Dissipative Hamiltonian Memory: A Physics-Informed Architecture for Stable Long-Context Sequence Modeling

Yueyi Wang (University of Cambridge)
yw562@cam.ac.uk

August 27, 2025

Abstract

Long-context modeling remains a central challenge in machine learning. Existing architectures—Transformers, RNNs, and modern State Space Models (SSMs)—each suffer from trade-offs between computational cost, stability, and interpretability. We introduce the **Hamiltonian Memory Layer with Dissipation (HML-D)**, a physics-informed recurrent architecture grounded in Hamiltonian dynamics with dissipation. By representing hidden states as points in phase space (q, p) and evolving them via a learnable Hamiltonian plus dissipative terms, HML-D provides *long-term gradient stability, intrinsic interpretability, and robustness against distributional shifts*.

We validate HML-D in both synthetic long-context tasks and financial time-series prediction, where chaotic dynamics and distribution shifts pose severe challenges. Even if predictive accuracy matches baselines, HML-D's *stability + interpretability* deliver unique value, positioning it as a promising direction for long-context and safety-critical AI systems.

1 Motivation

Why now?

- LLMs demand 100k+ token context, yet Transformers scale quadratically and lose robustness in ultra-long settings.
- RNNs offer linear scaling but collapse due to vanishing/exploding gradients.
- SSMs (e.g., S4, Mamba) improve scaling but rely on crafted transition matrices; long-horizon stability remains fragile.

Opportunity: Hamiltonian mechanics offers a natural inductive bias:

- **Energy conservation** → prevents uncontrolled drift.
- **Symplectic integration** → ensures stability over arbitrarily long trajectories.
- **Dissipation terms** → allow realistic memory decay.

We hypothesize that embedding this structure into sequence models yields a *stable, interpretable, and efficient memory primitive* for both AI and financial time-series.

2 Related Work

- **Long-context models:** Transformer-XL, Hyena, RetNet, Mamba (efficient memory mechanisms, but black-box).
- **Physics-inspired models:** Hamiltonian Neural Networks, Symplectic RNNs, Lagrangian NNs (applied in physics, not NLP/finance).

- **AI Safety & Interpretability:** Current methods audit black-box LLMs; our approach introduces *interpretability-by-design*.

Gap: No prior work has used Hamiltonian dynamics as a general-purpose memory layer for arbitrary sequence modeling.

3 Research Questions

Performance & Capability

1. Can HML-D replace attention/SSM blocks in sequence tasks while maintaining efficiency?
2. Does enforcing Hamiltonian + dissipative structure improve gradient stability at horizons of 10^5+ tokens?
3. Can recurrent updates be parallelized (scan/convolutional formulation)?

AI Safety & Interpretability

1. Do phase-space trajectories and energy diagnostics provide meaningful interpretability?
2. Does Hamiltonian structure improve robustness under adversarial perturbations and distribution shifts?
3. Can interpretable dynamics correlate with prediction confidence or factual consistency?

4 Methodology: HML-D Architecture

4.1 Phase Space Representation

Hidden state: $s_t = (q_t, p_t) \in \mathbb{R}^{2D}$, where q = positions (memory), p = momenta (temporal dynamics).

4.2 Learnable Hamiltonian

$$H(q, p; \theta) = \frac{1}{2} p^T M^{-1} p + V_{\text{NN}}(q; \theta)$$

Kinetic term: quadratic in p. Potential term: MLP.

4.3 Dissipation & Input Forcing

$$\dot{p} = -\nabla_q H(q, p) - \gamma p + F_\phi(x_t)$$

Dissipation γ introduces controlled memory decay.

4.4 Symplectic Integration (Leapfrog)

$$\begin{aligned} p_{t+1/2} &= p_t - \tfrac{\epsilon}{2} \nabla_q V(q_t) - \tfrac{\gamma\epsilon}{2} p_t + \tfrac{\epsilon}{2} F(x_t) \\ q_{t+1} &= q_t + \epsilon M^{-1} p_{t+1/2} \\ p_{t+1} &= p_{t+1/2} - \tfrac{\epsilon}{2} \nabla_q V(q_{t+1}) - \tfrac{\gamma\epsilon}{2} p_{t+1/2} + \tfrac{\epsilon}{2} F(x_t) \end{aligned}$$

Output: $y_t = \text{Proj}_\psi(q_t, p_t)$.

4.5 Efficiency

Sequential update by default (like RNN). We will explore parallelization via scan/FFT convolution.

5 Experimental Plan

Phase 1 (Months 1–3): Synthetic Validation

Copy task, arithmetic, chaotic signal reconstruction. Metrics: long-horizon gradient stability, error growth.

Phase 2 (Months 3–6): Benchmarks

Datasets: LRA, PG-19. Baselines: LSTM, Transformer, Mamba, Hyena.

Phase 3 (Months 6–9): Finance Proof-of-Concept

Datasets: EUR/USD tick data, CME Futures. Tasks: return prediction, volatility forecasting. Metrics: Sharpe ratio, AUC/F1, regime robustness. Interpretability: phase-space plots, energy diagnostics.

Phase 4 (Months 9–12): Safety & Interpretability

Energy vs. confidence calibration, robustness to adversarial perturbations, instability detection as safety signal.

6 Risks & Mitigation

Risk 1: Hyperparameter sensitivity. *Mitigation:* Provide default configs + auto-tuning.

Risk 2: Financial gains uncertain. *Mitigation:* Multi-metric success matrix; contribution justified via stability/interpretability.

Risk 3: Reviewer background mismatch. *Mitigation:* Frame as general ML contribution; finance as case study.

Plan C: If finance results weak, pivot narrative to stable sequence modeling.

7 Timeline

- Months 1–2: Prototype + synthetic tasks.
- Month 3: ArXiv preprint + NeurIPS Workshop.
- Months 4–6: LRA/PG-19 benchmarks.
- Months 6–9: Finance PoC.
- Months 9–12: Safety analysis + main submission.

8 Broader Impact

- **For ML:** Introduces physics-inspired inductive bias.
- **For AI Safety:** Architecture-level interpretability.
- **For Quant Finance:** Physics-informed ML for chaotic markets.
- **For Career:** Bridges QR and AI Safety.

9 Resource Needs

- Phase 1–2: 1–2 GPUs.
- Phase 3–4: Ideally 8x A100 cluster; or research cloud credits.
- Budget: $\sim \$20k$ compute.

References

[1] Greydanus et al., "Hamiltonian Neural Networks," NeurIPS 2019.

[2] Gu et al., "Efficiently Modeling Long Sequences with Structured State Spaces," ICLR 2022.