Then $C \geq 0$, $C = 0$ on the attractor, and $D = \int C = V(x) - V(\phi_T(x))$. Conversely, if $\nabla D \cdot f = -C$, then $D' = -C \leq 0$, so D is a Lyapunov function. \square

Interpretation: The persistence cost framework is mathematically equivalent to classical Lyapunov stability theory. For the connection to contraction analysis, see Lohmiller & Slotine (1998). For control Lyapunov functions, see Freeman & Kokotovic (1996). Entropy production is one physically meaningful realization of the cost function CC . For a detailed treatment of Lipschitz continuity of $D^\infty D^\infty$ under a Lipschitz-flow hypothesis, see Galida (2026a), Proposition 4.

4. Entropy Production as Persistence Cost

4.1 Entropy Balance

For an open system, the entropy balance equation is: $dS_{\text{system}}/dt = \sigma - \Phi$

where $\sigma \geq 0$ is the entropy production rate (always non-negative by the second law) and Φ is the entropy export rate to the environment. For foundational treatments of stochastic thermodynamics and entropy production, see Seifert (2012) and Sekimoto (2010).

For a system in a steady state: $dS_{\text{system}}/dt = 0 \implies \sigma = \Phi$

4.2 Excess Entropy Production

Define the **steady-state entropy production rate** σ_{ss} as the rate when the system is at its attractor.

Define the **excess entropy production rate**: $\sigma_{\text{excess}}(x) = \sigma(x) - \sigma_{\text{ss}}$

Assumption (Excess Entropy Decay): For all trajectories in the

basin, there exist constants $C < \infty$ and $\mu > 0$ such that:

$$\sigma_{\text{excess}}(\phi^t(x)) \leq C e^{-\mu t} \sigma_{\text{excess}}(x)$$

$$\sigma_{\text{excess}}(\phi^t(x)) \leq C e^{-\mu t} \sigma_{\text{excess}}(x)$$

for all $t \geq 0$. This ensures $D^\infty(x) < \infty$ and is the standard hypothesis under which the persistence functional and its associated bounds are well-defined, consistent with Galida (2026a, 2026b). The decay rate μ may be domain-specific and is empirically measurable.

Note on generalization: The exponential decay assumption is adopted here to ensure finiteness of D^∞ and to maintain consistency with the prior papers in this series. Generalization to L1 integrable decays (e.g., algebraic) is a priority for future work.

4.3 The Entropy Persistence Functional

Definition 3 (Cumulative Excess Entropy Functional): For a finite horizon $T > 0$:

$$D^T(x) = \int_0^T \sigma_{\text{excess}}(\phi^t(x)) dt$$

$$D^T(x) = \int_0^T \sigma_{\text{excess}}(\phi^t(x)) dt$$

For trajectories that converge to the attractor:

$$D^\infty(x) = \int_0^\infty \sigma_{\text{excess}}(\phi^t(x)) dt$$

$$D^\infty(x) = \int_0^\infty \sigma_{\text{excess}}(\phi^t(x)) dt$$

Interpretation: The persistence functional is the total excess entropy generated during reconfiguration and recovery.

4.4 Corrective Permeability

Definition 4 (Corrective Permeability): $\kappa = \inf_{x \in B} \delta(x) D^\infty(x)$

$$\kappa = \inf_{x \in B} \delta(x) D^\infty(x)$$

where $\delta(x) = d(x, A)$ is the distance to the attractor.

Interpretation: κ is the minimum excess entropy cost per unit distance. It measures the efficiency of reconfiguration: a system that returns with minimal excess entropy generation has

high κ ; a system that generates excess entropy has low κ .

4.5 Basin Depth

Proposition 1 (Properties of Basin Depth): Define $B = D^\infty(\text{saddle}) - D^\infty(A)$, where saddle is the lowest point on the basin boundary (the separatrix between attractors). For the connection to large-deviation theory and escape rates, see Freidlin & Wentzell (2012). Then:

1. $B \geq 0$, with equality iff the basin has no barrier (i.e., the boundary coincides with the attractor).
2. For gradient systems $\dot{x} = -\nabla V(x)$, $B = V(\text{saddle}) - V(A)$ (the classical energy barrier).
3. B is invariant under smooth coordinate changes (coordinate invariance).
4. B depends on the chosen persistence cost functional C ; different costs yield different barriers.

Proof: (1) follows from non-negativity of D^∞ . (2) follows from the transport equation $\nabla \cdot f = -C$ and the identity $f = -\nabla V$. (3) follows from the invariance of the integral under diffeomorphisms. (4) is self-evident.

5. Domain-Specific Realizations

5.1 Physical Systems: Thermodynamic

Excess Entropy

For a thermodynamic system, $S(x) = k_B \log \Omega(x)$, where $\Omega(x)$ is the number of microstates. For an isolated system, $\sigma_{ss} = 0$ (equilibrium), so $\sigma_{\text{excess}} = \sigma - S'$. $\kappa = \inf_x \delta(x) [S(A) - S(x)]$

Example: A gas returning to equilibrium after compression. The entropy generated is $\Delta S = nR \log(V_f/V_i)$.

5.2 Biological Systems: Metabolic Excess Entropy

For a biological system, $S(x)$ is the metabolic entropy. The baseline σ_{ss} is the resting metabolic rate (homeostasis). The excess is: $\sigma_{\text{excess}} = \text{metabolic rate} - \text{resting metabolic rate}$. $\kappa = \inf_x \delta(x) \int_0^\infty \sigma_{\text{excess}}(\phi_t(x)) dt$

Example: A cell returning to homeostasis after a nutrient shock. The excess entropy generated is the metabolic cost of restoring homeostasis above baseline. For the dissipative-structures framework underlying biological self-organization, see Nicolis & Prigogine (1989).

5.3 Cognitive Systems: Free Energy Dissipation

For a cognitive system, variational free energy $F = -\log p(y|x) + \text{DKL}[q(\cdot) \| p(\cdot|x)]$ is adopted here as one candidate persistence functional. We do not claim variational free energy is uniquely correct; it is adopted as the most developed existing candidate persistence functional for cognitive systems. Other candidates (Bayesian surprise, expected free energy, predictive information) are possible; this paper focuses

on FF due to its established role in the free-energy principle (Friston, 2010). For the thermodynamics of information and its connection to free-energy minimization, see Parrondo, Horowitz & Sagawa (2015) and Sagawa & Ueda (2008).

The baseline σ_{ss} is the baseline neural dissipation rate (resting brain activity). The excess is: $\sigma_{\text{excess}} = F' - F'_{ss}$
 $\kappa = \int_0^\infty \int \sigma_{\text{excess}}(\phi_t(x)) dt \delta(x)$

Example: A cognitive system updating its beliefs after a prediction error. The excess entropy generated is the free energy dissipated during belief updating above baseline.

5.4 Social Systems: Coordination Excess Entropy

For a social system, define the aggregate social entropy production rate as: $\sigma_{\text{social}}(t) = \sum_i (S'_i(t) - S'_{i\text{rest}})$

where $S'_i(t)$ is the total entropy production rate of individual i , and $S'_{i\text{rest}}$ is the individual's baseline entropy production rate in a resting, minimally socially constrained state. This is measured via physiological proxies such as basal metabolic rate, resting allostatic load, or cortisol baseline (McEwen, 1998; Sterling & Eyer, 1988).

Interpretation: σ_{social} measures the excess dissipation attributable to social constraints: the additional entropy generated by coordination, communication, conflict, norm enforcement, and institutional friction.

Non-Negativity: Unlike total entropy production $S'_i \geq 0$ (which follows from the second law), σ_{social} is not guaranteed to be non-negative. Division of labor, infrastructure, and specialization may reduce an individual's

metabolic burden relative to a solitary baseline. The hypothesis is that during recovery from social disruption, $\sigma_{\text{social}} \geq 0$ $\sigma_{\text{social}} \rightarrow 0$; in steady-state, $\sigma_{\text{social}} \rightarrow 0$ $\sigma_{\text{social}} \rightarrow 0$. This is an empirical claim, not a theorem.

The baseline σ_{ss} is the steady-state social entropy production rate (well-coordinated society). The excess is: $\sigma_{\text{excess}} = \sigma_{\text{social}} - \sigma_{\text{ss}}$

$$\kappa = \int_0^\infty \int_{\mathcal{X}} \sigma_{\text{excess}}(\phi_t(x)) dt \delta(x) = \int_0^\infty \int_{\mathcal{X}} \sigma_{\text{excess}}(\phi_t(x)) dt \delta(x)$$

Example: A society recovering from a shock (economic crisis, political upheaval). The excess entropy generated is the coordination cost of restructuring above baseline. A harmonious society has $\sigma_{\text{excess}} = 0$; a turbulent society has $\sigma_{\text{excess}} > 0$; a chronically turbulent society may have settled into a new attractor with a higher σ_{ss} . This illustrates the framework's central distinction: the attractor is the state of minimum entropy generation for that class of system.

6. The Unified Framework

6.1 Summary Table

Domain	Entropy Functional	Baseline σ_{ss}	Excess σ_{excess}	Recovery Rate κ
Physical	Thermodynamic entropy	0 (equilibrium)	$S - S'$	$\int_0^\infty \delta \Delta S dt$
Biological	Metabolic entropy	Resting metabolic rate	Metabolic rate - resting	$\int_0^\infty \delta \sigma_{\text{excess}} dt$
Cognitive	Free energy	Baseline neural dissipation	$F' - F'_{\text{ss}}$	$\int_0^\infty \delta \sigma_{\text{excess}} dt$
Social	Social entropy production	Steady-state social dissipation	$\sigma_{\text{social}} - \sigma_{\text{ss}}$	$\int_0^\infty \delta \sigma_{\text{excess}} dt$

6.2 The Universal Structure

Every domain follows the same mathematical structure:

Component	Expression
Excess entropy production	$\sigma_{\text{excess}}(x) = \sigma(x) - \sigma_{\text{ss}}$ $\sigma_{\text{excess}}[\cdot](x) = \sigma(x) - \sigma_{\text{ss}}[\cdot]$
Cumulative cost	$D^\infty(x) = \int_0^\infty \sigma_{\text{excess}}(\phi^t(x)) dt$ $D^\infty[\cdot](x) = \int_0^\infty \sigma_{\text{excess}}[\cdot](\phi^t[\cdot](x)) dt$
Recovery rate	$\kappa = \inf_x \delta(x) / D^\infty(x)$ $\kappa = \inf_x \delta[\cdot](x) / D^\infty[\cdot](x)$
Basin depth	$B = D^\infty(\text{saddle})$ $B = D^\infty[\cdot](\text{saddle})$
Transport equation	$\nabla D \cdot f = -\sigma_{\text{excess}}$ $\nabla D \cdot f = -\sigma_{\text{excess}}[\cdot]$

6.3 The Low-Energy Attractor Benchmark (Proposed Hypothesis)

We propose the following benchmark as an additional hypothesis: **the attractor is the state of minimum entropy generation for that class of system.**

Domain	Attractor	Entropy Generation at Attractor
Physical	Equilibrium	$\sigma = 0$ $\sigma = 0$
Biological	Homeostasis	$\sigma = \sigma_{\text{ss}} > 0$ $\sigma = \sigma_{\text{ss}}[\cdot] > 0$ (resting metabolism)
Cognitive	Settled Belief	$\sigma = \sigma_{\text{ss}} > 0$ $\sigma = \sigma_{\text{ss}}[\cdot] > 0$ (baseline neural dissipation)
Social	Coordinated Order	$\sigma = \sigma_{\text{ss}} > 0$ $\sigma = \sigma_{\text{ss}}[\cdot] > 0$ (baseline institutional friction)

Interpretation:

1. For **equilibrium systems** (gases, isolated systems), the attractor is the state where entropy generation reaches zero – the system has nowhere lower to go.

2. For **dissipative systems** (cells, brains, societies), the attractor is the state where entropy generation reaches its lowest *non-zero* steady-state value – the minimum entropy generation the system can sustain while maintaining its functional organization.

Important caveats:

- This is a proposed benchmark, not a derived theorem.
- For cognitive systems in particular, minimizing *entropy production rate* (a thermodynamic quantity) and minimizing *free energy/surprise* (the actual claim in the free-energy principle) are distinct minimization principles. The framework does not establish a bridge between them; this is an open question.
- The benchmark is an empirical hypothesis that requires domain-specific validation.

In all cases, the attractor is the **lowest entropy-generating state that system can have while remaining itself.**

7. Testable Predictions

7.1 Core Prediction

Prediction: The recovery rate κ is inversely proportional to the excess entropy generated during reconfiguration: $\kappa \propto 1/D_{\infty}$ $\kappa \propto D_{\infty}^{-1}$

Falsification: If a system returns to its attractor with high excess entropy generation but high recovery rate, the prediction is falsified.

7.2 Secondary Prediction

Prediction: Systems that maintain their attractor with minimal excess entropy generation are more “efficient.” Systems that generate excess entropy are “inefficient” or “stressed.”

Falsification: If an inefficient system has lower excess entropy generation than an efficient system, the prediction is falsified.

7.3 Domain-Specific Predictions

Domain	Prediction	Falsification
Physical	κ correlates with thermal efficiency	κ high but efficiency low
Biological	κ correlates with metabolic efficiency	κ high but metabolic cost high
Cognitive	κ correlates with learning efficiency	κ high but learning cost high
Social	κ correlates with institutional efficiency	κ high but coordination cost high

8. Experimental Design

8.1 Physical Systems

- **System:** Gas in a piston
- **Perturbation:** Compression
- **Measurement:** Excess entropy generation (heat measurement) and recovery time
- **Test:** Correlation between κ and $1/D^\infty$

8.2 Biological Systems

- **System:** Cell culture
- **Perturbation:** Nutrient shock
- **Measurement:** Metabolic rate above resting (oxygen consumption) and recovery time
- **Test:** Correlation between $\kappa\kappa$ and metabolic cost

8.3 Cognitive Systems

- **System:** Human participants in a learning task
- **Perturbation:** Prediction error
- **Measurement:** Free energy dissipation above baseline (EEG complexity, pupil dilation) and belief updating rate
- **Test:** Correlation between $\kappa\kappa$ and free energy dissipation

8.4 Social Systems

- **System:** Institutional response to shocks
 - **Perturbation:** Economic or political crisis
 - **Measurement:** Social entropy production above baseline (allostatic load, cortisol, institutional friction) and recovery time
 - **Test:** Correlation between $\kappa\kappa$ and social entropy production
-

9. Open Questions

Question	Status	Difficulty
Q1: Uniqueness of $S(x)$	Are there multiple valid entropy functionals for a given domain?	Hard
Q2: Variational principle	Is there a universal variational principle that yields $S(x)$?	Hard
Q3: Social second law	Does $\sigma_{\text{social}} \geq 0$ always hold during recovery?	Very Hard
Q4: Cross-level entropy	How does entropy generation at one level relate to entropy generation at another?	Hard
Q5: Measurement	Can we measure excess entropy generation in cognitive and social systems directly?	Moderate
Q6: Unification	Can all domain-specific entropy functionals be derived from a single universal functional?	Very Hard

10. Conclusion

Every dissipative system maintains its attractor through continuous reconfiguration. Reconfiguration requires work; work generates excess entropy. The recovery rate κ – corrective permeability – is the rate at which a system reconfigures to return to its attractor after perturbation. We have proposed that κ is a measure of excess entropy generation rate.

We developed an abstract persistence cost framework and proved its equivalence to Lyapunov theory. We then identified entropy production as a physical realization of this cost, deriving: $\kappa = \inf_{\delta} \int_0^{\infty} \sigma_{\text{excess}}(\phi_t(x)) dt$

$$(\phi t(x)) dt \delta(x)$$

where $\sigma_{\text{excess}} = \sigma - \sigma_{\text{ss}}$ is the excess entropy production rate above the system's steady-state baseline – thermodynamic entropy for physical systems, metabolic entropy for biological systems, free energy dissipation for cognitive systems, and social entropy production for social systems.

We proposed a unified benchmark: the attractor is the state of minimum entropy generation for that class of system – zero for equilibrium systems, non-zero steady-state for dissipative systems. This provides a unified criterion for identifying attractors across domains: an attractor is a state from which the system cannot reduce its entropy generation further without losing its defining structure or function.

This unifies physical, biological, cognitive, and social systems. In each domain, persistence requires reconfiguration; reconfiguration generates excess entropy; κ measures the entropy cost of that reconfiguration. The framework is grounded in the second law of thermodynamics and non-equilibrium steady-state thermodynamics, not analogy.

Social Application: The framework provides a thermodynamic interpretation of social dynamics: harmony is a low-entropy attractor state; turbulence is a high-entropy state generated by excess dissipation during reconfiguration. The recovery rate κ measures how efficiently a society transitions from turbulence back to harmony – that is, how quickly it reduces its excess entropy production to zero.

11. Limitations

This paper establishes an abstract persistence cost framework with a proposed thermodynamic realization. Several limitations

should be explicitly acknowledged:

1. **Uniqueness.** Entropy production is not proved to be the unique persistence cost. Many positive functionals $C(x)$ satisfy $\nabla D \cdot f = -C$. The identification of entropy production as the canonical cost is a physically motivated hypothesis, not a mathematical theorem.
 2. **Scope.** The framework does not imply that all domains obey thermodynamics literally. The cognitive and social realizations are proposed hypotheses requiring empirical validation.
 3. **Decay assumption.** Exponential decay of excess is a sufficient assumption to ensure finiteness of D^∞ , not a necessary one. Generalization to L^1 integrable decays (e.g., algebraic) is a priority for future work.
 4. **Basin depth.** Basin depth $B = D^\infty(\text{saddle})$ is defined in terms of the persistence cost functional. Its relationship to classical energy barriers is established only for gradient systems.
 5. **Empirical validation.** The predictions of the framework – particularly the inverse relationship between κ and D^∞ – remain to be tested empirically across domains.
 6. **Low-energy attractor benchmark.** The benchmark proposed in §6.3 is a hypothesis, not a derived theorem. For cognitive systems, it risks conflating thermodynamic entropy production with free-energy minimization – distinct principles whose relationship remains open.
-

References

- Aubin, J. P. (1991). *Viability Theory*. Birkhäuser.
- Boltzmann, L. (1877). "Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung." *Wiener Berichte*, 76, 373-435.
- Clausius, R. (1865). "Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie." *Annalen der Physik*, 125(7), 353-400.
- Freeman, R. A., & Kokotovic, P. V. (1996). *Robust Nonlinear Control Design: State-Space and Lyapunov Techniques*. Birkhäuser.
- Freidlin, M. I., & Wentzell, A. D. (2012). *Random Perturbations of Dynamical Systems* (3rd ed.). Springer.
- Friston, K. (2010). "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, 11(2), 127-138.
- Galida, R. (2026a). "The Persistence Functional: A Candidate Formal Foundation for the Attractor Framework." *Fantasy Attractor*.
- Galida, R. (2026b). "Deriving Corrective Permeability from the Cumulative Deviation Functional." *Fantasy Attractor*.
- Jaynes, E. T. (1957). "Information Theory and Statistical Mechanics." *Physical Review*, 106(4), 620-630.
- Khalil, H. K. (2002). *Nonlinear Systems* (3rd ed.). Prentice Hall.
- Kondepudi, D., & Prigogine, I. (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. Wiley.
- Lohmiller, W., & Slotine, J. J. E. (1998). "On contraction analysis for non-linear systems." *Automatica*, 34(6), 683-696.

Lyapunov, A. M. (1892). *The General Problem of the Stability of Motion*.

McEwen, B. S. (1998). "Stress, Adaptation, and Disease: Allostasis and Allostatic Load." *Annals of the New York Academy of Sciences*, 840(1), 33-44.

Nicolis, G., & Prigogine, I. (1989). *Exploring Complexity: An Introduction*. W. H. Freeman.

Parrondo, J. M. R., Horowitz, J. M., & Sagawa, T. (2015). "Thermodynamics of information." *Nature Physics*, 11(2), 131-139.

Prigogine, I. (1947). *Étude Thermodynamique des Phénomènes Irréversibles*. Dunod.

Prigogine, I., & Nicolis, G. (1977). *Self-Organization in Non-Equilibrium Systems*. Wiley.

Sagawa, T., & Ueda, M. (2008). "Second law of thermodynamics with discrete quantum feedback control." *Physical Review Letters*, 100(8), 080403.

Seifert, U. (2012). "Stochastic thermodynamics, fluctuation theorems and molecular machines." *Reports on Progress in Physics*, 75(12), 126001.

Sekimoto, K. (2010). *Stochastic Energetics*. Springer.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27(3), 379-423.

Sterling, P., & Eyer, J. (1988). "Allostasis: A New Paradigm to Explain Arousal Pathology." In *Handbook of Life Stress, Cognition and Health*, 629-649.

Suggested citation: Galida, R. S. (2026). Excess Entropy Production as a Candidate Universal Cost of Persistence: A

The Persistence Functional: A Candidate Formal Foundation for the Attractor Framework; Robert Galida (July 2026) [F]

Abstract

The attractor framework provides a domain-general vocabulary for describing persistence and change across physical, biological, cognitive, and social systems. However, its core variables— κ (corrective permeability), B (basin depth), and R (reality alignment)—have been defined inconsistently across application papers, and their formal relationships have remained implicit. This paper proposes a candidate mathematical formalization for the framework.

The central mathematical innovation of this paper is treating persistence as a functional defined over trajectories— $D_T(x) = \int \theta_T d(\phi_\tau(x), A)$ —rather than as a scalar property of states. We prove several mathematical properties of D_T , including non-negativity, monotonicity in T , additivity, Lipschitz continuity with respect to initial conditions, and a bound relating D_∞ to the recovery rate κ : $D_\infty(x) \leq \kappa d(x, A)$. We establish connections to dynamic programming and ergodic theory via occupation measures. We introduce a

complementary **topological** **persistence functional** $P_{\text{topo}}(t)P_{\text{topo}}(t)$, which measures the lifetime of topological features in the trajectory's state-space geometry, and the **topological evolution rate** $E(t)E(t)$.

We unify the framework's variable set: κ is the recovery rate (operationalized as $1/\tau$); γ is a proposed drift rate for persistent chaos, grounded in the literature on high-dimensional neural networks; B is the energy barrier (basin depth); \tilde{B} is a complementary persistence depth; R is the expected log predictive likelihood. We propose testable predictions linking $E(t)$ to κ and γ , and provide a falsifiable experimental protocol using neural network training and persistent homology.

The paper offers a candidate formal foundation, with explicit definitions, mathematical properties, and empirical grounding. All unverified sources are clearly labeled as such.

Keywords: attractor framework, persistence functional, cumulative deviation, topological persistence, corrective permeability, basin depth, reality alignment, persistent homology

1. Introduction

The attractor framework has been applied across physics (hydrogen decay, Jeans instability), biology (ECM mechanics, HRV), cognition (belief updating, performance attractors), and social systems (religious attractors, civilizational dynamics). A common vocabulary has emerged: κ (corrective permeability), B (basin depth), and R (reality alignment). However, these variables have been defined inconsistently across papers, and their formal relationships have remained implicit. This paper proposes a candidate mathematical

formalization that addresses these inconsistencies.

The central mathematical innovation of this paper is treating persistence as a functional defined over trajectories rather than as a scalar property of states. $DT(x) = \int_0^T d(\phi_\tau(x), A) d\tau$ can be understood as a type of **action functional** (carefully qualified). Like the classical action $\int L(q, \dot{q}) dt$, it assigns a scalar to an entire trajectory, is additive under concatenation, and suggests variational and optimal-control interpretations. However, it is not the mechanical action; it is a cumulative deviation functional that measures time away from equilibrium. This moves the framework into the domain of trajectory-level analysis, aligning it with modern dynamical systems and geometric control theory.

We introduce the **cumulative deviation functional** $DT(x)$ as this central object, and we establish its mathematical properties, including its relationship to the recovery rate κ . We introduce a complementary **topological persistence functional** $P_{\text{topo}}(t)$ and the **topological evolution rate** $E(t)$. We unify the framework's variable set with operational definitions and propose testable predictions with falsification criteria.

1.1 Scope and Status

This paper is a **candidate formalization**—it provides definitions, mathematical properties, and empirical hypotheses. It is not a completed empirical validation; that is the subject of future work. All claims are labeled as **definitions** (part of the formal structure), **propositions/theorems** (proved), **hypotheses** (testable predictions), or **heuristics** (suggestive connections not yet formalized). This distinction is maintained throughout.

2. Formal Definitions

Let X be a metric space with distance function $d(\cdot, \cdot)$. Let $\phi_\tau(x)$ be the flow of a dynamical system starting from state $x \in X$ at time $\tau = 0$. Let $A \subseteq X$ be an attractor set (a compact, invariant set to which trajectories converge). Assume the flow is continuous and measurable so that $d(\phi_\tau(x), A)$ is measurable. The flow ϕ_τ satisfies the semigroup property $\phi_{t+s} = \phi_t \circ \phi_s = \phi_s \circ \phi_t$ for all $t, s \geq 0$, with $\phi_0 = \text{id}$. We assume $d(\phi_\tau(x), A) \in L^1([0, T])$ for all finite T , so the integral defining DT is well-defined.

Define the distance from a point to the attractor: $d(x, A) = \inf_{a \in A} d(x, a)$

The definition applies to any metric space; for infinite-dimensional spaces, the usual measurability and integrability conditions are assumed.

2.1 Cumulative Deviation Functional

Definition 1 (Cumulative Deviation Functional): For a finite horizon $T > 0$, the cumulative deviation functional is: $DT(x) = \int_0^T d(\phi_\tau(x), A) d\tau$

Interpretation: $DT(x)$ is the total accumulated deviation from the attractor over the interval $[0, T]$. It measures integrated error, residence-time-weighted distance, or accumulated regret. This is **not** a path length; it measures time spent away from equilibrium, whereas path length $\int_0^T \|\dot{\phi}_\tau(x)\| d\tau$ measures distance traveled.

Domain generality: This definition applies to any system with a well-defined state space, a flow, and an attractor set. It does not require linearity, differentiability, or specific

functional forms.

Empirical note: $D_T D_T^\square$ is the fundamental object for empirical work; $D_\infty D_\infty^\square$ is primarily an analytical limit used for theoretical bounds.

Note: $D_T D_T^\square$ is not a Lyapunov function. A Lyapunov function is a scalar function of the current state; $D_T D_T^\square$ is a functional of the entire trajectory. It does not decrease monotonically along trajectories, and it does not provide pointwise stability information. Its purpose is to measure accumulated history, not instantaneous energy.

Occupation measure connection: Define the occupation measure of the trajectory up to time T as: $\mu_T(B) = \int_0^T \mathbf{1}_B(\phi_\tau(x)) d\tau$
 $(B) = \int_0^T \mathbf{1}_B(\phi_\tau(x)) d\tau$

for measurable $B \subseteq X$. Then: $D_T(x) = \int_X d(y, A) d\mu_T(y)$
 $D_T^\square(x) = \int_X d(y, A) d\mu_T^\square(y)$

Thus $D_T D_T^\square$ is the expected distance to the attractor under the occupation measure. This connects the functional directly to ergodic theory and occupation measure analysis. For foundational treatments of occupation measures and invariant measures, see Ruelle (1989) and Bowen (1975).

2.1.1 Why the L^1 Trajectory Functional?

The choice of the L^1 integral over alternatives is motivated by the following properties:

- **Linearity:** Each moment contributes equally; accumulation is additive over time.
- **Physical units:** For systems with a natural distance metric, $D_T D_T^\square$ has units of distance \times time, which is interpretable as accumulated deviation.

- **Simplicity:** It is the simplest nontrivial trajectory functional that is not a path length.
- **Analogy:** It mirrors cumulative regret and occupation measures in control theory and ergodic theory.
- **Avoidance of overweighting:** Unlike d_2 , it does not disproportionately weight large deviations; unlike \max , it is sensitive to the full trajectory.

This is one natural choice; other functionals (e.g., d_p , exponentially weighted integrals) could be substituted without changing the framework's structure.

2.2 Topological Persistence Functional

Let $X_\tau = \{\phi_s(x) : s \in [0, \tau]\}$, $X_{\tau'} = \{\phi_{s'}(x) : s' \in [0, \tau']\}$ be the trajectory segment up to time τ . Let $\text{PH}_k(X_\tau)$, $\text{PH}_k(X_{\tau'})$ be the k -dimensional persistent homology of the point cloud X_τ , $X_{\tau'}$ at scale ϵ . Each feature (component, loop, void) has a birth scale b and a death scale d , with persistence $d - b$. For foundational treatments of persistent homology, see Edelsbrunner & Harer (2010) or Carlsson (2009).

Definition 2 (Topological Persistence Functional): We define the following complementary topological persistence functional.

$$\text{For } t \geq 0: P_{\text{topo}}(t) = \int_0^t \sum_{k \geq 0} (b, d) \in \text{PH}_k(X_\tau) (d - b) d\tau$$

$$P_{\text{topo}}(t) = \int_0^t \sum_{k \geq 0} (b, d) \in \text{PH}_k(X_{\tau'}) (d - b) d\tau$$

The map $\tau \mapsto \text{PH}_k(X_\tau)$, $\tau \mapsto \text{PH}_k(X_{\tau'})$ is piecewise constant on intervals where the trajectory does not cross a homology-critical threshold. Assuming the trajectory crosses such thresholds at discrete times, the integral is well-defined as a sum of piecewise continuous segments. This is the standard assumption in time-varying persistent homology (see Carlsson & Zomorodian, 2009).

Interpretation: $P_{\text{topo}}(t)P_{\text{topo}}^{\square}(t)$ is the total lifetime of all topological features in the trajectory's state-space geometry up to time t . This is a separate mathematical object from $DTDT^{\square}$; the relationship between them is an empirical hypothesis. This is one possible choice among several topological summaries (e.g., persistence landscapes, persistence images) and is selected because it mirrors the cumulative interpretation of $DTDT^{\square}$, rather than because it is uniquely canonical. Other stable summaries—such as persistence landscapes, persistence images, or Betti curves—could be substituted for the present functional without changing the framework's structure.

Measurement: In practice, $P_{\text{topo}}(t)P_{\text{topo}}^{\square}(t)$ is computed by sampling the trajectory at discrete times, computing persistent homology on latent activation manifolds, and summing the persistence of all features using standard libraries (e.g., GUDHI, Ripser). Turner & Barak (2023) demonstrated that trained RNNs develop attractors sequentially during training; the topological structure of these attractors can be analyzed using persistent homology.

Falsification: If persistent homology features do not correlate with any behavioral or dynamical measure in a given system, $P_{\text{topo}}P_{\text{topo}}^{\square}$ is not a useful construct for that domain.

2.3 Topological Evolution Rate

Definition 3 (Topological Evolution Rate): For a learning system with time-dependent topological persistence, the topological evolution rate is defined as: $E(t) = \frac{d}{dt} P_{\text{topo}}(t)$ $E(t) = \frac{d}{dt} P_{\text{topo}}^{\square}(t)$

where $E(t)$ is differentiable, and $E(t)$ is experimentally as $E(t) \approx \frac{\Delta P_{\text{topo}}}{\Delta t}$ $E(t) \approx \frac{\Delta P_{\text{topo}}^{\square}}{\Delta t}$ over finite intervals.

Interpretation: $E(t)E(t)$ measures how quickly the system's topological complexity changes during learning. Negative $E(t)E(t)$ indicates topological simplification (compression); positive $E(t)E(t)$ indicates increasing complexity (expansion); $E(t) \approx 0, E(t) \approx 0$ indicates stagnation. Learning is one possible cause of topological change; random drift, noise, or chaotic wandering can also change topology.

Empirical anchor: Karuppiah, Nazreen Banu et al. (2026) examine the evolution of topological signatures during training. Turner & Barak (2023) show that RNNs develop attractors sequentially, which may correspond to phases of topological simplification. We hypothesize that successful learning corresponds to negative average values of $E(t)E(t)$ over defined phases, but this is a testable claim, not a definition.

3. Mathematical Properties of the Cumulative Deviation Functional

This section establishes the mathematical behavior of $DTDT_{\square}$, providing the foundation for its use in the framework.

3.1 Non-negativity

Proposition 1 (Non-negativity): For any $x \in X, x \in X$ and any $T \geq 0, T \geq 0: DT(x) \geq 0, DT_{\square}(x) \geq 0$

with equality iff $\phi_{\tau}(x) \in A, \phi_{\tau_{\square}}(x) \in A$ for almost all $\tau \in [0, T], \tau \in [0, T]$.

Proof: The integrand is a distance function $d(\phi_{\tau}(x), A), d(\phi_{\tau_{\square}}(x), A)$, which is non-negative by definition. The integral of a non-negative function is non-negative. Equality holds only if the integrand is zero almost everywhere.

3.2 Monotonicity in T

Proposition 2 (Monotonicity): For fixed x , $DT(x)$ is monotonically non-decreasing in T : $DT_2(x) \geq DT_1(x)$ for $T_2 \geq T_1$.
 $DT_2(x) \geq DT_1(x)$ for $T_2 \geq T_1$

Proof: For $T_2 \geq T_1$
 $DT_2(x) = \int_0^{T_2} d(\phi_\tau(x), A) d\tau + \int_{T_1}^{T_2} d(\phi_\tau(x), A) d\tau$
 $DT_1(x) = \int_0^{T_1} d(\phi_\tau(x), A) d\tau$
 $DT_2(x) - DT_1(x) = \int_{T_1}^{T_2} d(\phi_\tau(x), A) d\tau$

The second integral is non-negative by Proposition 1. Therefore $DT_2(x) \geq DT_1(x)$.

Corollary: If the trajectory converges exactly to the attractor at time $\tau_0 < T$, then: $DT(x) = D\tau_0(x)$ for all $T \geq \tau_0$.

3.3 Additivity

Proposition 3 (Additivity): For any $T, S \geq 0$: $D_{T+S}(x) = DT(x) + DS(\phi_T(x))$.

Proof: $D_{T+S}(x) = \int_0^{T+S} d(\phi_\tau(x), A) d\tau = \int_0^T d(\phi_\tau(x), A) d\tau + \int_T^{T+S} d(\phi_\tau(x), A) d\tau$
 $= DT(x) + \int_0^S d(\phi_{\tau+T}(x), A) d\tau$ (by the semigroup property) $= DT(x) + DS(\phi_T(x))$

This connects DT naturally to Bellman equations, dynamic programming, and occupation measures.

3.4 Heuristic Connection: Dynamic Programming

The additivity property $D(T+S)(x) = DT(x) + DS(\phi_T(x))$ suggests a natural connection to dynamic programming. For a controlled system $X' = f(X, u)$ with control $u \in U$, the value function $V(x) = \inf_u D_\infty(x)$ would formally satisfy the Hamilton-Jacobi-Bellman equation: $0 = \inf_u \{d(x, A) + \nabla V(x) \cdot f(x, u)\}$

This is a standard result for additive cost functionals. A full derivation for the specific functional DT is left for future work. This section is a heuristic connection, not a formal result.

3.5 Lipschitz Continuity with Respect to Initial Conditions

Proposition 4 (Lipschitz Continuity of DT): Suppose the flow ϕ_τ is Lipschitz continuous in x with constant L , i.e., $\|\phi_\tau(x) - \phi_\tau(y)\| \leq e^{L\tau} \|x - y\|$. Then for any x, y in the basin of A : $\|DT(x) - DT(y)\| \leq \int_0^T e^{L\tau} d\tau \|x - y\| = (e^{LT} - 1)L^{-1} \|x - y\|$

Proof: First, note that the distance function $d(\cdot, A)$ is 1-Lipschitz: for any $x, y \in X$, $\|d(x, A) - d(y, A)\| \leq \|x - y\|$

This follows from the triangle inequality and the definition of the infimum. Then, using the Lipschitz property of the flow: $\|DT(x) - DT(y)\| \leq \int_0^T \|d(\phi_\tau(x), A) - d(\phi_\tau(y), A)\| d\tau \leq \int_0^T \|\phi_\tau(x) - \phi_\tau(y)\| d\tau$

$$\begin{aligned}
 |y) \int_0^T e^{L\tau} |x-y| \, d\tau &= e^{LT} |x-y| \int_0^1 DT(\tau)(x) - DT(\tau)(y) \, d\tau \leq \int_0^T \theta \, d(\phi\tau(x), A) - d(\phi\tau(y), A) \, d\tau \\
 &\leq \int_0^T \theta \, |\phi\tau(x) - \phi\tau(y)| \, d\tau \leq \int_0^T \theta e^{L\tau} |x-y| \, d\tau = Le^{LT} |x-y|
 \end{aligned}$$

Interpretation: This proposition guarantees that empirical estimates of DT are robust under small perturbations of initial conditions and establishes that DT defines a continuous functional on the basin of attraction. This is essential for numerical estimation and experimental measurement.

3.6 Instantaneous Growth Rate

Remark 1 (Instantaneous Growth Rate): If the integrand $d(\phi\tau(x), A)$ is continuous in τ , then: $\frac{d}{d\tau} DT(x) = d(\phi T(x), A)$

This follows directly from the Fundamental Theorem of Calculus.

3.7 Ergodic Limit

Proposition 5 (Ergodic Limit): Suppose the normalized occupation measure $\nu_T = \mu_T / T$ converges weakly to an invariant probability measure μ as $T \rightarrow \infty$. Then: $\lim_{T \rightarrow \infty} \frac{1}{T} DT(x) = \int_X d(y, A) \, d\mu(y)$

Proof: From the occupation measure representation $DT(x) = \int d(y, A) \, d\mu_T(y) = T \int d(y, A) \, d\nu_T(y)$, weak convergence of ν_T to μ and boundedness/continuity of $d(\cdot, A)$ gives the result.

This is the pointwise ergodic theorem applied to the observable $d(\cdot, A)$. For the ergodic theory of dynamical systems, see Bowen (1975) and Ruelle (1989).

3.8 Bound under Exponential Stability

Theorem 2 (Bound under Exponential Stability): Suppose the flow $\phi_\tau(x)$ converges to the attractor A with exponential rate $\kappa > 0$: $d(\phi_\tau(x), A) \leq C e^{-\kappa \tau} d(x, A)$

for some constant $C < \infty$, for all $\tau \geq 0$. Then: $D^\infty(x) = \int_0^\infty d(\phi_\tau(x), A) d\tau \leq C \kappa d(x, A)$ $D^\infty(x) = \int_0^\infty d(\phi_\tau(x), A) d\tau \leq C \kappa d(x, A)$

Proof: $D^\infty(x) = \int_0^\infty d(\phi_\tau(x), A) d\tau \leq \int_0^\infty C e^{-\kappa \tau} d(x, A) d\tau = C d(x, A) \int_0^\infty e^{-\kappa \tau} d\tau = C d(x, A) \frac{1}{\kappa}$

Corollary: For linearly stable systems with recovery rate κ , $D^\infty(x) \leq \frac{1}{\kappa} d(x, A)$ (when $C=1$).

Important: Exponential stability implies $D^\infty < \infty$. The converse is not claimed; polynomial convergence can also yield finite D^∞ .

3.9 Recovery Rate Bound

Corollary 1 (Recovery Rate Bound): For a system satisfying the exponential stability hypothesis with constant C , the recovery rate κ satisfies: $\kappa \leq \frac{D^\infty(x)}{C d(x, A)}$

For systems with $C=1$ (e.g., normal/symmetric linearizations

with no transient overshoot), this reduces to: $\kappa \leq d(x, A) D^\infty(x) \kappa \leq D^\infty(x) d(x, A)$

Proof: From Theorem 2, we have $D^\infty(x) \leq C \kappa d(x, A) D^\infty(x) \leq \kappa C d(x, A)$. Rearranging gives $\kappa \leq C d(x, A) D^\infty(x) \kappa \leq D^\infty(x) C d(x, A)$. When $C=1$, this reduces to $\kappa \leq d(x, A) D^\infty(x) \kappa \leq D^\infty(x) d(x, A)$.

Interpretation: Small cumulative deviation implies rapid recovery (large κ). Large cumulative deviation implies slow recovery (small κ). This formalizes the intuitive link between DT and κ . The C factor accounts for possible transient overshoot in non-normal systems.

3.10 Finite Horizon Approximation

Proposition 6 (Finite Horizon): For any $\epsilon > 0$, there exists a finite T_ϵ such that for all $T > T_\epsilon$: $|DT(x) - D^\infty(x)| \leq \epsilon$

Proof: This follows directly from Theorem 2 under the exponential stability hypothesis. Since the integrand decays exponentially, the tail integral $\int_T^\infty d(\phi_\tau(x), A) d\tau$ can be made arbitrarily small by choosing T sufficiently large.

3.11 Summary of Properties

Property	Statement
Non-negativity	$DT(x) \geq 0$
Monotonicity	$DT_2(x) \geq DT_1(x)$ for $T_2 \geq T_1$
Additivity	$DT+S(x) = DT(x) + DS(\phi_T(x))$

Property	Statement		
Lipschitz continuity	($D_{-T}(x)$ -	$\leq \frac{e^{LT}}{L} x - y $
Instantaneous growth	$d(\phi_T(x), A) = d(\phi_{T-\Delta T}(x), A)$		
Ergodic limit	$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T d(\phi_t(x), A) dt = \int d(y, A) d\mu(y)$		
Exponential stability implies finite D_{∞}	$D_{\infty}(x) \leq C \kappa d(x, A)$		
Recovery bound (general)	$\kappa \leq C d(x, A) D_{\infty}(x) \leq C d(x, A) \kappa$		
Recovery bound (C=1)	$\kappa \leq d(x, A) D_{\infty}(x) \leq \kappa d(x, A)$		
Finite horizon approximation	$D_T(x) \rightarrow D_{\infty}(x)$ as $T \rightarrow \infty$		

4. The Unified Variable Set

The following variables are defined operationally. Where a variable is a proposal, that is stated explicitly.

4.1 Corrective Permeability (κ)

Definition 4 (Corrective Permeability): κ is the recovery rate of the system to its attractor after a small perturbation. Operationally estimated as $\kappa = 1/\tau$ under approximately exponential relaxation, where τ is the characteristic recovery time constant. This coincides with the exponential convergence exponent in the linearized regime and is consistent with the original definition in the attractor framework.

Relationship to DTD \square : From Corollary 1, for a system with initial deviation $d(x,A)d(x,A)$, $\kappa \leq C d(x,A)D^\infty(x)\kappa \leq D^\infty(x)Cd(x,A)\square$.

Note on κ 's status: In this paper, κ is treated as a primitive empirical regime parameter. A stronger theory would derive κ from DTD \square and system geometry; this remains an open direction for future work.

4.2 Drift Rate ($\gamma\gamma$) – A Proposed Distinction

Definition 5 (Drift Rate): We propose the following operational distinction between dynamical regimes, based on the dominant Lyapunov exponent $\lambda_{\max}\square\lambda_{\max}\square$:

Regime	$\lambda_{\max}\square\lambda_{\max}\square$	$\kappa\kappa$	$\gamma\gamma$	Behavior
Stable attractor	$<-0.01<-0.01$	$>0>0$	00	Converges to fixed point
Persistent chaos	$\approx 0 \approx 0$	$\approx 0 \approx 0$	$>0>0$	Wanders without convergence
Full chaos	$>0>0$	undefined	$>0>0$	Diverges

Thresholds: $\lambda_{\max}\square < -0.01$, $\square\lambda_{\max}\square \leq 0.01$, and $\lambda_{\max}\square > 0.01$ (pre-registered, measured in units of 1/epoch). These numerical thresholds are illustrative defaults rather than theoretically privileged constants.

Grounding: This distinction is inspired by the literature on chaos in high-dimensional neural networks (Engelken, Wolf & Abbott, 2023; Sompolinsky, Crisanti & Sommers, 1988; Clark, Abbott & Litwin-Kumar, 2023; Fournier & Urbani, 2023). For the treatment of stochastic and random perturbations, see Arnold (1998).

Falsification: If $\kappa\kappa$ and $\gamma\gamma$ are perfectly correlated (i.e.,

systems with small κ always have small γ), the distinction is not useful.

4.3 Basin Depth (BB) and Persistence Depth (B~B~)

Definition 6a (Basin Depth – Energy Barrier): BB is the energy barrier required to escape the basin, measured as the potential difference between the attractor and the saddle point on the basin boundary:
 $B = V(\text{saddle}) - V(\text{attractor})$
 $B \sim B \sim = V(\text{saddle}) - V(\text{attractor})$

This preserves the original definition from earlier papers.

Definition 6b (Persistence Depth): As a complementary measure, we define:
 $B \sim = \min_{x \in \partial B} \int_{DT} \dot{D}T(x)$

This is the cumulative deviation required to reach the basin boundary. The relationship between BB and $B \sim B \sim$ remains an open mathematical question.

Operational alternative: In practice, the basin boundary may not be well-defined. Estimate BB via the Arrhenius relationship $P_{\text{escape}} \propto e^{-B/T}$, where T is the noise level.

4.4 Reality Alignment (RR)

Definition 7 (Reality Alignment): RR is the expected log predictive likelihood:
 $R = E[\log p(y|X)]$

where $p(y|X)$ is the system's predictive distribution over outcomes y given state X . Higher RR indicates better predictive accuracy. This is a standard measure of predictive

performance; the label “reality alignment” is a philosophical interpretation.

Direction-dependence: The framework interprets RR as potentially direction-dependent: $RA \rightarrow B \neq RB \rightarrow A$ and $RA \rightarrow B \neq RB \rightarrow A$. This captures the asymmetry found in Berglund et al. (2024), where models trained on “A is B” fail to generalize to “B is A.” This interpretation is a framework-level claim.

Note on integration: Among the core variables, RR is the least integrated with the trajectory-based formalism. Unlike κ , BB , and $B \sim B$, which are directly derived from or related to $DTDT$, RR is imported from Bayesian statistics. A more complete theoretical derivation of RR from the same dynamical principles—perhaps as an information-theoretic functional of the occupation measure—remains an open direction for future work.

5. Theoretical Framework

5.1 Relationship Between $DTDT$, P_{topo} , and $E(t)$

Functional	What It Measures	Regime
$DT(x)$	Cumulative deviation from attractor	All systems
$P_{topo}(t)$	Topological feature lifetime	Systems with topological structure
$E(t)$	Rate of topological change	Learning systems

Hypothesis: In learning systems, $DTDT$ and P_{topo} are positively correlated early in learning and negatively

correlated late in learning. Turner & Barak (2023) demonstrate that RNNs develop attractors sequentially during training, which may correspond to phases of topological simplification. This is a testable prediction.

5.2 Relationship Between κ , γ , and $E(t)$

Hypothesis: In a learning system, the topological evolution rate $E(t)$ is monotonically related to κ only if the system is not in persistent chaos: $\partial E/\partial \kappa > 0$ (with E and κ measured on appropriate scales) in convergent regimes. In persistent chaos, $E(t)$ is monotonically related to γ : $\partial E/\partial \gamma > 0$. Correlation analysis provides a statistical test of these monotonicity relationships.

5.3 Adaptive Landscape (Heuristic Note)

The adaptive landscape $V(X, t)$ evolves as: $\dot{V} = g(X, V) - \lambda V + \xi(t)$

For gradient systems with $X' = -\nabla_X V(X)$, and assuming the dynamics remain within the basin where higher-order nonlinearities are negligible, the cumulative deviation functional can be approximated as: $DT(x) \approx \int_0^T \nabla_X V(\phi_\tau(x), \tau) d\tau$

This is a local heuristic. A full derivation and integration into the core formalism is left for future work.

6. Testable Predictions

6.1 Core Prediction

Prediction: In a learning system, $E(t)$ is monotonically related to κ in convergent regimes: $\partial E/\partial \kappa > 0$ (with E and κ measured on appropriate scales), and $\partial E/\partial \gamma > 0$ in persistent chaos. Correlation analysis provides a statistical test of this monotonicity: $\text{Corr}(E(t), \kappa) > 0 \implies \lambda_{\max} < 0$, $\text{Corr}(E(t), \gamma) > 0 \implies \lambda_{\max} \approx 0$.

Falsification: If $E(t)$ correlates with κ in all regimes, or with γ in all regimes, the prediction is falsified.

6.2 Secondary Prediction

Prediction: In systems with high RR , $DTDT$ and P_{topo} are negatively correlated late in learning; in systems with low RR , they are uncorrelated or positively correlated.

Falsification: If $DTDT$ and P_{topo} are negatively correlated in both high- R and low- R systems, the prediction is falsified.

6.3 Boundary Condition and Global Falsifier

Conjecture: We conjecture that the framework applies to any system satisfying:

- A. Well-defined state space.
- B. Subject to perturbations.

- C. Exhibits at least one identifiable attractor.
- D. Dynamics are observable and measurable.

Global Falsifier: The unified ontology claim collapses if a system is found where $DTDT$, κ , and topological persistence are mutually independent across all regimes, and where RR cannot be expressed as a functional of the trajectory or occupation measure. If such a system exists, the framework's claim to unify persistence, stability, and reality alignment would be falsified.

7. Experimental Design

7.1 System Choice

Train a CNN on MNIST or CIFAR-10. Use latent activation manifolds for topological analysis.

Justification: Karuppiah, Nazreen Banu et al. (2026) demonstrate the use of persistent homology on activations to study feature learning and generalization. Turner & Barak (2023) show that RNNs develop attractors sequentially, providing a controlled setting for studying topological evolution during learning.

7.2 Variable Measurement

Variable	Protocol
$DT(x)DT(x)$	Sample weights; compute distance to final attractor; integrate.
$P_{topo}(t)P_{topo}(t)$	Compute persistent homology on latent activations; sum feature lifetimes.
$E(t)E(t)$	Finite differences of $P_{topo}(t)P_{topo}(t)$.

Variable	Protocol
κ	Perturb weights; measure recovery time τ ; $\kappa=1/\tau$.
γ	Compute average drift rate during training.
RR	Cross-domain generalization accuracy.

7.3 Statistical Analysis

- Correlate $E(t)E(t)$ with κ and γ conditional on regime.
- Pre-register thresholds and sample size.

Note on future empirical work: A full empirical validation would require pre-registration with specified sample size, significance thresholds, power analysis, and robustness checks. These are planned for subsequent work.

8. Discussion

8.1 Implications

The paper provides a candidate formalization with defined variables, mathematical properties, and testable predictions. The mathematical properties of $DTDT$ establish its relationship to κ and provide a foundation for the framework's core claims.

8.2 Limitations

- P_{topo} is computationally expensive.
- The framework is a meta-theory, not a complete domain-specific theory.
- Variables may be confounded; causal inference requires

controlled experiments.

- The κ/γ regime distinction is proposed and requires empirical validation.

8.3 Future Work

- Empirical validation of predictions.
 - Formal derivation of relationships from first principles.
 - Extension to other domains.
 - Computational efficiency improvements.
-

9. Conclusion

This paper proposes a candidate formalization for the attractor framework. The central mathematical innovation is treating persistence as a functional defined over trajectories— $DT(x) = \int_0^T d(\phi_\tau(x), A) d\tau$ —rather than as a scalar property of states. We defined the cumulative deviation functional DT , the topological persistence functional P_{topo} , and the topological evolution rate $E(t)$. We proved several mathematical properties of DT , including non-negativity, monotonicity, additivity, Lipschitz continuity, and a bound relating D^∞ to κ : $D^\infty(x) \leq C\kappa d(x, A)$. We established connections to dynamic programming and ergodic theory. We unified the variable set with operational definitions. We derived testable predictions and provided a falsifiable experimental protocol.

The framework now admits formal definitions, operational variables, and empirical tests. The next step is empirical validation.

Appendix A: Possible Extensions from Larose (2025) – Unverified Source

Note: The following source has not been independently verified. It is included for completeness and as a potential direction for future exploration, but should not be treated as established.

Larose (2025) develops a framework for recursive deformation systems. Two constructs are potentially relevant:

Constraint

Functional: $C(X) = \int_{\text{trajectory}} \|\nabla\Phi\| d\tau$, measuring cumulative irreversible deformation.

Persistence Invariant: $I_p = \int_{\mathbb{R}} d\Phi I_p = \int R d\Phi$, a topological invariant.

These are not yet integrated into the core framework and are presented here for completeness and future exploration. They should be treated as unverified candidate extensions.

References

Arnold, L. (1998). *Random Dynamical Systems*. Springer.

Berglund, L., et al. (2024). "The Reversal Curse: LLMs Trained on 'A is B' Fail to Learn 'B is A'." *arXiv:2309.12288*.

Bowen, R. (1975). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*. Springer.

- Carlsson, G. (2009). "Topology and data." *Bulletin of the American Mathematical Society*, 46(2), 255-308.
- Carlsson, G., & Zomorodian, A. (2009). "The theory of multidimensional persistence." *Discrete & Computational Geometry*, 42(1), 71-93.
- Clark, D. G., Abbott, L. F., & Litwin-Kumar, A. (2023). "Dimension of activity in random neural networks." *Physical Review Letters*, 131, 118401.
- Edelsbrunner, H., & Harer, J. (2010). *Computational Topology: An Introduction*. American Mathematical Society.
- Engelken, R., Wolf, F., & Abbott, L. F. (2023). "Lyapunov spectra of chaotic recurrent neural networks." *Physical Review Research*, 5, 043044.
- Fournier, S. J., & Urbani, P. (2023). "Statistical physics of learning in high-dimensional chaotic systems." *Journal of Statistical Mechanics: Theory and Experiment*, 2023(11), 113301.
- Karuppiah, K., Nazreen Banu, M., et al. (2026). "Topological Data Analysis (TDA) as a Framework for Understanding Deep Learning Behavior." *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)*, Indore, India, December 12-13, 2025. IEEE Xplore. DOI: 10.1109/ICTBIG68706.2025.11323998.
- Larose, H. (2025). "A Mathematical Theory of Frame-Independent Persistence." [Academia.edu](https://www.academia.edu). [Unverified source.]
- Ruelle, D. (1989). *Chaotic Evolution and Strange Attractors*. Cambridge University Press.
- Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). "Chaos in Random Neural Networks." *Physical Review Letters*, 61(3), 259-262.

Turner, E., & Barak, O. (2023). "The Simplicity Bias in Multi-Task RNNs: Shared Attractors, Reuse of Dynamics, and Geometric Representation." *Advances in Neural Information Processing Systems (NeurIPS)*.

Suggested citation: Galida, R. S. (2026). The Persistence Functional: A Candidate Formal Foundation for the Attractor Framework (Foundational Edition). *Fantasy Attractor*.