NEURIPS 2026 · SYDNEY

Phase Space Attention

A Hairer Lift Circumvents the Single-Layer Induction Obstruction

arXiv:2609.32319

Kingsuk Maitra · Shagun Sood · Morteza Hosseini · Suman Gunnala · Vikram Gupta

Qualcomm Cloud AI Division, Santa Clara

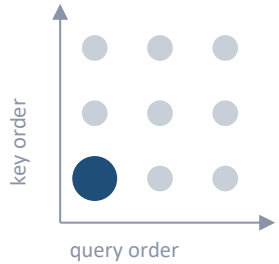
Code: 41 notebooks in the supplementary

THE OBSTRUCTION

Single-layer induction is a filter-order gap

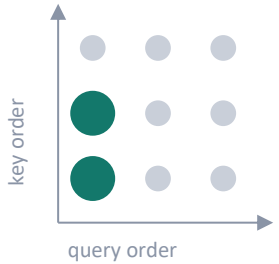
Standard attention

order (0, 0)



What induction needs

key-side order ≥ 1



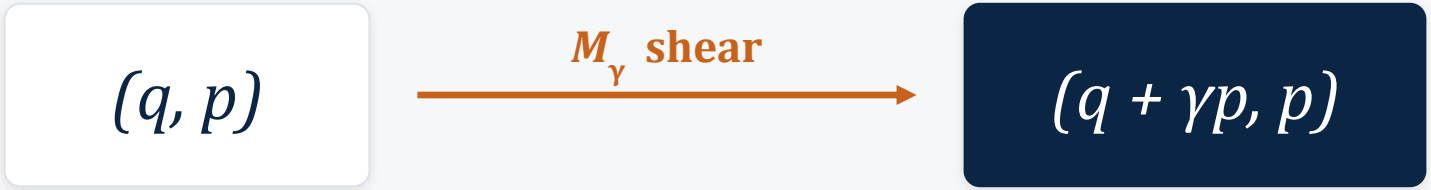
1. Induction must match the token *after* the previous occurrence: each key has to expose its predecessor.
2. Standard attention is order (0, 0): one layer is capped at the order-zero ceiling **1/(T – 2)**.
3. Order on the query side alone, (1, 0), adds **nothing**.

Why it matters: below ~100 M parameters, every layer is expensive.

THE HAIRER LIFT

Lift to phase space: two steps become one

Störmer–Verlet is a two-step recurrence in q alone. Carry a momentum, $p = q_n - q_{n-1}$, and it becomes one-step.



One-step, linear, causal

Uses only the current and previous token.

Exact γ-flow

of the free Hamiltonian $H = \tfrac{1}{2}\|p\|^2$: no truncation error.

Symplectic, but…

so is every variant considered. It is not what selects the shear.

PHASE SPACE ATTENTION

Shear the post-RoPE queries and keys

$$\hat{q} = q + \gamma\, p_q \qquad \hat{k} = k + \gamma\, p_k \qquad A = \mathrm{softmax}(\, \hat{q}^{\top}\hat{k} \,/\, \sqrt{d_k}\,)$$

T1

$q^{\top}k$

standard score

T2

$\gamma\, q^{\top}p_k$

ghost key: sees each key's predecessor

T3

$\gamma\, p_q{}^{\top}k$

query-side mirror

T4

$\gamma^2\, p_q{}^{\top}p_k$

coherence channel

2 lines of code

1 scalar γ

0 new weight matrices

WHAT WE PROVE

Four results that pin the operator down

Uniqueness Theorem 4

In the factorised subclass, causality, high-pass, DC-neutrality and q ↔ k swap symmetry single out the shear, up to γ.

Structure Proposition 5

Asymmetric shears are symplectic too. What symmetry buys is the transport identity.

Invariance Theorem 6

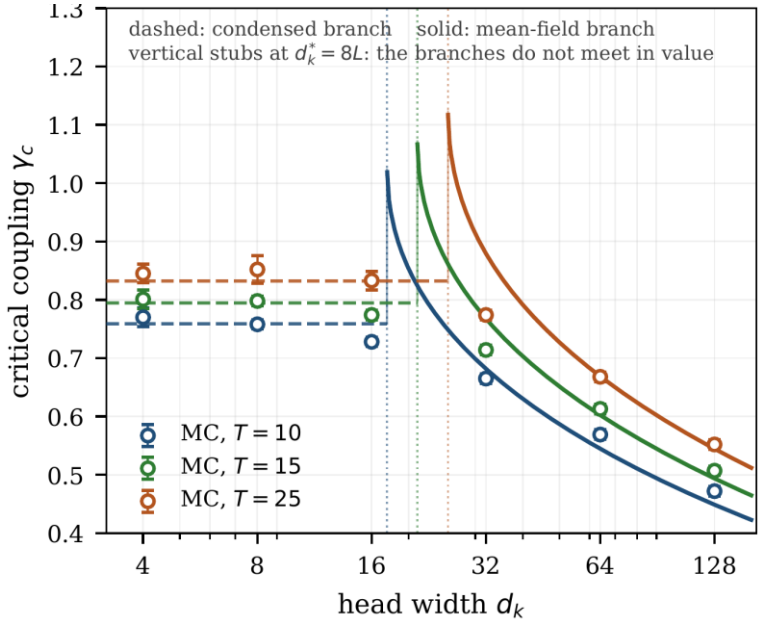
An operator-level invariance. No long-time stability claim: the map is applied once.

Placement Corollary 7

Shear after RoPE. Differencing before rotation leaves a Coriolis term $2\gamma|\sin(\theta/2)|\|q\|$.

A CLOSED-FORM TRANSITION

Theory predicts where the lift switches on



Two-branch law (lines) vs Monte Carlo (points)

$$\mathrm{SNR} = \sqrt{d_k} / 2$$

$$\gamma_c^2 = 2L / (\sqrt{d_k} + \sqrt{(d_k - 8L)}) \quad L = \log(T - 1)$$

r = 0.9876

9 Monte Carlo points · 0 fitted parameters

Trained models: γ_c falls with width: 1.29 → 0.12 as d_k goes 8 → 64.

Open gap: fitted exponent −1.13 vs −¼ predicted.

EVIDENCE

One layer now does induction

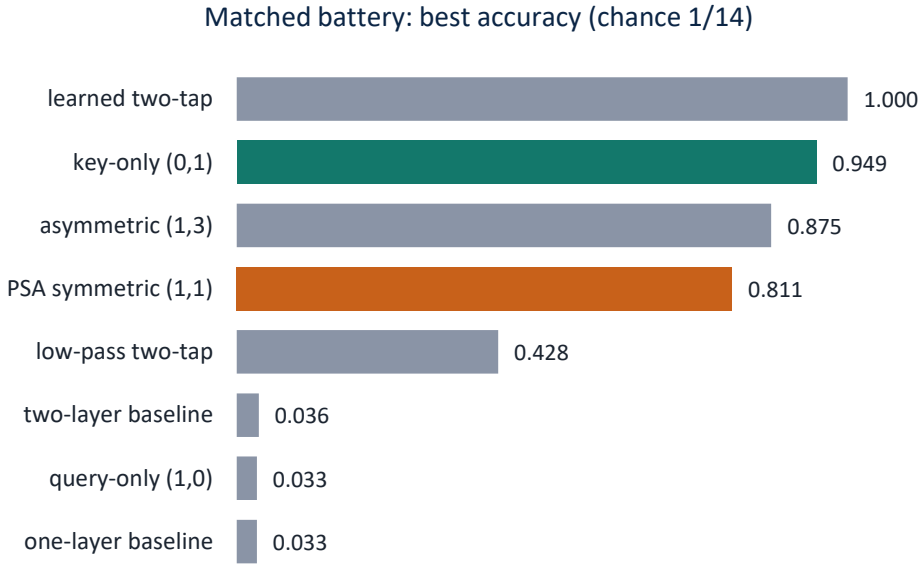
69.1%

± 8.2, five seeds, single layer

1.6%

chance, where the standard single layer sits

Accuracy only? Use key-only. Symmetric PSA carries the structure.



Query-only stays at baseline, as the theory predicts.

TAKE-AWAYS

A two-line change closes a structural gap

1. The single-layer induction obstruction is a filter-order gap.
2. A one-parameter Hairer lift closes it: unique in its class, placed after RoPE.
3. A closed-form law predicts where it switches on.

arXiv:2609.32319

Scope: synthetic tasks; no claim at language-model scale.

Reproducible: 41 notebooks, outputs embedded: verify every figure without a GPU.

Poster session Wednesday 9 December · 5–8 PM AEDT · Sydney