NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2026 · Atlanta

Beyond Unit-Circle Eigenvalues: Invariant Bases for Stable State Space Dynamics

Jiaqian Zhu[1,†] Yang Zhang[2,*] Junhua Ding[3] Xiaowei Yu[1,*]

[1]Missouri University of Science and Technology [2]University of North Texas [3]University of Wyoming

[*]Corresponding authors [†]Project leader

Code Website WeChat

Open to co-founders & investors

jiaqian-janelle.github.io

1 Problem: Stability Is Not Conservation

A diagonal state space model applies **the same transition** at every rollout step, so any quantity the transition does not conserve **drifts with the horizon**.

Gap. A unit-circle spectrum preserves the **norm**, yet total momentum drifts by **141%**. Prior work constrains the **spectrum** or penalises a named quantity in the **loss**. Neither fixes the **modal basis**, which determines what is conserved.

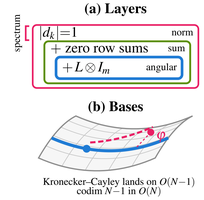


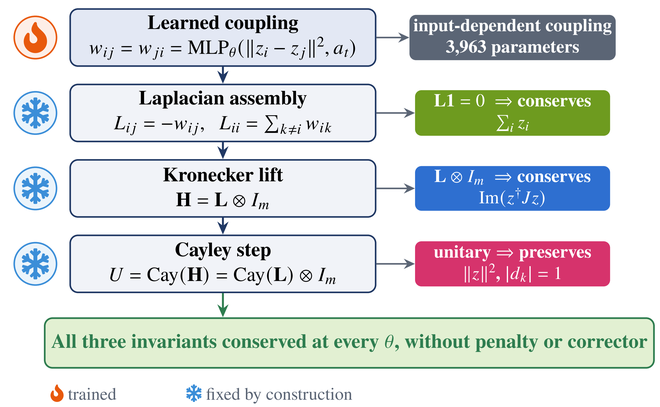
Figure 1. Conserved quantities are layered. (a) Each layer requires a strictly stronger condition. (b) Conserving pairs of basis and spectrum form a lower-dimensional set, and a basis trained under a penalty stays near its initialisation.

Table 1. What each design constrains, which invariants it conserves, and what it leaves free.

Design	Norm	Sum	Ang.	Diag.	All θ	Leaves free
Spectrum: constrain the eigenvalues						
Diagonal SSM	✗	✗	✗	✓	✗	eigenvectors
LRU	✓[p]	✗	✗	✓	✓	eigenvectors
Unitary RNN	✓	✗	✗	✗	✓	designated qty.
Generator: name the quantity						
Hamiltonian, SymODEN	✗	✓[p]	✓[p]	✗	✓	diagonal form
Deep conservation	✓[p]	✓[p]	✓[p]	✗	✓	diagonal form
Operator learning: correct the output						
FNO	✗	✗	✗	✗	✗	conserved qty.
CPFNO	✗	✓[p]	✗	✗	✓[p]	conserved qty.
KC (ours)	✓	✓	✓	✓	✓	coupling values

SSM: state space model. LRU: Linear Recurrent Unit. RNN: recurrent neural network. SymODEN: Symplectic ODE-Net. FNO: Fourier Neural Operator. CPFNO: FNO with the adaptive conservation corrector of Liu et al. (2025). KC: Kronecker–Cayley transition. Ang.: angular quantity. Diag.: diagonal form. All θ: conserved at every parameter value. qty.: quantity. ✓[p]: the norm is only bounded, the quantity is conserved only when named, or it is restored by a corrector while the prediction remains within reach.

2 Method: Conservation by Construction, the Kronecker–Cayley Transition



Theorem 1 (conserved quadratic forms). For $\mathbf{H} = \mathbf{H}^\dagger$, $U = \mathrm{Cay}(\mathbf{H})$ and Hermitian G,

$$z^\dagger U^\dagger G\, U z = z^\dagger G z \ \ \forall z \iff [\mathbf{H}, G] = 0.$$

Corollary 3 (layered conservation). Each condition implies the one to its left, and no converse holds:

$$\underbrace{|d_k| = 1}_{\|z\|^2} \Leftarrow \underbrace{\mathbf{L1} = 0}_{+\sum_i z_i} \Leftarrow \underbrace{\mathbf{H} = \mathbf{L}\otimes I_m}_{+\mathrm{Im}(z^\dagger J z)}$$

Theorem 2 (Kronecker–Cayley decomposition). With $\mathbf{L} = P\Lambda P^\top$ and $\tilde{z} = (P^\top \otimes I_m)z$,

$$\tilde{z}^{(k)}_{t+1} = d_k\, \tilde{z}^{(k)}_t, \qquad d_k = \frac{1 - i\lambda_k/2}{1 + i\lambda_k/2}, \quad |d_k| = 1, \quad d_{k^*} = 1.$$

Corollary 4 (invariant family). Over all symmetric $\mathbf{L}$ with zero row sums, the conserved quadratic forms are exactly $z^\dagger(I_N \otimes C_1 + \mathbf{1}\mathbf{1}^\top \otimes C_2)z$, with C_1, C_2 Hermitian, a $2m^2$-dimensional family.

Proposition 6 (unitary bound). For every unitary U, $\|Uz_t - z_{t+1}\| \geq \big|\|z_{t+1}\| - \|z_t\|\big|$, so a contracting system cannot be fit exactly.

Proposition 7 (summed quantity). $U = (Q\,\mathrm{diag}(d)\,Q^\top)\otimes I_m$ conserves $\sum_i z_i$ iff $(d_k - 1)c_k = 0$ for every k, with $c = Q^\top \mathbf{1}/\sqrt{N}$. KC places $\mathbf{1}/\sqrt{N}$ in the null mode, so the other $N-1$ modes rotate freely.

3 Experiments | *The structure of the generator, not the spectrum or the loss, decides what is conserved*

Q1 Is the invariant conserved?

Setup. Dam break and advection (PDEBench), shallow water on the sphere (PDEArena), conservative Allen–Cahn. One shared encoder and decoder.

Table 2. Mass drift $|\Delta M|\downarrow$ at the final horizon, geometric mean.

Model	Dam	Adv.	A–C	Sphere
reference	0	10^{-6}	10^{-16}	30.8
Diagonal recurrences				
level B	27.7	477	4.70	2.7e4
level C	41.9	409	3.28	2.4e4
free	286	421	53.9	1.9e11
LRU	2.3e3	1.8e3	18.6	6.7e4
Neural operators				
FNO	3.6e6[†]	9.3e3	61.0	2.4e7[†]
CPFNO, corr.	1.3e6[†]	**2.8e−4**	2.7e−4	1.8e4[†]
KC (ours)	**8.3e−3**	2.6e−3	**2.0e−5**	**483**

Dam: dam break. Adv.: advection. A–C: Allen–Cahn. level B: $|d_k|$=1. level C: level B with one mode per object fixed at 1. free: unconstrained spectrum. corr.: corrected. [†]: seed 42 only. Others as in Table 1.

KC conserves mass to the **round-off level** of the float32 rollout on the three flat benchmarks, whereas every diagonal baseline loses it by **two to nine orders of magnitude**.

Q2 Can a corrector substitute?

Setup. Dam break, five seeds. The learned corrector of CPFNO and a fixed uniform corrector on FNO. A seed counts as diverged once its normalised error exceeds 100.

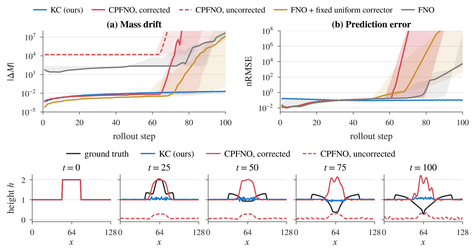


Figure 2. Top: mass drift and prediction error, median of five seeds. Bottom: water height along the central row, seed 42.

A corrector restores the **conserved scalar, not the field**, and diverges with the prediction in **4 of 5** CPFNO seeds and in **every** fixed-corrector seed. Only KC remains bounded on every seed.

Q3 Is each condition necessary?

Setup. Visual N-body. We perturb the trained $\mathbf{L}$ with and without restoring its zero row sums, and compare four norm-preserving operators.

Table 3. Drift (%), where $\sim 10^{-6}$ is the float32 floor.

Transition	$\lvert d_k\rvert$=1	L1=0	L⊗I	energy	mom.
Perturbed trained Laplacian, $\varepsilon = 0.5$					
row sums kept	✓	✓	✓	**1.8e−6**	**3.5e−5**
row sums broken	✓	✗	✓	**2.0e−6**	110
Norm-preserving operators, 2000 steps					
Haar unitary	✓	✗	✗	3.5e−3	169.7
Cayley, dense	✓	✗	✗	1.2e−3	170.7
diag. unitary	✓	✗	✗	2.5e−4	172.5
KC (ours)	✓	✓	✓	**3.7e−5**	**4.0e−3**

mom.: momentum. diag.: diagonal. Haar unitary: unitary drawn from the Haar measure. ε: perturbation scale.

Unit modulus preserves energy yet lets momentum drift by **141%**. Breaking only the row sums costs **110%** in momentum while energy remains at the float32 floor.

Q4 Can training find the condition?

Setup. A free orthogonal basis trained with a momentum penalty on 16 seeds, against Hamiltonian and graph networks trained to convergence.

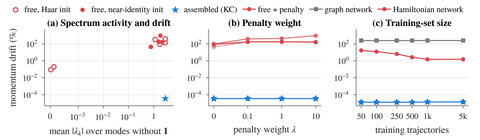


Figure 3. (a) Drift against spectrum activity. (b) Penalty weight. (c) Training-set size.

Table 4. Momentum drift (%) at step 25 against training-set size.

Training trajectories	KC	HNN	GNN
50	**1.29e−5**	16.9	247
250	**1.30e−5**	6.46	246
1000	**1.38e−5**	1.49	248
5000	**1.41e−5**	1.52	248

HNN: Hamiltonian neural network. GNN: graph neural network.

14 of 16 seeds retain a rotating spectrum and drift by **44–932%**. The other two conserve only by freezing it (mean $|d_k| < 10^{-3}$). KC conserves with an **active spectrum**.

Q5 How far does it transfer?

Setup. EGNN N-body (charged, gravitational, spring) over 1000 steps, a stiff reaction network, and four systems from other fields.

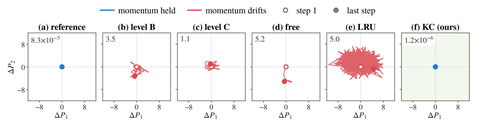


Figure 4. Charged N-body, total linear momentum over 1000 steps, seed 42. Only the reference and KC stay at a point.

Table 5. Linear-momentum drift at $T = 1000$, geometric mean over seeds 42–44.

System	B	C	free	LRU	KC
charged	3.29	2.66	1.9e4	6.41	**2.2e−6**
gravity	4.06	3.92	4.7e3*	4.70	**3.6e−6**
spring	1.25	1.36	5.64	3.79	**1.6e−6**

EGNN: E(n) Equivariant Graph Neural Network. B, C: levels as in Table 2. *: the free diagonal diverged on one seed.

KC conserves linear momentum to 10^{-6}. Angular momentum, bilinear in the state, improves 3–12× but stays above the floor, and a contracting system is fit less accurately, as Proposition 6 predicts.

4 Applications: Conservation Laws in Real-World Dynamical Systems

Molecular dynamics
Conserved quantity: **total linear momentum**
1.4×10^{-4} % mean drift over **2000 steps**
100.5% with row sums broken

Chemical kinetics
Conserved quantity: **total mass**
6.3×10^{-3} % mean drift over **2000 steps**
126.9% with row sums broken

Pharmacokinetics
Conserved quantity: **total drug amount**
1.0×10^{-3} % mean drift over **2000 steps**
127.8% with row sums broken

Power-grid swing dynamics
Conserved quantity: **angular momentum**
5.3×10^{-3} % mean drift over **2000 steps**
127.0% with row sums broken

Outlook: sequence models
Diagonal SSMs such as S4D, S5 and Mamba constrain the spectrum for stability. Our results suggest that what a long rollout conserves is decided by the **modal basis**, a design axis these models leave free.