FRC 840.101: The Phase–Attention Boundary

Author: Hadi Servat

Affiliation: Independent Researcher

Contact: publish@fractalresonance.com

Website: fractalresonance.com

DOI: 10.5281/zenodo.20416012 · **Version:** v2.0
Permanent record: https://doi.org/10.5281/zenodo.20416012 · PDF: https://zenodo.org/records/20416012

Abstract

This paper formulates the Phase–Attention Boundary: a structural separation between continuous phase-state architectures and discrete attention-based architectures. Within the Fractal Resonance Cognition (FRC) program, the Large Lambda-Tensor Model (LLTM) was developed as a continuous recurrent phase-coupled architecture inspired by Kuramoto dynamics and low-rank coherence fields. Controlled comparisons against Transformer baselines revealed a fundamental limitation: continuous state compression blends historical information into a finite evolving state, producing recall smearing. We prove a formal Recall Smearing Theorem: under gamma-contractive recurrence, mutual information about a past token decays exponentially with distance. This bound is derived from the Data Processing Inequality and

> *applies universally to fixed-state recurrent systems, including state-space models like S4, Mamba, RWKV, Griffin, and xLSTM. We show that data-dependent selectivity can reduce the rate of smearing but cannot eliminate it; only explicit key-value addressability achieves zero-smearing recall.*

1. Purpose and Scope

FRC 840.101 is a boundary paper. It does not claim that continuous phase-state models are universally inferior to attention, nor that attention is universally sufficient for cognition. Its purpose is narrower:

> ***To identify the mathematical and architectural boundary between continuous state compression and discrete addressable retrieval.***

The paper emerges from the Large Lambda-Tensor Model (LLTM) experimental sequence in which a phase-coupled recurrent architecture was developed, scaled, benchmarked, and compared against Transformer baselines. The negative result in language modeling is the central scientific contribution: the LLTM did not close the validation-loss gap with attention-based Transformers when the task demanded exact symbolic recall. This failure revealed a deeper principle.

In FRC language, the boundary can be summarized as follows:

Continuous Phase-State Cognition	**Discrete Attention Cognition**
Coherence, smoothing, memory as state, attractor dynamics, analog continuity.	Routing, retrieval, memory as addressable tokens, symbolic precision.

The claim is structural, not merely empirical. Continuous systems compress history into the present. Attention systems keep history explicitly addressable.

---

2. Background: LLTM as a Phase-State Architecture

The Large Lambda-Tensor Model (LLTM) was designed as a continuous phase-coupled alternative to attention. In its simplest form, a token sequence is embedded into a recurrent oscillator state. The latent state is represented by phases:

\[\bm{\theta}_t = (\theta_{1,t},\theta_{2,t},\ldots,\theta_{N,t})\]

or equivalently by complex phase variables:

$$z_{i,t}=e^{i\theta_{i,t}}$$

A simplified phase update takes the form:

$$\theta_{i,t+1} = \theta_{i,t} + \Delta t \left[\omega_i(x_t) + \frac{K}{N} \sum_{j=1}^{N} A_{ij} \sin(\theta_{j,t}-\theta_{i,t}) \right]$$

where x_t is the current input, $\omega_i(x_t)$ is an input-dependent driving frequency, K is coupling strength, and A_{ij} is the phase-coupling matrix.

The global coherence of the state is measured by the Kuramoto order parameter:

$$R_t e^{i\psi_t} = \frac{1}{N} \sum_{j=1}^{N} e^{i\theta_{j,t}}$$

where $R_t \in [0,1]$ measures phase alignment and ψ_t is the mean phase.

The LLTM hypothesis was that a sufficiently expressive phase-state network could compress sequence history into a coherent dynamical state while avoiding the quadratic context cost of attention.

---

3. The Low-Rank Scaling Insight

A dense coupling matrix $A \in \mathbb{R}^{N\times N}$ scales quadratically in the hidden dimension. The low-rank LLTM variant factorizes coupling as:

$$A = UV^\top$$

where $U,V \in \mathbb{R}^{N\times r}$ and $r \ll N$. This changes the interaction parameterization from $\mathcal{O}(N^2)$ to $\mathcal{O}(Nr)$.

In the LLTM experiments, low-rank coupling improved stability and often outperformed dense coupling at comparable or lower parameter counts. This was an important positive result: it showed that phase-coupled systems do not require full dense all-to-all interaction to express useful structure. Low-rank channels act as global resonance bottlenecks.

This supports an FRC architectural principle:

> ***Coherence can be mediated through low-dimensional coupling channels rather than exhaustive dense interaction.***

3.1 The Hierarchy Illusion: Physical Space vs. Mathematical Space

A crucial design question arose during scaling: Should the phase coupling be organized into biological-like nested hierarchies, or should it remain flat?

In biological brains, cognition is famously hierarchical and modular. Local clusters of neurons synchronize at fast frequencies (e.g., gamma waves), and these clusters are globally coordinated by slower macroscopic frequencies (e.g., theta waves).

We hypothesized that explicitly engineering this biological, nested hierarchy into the LLTM would improve language modeling by allowing different blocks of oscillators to specialize and communicate through macroscopic "order parameters." We designed the **Hierarchical LLTM**, which divided oscillators into isolated blocks that could only communicate by emitting a single average phase (Φ_m) to a global routing network.

We tested this against a completely unconstrained, flat **Low-Rank LLTM**, where every oscillator could project its individual state into a global r-dimensional bottleneck.

The Hierarchy Illusion

Figure 1: The Hierarchy Illusion. Flat Low-Rank coupling (left) vs. Hierarchical block coupling (right). The flat model projects all oscillators into shared low-rank channels, while the hierarchical model forces communication through a single average phase per block.

The Hierarchy Shootout Results

The flat `LowRank` model absolutely crushed the hierarchical models in both validation loss and generation speed.

Model	Params	Best Val Loss	Gen Speed (2048 ctx)
Flat LowRank ($r=8$)	216,129	**2.2778**	**166 tok/s**
Hierarchical ($M=8$)	208,001	2.3703	74 tok/s
Hierarchical ($M=16$)	208,193	2.4228	—

Why? Biological brains use nested modular hierarchy because they are constrained by 3D physical space and metabolic wiring costs (white matter tracts). A GPU has no such physical constraints. In the Hierarchical model, 128 continuous phase variables were mathematically forced to compress their entire state into a *single* average angle (Φ_m) to communicate. This created a severe, rigid architectural bottleneck.

When spatial wiring costs are removed, a soft mathematical bottleneck (Flat Low-Rank factorization) is functionally and representationally superior to a hard architectural bottleneck (Hierarchical Blocks). **Flat low-rank is the optimal topology for simulated physical systems.**

However, low-rank scaling did not remove the recall boundary. It improved global state coherence, but it did not grant exact addressability of arbitrary past tokens.

---

4. The Recall Smearing Constraint

Let a continuous recurrent model update its hidden state by:

$$\bm{h}_{t+1}=F_\theta(\bm{h}_t,x_t)$$

After a sequence $x_{1:T}$, the state is:

$$\bm{h}_T = F_\theta(F_\theta(\cdots F_\theta(\bm{h}_0,x_1), x_2),\ldots,x_T)$$

All past inputs are represented only through the current state $\bm{h}_T$. This is efficient, but it creates an irreversible compression unless the state dimension grows with sequence length or the transition is explicitly invertible with sufficient capacity.

We define **recall smearing** as:

> ***The degradation of recoverable information about a specific past input caused by continuous blending of all prior inputs into a finite evolving state.***

Formally, let $I(x_\tau;\bm{h}_T)$ denote the mutual information between a past token x_τ and the current state $\bm{h}_T$, with $\tau < T$. Under ordinary noisy, dissipative, or contractive recurrence, one expects:

$$I(x_\tau;\bm{h}_T) \leq I(x_\tau;\bm{h}_{T-1}) \leq \cdots \leq I(x_\tau;\bm{h}_{\tau})$$

with degradation increasing as $T-\tau$ grows.

This is not a flaw for all domains. In many continuous systems, smearing is desirable. It filters noise, integrates momentum, and tracks regime state. But in language and code, where the system may need a specific symbol from many steps ago, smearing is harmful.

4.1 Formal Recall Smearing Theorem

We now formalize the decay bound. The key tools are the *Data Processing Inequality* (DPI) and contractivity of the recurrent map.

Definition (γ-Contractivity) A parametric transition $F_\theta:\mathbb{R}^d\times\mathcal{X}\to\mathbb{R}^d$ is *γ-contractive* (with $0<\gamma<1$) if for every input $x\in\mathcal{X}$ and every pair of states $\bm{h},\bm{h}'\in\mathbb{R}^d$:

$$\|F_\theta(\bm{h},x) - F_\theta(\bm{h}',x)\| \leq \gamma\,\|\bm{h}-\bm{h}'\|$$

Contractivity implies that the recurrent map is a *noisy channel* with bounded capacity per step. Each application of F_θ forms a Markov kernel:

$$x_\tau \to \bm{h}_t \to \bm{h}_{t+1} \to \cdots \to \bm{h}_T$$

Theorem (Recall Smearing Bound) Let F_θ be a γ-contractive recurrent transition with $0<\gamma<1$. Then for any past input x_τ at distance $T-\tau$ from the current state:

$$I(x_\tau;\bm{h}_T) \leq \gamma^{2(T-\tau)}\cdot I(x_\tau;\bm{h}_\tau)$$

Proof For each step $t>\tau$, the state $\bm{h}_t$ is obtained from $\bm{h}_{t-1}$ and input x_t via $\bm{h}_t=F_\theta(\bm{h}_{t-1},x_t)$. The triple $(x_\tau, \bm{h}_{t-1}, \bm{h}_t)$ forms a Markov chain $x_\tau \to \bm{h}_{t-1} \to \bm{h}_t$ conditioned on x_t, because $\bm{h}_t$ depends on x_τ only through $\bm{h}_{t-1}$.

By the *Data Processing Inequality*:

$$I(x_\tau;\bm{h}_t) \leq I(x_\tau;\bm{h}_{t-1})$$

This establishes monotonic non-increasing decay. For the exponential rate, we use the fact that γ-contractivity bounds the Fisher information of the channel. Specifically, if $F_\theta(\cdot,x_t)$ is γ-contractive, then the conditional distribution $p(\bm{h}_t \mid \bm{h}_{t-1},x_t)$ satisfies a strong data processing inequality (SDPI):

$$I(x_\tau;\bm{h}_t\mid x_t) \leq \gamma^2\, I(x_\tau;\bm{h}_{t-1})$$

Since x_τ and x_t are independent for $t>\tau$ (in a standard autoregressive setup where inputs are drawn sequentially), we have $I(x_\tau;\bm{h}_t) \leq I(x_\tau;\bm{h}_t\mid x_t) + I(x_\tau;x_t)$, and the second term vanishes by independence. Therefore:

$$I(x_\tau;\bm{h}_t) \leq \gamma^2\, I(x_\tau;\bm{h}_{t-1})$$

Iterating from τ to T:

$$I(x_\tau;\bm{h}_T) \leq \gamma^2\,I(x_\tau;\bm{h}_{T-1}) \leq \gamma^4\,I(x_\tau;\bm{h}_{T-2}) \leq \cdots \leq \gamma^{2(T-\tau)}\,I(x_\tau;\bm{h}_\tau)$$

$\square$

Corollary (Exponential Recall Horizon) For a γ-contractive recurrence, the effective recall horizon—the maximum distance $T-\tau$ at which a past token can be recovered above a threshold $I_{\min}$—is bounded by:

\[T-\tau \leq \frac{\ln(I(x_\tau;\bm{h}_\tau)/I_{\min})}{2|\ln\gamma|}\]

Beyond this horizon, the past token's information content in the state falls below the recovery threshold.

Note on non-contractive dynamics: The theorem assumes strict contractivity ($\gamma < 1$). Hamiltonian or volume-preserving dynamics ($\gamma = 1$) do not exhibit exponential decay, but the DPI still guarantees monotonic non-increase. In practice, all trained neural recurrences include nonlinearities, normalization, and noise that break strict measure preservation, making $\gamma < 1$ the generic case.

---

5. Attention as Addressable Memory

In a Transformer, past states are not only compressed into the current state. They remain addressable through key–value pairs:

\[K = (k_1,\ldots,k_T),\qquad V=(v_1,\ldots,v_T)\]

A query q_t retrieves a weighted combination of previous values:

\[y_t = \sum_{\tau=1}^{T} \mathrm{softmax} \left(\frac{q_t k_\tau^\top}{\sqrt{d}} \right) v_\tau\]

The crucial difference is that token τ remains externally addressable. The model can retrieve a specific past representation by similarity search rather than relying only on its compressed effect on the current state.

Thus attention implements a different memory geometry:

Continuous Recurrence	**Attention**
History is folded into current state.	History remains addressable as memory slots.
Retrieval is indirect and degraded by later inputs.	Retrieval is direct through query–key matching.
Memory cost can be $\mathcal{O}(1)$ in sequence length.	Memory cost grows with sequence length.
Subject to Theorem 1 decay.	No smearing; $I(x_\tau;y_t)$ depends only on routing accuracy.

5.5 Selective State Spaces and the Contractivity Spectrum

The recall smearing bound (Theorem 1) applies to all fixed-state recurrent systems. However, modern state-space models occupy different positions along the contractivity spectrum:

- **S4** [Gu et al., 2021]: Uses a fixed, data-independent state matrix A. The contraction rate $ \gamma$ is determined by the eigenvalues of A and does not adapt to the input. S4 exhibits *fixed-rate* smearing: all tokens fade at the same exponential rate regardless of content.
- **Mamba** [Gu & Dao, 2023]: Introduces *data-dependent* selectivity where matrices $A(x_t)$, $B(x_t)$, $C(x_t)$ are functions of the current input. When a salient token arrives, the gate can widen (reducing γ locally) to preserve it. Non-informative tokens trigger narrowing (increasing γ) to allow faster forgetting. This yields *adaptive-rate* smearing. However, the state $\bm{h}_t \in \mathbb{R}^d$ remains finite-dimensional and the DPI chain still applies; Mamba reduces the *rate* of smearing but cannot eliminate it.
- **RWKV** [Peng et al., 2023]: Uses exponential time-decay with per-channel decay rates plus a token-shift mechanism, displaying fixed-rate exponential smearing with learnable per-channel rates.
- **Griffin** [De et al., 2024]: Combines gated linear recurrence with *local* attention windows. Within each window, exact recall is available. Across windows, recurrent smearing applies.
- **xLSTM** [Beck et al., 2024]: Introduces exponential gating and memory mixing cells, yielding adaptive-rate smearing analogous to Mamba.
- **LLTM (this work)**: Kuramoto phase coupling with fixed coupling strength K and coupling matrix A_{ij}. The contraction rate is set by the physics, yielding fixed-rate smearing.

Architecture	State Update	Smearing Rate	Escape?
S4	Fixed A, linear scan	Fixed γ	No
RWKV	Exp. decay + token shift	Fixed per channel	No
LLTM	Kuramoto phase coupling	Fixed (coupling K)	No
Mamba	Data-dependent $A(x_t)$, $B(x_t)$	Adaptive $\gamma(x_t)$	Partial
xLSTM	Exp. gating + memory mix	Adaptive	Partial
Griffin	Gated recurrence + local attn.	Mixed	Partial*
Transformer	KV cache + global attention	**None**	**Yes**

\Griffin escapes within local attention windows but not across them.*

The key insight is that the Phase–Attention Boundary is not a binary divide but a *spectrum*. Recurrent architectures trade recall precision for computational efficiency ($\mathcal{O}(1)$ per

step), whereas Transformers trade computational cost ($\mathcal{O}(T)$ memory) for exact addressability.

---

6. The Phase–Attention Boundary

We define the **Phase–Attention Boundary** as the point at which a task's demand for exact addressable recall exceeds the capacity of continuous state compression.

Let $\mathcal{D}_{\mathrm{recall}}$ denote the task's discrete recall demand and $\mathcal{C}_{\mathrm{state}}$ denote the recoverable capacity of the continuous state. A continuous phase-state model is expected to perform well when:

$$\mathcal{D}_{\mathrm{recall}} \ll \mathcal{C}_{\mathrm{state}}$$

or when exact recall is unnecessary and a smoothed state is sufficient.

Attention becomes structurally advantageous when:

$$\mathcal{D}_{\mathrm{recall}} \gtrsim \mathcal{C}_{\mathrm{state}}$$

This boundary is not simply about sequence length. It is about the nature of the information that must be preserved:

- If the task requires trend, rhythm, phase, momentum, or regime state, continuous recurrence may be ideal.
- If the task requires exact retrieval of a previous token, variable, quote, syntax marker, or code dependency, attention is superior.

---

7. Empirical LLTM Findings

The LLTM experiments were conducted on the Tiny Shakespeare dataset (character-level tokenization) using an Apple M2 GPU (MPS backend) with AdamW optimization and learning rate scheduling over 1,500 training iterations. We report results from two controlled shootouts.

7.1 The Deep Residual Shootout: Phase vs. Attention

Having established Flat Low-Rank as the definitive physical architecture, we posed the ultimate question: *Can a deep continuous state-space model (LLTM) close the validation-loss gap with a Transformer?*

We gave the LLTM the exact same structural advantages as a modern GPT architecture: Depth (4 layers), LayerNorms, and alternating Multi-Layer Perceptrons (MLPs) to inject massive raw memorization capacity. We precisely tuned the rank so that the Deep LLTM had ~2.93 million parameters, roughly matching the 4-layer Transformer baseline.

The Phase-Attention Boundary

Figure 2: The Phase-Attention Boundary. Validation loss over 1,500 training iterations for the parameter-matched 4-layer Deep LLTM ($\sim$2.93M params) vs. the 4-layer Transformer ($ \sim$3.71M params).

Deep Residual Shootout Results

Model	Layers	Params	Val Loss	Train Loss	Gen
Transformer (4H, $d=256$)	4	3,714,113	**1.5244**	**1.23**	**107 tok/s**
Deep LLTM (Dense, $d=256$)	4	2,927,681	2.1693	2.05	75 tok/s

The Transformer achieved a massive drop in validation loss (`1.52`), proving that its depth and MLPs were utilized perfectly. The Deep LLTM only dropped to `2.16`, significantly underfitting the training data.

This failure is not an implementation bug; it reveals a fundamental mathematical constraint of physics-based continuous state systems, leading to the **Recall Smearing** problem described in Section 4.

7.2 Core Findings

1. **Phase models can learn local sequence statistics:** Tiny LLTM models trained on character-level language data learned basic structure (spaces, pseudo-words, and dialogue-like formatting), demonstrating that phase-coupled recurrence can carry trainable sequence information.
2. **Low-rank coupling is viable:** Low-rank coupling improved scalability and regularized the phase dynamics. Dense coupling often overfit or destabilized, while low-rank coupling forced the system into useful shared channels.
3. **Phase recurrence has strong cached inference dynamics:** Because LLTM carries state recurrently, generation proceeds without reprocessing the entire context. This gives strong speed and memory advantages in the small-model regime and suggests value for streaming applications.
4. **LLTM does not close the language gap with attention:** When depth, residual structure, and parameter-matched baselines were introduced, Transformers retained a significant validation-

loss advantage on language. The attention mechanism is structurally better suited to exact symbolic retrieval.

---

8. Domain Taxonomy

The boundary suggests the following taxonomy.

Domain	Better suited to phase-state models	Better suited to attention
Language	Prosody, rhythm, discourse mood, long-term style	Exact token recall, syntax, code, quotation
Markets	Volatility regime, phase transition, smooth risk state	Discrete event/news retrieval, exact historical lookup
Robotics	Sensorimotor flow, balance, control dynamics	Episodic memory, task instructions
Biology	Oscillations, morphogenesis, physiological rhythms	Dynamic annotation, protocol recall
Audio	Wave phase, timbre, continuous dynamics	Lyrics, transcript alignment
Weather	Regimes, fields, continuous dynamics	Event metadata, discrete records

---

9. Connection to FRC 566.001 and FRC 100.009

[[FRC-566-001]] introduces entropy–coherence reciprocity:

\[dS+k^*d\ln C=0\]

[[FRC-100-009]] localizes this relationship into open-system field dynamics by introducing reciprocity storage, flux, defect, and activation. In that context, coherent activation emerges where local reciprocity permits transport and transformation.

FRC 840.101 adds an architectural boundary:

> ***Coherence trades detail for unity.***

A continuous state can increase global coherence by integrating many inputs into one field. But the same operation reduces separability of micro-detail. This is the architectural expression of the entropy–coherence trade-off.

In language modeling, too much unity harms exact recall. In field systems, unity is often the point.

---

10. Implications for Hybrid Architectures

The Phase–Attention Boundary implies that future cognitive architectures should not force one memory geometry to handle all tasks.

A practical hybrid architecture should include:

1. **Attention memory** for addressable symbolic retrieval.
2. **Phase-state recurrence** for continuous coherence, streaming state, and regime tracking.
3. **Gating mechanisms** to switch between discrete recall and continuous smoothing.
4. **Low-rank coherence channels** for efficient global coupling.
5. **External memory or retrieval** when the task requires exact records.

A minimal hybrid update can be written as:

$$\bm{y}_t = g_t \, \mathrm{Attention}(q_t,K,V) + (1-g_t)\, \mathrm{PhaseState}(\bm{h}_{t-1},x_t)$$

where $g_t \in [0,1]$ is a learned or rule-based routing gate.

- When exact recall is required, $g_t \to 1$.
- When continuous field tracking is required, $g_t \to 0$.

---

11. Interpretation: Failure as Boundary Discovery

The LLTM language results should be treated as a positive scientific outcome because they prevented overextension of the FRC architecture.

A weak research program attempts to win everywhere. A strong research program identifies its boundary conditions.

FRC 840.101 therefore establishes a key design law:

> ***Continuous phase-state models are not replacements for attention. They are complementary engines for domains where coherence, smoothing, and dynamics matter more than exact symbolic recall.***

This protects future FRC work from inflated claims and points development toward the correct domains: markets as field dynamics, biological morphogenesis, sensorimotor control, audio, EEG, climate, plasma, robotics, and streaming world models.

---

12. Limitations

This paper establishes an architectural boundary through formal analysis and small-scale experiments. Several limitations remain:

- The LLTM experiments were conducted at small scale ($\sim$3M parameters on Tiny Shakespeare) and should be reproduced with larger models and standard benchmarks (WikiText-103, The Pile).
- Theorem 1 provides exponential decay bounds under strict contractivity. The bound may be tighter or looser depending on the spectral radius of the Jacobian and data distribution. Extensions to non-contractive or measure-preserving dynamics remain open.
- The contractivity spectrum (Table 4) classifies existing architectures qualitatively. Precise measurement of effective γ values for trained Mamba and xLSTM models would strengthen the classification.
- Hardware-level optimization may change practical performance but does not remove the fundamental memory-geometry distinction.
- Some continuous systems can be made more addressable through external memory, gating, or state augmentation, partially circumventing the bound.

These limitations do not weaken the boundary; they define the next research program.

---

13. Predictions and Testable Consequences

The Phase–Attention Boundary predicts:

1. On tasks requiring exact copy or retrieval from long context, attention-based models will outperform fixed-state phase recurrence.
2. On noisy continuous regime tasks, phase-state models may outperform attention when smoothing and coherence tracking are more important than discrete retrieval.

3. Hybrid models should outperform either pure mechanism on mixed tasks.
4. Increasing continuous state dimension improves capacity but does not fully solve addressability unless memory grows with sequence length or becomes externally indexed.
5. Low-rank phase coupling will improve generalization in continuous domains by forcing global coherence channels.
6. Mamba-style adaptive gating will close but not eliminate the recall gap with attention, with the residual gap growing logarithmically with context length.

---

14. Conclusion

FRC 840.101 defines the Phase–Attention Boundary. The LLTM experiments showed that continuous phase-state architectures can learn, scale through low-rank coupling, and generate coherent local structure. They also showed that such architectures cannot match attention on tasks requiring exact symbolic recall.

We formalized this observation as the Recall Smearing Theorem: under γ-contractive recurrence, mutual information about a past token decays exponentially at rate $\gamma^{2(T-\tau)}$. This bound applies across the spectrum of recurrent architectures, from fixed-rate models (S4, RWKV, LLTM) to adaptive-rate models (Mamba, xLSTM), and is escaped only by explicit key–value addressability (Attention).

This is not a failure of FRC. It is a sharpening of FRC.

The boundary can be stated simply:

> ***Attention is for addressable symbolic memory. Phase-state recurrence is for continuous coherence fields.***

The next generation of architectures should therefore combine both. FRC does not require one universal mechanism. It requires that each domain use the memory geometry appropriate to its structure.

- Continuous cognition belongs where the world behaves like a field.
- Attention belongs where the world behaves like an archive.

Appendix A: Minimal Recall Smearing Toy Task

A simple toy task can illustrate the boundary. Let a model receive a sequence $x_1,x_2,\ldots,x_T$ and be asked to output x_τ for a randomly chosen $\tau < T$.

- A continuous recurrent model must recover x_τ from $\bm{h}_T$, where $\bm{h}_T = F_\theta(\bm{h}_{T-1},x_T)$. If $T-\tau$ is large and the state is finite-dimensional, recovery degrades.
- An attention model can form a query to the stored key k_τ and retrieve v_τ. Therefore accuracy should remain higher as $T-\tau$ grows, provided memory is retained.

The expected result is:

Model Type	Expected Behavior
Continuous recurrence	Accuracy decays with distance due to smearing.
Attention	Accuracy remains stable when key–value routing succeeds.
Hybrid	Uses recurrence for state and attention for exact recall.

References

1. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*.
2. Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. Springer.
3. Strogatz, S. H. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143(1–4), 1–20.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
5. Gu, A., Goel, K., \& R\'e, C. (2021). Efficiently Modeling Long Sequences with Structured State Spaces. *International Conference on Learning Representations*.
6. Gu, A., \& Dao, T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752*.
7. Peng, B., Alcaide, E., Anthony, Q., et al. (2023). RWKV: Reinventing RNNs for the Transformer Era. *arXiv preprint arXiv:2305.13048*.
8. Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory*. 2nd ed. Wiley-Interscience.
9. Raginsky, M. (2016). Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6), 3355–3389.

10. De, S., Smith, S. L., Fernando, A., et al. (2024). Griffin: Mixing Gated Linear Recurrences with Local Attention for Efficient Language Models. *arXiv preprint arXiv:2402.19427*.
11. Beck, M., P\"ock, K., Eidinger, H., et al. (2024). xLSTM: Extended Long Short-Term Memory. *arXiv preprint arXiv:2405.04517*.
12. Lieber, O., Lenz, B., Bata, H., et al. (2024). Jamba: A Hybrid Transformer-Mamba Language Model. *arXiv preprint arXiv:2403.19887*.
13. Servat, H. (2025). FRC 566.001: Entropy–Coherence Reciprocity. *Fractal Resonance Cognition Research Series*.
14. Servat, H. (2026). FRC 100.009: Local Reciprocity Dynamics: Open-System Activation, Adaptive Resonance, and Self-Starting Coherence Fields. *Fractal Resonance Cognition Research Series*. Zenodo DOI: 10.5281/zenodo.20328893.
15. Servat, H. (2025). FRC 100.007: The Lambda-Field: Scalar Completion of the FRC Framework. *Fractal Resonance Cognition Research Series*.